第 10 章 方差分析

10.1 内容提要

10.1.1 单因素试验的方差分析

1. 单因素试验

为了考察某个因素 A 对所研究的随机变量 X 的影响, 我们在试验中让其他因素保持不变, 而仅让因素 A 改变, 这样的试验称为单因素试验.

2. 数学模型

设因素 A 有 r 个不同的水平 $A_1, A_2, \cdots, A_r$, 在水平 A_j 下的总体记为 x_j, $j=1,2,\cdots,r$, 并设 $x_1, x_2, \cdots, x_r$ 相互独立且

$$x_j \sim N(\mu_j, \sigma^2), j=1,2,\cdots,r .$$

在水平 A_j 下进行 n_j 次试验, 得到取自总体 x_j 的容量为 n_j 的样本 $x_{1j}, \cdots, x_{n_j j}$ $(j=1, 2, \cdots, r)$. 于是有

$$x_{ij} \sim N(\mu_j, \sigma^2), j=1,2,\cdots,r; i=1,2,\cdots,n_j ,$$

并且所有的 x_{ij} 相互独立.

令 $\varepsilon_{ij} = x_{ij} - \mu_j (j=1,2,\cdots,r; i=1,2,\cdots,n_j)$, 则 ε_{ij} 是在水平 A_j 下做第 i 次观察时由于随机因素的影响而产生的随机误差, 且 ε_{ij} 相互独立. 所以可得如下的数据结构:

$$\begin{cases} x_{ij} = \mu_j + \varepsilon_{ij}, j=1,2,\cdots,r; i=1,2,\cdots,n_j \\ \varepsilon_{ij} \sim N(0,\sigma^2), \end{cases}$$

上式称为单因素方差分析的数学模型. 单因素方差分析的主要任务主要可归结为以下两个:

(1) 在给定的显著水平 α 下检验假设:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_r; \leftrightarrow H_1 : \mu_1, \mu_2, \cdots, \mu_r \text{ 不全相等};$$

(2) 估计参数 $\mu_1, \cdots, \mu_r, \sigma^2$.

3. 统计分析

总平方和 $S_T = \sum_{j=1}^{r}\sum_{i=1}^{n_j}(x_{ij} - \overline{x})^2$, 其中 $\overline{x} = \frac{1}{n}\sum_{j=1}^{r}\sum_{i=1}^{n_j} x_{ij}$.

误差平方和 $S_E=\sum_{j=1}^{r}\sum_{i=1}^{n_j}(x_{ij}-\overline{x}_{\cdot j})^2$, 其中 $\overline{x}_{\cdot j}=\frac{1}{n_j}\sum_{i=1}^{n_j}x_{ij}$.

因素 A 的效应平方和 $S_A=\sum_{j=1}^{r}\sum_{i=1}^{n_j}(\overline{x}_{\cdot j}-\overline{x})^2$.

4. 假设检验

当 H_0 成立时, 则 $\frac{S_E}{\sigma^2}\sim\chi^2(n-r)$, $\frac{S_A}{\sigma^2}\sim\chi^2(r-1)$, 且 S_E 与 S_A 相互独立. 从而

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

对给定的检验水平 α, 若 $F\geq F_\alpha(r-1,n-r)$, 则拒绝原假设 H_0, 即认为因素 A 影响显著.

5. 参数估计

(1) $\hat{\mu}=\overline{x}$ 是 μ 的无偏估计;

(2) $\hat{\mu}_j=\overline{x}_{\cdot j}$ 是 μ_j 的无偏估计;

(3) $\hat{\alpha}_j=\overline{x}_{\cdot j}-\overline{x}$ 是 α_j 的无偏估计;

(4) $\hat{\sigma}^2=\frac{S_E}{n-r}$ 是 σ^2 的无偏估计;

(5) $\mu_j-\mu_k$ 的置信度为 $1-\alpha$ 的置信区间为

$$\left(\overline{x}_{\cdot j}-\overline{x}_{\cdot k}-t_{\alpha/2}(n-r)\sqrt{\frac{S_E}{n-s}(\frac{1}{n_j}+\frac{1}{n_k})},\overline{x}_{\cdot j}-\overline{x}_{\cdot k}+t_{\alpha/2}(n-r)\sqrt{\frac{S_E}{n-s}(\frac{1}{n_j}+\frac{1}{n_k})}\right).$$

10.1.2 双因素等重复试验的方差分析

1. 数学模型

设有两个 A, B 作用于试验的指标, 因素 A 有 r 个水平 $A_1,A_2,\cdots,A_r$, 因素 B 有 s 个水平 $B_1,B_2,\cdots,B_s$, 对因素 A,B 的水平的每对组合 $(A_i,B_j),i=1,2,\cdots,r;\ j=1,2,\cdots,s$ 都作 $t(t\geq 2)$ 次试验, 试验数记为 x_{ijk}, 设 $x_{ijk}\sim N(\mu_{ij},\sigma^2),i=1,2,\cdots,r;\ j=1,2,\cdots,s;\ k=1,2,\cdots,t$. 令 $\varepsilon_{ijk}=x_{ijk}-\mu_{ij}(i=1,2,\cdots,r;\ j=1,2,\cdots,s;k=1,2,\cdots,t)$, 则 ε_{ijk} 相互独立, 且 $\varepsilon_{ijk}\sim N(0,\sigma^2)$.
于是数据就有如下结构

$$\begin{cases}x_{ijk}=\mu_{ij}+\varepsilon_{ijk},i=1,\cdots,r;\ j=1,\cdots,s;k=1,\cdots,t\\ \varepsilon_{ijk}\sim N(0,\sigma^2)\end{cases},$$

上式就是双因素方差分析的数学模型.

引入如下记号:

$$\mu=\frac{1}{rs}\sum_{i=1}^{r}\sum_{j=1}^{s}\mu_{ij}\text{；}\qquad \mu_{i\cdot}=\frac{1}{s}\sum_{j=1}^{s}\mu_{ij},\alpha_i=\mu_{i\cdot}-\mu,i=1,2,\cdots,r\text{；}$$

$$\mu_{\cdot j}=\frac{1}{r}\sum_{i=1}^{r}\mu_{ij},\beta_i=\mu_{\cdot j}-\mu,j=1,2,\cdots,s.$$

称 μ 为总平均, 称 α_i 为水平 A_i 的效应, β_i 为水平 B_j 的效应, 记

$$\gamma_{ij}=\mu_{ij}-\mu_{i\cdot}-\mu_{\cdot j}+\mu,i=1,\cdots,r;j=1,\cdots,s,$$

则 $\mu_{ij}=\mu+\alpha_i+\beta_j+\gamma_{ij}$, 称 γ_{ij} 为水平 A_i 和水平 B_j 的交互效应, 于是数据的结构可以写成如下的数学模型:

$$\begin{cases} x_{ijk}=\mu+\alpha_i+\beta_j+\gamma_{ij}+\varepsilon_{ijk}, & i=1,\cdots,r,\\ \varepsilon_{ijk}\sim N(0,\sigma^2), & j=1,\cdots,s,\\ \sum_{i=1}^{r}\alpha_i=0,\sum_{j=1}^{s}\beta_j=0,\sum_{i=1}^{r}\gamma_{ij}=\sum_{j=1}^{s}\gamma_{ij}=0, & k=1,\cdots,t \end{cases}$$

其中 μ, α_i, β_j, γ_{ij}, σ^2 都是未知参数.

这样假设检验问题可以表述成如下的三个假设检验问题:

$H_{01}:\alpha_1=\alpha_2=\cdots=\alpha_r=0;\leftrightarrow H_{11}:\alpha_1,\cdots,\alpha_r$ 不全为零;

$H_{02}:\beta_1=\beta_2=\cdots=\beta_s=0;\leftrightarrow H_{12}:\beta_1,\cdots,\beta_s$ 不全为零;

$H_{03}:\gamma_{ij}=0,i=1,\cdots,r,j=1,\cdots,s;\leftrightarrow H_{13}:\gamma_{ij}$ 不全为零.

2. 统计分析

总平方和 $S_T=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(x_{ijk}-\overline{x})^2$, 其中 $\overline{x}=\frac{1}{rst}\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}x_{ijk}$,

因素 A 的效应平方和 $S_A=st\sum_{i=1}^{r}(\overline{x}_{i\cdot\cdot}-\overline{x})^2$, 其中 $\overline{x}_{i\cdot\cdot}=\frac{1}{st}\sum_{j=1}^{s}\sum_{k=1}^{t}x_{ijk},i=1,\cdots,r$,

因素 B 的效应平方和 $S_B=rt\sum_{j=1}^{s}(\overline{x}_{\cdot j\cdot}-\overline{x})^2$, 其中 $\overline{x}_{\cdot j\cdot}=\frac{1}{rt}\sum_{i=1}^{r}\sum_{k=1}^{t}x_{ijk},j=1,\cdots,s$,

误差平方和 $S_E=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(x_{ijk}-\overline{x}_{ij\cdot})^2$, 其中 $\overline{x}_{ij\cdot}=\frac{1}{t}\sum_{k=1}^{t}x_{ijk},i=1,\cdots,r;j=1,\cdots,s$,

A, B 的交互效应平方和 $S_{A\times B}=t\sum_{i=1}^{r}\sum_{j=1}^{s}(\overline{x}_{ij\cdot}-\overline{x}_{i\cdot\cdot}-\overline{x}_{\cdot j\cdot}+\overline{x})^2$.

3. 假设检验

(1) $\dfrac{S_E}{\sigma^2} \sim \chi^2(rs(t-1))$,

(2) 当H_{01}为真时, $\dfrac{S_A}{\sigma^2} \sim \chi^2(r-1)$, 从而 $F_A = \dfrac{S_A/(r-1)}{S_E/(rs(t-1))} \sim F(r-1, rs(t-1))$;

(3) 当H_{02}为真时, $\dfrac{S_B}{\sigma^2} \sim \chi^2(s-1)$, 从而 $F_B = \dfrac{S_B/(s-1)}{S_E/(rs(t-1))} \sim F(s-1, rs(t-1))$;

(4) 当H_{03}为真时, $\dfrac{S_{A\times B}}{\sigma^2} \sim \chi^2((r-1)(s-1))$, 从而

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

对于显著性水平α, H_{01}拒绝域为 $F_A = \dfrac{S_A/(r-1)}{S_E/(rs(t-1))} \ge F_\alpha(r-1, rs(t-1))$,

H_{02}拒绝域为 $F_B = \dfrac{S_B/(s-1)}{S_E/(rs(t-1))} \ge F_\alpha(s-1, rs(t-1))$,

H_{03}拒绝域为 $F_{A\times B} = \dfrac{S_{A\times B}/[(r-1)(s-1)]}{S_E/(rs(t-1))} \ge F_\alpha((r-1)(s-1), rs(t-1))$.

10.1.3 双因素无重复试验的方差分析

1. 数学模型

设有两个因素A, B, 因素A有r个水平$A_1, A_2, \cdots, A_r$, 因素 B 有s个水平$B_1, B_2, \cdots, B_s$, 对因素A, B的每对组合$(A_i, B_j), i=1,2,\cdots,r, j=1,2,\cdots,s$, 作一次试验, 试验数据记为$x_{ij}$, 设$x_{ij}$相互独立, 且设$x_{ij} \sim N(\mu_{ij}, \sigma^2)$. 此时数据结构的数学模型可以写成如下形式:

$$\begin{cases} x_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, i=1,2,\cdots,r; j=1,2,\cdots,s \\ \varepsilon_{ij} \sim N(0,\sigma^2), \sum_{i=1}^{r}\alpha_i = 0, \sum_{j=1}^{s}\beta_j = 0 \end{cases},$$

其中μ为总平均, α_i为水平A_i的效应, β_j为水平B_j的效应, 且μ, α_i, β_j, σ^2均为未知参数. 此时要检验假设有以下两个:

$$H_{01}: \alpha_1 = \alpha_2 = \cdots = \alpha_r = 0; \leftrightarrow H_{11}: \alpha_1, \cdots, \alpha_r \text{不全为零};$$

$$H_{02}: \beta_1 = \beta_2 = \cdots = \beta_s = 0; \leftrightarrow H_{12}: \beta_1, \cdots, \beta_s \text{不全为零}.$$

与双因素等重复试验方差分析的讨论过程类似, 可以得到方差分析表:

方差来源	平方和	自由度	均方	F比
因素A	S_A	$r-1$	$S_A/(r-1)$	$F_A = \dfrac{S_A/(r-1)}{S_E/((r-1)(s-1))}$

因素 B	S_B	$s-1$	$S_B/(s-1)$	$F_A=\dfrac{S_B/(s-1)}{S_E/((r-1)(s-1))}$
误差	S_E	$(r-1)(s-1)$	$S_E/(r-1)(s-1)$	
总和	S_T	$rs-1$		

取显著性水平为α，得$H_{01}\leftrightarrow H_{11}$的拒绝域为

$$F_A=\frac{S_A/(r-1)}{S_E/((r-1)(s-1))}\ge F_\alpha(r-1,(r-1)(s-1)),$$

得$H_{02}\leftrightarrow H_{12}$的拒绝域为

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

10.2 习题详解

1. 三台机器制造同一种产品，记录五天的产量如下：

机器	A_1	A_2	A_3
日产量	138	163	155
	144	148	144
	135	152	159
	149	146	147
	143	157	153

试在显著性$\alpha=0.05$下检验这三台机器的日产量是否有显著差异.

解 对假设检验问题$H_0:\mu_1=\mu_2=\mu_3 \quad \leftrightarrow \quad H_1:\mu_1,\mu_2,\mu_3$不全相等,

取检验统计量$F=\dfrac{S_A/(r-1)}{S_E/(n-r)}$，则在零假设$H_0$下有

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

拒绝域为$F=\dfrac{S_A/(r-1)}{S_E/(n-r)}\ge F_\alpha(r-1,n-r)$，$\alpha$为显著性水平,

其中$S_A=\sum\limits_{j=1}^{r}\sum\limits_{i=1}^{n_j}(\overline{x}_{.j}-\overline{x})^2=\sum\limits_{j=1}^{r}n_j\overline{x}_{.j}^2-n\overline{x}^2$，$S_E=\sum\limits_{j=1}^{r}\sum\limits_{i=1}^{n_j}(x_{ij}-\overline{x}_{.j})^2$，$r=3$，$n=15$.

计算分析得

组	观测数	求和	平均	方差
列 1	5	709	141.8	29.7
列 2	5	766	153.2	47.7
列 3	5	758	151.6	36.8

差异源	SS	df	MS	F	P-value	F crit
组间	380.9333	2	190.4667	5.003503	0.026286	3.885294
组内	456.8	12	38.06667			
总计	837.7333	14				

查表得$F=5.003503$，因为$F_{0.05}(2,12)=3.89$，所以$F\geq F_{0.05}(2,12)$，拒绝原假设，因此三台机器的日产量有显著差异.

2. 下列数据给出了对灯泡光通量的试验结果(单位:流明/瓦特)

工厂	测量值					
1	9.47	9.00	9.12	9.27	9.27	9.25
2	10.80	11.28	11.15			
3	10.37	10.42	10.28			
4	10.65	10.33				
5	9.54	8.62				

试在显著性水平$\alpha=0.05$下检测不同工厂生产的灯泡光通量有无显著差别？

解 $H_0:\mu_1=\mu_2=\mu_3=\mu_4=\mu_5 \quad \leftrightarrow \quad H_1:\mu_1,\mu_2,\mu_3,\mu_4,\mu_5$不全相等,

构造检验统计量$F=\dfrac{S_A/(r-1)}{S_E/(n-r)}$,则在零假设$H_0$下有

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

拒绝域为

$$F=\frac{S_A/(r-1)}{S_E/(n-r)}\geq F_\alpha(r-1,n-r),\ \alpha\text{ 为显著性水平},$$

其中$S_A=\sum_{j=1}^{r}\sum_{i=1}^{n_j}(\overline{x}_{.j}-\overline{x})^2=\sum_{j=1}^{r}n_j\overline{x}_{.j}^2-n\overline{x}^2$，$S_E=\sum_{j=1}^{r}\sum_{i=1}^{n_j}(x_{ij}-\overline{x}_{.j})^2$，$r=5$，$n=16$.

计算分析得

组	观测数	求和	平均	方差
列 1	6	55.38	9.23	0.02524
列 2	3	33.23	11.07667	0.061633
列 3	3	31.07	10.35667	0.005033
列 4	2	20.98	10.49	0.0512
列 5	2	18.16	9.08	0.4232

差异源	SS	df	MS	F	P-value	F crit
组间	9.502642	4	2.37566	35.60577	3.1E-06	3.35669
组内	0.733933	11	0.066721			
总计	10.23658	15				

计算得 $F=35.60577$，$F_{0.05}(4,11)=3.36$，$F \ge F_{0.05}(4,11)$，拒绝原假设，因此不同工厂生产的灯泡光通量有显著差别.

3. 将抗生素注入人体会产生抗生素与血浆蛋白质结合的现象,以致减少了药效. 下表列出 5 种常用的抗生素注入到牛的体内时,抗生素与血浆蛋白质结合的百分比. 试在显著性水平 $\alpha=0.05$ 检验这些百分比的均值有无显著的差异. 设各总体服从正态分布,且方差相同.

抗生素	青霉素	四环素	链霉素	红霉素	氯霉素
测量值	29.6	27.3	5.8	21.6	29.2
	24.3	32.6	6.2	17.4	32.8
	28.5	30.8	11.0	18.3	25.0
	32.0	34.8	8.3	19.0	24.2

解 $H_0: \mu_1=\mu_2=\mu_3=\mu_4=\mu_5 \quad \leftrightarrow \quad H_1: \mu_1,\mu_2,\mu_3,\mu_4,\mu_5$ 不全相等,

构造检验统计量 $F=\dfrac{S_A/(r-1)}{S_E/(n-r)}$,则在零假设 H_0 下有

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

拒绝域为

$$F=\frac{S_A/(r-1)}{S_E/(n-r)} \ge F_\alpha(r-1,n-r),\ \alpha \text{ 为显著性水平},$$

其中 $S_A=\sum_{j=1}^{r}\sum_{i=1}^{n_j}(\overline{x}_{.j}-\overline{x})^2=\sum_{j=1}^{r}n_j\overline{x}_{.j}^2-n\overline{x}^2$, $S_E=\sum_{j=1}^{r}\sum_{i=1}^{n_j}(x_{ij}-\overline{x}_{.j})^2$, $r=5$, $n=20$.

计算分析得

组	观测数	求和	平均	方差
列 1	4	114.4	28.6	10.35333
列 2	4	125.5	31.375	10.05583
列 3	4	31.3	7.825	5.6825
列 4	4	76.3	19.075	3.2625
列 5	4	111.2	27.8	15.92

差异源	SS	df	MS	F	P-value	F crit
组间	1480.823	4	370.2058	40.88488	6.74E-08	3.055568
组内	135.8225	15	9.054833			
总计	1616.646	19				

计算得 $F=40.88488$, $F_{0.05}(4,15)=3.06$, $F\geq F_{0.05}(4,15)$, 拒绝原假设, 即表明均值有显著差异.

4. 一个年级有三个班, 他们进行了一次数学考试, 现从各个班级随机地抽取了一些学生, 记录其成绩如下:

1 班			2 班			3 班		
73	66	89	88	77	78	68	41	79
60	82	45	31	48	78	59	56	68
80	43	93	91	62	51	91	53	71
36	73	77	76	85	96	79	71	15
			74	80	56		87	

试在显著性水平 $\alpha=0.05$ 下检验各班级的平均分数有无显著性差异. 设各个总体服从正态分布, 且方差相同.

解 $H_0:\mu_1=\mu_2=\mu_3 \quad \leftrightarrow \quad H_1:\mu_1,\mu_2,\mu_3$ 不全相等,

构造检验统计量 $F=\dfrac{S_A/(r-1)}{S_E/(n-r)}$, 则在零假设 H_0 下有

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

拒绝域为

$$F=\frac{S_A/(r-1)}{S_E/(n-r)}\ge F_\alpha(r-1,n-r)\text{，}\boldsymbol{\alpha}\text{ 为显著性水平，}$$

其中 $S_A=\sum_{j=1}^{r}\sum_{i=1}^{n_j}(\overline{x}_{.j}-\overline{x})^2=\sum_{j=1}^{r}n_j\overline{x}_{.j}^2-n\overline{x}^2$，$S_E=\sum_{j=1}^{r}\sum_{i=1}^{n_j}(x_{ij}-\overline{x}_{.j})^2$，$r=3$，$n=30$．

计算分析得

组	观测数	求和	平均	方差
列 1	12	817	68.08333	343.9015
列 2	15	1071	71.4	327.9714
列 3	13	838	64.46154	414.6026

差异源	SS	df	MS	F	P-value	F crit
组间	335.3526	2	167.6763	0.46473	0.631923	3.251924
组内	13349.75	37	360.804			
总计	13685.1	39				

计算得 $F=0.46473$，$F_{0.05}(2,37)=3.25$，$F\le F_{0.05}(2,37)$，接受原假设，即表明平均分数没有显著差异.

5. 为了寻找适应某地区的高产水稻品种，今选取五个不同品种的种子进行试验，每一品种在四种试验田上试种. 假定这 20 块土地面积与其他条件基本上相同，观测到各块土地的产量 (kg) 如下：

种子品种 A	田号			
	1	2	3	4
A_1	67	67	55	42
A_2	68	96	90	66
A_3	60	69	50	55
A_4	79	64	81	70
A_5	90	70	79	88

试检验：(1) 种子品种对水稻高产有无显著影响 $(\alpha=0.01)$，(2) 第 2，5 号种子对水稻高产的影响有无显著差异 $(\alpha=0.05)$．

解 (1) $H_0:\alpha_1=\alpha_2=\alpha_3=\alpha_4=\alpha_5=0 \leftrightarrow H_1:\alpha_1,\alpha_2,\alpha_3,\alpha_4,\alpha_5$ 不全为 0，

构造检验统计量 $F_A=\dfrac{S_A/(r-1)}{S_E/((r-1)(s-1))}$,则在零假设 H_0 下有

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

拒绝域为 $F_A=\dfrac{S_A/(r-1)}{S_E/((r-1)(s-1))}\ge F_\alpha(r-1,(r-1)(s-1))$,

其中 $S_A=st\sum_{i=1}^{r}(\overline{x}_{i..}-\overline{x})^2$, $S_E=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(x_{ijk}-\overline{x}_{ij.})^2$, 该题中 $t=1$.

SUMMARY	观测数	求和	平均	方差
行 1	4	231	57.75	142.25
行 2	4	320	80	232
行 3	4	234	58.5	65.66667
行 4	4	294	73.5	63
行 5	4	327	81.75	84.25
列 1	5	364	72.8	138.7
列 2	5	366	73.2	167.7
列 3	5	355	71	305.5
列 4	5	321	64.2	295.2

差异源	SS	df	MS	F	P-value	F crit
行	2128.7	4	532.175	4.258252	0.022536	5.411951
列	261.8	3	87.26667	0.698273	0.570851	5.952545
误差	1499.7	12	124.975			
总计	3890.2	19				

计算得 $F_A=4.258252$，$F_{0.01}(4,12)=5.411951$，$F_A\le F_{0.01}(4,12)$, 接受原假设, 即表明种子品种对水稻高产无显著差异.

(2) $H_0:\alpha_2=\alpha_5=0 \quad \leftrightarrow \quad H_1:\alpha_2,\alpha_5$ 不全为 0 ,

构造检验统计量 $F_A=\dfrac{S_A/(r-1)}{S_E/((r-1)(s-1))}$,则在零假设 H_0 下有

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

拒绝域为 $F_A=\dfrac{S_A/(r-1)}{S_E/((r-1)(s-1))}\ge F_\alpha(r-1,(r-1)(s-1))$,

其中 $S_A = st\sum_{i=1}^{r}(\overline{x}_{i..} - \overline{x})^2$, $S_E = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(x_{ijk} - \overline{x}_{ij.})^2$, 该题中 $t = 1$.

SUMMARY	观测数	求和	平均	方差
行 1	4	320	80	232
行 2	4	327	81.75	84.25
列 1	2	158	79	242
列 2	2	166	83	338
列 3	2	169	84.5	60.5
列 4	2	154	77	242

差异源	SS	df	MS	F	P-value	F crit
行	6.125	1	6.125	0.020967	0.89405	10.12796
列	72.375	3	24.125	0.082585	0.965061	9.276628
误差	876.375	3	292.125			
总计	954.875	7				

计算得 $F_A = 0.020967$，$F_{0.05}(1,3) = 10.12796$，$F_A \le F_{0.05}(1,3)$，接受原假设,即无显著差异.

6. 下面记录了三位操作工分别在四种不同机器上操作三天的日产量:

机器 A	操作工 B								
	B_1			B_2			B_3		
A_1	15	15	17	19	19	16	16	18	21
A_2	17	17	17	15	15	15	19	22	22
A_3	15	17	16	18	17	16	18	18	18
A_4	18	20	22	15	16	17	17	17	17

试在显著性水平 $\alpha = 0.05$ 下检验操作工人之间的差异是否显著?机器之间差异是否显著?交互影响是否显著?

解 $H_0 : \alpha_{11} = \alpha_2 = \alpha_3 = \alpha_4 = 0 \quad \leftrightarrow \quad H_1 : \alpha_{11}, \alpha_2, \alpha_3, \alpha_4$ 不全为 0 ,

$H_{02} : \beta_1 = \beta_2 = \beta_3 = 0 \quad \leftrightarrow \quad H_{12} : \beta_1, \ \beta_2 \ \ \beta_3$ 不全为 0 ,

$H_{03} : \gamma_{ij} = 0, i = 1,2,3,4; j = 1,2,3 \quad \leftrightarrow \quad H_{13} : \gamma_{ij}$ 不全为 0 .

构造检验统计量,并分别得到零假设下分布

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

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

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

拒绝域分别为

$$F_A=\frac{S_A/(r-1)}{S_E/(rs(t-1))}\geq F_\alpha(r-1,rs(t-1)),$$

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

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

观测数	3	3	3	9
求和	47	54	55	156
平均	15.66667	18	18.33333	17.33333
方差	1.333333	3	6.333333	4.25
观测数	3	3	3	9
求和	51	45	63	159
平均	17	15	21	17.66667
方差	0	0	3	7.75
观测数	3	3	3	9
求和	48	51	54	153
平均	16	17	18	17
方差	1	1	0	1.25
观测数	3	3	3	9
求和	60	48	51	159
平均	20	16	17	17.66667
方差	4	1	0	4.5
总计				
观测数	12	12	12	
求和	206	198	223	

平均	17.16667	16.5	18.58333
方差	4.333333	2.272727	4.083333

差异源	SS	df	MS	F	P-value	F crit
样本	2.75	3	0.916667	0.532258	0.664528	3.008787
列	27.16667	2	13.58333	7.887097	0.00233	3.402826
交互	73.5	6	12.25	7.112903	0.000192	2.508189
内部	41.33333	24	1.722222			
总计	144.75	35				

由计算得 $F_A = 0.532258$，$F_B = 7.887097$，$F_{A\times B} = 7.112903$，又有 $F_{0.05}(3,24) = 3.01$，$F_{0.05}(2,24) = 3.4$，$F_{0.05}(6,24) = 2.51$，所以机器间无显著差异，工人间有显著差异，交互影响有显著差异.

7. 在化工生产中为了提高得率，选了三种不同浓度，四种不同温度情况做试验. 为了考虑浓度与温度的交互作用，在浓度 (%) 与温度 (^0C) 的每一种水平组合下各做两次试验，其得率数据如下面的表所示(数据均已减去 75)

	$B_1=10$		$B_2=24$		$B_3=38$		$B_4=52$	
$A_1=2$	14	10	11	11	13	9	10	12
$A_2=4$	9	7	10	8	7	11	6	10
$A_3=6$	5	11	13	14	12	13	14	10

试在显著性水平 $\alpha = 0.05$ 下检验不同浓度，不同温度以及它们之间的交互作用对得率有无显著影响？

解 $H_{01}:\alpha_1=\alpha_2=\alpha_3=0 \ \leftrightarrow \ H_{11}:\alpha_1,\alpha_2,\alpha_3$ 不全为 0，

$H_0 \ :\beta_{12}=\beta_2=\beta_3=\beta_4=0 \ \leftrightarrow H_{12}:\beta_1,\ \beta_2,\ \beta_3\ \beta_4$ 不全为 0，

$H_{03}:\gamma_{ij}=0, i=1,2,3; j=1,2,3,4. \ \leftrightarrow \ H_{13}:\gamma_{ij}$ 不全为 0.

构造检验统计量,并分别得到零假设下分布

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

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

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

则拒绝域分别为 $F_A=\dfrac{S_A/(r-1)}{S_E/(rs(t-1))}\geq F_\alpha(r-1,rs(t-1))$,

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

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

观测数	2	2	2	2	8
求和	24	22	22	22	90
平均	12	11	11	11	11.25
方差	8	0	8	2	2.785714

观测数	2	2	2	2	8
求和	16	18	18	16	68
平均	8	9	9	8	8.5
方差	2	2	8	8	3.142857

观测数	2	2	2	2	8
求和	16	27	25	24	92
平均	8	13.5	12.5	12	11.5
方差	18	0.5	0.5	8	8.857143

总计

观测数	6	6	6	6
求和	56	67	65	62
平均	9.333333	11.16667	10.83333	10.33333
方差	9.866667	4.566667	5.766667	7.066667

差异源	SS	df	MS	F	P-value	F crit
样本	44.33333	2	22.16667	4.092308	0.044153	3.885294
列	11.5	3	3.833333	0.707692	0.565693	3.490295
交互	27	6	4.5	0.830769	0.568369	2.99612
内部	65	12	5.416667			

总计	147.8333	23

由计算得 $F_A = 4.092308$，$F_B = 0.707692$，$F_{A\times B} = 0.830769$，又有 $F_{0.05}(2,12) = 3.885294$，$F_{0.05}(3,12) = 3.490295$，$F_{0.05}(6,12) = 2.99612$，所以浓度间有显著影响，温度无显著影响，交互作用无显著影响.

8. 考察合成纤维弹性，影响因素为：收缩率 A 和总的拉伸倍数 B. 试验结果如下表：

	$A_1=0$		$A_2=4$		$A_3=8$		$A_4=12$	
$B_1=460$	71	73	73	75	76	73	75	73
$B_2=520$	72	73	76	74	79	77	73	72
$B_3=580$	75	73	78	77	74	75	70	71
$B_4=640$	77	75	74	74	74	73	69	69

试在显著性水平 $\alpha = 0.05$ 下检验因素 A，B 及它们的交互作用对试验结果是否有显著性影响差异？

解 $H_0 : \alpha_{11} = \alpha_2 = \alpha_3 = \alpha_4 = 0 \quad \leftrightarrow \quad H_1 : \alpha_{11}, \alpha_2, \alpha_3, \alpha_4$ 不全为 0，

$H_0 : \beta_{12} = \beta_2 = \beta_3 = \beta_4 = 0 \quad \leftrightarrow \quad H_{12} : \beta_1,\ \beta_2,\ \beta_3\ \beta_4$ 不全为 0，

$H_{03} : \gamma_{ij} = 0, i = 1,2,3,4; j = 1,2,3,4 \quad \leftrightarrow \quad H_{13} : \gamma_{ij}$ 不全为 0.

构造检验统计量,并分别得到零假设下分布

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

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

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

拒绝域分别为

$$F_A = \frac{S_A/(r-1)}{S_E/(rs(t-1))} \ge F_\alpha(r-1, rs(t-1)),$$

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

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

观测数	2	2	2	2	8
求和	144	145	148	152	589
平均	72	72.5	74	76	73.625
方差	2	0.5	2	2	3.696429

观测数	2	2	2	2	8
求和	148	150	155	148	601
平均	74	75	77.5	74	75.125
方差	2	2	0.5	0	2.982143

观测数	2	2	2	2	8
求和	149	156	149	147	601
平均	74.5	78	74.5	73.5	75.125
方差	4.5	2	0.5	0.5	4.410714

观测数	2	2	2	2	8
求和	148	145	141	138	572
平均	74	72.5	70.5	69	71.5
方差	2	0.5	0.5	0	4.571429

总计

观测数	8	8	8	8
求和	589	596	593	585
平均	73.625	74.5	74.125	73.125
方差	2.553571	6.571429	7.553571	7.839286

差异源	SS	df	MS	F	P-value	F crit
样本	70.59375	3	23.53125	17.51163	2.62E-05	3.238872
列	8.59375	3	2.864583	2.131783	0.136299	3.238872
交互	79.53125	9	8.836806	6.576227	0.000591	2.537667
内部	21.5	16	1.34375			
总计	180.2188	31				

由计算得 $F_A = 17.51163$，$F_B = 2.131783$，$F_{A\times B} = 6.576227$，又有 $F_{0.05}(3,16) = 3.24$，

$F_{0.05}(3,16) = 3.24$，$F_{0.05}(9,16) = 2.54$，所以 A 的影响显著，B 的影响不显著，交互作用影响显著.

9. 进行农业试验，选择四个不同品种的小麦及三块试验田，每块试验田分成四块面积相等的小块，各种植一个品种的小麦，收获量 (kg) 如下：

	B_1	B_2	B_3	B_4
A_1	26	30	22	20
A_2	25	23	21	21
A_3	24	25	20	19

试在显著性水平 $\alpha = 0.05$ 下检验小麦品种及试验田对收获量是否有显著影响？

解 $H_{01}:\alpha_1 = \alpha_2 = \alpha_3 = 0 \ \leftrightarrow\ H_{11}:\alpha_1,\alpha_2,\alpha_3$ 不全为 0，

$H_0\ :\beta_{12}= \beta_2 = \beta_3 = \beta_4 = 0 \ \leftrightarrow\ H_{12}:\beta_1,\ \beta_2,\ \beta_3\ \beta_4$ 不全为 0，

构造检验统计量,并分别得到零假设下分布

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

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

拒绝域分别为

$$F_A = \frac{S_A/(r-1)}{S_E/((r-1)(s-1))} \ge F_\alpha(r-1,(r-1)(s-1)),$$

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

SUMMARY	观测数	求和	平均	方差
行 1	4	98	24.5	19.66667
行 2	4	90	22.5	3.666667
行 3	4	88	22	8.666667
列 1	3	75	25	1
列 2	3	78	26	13
列 3	3	63	21	1
列 4	3	60	20	1

差异源	SS	df	MS	F	P-value	F crit
行	14	2	7	2.333333	0.177979	5.143253
列	78	3	26	8.666667	0.013364	4.757063
误差	18	6	3			
总计	110	11				

则计算得 $F_A = 2.3333$，$F_B = 8.6667$，又有 $F_{0.05}(2,6) = 5.14$，$F_{0.05}(3,6) = 4.757$，所以试验田对收获量无显著影响，小麦品种对收获量有显著差异.

10. 在橡胶生产过程中，选择四种不同的配料方案 A 及五种不同的硫化时间 B，测得产品的抗压强度 (kg/cm^3) 如下：

	B_1	B_2	B_3	B_4	B_5
A_1	151	157	144	134	136
A_2	144	162	128	138	132
A_3	134	133	130	122	125
A_4	131	126	124	126	121

试分别在显著性水平 $\alpha = 0.05$，0.01 下检验配料方案及硫化时间对产品的抗压强度是否有显著影响？

解 （1）显著性水平 $\alpha = 0.05$ 情况下：

$H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0 \quad \leftrightarrow \quad H_1: \alpha_1, \alpha_2, \alpha_3, \alpha_4$ 不全为 0，

$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0 \quad \leftrightarrow \quad H_1: \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ 不全为 0，

构造检验统计量，并分别得到零假设下分布

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

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

拒绝域分别为

$$F_A = \frac{S_A/(r-1)}{S_E/((r-1)(s-1))} \ge F_\alpha(r-1,(r-1)(s-1)),$$

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

SUMMARY	观测数	求和	平均	方差
行 1	5	722	144.4	95.3
行 2	5	704	140.8	177.2

行 3	5	644	128.8	26.7
行 4	5	628	125.6	13.3
列 1	4	560	140	84.66667
列 2	4	578	144.5	312.3333
列 3	4	526	131.5	75.66667
列 4	4	520	130	53.33333
列 5	4	514	128.5	45.66667

差异源	SS	df	MS	F	P-value	F crit
行	1243.8	3	414.6	10.55857	0.001103	3.490295
列	778.8	4	194.7	4.958404	0.013595	3.259167
误差	471.2	12	39.26667			
总计	2493.8	19				

则计算得 $F_A = 10.55857$，$F_B = 4.958404$，因为 $F_{0.05}(3,12) = 3.49$，$F_{0.05}(4,12) = 3.259$，所以两者均有显著影响.

(2) 显著性水平 $\alpha = 0.01$ 情况下:

$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0 \quad \leftrightarrow H_1 : \alpha_1, \alpha_2, \alpha_3, \alpha_4$ 不全为 0，

$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0 \quad \leftrightarrow \quad H_1 : \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ 不全为 0，

构造检验统计量,并分别得到零假设下分布

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

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

则拒绝域分别为

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

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

SUMMARY	观测数	求和	平均	方差
行 1	5	722	144.4	95.3
行 2	5	704	140.8	177.2

行 3	5	644	128.8	26.7		
行 4	5	628	125.6	13.3		
列 1	4	560	140	84.66667		
列 2	4	578	144.5	312.3333		
列 3	4	526	131.5	75.66667		
列 4	4	520	130	53.33333		
列 5	4	514	128.5	45.66667		

差异源	SS	df	MS	F	P-value	F crit
行	1243.8	3	414.6	10.55857	0.001103	5.952545
列	778.8	4	194.7	4.958404	0.013595	5.411951
误差	471.2	12	39.26667			
总计	2493.8	19				

则计算得 $F_A = 10.55857$，$F_B = 4.958404$, 因为 $F_{0.05}(3,12) = 5.9525$，$F_{0.05}(4,12) = 5.412$，所以硫化时间无显著差异, 表明配料方案有显著影响.