MATLAB统计分析与应用

——方差分析

主讲人：谢中华

主要内容

- 单因素一元方差分析
- 双因素一元方差分析
- 多因素一元方差分析
- 非参数方差分析

第一节 单因素一元方差分析

方差分析就是多个总体均值的比较检验

一、问题引入与原理简介

【例11.1-1】 某饮料生产企业研制出一种新型饮料。饮料的颜色共有四种，分别为橘黄色、粉色、绿色和无色透明。这四种饮料的营养含量、味道、价格、包装等可能影响销售量的因素全部相同。现从地理位置相似、经营规模相仿的五家超级市场上收集了前一时期该饮料的销售情况，见表1。试分析饮料的颜色对销售量是否产生影响。

表1　该饮料在五家超市的销售情况

超市	无色	粉色	橘黄色	绿色
1	26.5	31.2	27.9	30.8
2	28.7	28.3	25.1	29.6
3	25.1	30.8	28.5	32.4
4	29.1	27.9	24.2	31.7
5	27.2	29.6	26.5	32.8

设有k组数据，每组有 n_i个观察值

组别	观测值(x_{ij} , j = 1,…,n_i ; i = 1,…,k)						总和	平均	均方
π_1	x_{11}	x_{12}	…	x_{1j}	…	x_{1n_1}	T_1	$\overline{x}_1$	s_1^2
π_2	x_{21}	x_{22}	…	x_{2j}	…	x_{2n_2}	T_2	$\overline{x}_2$	s_2^2
…	…	…	…		…	…	…	…	…
π_i	x_{i1}	x_{i2}	…	x_{ij}	…	x_{in_i}	T_i	$\overline{x}_i$	s_i^2
…	…	…	…		…	…	…	…	…
π_k	x_{k1}	x_{k2}	…	x_{kj}	…	x_{kn_k}	T_k	$\overline{x}_k$	s_k^2
	$n=\sum_{i=1}^{k} n_i$			$\overline{x}_i=\frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$				$\overline{x}$	

1. 平方和与自由度的分解

观测数据总变异的平方和记为

$$SS_T = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\overline{x})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\overline{x}_i+\overline{x}_i-\overline{x})^2$$

$$= \sum_{i=1}^{k} n_i(\overline{x}_i-\overline{x})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\overline{x}_i)^2$$

$$= SS_A + SS_E$$

SS_T反映了全部数据总的偏差程度，称为总离差平方和；

SS_E反映了随机误差的大小，称为组内离差平方和；

SS_A反映了系统误差的大小，称为组间离差平方和。

2. 检验规则：

检验统计量

组间均方

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

组内均方

当F的观测值超过某个临界值时就可认为组间差异显著

3. 一元方差分析表

方差来源	SS（平方和）	DF（自由度）	MS（均方）	F值	F临界值
组间（SS_A）	SS_A	$k-1$	MS_A	$F=\dfrac{MS_A}{MS_E}$	$F_\alpha(k-1,n-k)$
组内（SS_E）	SS_E	$n-k$	MS_E		
总计（SS_T）	SS_T	$n-1$			

二、单因素一元方差分析的MATLAB实现

1. anova1函数

功能：单因素一元方差分析

调用方式：

```
p = anova1(X)
p = anova1(X,group)
p = anova1(X,group,displayopt)
[p,table] = anova1(...)
[p,table,stats] = anova1(...)
```

2. multcompare函数

功能：多重比较检验

调用方式：

```
c = multcompare(stats)
c = multcompare(stats,param1,val1,param2,val2,...)
[c,m] = multcompare(...)
[c,m,h] = multcompare(...)
[c,m,h,gnames] = multcompare(...)
```

【例11.1-1】 分析饮料的颜色对销售量是否产生影响。

```
x = [26.5    31.2    27.9    30.8
     28.7    28.3    25.1    29.6
     25.1    30.8    28.5    32.4
     29.1    27.9    24.2    31.7
     27.2    29.6    26.5    32.8];
[p,table] = anova1(x)
```

ANOVA Table

Source	SS	df	MS	F	Prob>F
Columns	76.845	3	25.6152	10.49	0.0005
Error	39.084	16	2.4427		
Total	115.929	19			

【例11.1-1】 分析饮料的颜色对销售量是否产生影响。

一元方差分析表

方差来源	SS	DF	MS	F	P值	F临界值
组间	76.845	3	25.6152	10.49	0.0005	3.2398
组内	39.084	16	2.4427			
总计	115.929	19				

这里取显著性水平为0.05，由于F 值> F临界值，故拒绝H_0，即认为饮料的颜色对销售量有显著影响。

【例11.1-2】 研究6种氮肥施用法对小麦的效应，每种施肥法种5盆小麦，完全随机设计。最后测定它们的含氮量（mg），试作方差分析

施氮法					
1	2	3	4	5	6
12.9	14.0	12.6	10.5	14.6	14.0
12.3	13.8	13.2	10.8	14.6	13.3
12.2	13.8	13.4	10.7	14.4	13.7
12.5	13.6	13.4	10.8	14.4	13.5
12.7	13.6	13.0	10.5	14.4	13.7
12.52	13.76	13.12	10.66	14.48	13.64

MATLAB程序

```
x = [12.9      14.0   12.6   10.5   14.6   14.0
12.3   13.8   13.2   10.8   14.6   13.3
12.2   13.8   13.4   10.7   14.4   13.7
12.5   13.6   13.4   10.8   14.4   13.5
12.7   13.6   13.0   10.5   14.4   13.7];
[p,table] = anova1(x)
```

ANOVA Table

Source	SS	df	MS	F	Prob>F
Columns	44.463	5	8.8926	164.17	9.62346e-018
Error	1.3	24	0.05417		
Total	45.763	29			

【例11.1-3】 某灯泡厂用四种不同配料方案制成的灯丝生产四批灯泡，在每一批中取若干个作寿命试验，得如下数据（单位：小时）：问灯丝的不同配料方案对灯泡寿命有无显著影响？

灯泡品种	A_1	1600,1610,1650,1680,1700,1720,1800
	A_2	1580,1640,1640,1700,1750
	A_3	1460,1550,1600,1620,1640,1660,1740,1820
	A_4	1510,1520,1530,1570,1600,1680

MATLAB程序

```
x1 = [1600,1610,1650,1680,1700,1720,1800];
x2 = [1580,1640,1640,1700,1750];
x3 = [1460,1550,1600,1620,1640,1660,1740,1820];
x4 = [1510,1520,1530,1570,1600,1680];
x = [x1 x2 x3 x4];
group =
[ones(1,7),2*ones(1,5),3*ones(1,8),4*ones(1,6)];
[p,table] = anova1(x,group)
```

ANOVA Table

Source	SS	df	MS	F	Prob>F
Groups	44360.7	3	14786.9	2.15	0.1229
Error	151350.8	22	6879.6		
Total	195711.5	25			

【例11.1-4】现有某高校2005-2006学年第1学期2077名同学的《高等数学》课程的考试成绩，共涉及6个学院的69个班级，试根据全部2077名同学的考试成绩，分析不同学院的学生的考试成绩有无显著差别。

学　　号	姓　　名	班　　级	学　　院	学院编号	总成绩
05010101	郭强	050101	机械	1	87
05010102	张旭鹏	050101	机械	1	71
05010103	李桂艳	050101	机械	1	75
…	…	…	…	…	…

➢ 正态性检验

在调用anova1函数作方差分析之前，应先检验数据是否满足方差分析的基本假定，即检验正态性和方差齐性。

```
>> [x,y]=xlsread('AnovaData1.xls');
>> score = x(:,2);
>> college = y(2:end,4);
>> college_id = x(:,1);
% 分别对6个学院的考试成绩进行正态性检验
>> for i = 1:6
     scorei = score(college_id == i);
     [h,p] = lillietest(scorei);   % 正态性检验
     result(i,:) = p;
   end
% 查看正态性检验的p值
>> result
```

➢ 方差齐性检验

下面调用vartestn函数检验6个学院的学生的考试成绩是否服从方差相同的正态分布。

```
% 调用vartestn函数进行方差齐性检验
>> [p,stats] = vartestn(score,college)
p =
   0.7138
stats =
   chisqstat: 2.9104
          df: 5
```

➤ 方差分析

经检验，认为6个学院学生的考试成绩服从同方差的正态分布，下面进行单因素一元方差分析，检验不同学院的成绩有无显著差别。

```
>> [p,table,stats] = anova1(score,college)
```

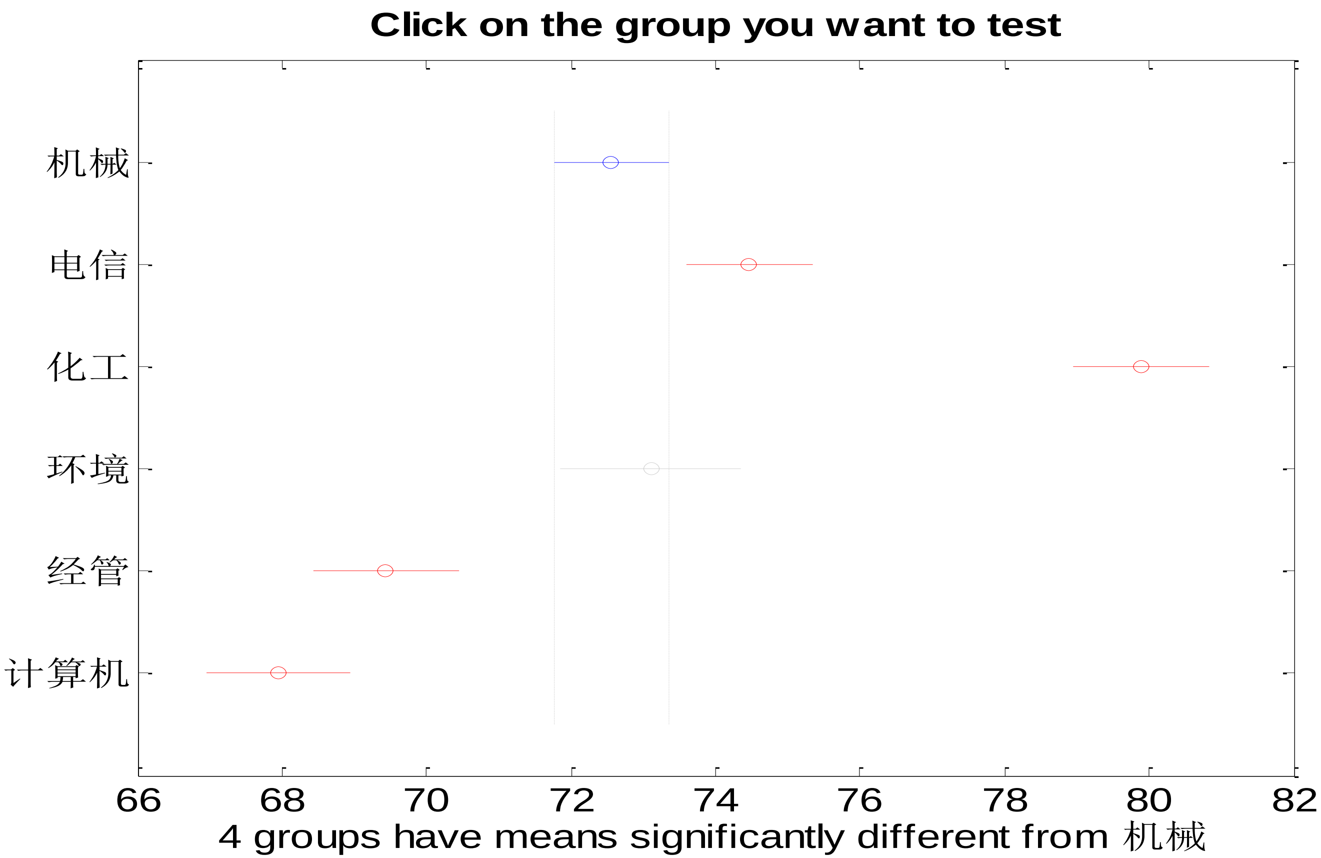
ANOVA Table

Source	SS	df	MS	F	Prob>F
Groups	29191.9	5	5838.38	76.74	0
Error	157560.8	2071	76.08		
Total	186752.7	2076			

➢ 多重比较检验

方差分析的结果已表明不同学院的学生的考试成绩有非常显著的差别，但这并不意味着任意两个学院学生的考试成绩都有显著的差别，因此还需要进行两两的比较检验，即多重比较，找出考试成绩存在显著差别的学院。

```
>> [c,m,h,gnames] = multcompare(stats);
>> head = {'组序号','组序号','置信下限','组均值差','置信上限'};
>> [head; num2cell(c)]
```



Click on the group you want to test
机械
电信
化工
环境
经管
计算机
66
68
70
72
74
76
78
80
82
4 groups have means significantly different from 机械

第二节 双因素一元方差分析

一、问题引入

例11.2-1 在某种橡胶的配方中，考虑了三种不同的促进剂，四种不同分量的氧化锌。各种配方试验一次，测得300%定强如下：问不同促进剂、不同分量氧化锌分别对定强有无显著影响？

氧化锌B / 促进剂A	B_1	B_2	B_3	B_4
A_1	32	35	35.5	38.5
A_2	33.5	36.5	38	39.5
A_3	36	37.5	39.5	43

例11.2-2 考察合成纤维中对纤维弹性有影响的两个因素：收缩率A和总拉伸倍数B。A和B各取4种水平，整个试验重复一次，试验结果如下：试问收缩率和总拉伸倍数分别对纤维弹性有无显著影响，并问两者对纤维弹性有无显著交互作用？

因子B / 因子A	460 (B_1)	520 (B_2)	580 (B_3)	640 (B_4)
0 (A_1)	71, 73	72, 73	75, 73	77, 75
4 (A_2)	73, 75	76, 74	78, 77	74, 74
8 (A_3)	76, 73	79, 77	74, 75	74, 73
12 (A_4)	75, 73	73, 72	70, 71	69, 69

二、无重复试验的二因素方差分析（无交互作用）

设因素A有r个不同的水平： $A_1, A_2, \cdots, A_r,$

因素B有s个不同的水平： $B_1, B_2, \cdots, B_s,$

对每个可能的搭配 $A_i \times B_j$ 进行一次独立试验，

共获得 rs个试验结果 $x_{ij}, i=1,2,\cdots,r, \; j=1,2,\cdots s,$

A \ B	B_1	B_2	$\cdots$	B_s	平均值$\bar{x}_{i\cdot}$
A_1	x_{11}	x_{12}	$\cdots$	x_{1s}	$\bar{x}_{1\cdot}$
A_2	x_{21}	x_{22}	$\cdots$	x_{2s}	$\bar{x}_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_r	x_{r1}	x_{r2}	$\cdots$	x_{rs}	$\bar{x}_{r\cdot}$
平均值$\bar{x}_{\cdot j}$	$\bar{x}_{\cdot 1}$	$\bar{x}_{\cdot 2}$	$\cdots$	$\bar{x}_{\cdot s}$	$\bar{x}$

1. 平方和与自由度的分解

$$SS_T=\sum_{i=1}^{r}\sum_{j=1}^{s}(x_{ij}-\overline{x})^2$$

$$=s\sum_{i=1}^{r}(\overline{x}_{i\cdot}-\overline{x})^2+r\sum_{j=1}^{s}(\overline{x}_{\cdot j}-\overline{x})^2$$

$$+\sum_{i=1}^{r}\sum_{j=1}^{s}(x_{ij}-\overline{x}_{i\cdot}-\overline{x}_{\cdot j}+\overline{x})^2$$

$$=SS_A+SS_B+SS_E$$

SS_T反映了全部数据总的偏差程度，称为总离差平方和；

SS_A称为因子A引起的离差平方和；

SS_B称为因子B引起的离差平方和；

SS_E称为随机误差引起的离差平方和。

2. 检验规则

对A因素检验，检验统计量

$$F_A = \frac{SS_A/(r-1)}{SS_E/\left((r-1)(s-1)\right)} = \frac{MS_A}{MS_E} \sim F\left(r-1,\ (r-1)(s-1)\right)$$

对B因素检验，检验统计量

$$F_B = \frac{SS_B/(s-1)}{SS_E/\left((r-1)(s-1)\right)} = \frac{MS_B}{MS_E} \sim F\left(s-1,\ (r-1)(s-1)\right)$$

检验规则：

- 当$F_A \geq F_\alpha\left(r-1,\ (r-1)(s-1)\right)$时，认为A因素影响显著；
- 当$F_A < F_\alpha\left(r-1,\ (r-1)(s-1)\right)$时，认为A因素影响不显著 。
- 当$F_B \geq F_\alpha\left(s-1,\ (r-1)(s-1)\right)$时，认为B因素影响显著；
- 当$F_B < F_\alpha\left(s-1,\ (r-1)(s-1)\right)$时，认为B因素影响不显著。

3. 无交互作用的二因素方差分析表

方差来源	平方和 SS	自由度 df	均方 MS	F值	临界值F_α
因素A	SS_A	r-1	MS_A	F_A	$F_\alpha(r-1, (r-1)(s-1))$
因素B	SS_B	s-1	MS_B	F_B	$F_\alpha(s-1, (r-1)(s-1))$
误差	SS_E	(r-1)(s-1)	MS_E		
总和	SS_T	rs-1			

三、双因素一元方差分析的MATLAB实现

1. anova2函数

功能：双因素一元方差分析

调用方式：

p = anova2(X,reps)

p = anova2(X,reps,displayopt)

[p,table] = anova2(...)

[p,table,stats] = anova2(...) **reps = 2 情形示例**

$$
X = \begin{array}{cc} A=1 & A=2 \end{array}
$$

$$
X = \left[\begin{array}{cc} x_{111} & x_{121} \\ x_{112} & x_{122} \\ x_{211} & x_{221} \\ x_{212} & x_{222} \\ x_{311} & x_{321} \\ x_{312} & x_{322} \end{array}\right] \begin{array}{l} \left.\vphantom{\begin{array}{c} x \\ x \end{array}}\right\} B=1 \\ \left.\vphantom{\begin{array}{c} x \\ x \end{array}}\right\} B=2 \\ \left.\vphantom{\begin{array}{c} x \\ x \end{array}}\right\} B=3 \end{array}
$$

例11.2-1 在某种橡胶的配方中，考虑了三种不同的促进剂，四种不同分量的氧化锌。各种配方试验一次，测得300%定强如下：问不同促进剂、不同分量氧化锌分别对定强有无显著影响？

氧化锌B / 促进剂A	B_1	B_2	B_3	B_4
A_1	32	35	35.5	38.5
A_2	33.5	36.5	38	39.5
A_3	36	37.5	39.5	43

MATLAB程序

```
x = [32   35   35.5  38.5
     33.5 36.5 38    39.5
     36   37.5 39.5  43]';
[p,table] = anova2(x)
```

ANOVA Table

Source	SS	df	MS	F	Prob>F
Columns	28.2917	2	14.1458	35.74	0.0005
Rows	66.0625	3	22.0208	55.63	0.0001
Error	2.375	6	0.3958		
Total	96.7292	11			

四、等重复试验的二因素方差分析（考虑交互作用）

设因素A有r个不同的水平： $A_1, A_2, \cdots, A_r,$

因素B有s个不同的水平： $B_1, B_2, \cdots, B_s,$

对每个可能的搭配 $A_i \times B_j$ 进行c次独立试验，

共获得 rsc 个试验结果 $x_{ijk},\ i=1,2,\cdots,r,\ j=1,2,\cdots s, k=1,\cdots,c$

因子A \ 因子B	B_1	B_2	$\cdots$	B_s
A_1	$x_{111},\cdots,x_{11c}$	$x_{121},\cdots,x_{12c}$	$\cdots$	$x_{1s1},\cdots,x_{1sc}$
A_2	$x_{211},\cdots,x_{21c}$	$x_{221},\cdots,x_{22c}$	$\cdots$	$x_{2s1},\cdots,x_{2sc}$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
A_r	$x_{r11},\cdots,x_{r1c}$	$x_{r21},\cdots,x_{r2c}$	$\cdots$	$x_{rs1},\cdots,x_{rsc}$

1. 平方和与自由度的分解

$$SS_T=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{c}(x_{ijk}-\bar{x})^2$$

$$=sc\sum_{i=1}^{r}(\bar{x}_{i\cdot\cdot}-\bar{x})^2+rc\sum_{j=1}^{s}(\bar{x}_{\cdot j\cdot}-\bar{x})^2$$

$$+c\sum_{i=1}^{r}\sum_{j=1}^{s}(\bar{x}_{ij\cdot}-\bar{x}_{i\cdot\cdot}-\bar{x}_{\cdot j\cdot}+\bar{x})^2$$

$$+\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{c}(x_{ijk}-\bar{x}_{ij\cdot})^2$$

$$=SS_A+SS_B+SS_{A\times B}+SS_E$$

SS_T 反映了全部数据总的偏差程度，称为总离差平方和；

SS_A 称为因子A引起的离差平方和；

SS_B 称为因子B引起的离差平方和；

$SS_{A\times B}$ 称为因子A和B的交互作用引起的离差平方和；

SS_E 称为随机误差引起的离差平方和。

2. 检验统计量

对A因素检验，检验统计量

$$F_A = \frac{SS_A/(r-1)}{SS_E/\left(rs(c-1)\right)} = \frac{MS_A}{MS_E}$$

对B因素检验，检验统计量

$$F_B = \frac{SS_B/(s-1)}{SS_E/\left(rs(c-1)\right)} = \frac{MS_B}{MS_E}$$

对交互效应检验，检验统计量

$$F_{A\times B} = \frac{SS_{A\times B}/\left((r-1)(s-1)\right)}{SS_E/\left(rs(c-1)\right)} = \frac{MS_{A\times B}}{MS_E}$$

3. 检验规则

- 当$F_A \geq F_\alpha(r-1,\ rs(c-1))$时，认为A影响显著；
- 当$F_A < F_\alpha(r-1,\ rs(c-1))$时，认为A影响不显著。

- 当$F_B \geq F_\alpha(s-1,\ rs(c-1))$时，认为B影响显著；
- 当$F_B < F_\alpha(s-1,\ rs(c-1))$时，认为B影响不显著。

- 当$F_{A\times B} \geq F_\alpha((r-1)(s-1),\ rs(c-1))$时，认为交互效应显著；
- 当$F_{A\times B} < F_\alpha((r-1)(s-1),\ rs(c-1))$时，认为交互效应不显著。

4. 考虑交互作用的二因素方差分析表

方差来源	平方和 SS	自由度 df	均方 MS	F值	临界值F_α
因素A	SS_A	r-1	MS_A	F_A	$F_\alpha(r-1,\ rs(c-1))$
因素B	SS_B	s-1	MS_B	F_B	$F_\alpha(s-1,\ rs(c-1))$
$A\times B$	$SS_{A\times B}$	(r-1)×(s-1)	$MS_{A\times B}$	$F_{A\times B}$	$F_\alpha((r-1)(s-1),\ rs(c-1))$
误差	SS_E	rs(c-1)	MS_E		
总和	SS_T	rsc-1			

例11.2-2 考察合成纤维中对纤维弹性有影响的两个因素：收缩率A和总拉伸倍数B。A和B各取4种水平，整个试验重复一次，试验结果如下：试问收缩率和总拉伸倍数分别对纤维弹性有无显著影响，并问两者对纤维弹性有无显著交互作用？

因子A \ 因子B	460 (B_1)	520 (B_2)	580 (B_3)	640 (B_4)
0 (A_1)	71, 73	72, 73	75, 73	77, 75
4 (A_2)	73, 75	76, 74	78, 77	74, 74
8 (A_3)	76, 73	79, 77	74, 75	74, 73
12 (A_4)	75, 73	73, 72	70, 71	69, 69

MATLAB程序

```
x = [71 73   72 73   75 73   77 75
     73 75   76 74   78 77   74 74
     76 73   79 77   74 75   74 73
     75 73   73 72   70 71   69 69]';
[p,table] = anova2(x,2)
```

ANOVA Table

Source	SS	df	MS	F	Prob>F
Columns	70.594	3	23.5313	17.51	0
Rows	8.594	3	2.8646	2.13	0.1363
Interaction	79.531	9	8.8368	6.58	0.0006
Error	21.5	16	1.3438		
Total	180.219	31			

第三节 多因素一元方差分析

一、多因素一元方差分析的原理

把样本数据的总离差平方和分解为：各因素主效应离差平方和、交互效应离差平方和、随机因素的离差平方和。根据平方和及自由度分解方式构造检验统计量，对各因素主效应、交互效应作出检验。

二、多因素一元方差分析的MATLAB实现

- anovan函数

功能：多因素一元方差分析

调用方式：

```
p = anovan(y,group)
p = anovan(y,group,param1,val1,param2,val2,...)
[p,table] = anovan(...)
[p,table,stats] = anovan(...)
[p,table,stats,terms] = anovan(...)
```

三、案例分析

【例11.3-1】 某养鸡场的蛋鸡育成期的配合饲料主要由5种成分（玉米、麸皮、豆饼、鱼粉和食盐）组成，分别记为A,B,C,D, E ，为了研究饲料配方对鸡产蛋效果的影响，对各成分均选取3个水平，进行5因素3水平的正交试验，通过试验找出饲料的最佳配方，试验要求考虑交互作用AB、AC 和AE . 5个因素（成分）的水平如下表所列。

水 平	因 素				
	A（玉米）	*B*（麸皮）	*C*（豆饼）	*D*（鱼粉）	*E*（食盐）
1	61.5	6.5	6.0	3.0	0.0
2	66.0	8.0	9.0	5.0	0.1
3	70.6	14.0	15.0	9.0	0.25

对这样的 5 因素 3 水平试验，可以选取正交表 $L_{27}(3^{13})$ 安排试验，表头设计为：将 A,B,C,E,D 依次放在第 1、2、5、8、11 列上，通过正交表的交互作用表可以查出，AB、AC 和 AE 应分别安排在(3，4)、(6，7)和(9，10)列上。按照这样的安排进行试验，得到试验数据如下表所列。试根据这些数据分析因素 A,B,C,D,E 以及交互作用 AB、AC 和 AE 对产蛋量 y 是否有显著影响，并找出饲料的最佳配方。

试验号	A	B	C	E	D	产蛋量 y
1	1	1	1	1	1	569
2	1	1	2	2	2	554
3	1	1	3	3	3	637
4	1	2	1	2	3	566
5	1	2	2	3	1	565
6	1	2	3	1	2	648
7	1	3	1	3	2	581
8	1	3	2	1	3	568
9	1	3	3	2	1	535
10	2	1	1	1	1	593
11	2	1	2	2	2	615
12	2	1	3	3	3	620
13	2	2	1	2	3	586
14	2	2	2	3	1	597
15	2	2	3	1	2	617
16	2	3	1	3	2	599
17	2	3	2	1	3	613
18	2	3	3	2	1	580
19	3	1	1	1	1	569
20	3	1	2	2	2	615
21	3	1	3	3	3	591
22	3	2	1	2	3	586
23	3	2	2	3	1	616
24	3	2	3	1	2	630
25	3	3	1	3	2	566
26	3	3	2	1	3	638
27	3	3	3	2	1	573

方差分析

```
>> ydata = xlsread('AnovaData2.xls');
>> y = ydata(:,7);
>> A = ydata(:,2);
>> B = ydata(:,3);
>> C = ydata(:,4);
>> D = ydata(:,6);
>> E = ydata(:,5);
>> varnames = {'A','B','C','D','E'};  % 定义因素名称
% 定义模型的效应项矩阵，考虑主效应：A,B,C,D,E，交互效应：AB,AC,AE
>> model = [eye(5);1 1 0 0 0;1 0 1 0 0;1 0 0 0 1]
% 调用anovan函数作多因素一元方差分析
>> [p,table] = anovan(y,{A,B,C,D,E},'model',model,'varnames',varnames)
```

% 定义模型的效应项矩阵，考虑主效应：A,B,C,D,E，交互效应：AC

```
>> model = [eye(5);1 0 1 0 0]
```

% 调用anovan函数作多因素一元方差分析

```
>> [p,table,stats] =
anovan(y,{A,B,C,D,E},'model',model,...
'varnames',varnames);
```

➢ 多重比较检验

```
>> [c,m,h,gnames] = multcompare(stats,'dimension',[1 2 3 4 5]);
% 将各处理的均值从小到大进行排序
>> [mean,id] = sort(m(:,1));
>> gnames = gnames(id);
% 显示排序后的后20个处理的名称及相应的均值
>> [{'处理','均值'};gnames(end-19:end),num2cell(mean(end-19:end))]
ans =
    '处理'                '均值'
    ......
    'A=3,B=1,C=2,D=2,E=3'    [637.7778]
    'A=3,B=2,C=2,D=3,E=3'    [640.8889]
    'A=3,B=2,C=2,D=2,E=3'    [643.1111]
    'A=3,B=1,C=2,D=3,E=1'    [643.6667]
    'A=3,B=1,C=2,D=2,E=1'    [645.8889]
    'A=3,B=2,C=2,D=3,E=1'    [649.0000]
    'A=3,B=2,C=2,D=2,E=1'    [651.2222]
```

结论

上面通过对各处理的均值从小到大进行排序，只显示了产蛋量的均值最大的后 20 个处理的名称及相应的均值。这 20 个处理之间没有显著差异，可以结合成本从中选择饲料的最佳配方，例如可以选取玉米 A_3（70.6）、麸皮 B_2（8.0）、豆饼 C_2（9.0）、鱼粉 D_2（5.0）和食盐 E_1（0.0），或者选取玉米 A_3（70.6）、麸皮 B_2（8.0）、豆饼 C_2（9.0）、鱼粉 D_2（5.0）和食盐 E_3（0.25）。

第四节 非参数方差分析

前面介绍的方差分析均要求样本来自于正态总体，并且这些正态总体应具有相同的方差，在这样的基本假定（正态性假定和方差齐性假定）下检验各总体均值是否相等，这属于参数检验。当数据不满足正态性和方差齐性假定时，参数检验可能会给出错误的结果，此时应采用基于秩的非参数检验。

一、Kruskal-Wallis检验（单因素非参数方差分析）

1. 数据表

	总体1	总体2	$\cdots$	总体k
重复	x_{11} x_{21} $\vdots$ $x_{n_1 1}$	x_{12} x_{22} $\vdots$ $x_{n_2 2}$	$\cdots$ $\cdots$ $\vdots$ $\cdots$	x_{1k} x_{2k} $\vdots$ $x_{n_k k}$

2. Kruskal-Wallis检验原理

- 原假设与备择假设

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

H_1: 至少存在一对 $\mu_i \neq \mu_j$

- 检验统计量

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} n_i (\bar{R}_i - \bar{R})^2$$

$$= \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1) \overset{\text{近似}}{\sim} \chi^2(k-1) \, .$$

- 拒绝域 $W = \{H \geq \chi_\alpha^2(k-1)\}$

3. Kruskal-Wallis检验的MATLAB实现

- kruskalwallis函数

调用方式：

```
p = kruskalwallis(X)
p = kruskalwallis(X,group)
p = kruskalwallis(X,group,displayopt)
[p,table] = kruskalwallis(...)
[p,table,stats] = kruskalwallis(...)
```

4. Kruskal-Wallis检验案例分析

【例11.4-1】 某灯泡厂用四种不同配料方案制成的灯丝生产四批灯泡，在每一批中取若干个作寿命试验，得如下数据（单位：小时）：问灯丝的不同配料方案对灯泡寿命有无显著影响？

灯泡品种	A_1	1600,1610,1650,1680,1700,1720,1800
	A_2	1580,1640,1600,1650,1660
	A_3	1460,1550,1600,1620,1640,1610,1540,1620
	A_4	1510,1520,1530,1570,1600,1680

➢ 利用Matlab求解

◘ 方差分析

```
>> A1 = [1600, 1610, 1650, 1680, 1700, 1720, 1800]';
>> g1 = repmat({'A1'},size(A1));
>> A2 = [1580, 1640, 1600, 1650, 1660]';
>> g2 = repmat({'A2'},size(A2));
>> A3 = [1460, 1550, 1600, 1620, 1640, 1610, 1540, 1620]';
>> g3 = repmat({'A3'},size(A3));
>> A4 = [1510, 1520, 1530, 1570, 1600, 1680]';
>> g4 = repmat({'A4'},size(A4));
>> life = [A1;A2;A3;A4];
>> group = [g1;g2;g3;g4];
% 调用kruskalwallis函数作Kruskal-Wallis检验
>> [p,table,stats] = kruskalwallis(life,group)
```

方差分析表

'Source'	'SS'	'df'	'MS'	'Chi-sq'	'Prob>Chi-sq'
'Groups'	[564.7908]	[3]	[188.2636]	[9.7043]	[0.0213]
'Error'	[890.2092]	[22]	[40.4641]	[]	[]
'Total'	[1455]	[25]	[]	[]	[]

多重比较

```
% 调用multcompare对不同配料方案下灯泡的寿命进行多重比较
>> [c,m,h,gnames] = multcompare(stats);
>> c  % 查看多重比较的结果矩阵c
c =1.0000    2.0000    -6.6331    4.8429    16.3188
   1.0000    3.0000    -0.5630    9.5804    19.7237
   1.0000    4.0000     1.0724   11.9762    22.8800
   2.0000    3.0000    -6.4356    4.7375    15.9106
   2.0000    4.0000    -4.7344    7.1333    19.0010
   3.0000    4.0000    -8.1888    2.3958    12.9804
```

二、Friedman检验

1. 数据表

		样本1	样本2		样本k
环境	区组1	x_{11}	x_{12}	$\cdots$	x_{1k}
	区组2	x_{21}	x_{22}	$\cdots$	x_{2k}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	区组b	x_{b1}	x_{b2}	$\cdots$	x_{bk}

2. Friedman检验原理

- 原假设与备择假设

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

H_1：至少存在一对 $\mu_i \neq \mu_j$

- 检验统计量

$$Q = \frac{12}{bk(k+1)} \sum_{i=1}^{k} \left(R_i - \frac{b(k+1)}{2} \right)^2$$

$$= \frac{12}{bk(k+1)} \sum_{i=1}^{k} R_i^2 - 3b(k+1) \overset{\text{近似}}{\sim} \chi^2(k-1)$$

- 拒绝域 $W = \left\{ Q \geq \chi_\alpha^2(k-1) \right\}$

3．Friedman检验的MATLAB实现

➢ friedman函数

$$
X = \begin{array}{c} \begin{array}{ccc} A=1 & A=2 & A=3 \end{array} \\ \left[\begin{array}{ccc} x_{111} & x_{121} & x_{131} \\ x_{112} & x_{122} & x_{132} \\ x_{211} & x_{221} & x_{231} \\ x_{212} & x_{222} & x_{232} \end{array}\right] \end{array} \begin{array}{l} \left.\begin{array}{c} \\ \\ \end{array}\right\} B=1 \\ \left.\begin{array}{c} \\ \\ \end{array}\right\} B=2 \end{array}
$$

reps = 2 情形示例

调用方式：

p = friedman(X,reps)

p = friedman(X,reps,displayopt)

[p,table] = friedman(...)

[p,table,stats] = friedman(...)

4. Friedman检验案例分析

【例11.4-2】设有来自A,B,C,D 四个地区的四名厨师制作名菜：京城水煮鱼，想比较它们的品质是否相同。四位美食评委对四名厨师的菜品分别作出了评分，如下表所列。试根据表中的数据检验四个地区制作的京城水煮鱼这道菜的品质有无区别。

美食评委	地区			
	A	*B*	*C*	*D*
1	85	82	82	79
2	87	75	86	82
3	90	81	80	76
4	80	75	81	75

➢ 利用Matlab求解

方差分析

```
>> x = [85    82    82    79
   87   75    86    82
   90   81    80    76
   80   75    81    75];
>> [p,table,stats] = friedman(x)
```

方差分析表

'Source'	'SS'	'df'	'MS'	'Chi-sq'	'Prob>Chi-sq'
'Columns'	[12.8750]	[3]	[4.2917]	[8.1316]	[0.0434]
'Error'	[6.1250]	[9]	[0.6806]	[]	[]
'Total'	[19]	[15]	[]	[]	[]

□ 多重比较

% 调用multcompare函数对四个地区制作的京城水煮鱼这道菜的品质进行多重比较

```
>> [c,m] = multcompare(stats);
>> c  % 查看多重比较的结果矩阵c
c =
    1.0000    2.0000   -0.5358    1.7500    4.0358
    1.0000    3.0000   -1.4108    0.8750    3.1608
    1.0000    4.0000    0.0892    2.3750    4.6608
    2.0000    3.0000   -3.1608   -0.8750    1.4108
    2.0000    4.0000   -1.6608    0.6250    2.9108
    3.0000    4.0000   -0.7858    1.5000    3.7858
```

Thank You