方差分析

方差分析

- 方差分析(analysis of variance，ANOVA)是英国遗传学家、统计学家Ronald Fischer提出的一种分析方法
- 方差分析技术是假设检验的拓展

$$\mathscr{H}_0 : \mu_1 = \mu_2 = \cdots = \mu_N$$

  - 单因子方差分析
  - 双因子方差分析
  - 多因子方差分析

例9-42 单元素方差分析

- 有5种药物比较疗效
  - 将30个病人随机地分成5组
  - 每组使用同一种药物，并记录病人治疗时间
  - 评价疗效——5种药物疗效是否有显著不同？

patient number	medicine numbers					patient number	medicine numbers				
	1	2	3	4	5		1	2	3	4	5
1	5	4	6	7	9	2	8	6	4	4	3
3	7	6	4	6	5	4	7	3	5	6	7
5	10	5	4	3	7	6	8	6	3	5	6

单因子方差分析

- 单因子方差分析就是指对一些观察来说，只有一个外界因素可能对观测的现象产生影响
- 求解单因子方差分析的函数调用格式

$$[p, \mathtt{tab}, \mathtt{stats}] = \mathtt{anova1}(\boldsymbol{X})$$

- 其中，X 为需要分析的数据
- 若 $p < \alpha$, 则拒绝假设

$$\mathscr{H}_0 : \mu_1 = \mu_2 = \cdots = \mu_N$$

单因子方差分析表

➢ANOVA表

source	sum of squares	DOF	mean squares	F	probability p
groups	$\mathrm{SSA}=\sum_i n_i\bar{y}_{i,:}^2-N\bar{y}_{:,:}^2$	$I-1$	$\mathrm{MSSA}=\frac{\mathrm{SSA}}{I-1}$	$\frac{\mathrm{MSSA}}{\mathrm{MSSE}}$	$p=P(F_{I-1,N-I}>c)$
error	$\mathrm{SSE}=\sum_i\sum_k y_{i,k}^2-\sum_i n_i\bar{y}_{i,:}^2$	$N-I$	$\mathrm{MSSE}=\frac{\mathrm{SSE}}{N-I}$		
total	$\mathrm{SST}=\sum_i\sum_k y_{i,k}^2-N\bar{y}_{:,:}^2$	$N-1$			

➢p 的值是不是很小

➢盒子图的观察

例9-42 单元素方差分析

- 有5种药物比较疗效
  - 将30个病人随机地分成5组
  - 每组使用同一种药物，并记录病人治疗时间
  - 评价疗效——5种药物疗效是否有显著不同？

patient number	medicine numbers					patient number	medicine numbers				
	1	2	3	4	5		1	2	3	4	5
1	5	4	6	7	9	2	8	6	4	4	3
3	7	6	4	6	5	4	7	3	5	6	7
5	10	5	4	3	7	6	8	6	3	5	6

例9-42的求解方法

- 方差分析——MATLAB求解语句

```
>> A=[5,4,6,7,9; 8,6,4,4,3;...
     7,6,4,6,5; 7,3,5,6,7;...
     10,5,4,3,7; 8,6,3,5,6];
   mean(A),
   [p,tbl,stats]=anova1(A)
```

- 盒子图的直观观察

双因子方差分析

- 如果有两种因子可能影响到某现象的统计规律，则应该引入双因子方差分析的概念
- 观测量 y 可以表示为一个三维数组 $y_{i,j,k}$，表示第 1 个因子取第 i 个水平，第 2 个因子取第 j 个水平时，组内第k个对象的观测指标

例9-43 树的生长地、树种的影响

- 比较3种松树在4个不同地区的生长情况有无差别
  - 在每个地区对每种松树随机地选择5株
  - 测量它们的胸径

pine species	living conditions																			
	1					2					3					4				
1	23	15	26	13	21	25	20	21	16	18	21	17	16	24	27	14	17	19	20	24
2	28	22	25	19	26	30	26	26	20	28	19	24	19	25	29	17	21	18	26	23
3	18	10	12	22	13	15	21	22	14	12	23	25	19	13	22	16	12	23	22	19

双因子方差分析的三个假设

- 三个假设:
  - α_i 为第一因子单独作用对现象没有影响

$$\mathscr{H}_1: \alpha_1 = \alpha_2 = \cdots = \alpha_I$$

  - β_j 为第二因子单独作用对现象没有影响

$$\mathscr{H}_2: \beta_1 = \beta_2 = \cdots = \beta_J$$

  - γ_k 为两个因子同时作用的效应

$$\mathscr{H}_3: \gamma_1 = \gamma_2 = \cdots = \gamma_{IJ}$$

三个概率的定义及意义

- 若 $p_A < c_1$ 则拒绝假设 $\mathscr{H}_1$

$$p_A = P\left(F_{[I-1,IJ(K-1)]} > c_1\right)$$

- 若 $p_B < c_2$ 则拒绝假设 $\mathscr{H}_2$

$$p_B = P\left(F_{[J-1,IJ(K-1)]} > c_2\right)$$

- 若 $p_{AB} < c_3$ 则拒绝假设 $\mathscr{H}_3$

$$p_{AB} = P\left(F_{[(I-1)(J-1),IJ(K-1)]} > c_3\right)$$

- 双因子方差分析 [p,tab,stats]=anova2($\boldsymbol{X}$)

双因子方差表

➢双因子方差表

source	square of sums	DOF	mean squared error	F	p
factor A	$\mathrm{SSA}=JK\sum_i \bar{y}_{i,:,:}^2 - IJK\bar{y}_{:,:,:}^2$	$I-1$	$\mathrm{MSSA}=\frac{\mathrm{SSA}}{I-1}$	$\frac{\mathrm{MSSA}}{\mathrm{MSSE}}$	p_A
factor B	$\mathrm{SSB}=IK\sum_i \bar{y}_{:,j,:}^2 - IJK\bar{y}_{:,:,:}^2$	$J-1$	$\mathrm{MSSB}=\frac{\mathrm{SSB}}{J-1}$	$\frac{\mathrm{MSSB}}{\mathrm{MSSE}}$	p_B
interaction	SSAB	$(I-1)(J-1)$	$\mathrm{MSSAB}=\frac{\mathrm{SSAB}}{(I-1)(J-1)}$	$\frac{\mathrm{MSSAB}}{\mathrm{MSSE}}$	p_{AB}
errors	$\mathrm{SSE}=\sum_{ijk} y_{i,j,k}^2 - K\sum_i\sum_j \bar{y}_{i,j,:}^2$	$IJ(K-1)$	$\mathrm{MSSE}=\frac{\mathrm{SSE}}{IJ(K-1)}$		
total	$\mathrm{SST}=\sum_{ijk} y_{i,j,k}^2 - IJK\bar{y}_{:,:,:}^2$	$IJK-1$			

$$\mathrm{SSAB} = K\sum_{ij}\bar{y}_{i,j,:}^2 - JK\sum_i \bar{y}_{i,:,:}^2 - IK\sum_j \bar{y}_{:,j,:}^2 + IJK\bar{y}_{:,:,:}^2$$

例9-43 树的生长地、树种的影响

- 比较3种松树在4个不同地区的生长情况有无差别
  - 在每个地区对每种松树随机地选择5株
  - 测量它们的胸径

pine species	living conditions																			
	1					2					3					4				
1	23	15	26	13	21	25	20	21	16	18	21	17	16	24	27	14	17	19	20	24
2	28	22	25	19	26	30	26	26	20	28	19	24	19	25	29	17	21	18	26	23
3	18	10	12	22	13	15	21	22	14	12	23	25	19	13	22	16	12	23	22	19

例9-43的求解

- MATLAB 求解

```
>> B=[23,15,26,13,21,25,20,21,16,18,21,17,16,24,27,14,17,19,20,24;
      28,22,25,19,26,30,26,26,20,28,19,24,19,25,29,17,21,18,26,23;
      18,10,12,22,13,15,21,22,14,12,23,25,19,13,22,16,12,23,22,19];
   anova2(B',5);
```

- 均值计算

```
>> C=[];
   for i=1:3, for j=1:4
      C(i,j)=mean(B(i,[1:5]+(j-1)*5));
   end, end
   C=[C; mean(C)];  C=[C mean(C')']
```

多因子方差分析

- MATLAB语言的统计学工具箱还可以进行三因子甚至多因子的方差分析，可以采用manova1()函数进行多因子方差分析