医学统计学的基本方法(四)

潘子昂　王石麟　李占魁

四　方差分析

方差分析是统计学的基本方法之一，也是科学研究中分析数据的一种重要方法。一个事物，其中有许多因素是互相制约、互相依存的。根据方差加和性原理，将试验结果总方差分解为各因素的方差和误差的方差，然后借助 F 检验，对各因素效应大小及各因素间的交互作用作出科学估计，这种分析试验结果的方法称为方差分析。同时这也是方差分析所要解决的主要问题。

在方差分析中，影响试验结果的条件称为因素，因素所处的状态称为水平，一定条件下的样本值与平均值之差的平方和简称为差方和。

1、单因素试验的方差分析

单因素试验方差分析是方差分析中最简单的一种，也是分析数据中应用较多的一种方法。

如果因素 A 有 a 个水平，每个水平进行 b 次试验，试验结果为 x_{ij}。它表示在 A_i 条件下进行的第 j 次试验的结果。由于因素水平的不同以及存在的试验误差，每次试验结果是不同的。

计算时列如下的表格，见表1。

表1　单因素试验安排及试验结果

	A_1	A_2	…	A_i	…	A_a	
1	x_{11}		…	x_{i1}	…	x_{a1}	
2	x_{12}	x_{21}	…	x_{i2}	…	x_{a2}	
⋮	⋮	x_{22}	…	⋮	…	⋮	
j	x_{1j}	x_{2j}	…	x_{ij}	…	x_{aj}	
⋮	⋮	x_{2b}	…	⋮	…	⋮	
b	x_{1b}		…	x_{ib}	…	x_{ab}	$\sum$
$\sum$	$\sum_{j=1}^{b} x_{1j}$	$\sum_{j=1}^{b} x_{2j}$	…	$\sum_{j=1}^{b} x_{ij}$	…	$\sum_{j=1}^{b} x_{aj}$	$\sum_{i=1}^{a}\sum_{j=1}^{b} x_{ij}$
$(\sum)^2$	$(\sum_{j=1}^{b} x_{1j})^2$	$(\sum_{j=1}^{b} x_{2j})^2$	…	$(\sum_{j=1}^{b} x_{ij})^2$	…	$(\sum_{j=1}^{b} x_{aj})^2$	$\sum_{i=1}^{a}(\sum_{j=1}^{b} x_{ij})^2$
$\sum^2$	$\sum_{j=1}^{b} x_{1j}^2$	$\sum_{j=1}^{b} x_{2j}^2$	…	$\sum_{j=1}^{b} x_{ij}^2$	…	$\sum_{j=1}^{b} x_{aj}^2$	$\sum_{i=1}^{a}\sum_{j=1}^{b} x_{ij}^2$

总的平均值 $\bar{x}=\frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}x_{ij}$，总的差方和 Q_T 为各次试验结果与它们总平均值 $\bar{x}$ 之差平方的总和，

$$Q_T=\sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij}-\bar{x})^2=\sum_{i=1}^{a}\sum_{j=1}^{b}x_{ij}^2-\frac{1}{ab}\left(\sum_{i=1}^{a}\sum_{j=1}^{b}x_{ij}\right)^2$$

根据方差加和原理，如果因素 A 水平变化引起的差方和为 Q_A，试验误差的差方和为 Q_E，则

$$Q_T=Q_A+Q_E$$

其中 $$Q_A=\frac{1}{a}\sum_{i=1}^{a}\left(\sum_{j=1}^{b}x_{ij}\right)^2-\frac{1}{ab}\left(\sum_{i=1}^{a}\sum_{j=1}^{b}x_{ij}\right)^2$$

$$Q_E=\sum_{i=1}^{a}\sum_{j=1}^{b}x_{ij}^2-\frac{1}{a}\sum_{i=1}^{a}\left(\sum_{j=1}^{b}x_{ij}\right)^2$$

Q_A 也称为组间差方和，Q_E 称为组内差方和。组

作者单位：100029，中日友好医院临床医学研究所

间差方和 Q_A，其自由度 $f_A=a-1$，组内差方和 Q_E，其自由度 $f_E=a(b-1)$，总差方和 Q_T，其自由度 $f_t=ab-1$，它们的方差估计值为

$$S_A^2=Q_A/f_A$$

$$S_E^2=Q_E/f_E。$$

$$S_T^2=Q_T/f_T$$

方差可反映试验结果波动大小，在单因素方差分析中，F 检验所使用的统计量为

$$F=S_A^2/S_E^2$$

F 值越大，说明因素水平变化的影响越显著。对于选定的显著性水平 α、如果 $F>F(\alpha,f_1,f_2)$ 表示分组因素即因素水平变化的影响是显著的，否则，表示对试验结果无显著影响。在 $F(\alpha,f_1,f_2)$ 中，α 为显著性水平，f_1、f_2 相应于因素和误差效应的自由度。$F(\alpha,f_1,f_2)$ 可从统计书的附表中查到。

在方差分析中，将各项计算列成方差分析表。

表 2　单因素方差分析表

方差来源	差方和	自由度	方差估计值	F 值	显著性
因素 A 变化（组间）	Q_A	$a-1$	$S_A^2=Q_A/f_A$	S_A^2/S_E^2	
试验误差（组内）	Q_E	$a(b-1)$	$S_E^2=Q_E/f_E$		
总　和	Q_T	$ab-1$	$S_T^2=Q_T/f_T$		

2. 两因素交叉分组全面试验的方差分析

在实际工作中，影响一个物理量的因素往往是很多的。在多因素试验中，两因素交叉分组全面试验是最为简单的。全面试验是将因素的每一个水平在一切可能组合的条件下进行试验，交叉分组是把因素 A 和 B 处于完全平等的地位。通过方差分析，来判断什么因素是主要的，因素之间有无交互作用。

试验安排如表 3。

表 3　两因素交叉分组全面试验表（重复试验次数为 C）

	B_1	B_2	$\cdots$	B_j	B_b
A_1	$x_{111}x_{112}\cdots x_{11k}\cdots x_{11c}$	$x_{121}x_{122}\cdots x_{12k}\cdots x_{12c}$	$\cdots$	$x_{1j1}x_{1j2}\cdots x_{1jk}\cdots x_{1jc}\cdots$	$x_{1b1}x_{1b2}\cdots x_{1bk}\cdots x_{1bc}$
A_2	$x_{211}x_{212}\cdots x_{21k}\cdots x_{21c}$	$x_{221}x_{222}\cdots x_{22k}\cdots x_{22c}$	$\cdots$	$x_{2j1}x_{2j2}\cdots x_{2jk}\cdots x_{2jc}\cdots$	$x_{2b1}x_{2b2}\cdots x_{2bk}\cdots x_{2bc}$
$\vdots$					
A_i	$x_{i11}x_{i12}\cdots x_{i1k}\cdots x_{i1c}$	$x_{i21}x_{i22}\cdots x_{i2k}\cdots x_{i2c}$	$\cdots$	$x_{ij1}x_{ij2}\cdots x_{ijk}\cdots x_{ijc}\cdots$	$x_{ib1}x_{ib2}\cdots x_{ibk}\cdots x_{ibc}$
$\vdots$					
A_a	$x_{a11}x_{a12}\cdots x_{a1k}\cdots x_{a1c}$	$x_{a21}x_{a22}\cdots x_{a2k}\cdots x_{a2c}$	$\cdots$	$x_{aj1}x_{aj2}\cdots x_{ajk}\cdots x_{ajc}\cdots$	$x_{ab1}x_{ab2}\cdots x_{abk}\cdots x_{abc}$

两因素方差分析的基本原理与单因素相同，总的差方和

$$Q_T=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(x_{ijk}-\bar{x})^2=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}x_{ijk}^2-\frac{T^2}{abc}$$

因素 A 和因素 B 的水平变化所引起的差方和分别为

$$Q_A=bc\sum_{i=1}^{a}(\bar{x}_i-\bar{x})^2=\frac{1}{bc}\sum_{i=1}^{a}T_i^2-\frac{T^2}{abc}$$

$$Q_B=ac\sum_{j=1}^{b}(\bar{x}_j-\bar{x})^2=\frac{1}{ac}\sum_{j=1}^{b}T_j^2-\frac{T^2}{abc}$$

交叉效应是因素 A、B 联合产生的影响，其差方和是在 A_iB_j 条件下的样本平均值 $\bar{x}_{ij}$ 与总平均值 $\bar{x}$ 之差的平方和减去因素 A 和因素 B 的差方和

$$Q_{AB}=c\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{ij}-\bar{x})^2-Q_A-Q_B$$

$$=\frac{1}{c}\sum_{i=1}^{a}\sum_{j=1}^{b}T_{ij}^2-\frac{1}{bc}\sum_{i=1}^{a}T_i^2-\frac{1}{ac}\sum_{j=1}^{b}T_j^2+\frac{T^2}{abc}$$

误差效应的差方和是由相同条件 A_iB_j 下单次测量值与该条件下的平均值之差的平方和构成。

$$Q_E=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(x_{ijk}-\bar{x}_{ij})^2$$

$$=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}x_{ijk}^2-\frac{1}{c}\sum_{i=1}^{a}\sum_{j=1}^{b}T_{ij}^2$$

上述式中，

$$T=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}x_{ijk}$$

$$T_i=\sum_{j=1}^{b}\sum_{k=1}^{c}x_{ijk},$$

$$T_j=\sum_{i=1}^{a}\sum_{k=1}^{c}x_{ijk},$$

$$T_{ij}=\sum_{k=1}^{c}x_{ijk}.$$

表4 两因素方差分析表

方差来源	差方和	自由度	方差估计值	$F_{固}$	$F_{随}$	$F_{混}$
A	Q_A	$f_A=a-1$	$S_A^2=Q_A/f_A$	S_A^2/S_E^2	S_A^2/S_{AB}^2	S_A^2/S_E^2
B	Q_B	$f_B=b-1$	$S_B^2=Q_B/f_B$	S_B^2/S_E^2	S_B^2/S_{AB}^2	S_B^2/S_{AB}^2
$A\times B$	Q_{AB}	$f_{AB}=(a-1)(b-1)$	$S_{AB}^2=Q_{AB}/f_{AB}$	S_{AB}^2/S_E^2	S_{AB}^2/S_E^2	S_{AB}^2/S_E^2
误差	Q_E	$f_E=ab(c-1)$	$S_E^2=Q_E/f_E$			
总和	Q_T	$f_T=abc-1$				

方差分析表见表4。

在多因素试验中，因素的性质可分为两大类，即固定因素和随机因素。固定因素是指完全能人为控制的因素。对固定因素进行试验所得的结论，只适用于所试验的因素水平范围，不能外推到试验范围以外。而随机因素指因素的水平不为人们所完全控制的，或试验的个体是随机选择的。对随机因素试验结果，可在试验样本所源自的总体内推广。

根据因素的不同性质，F 检验可由三种模型，即固定模型、随机模型和混合模型，在混合模型中因素 A 为随机因素，B 为固定因素。这三种模型中，统计量 F 的计算方法不相同，见表4。

(上接第47页)

则无明显差异($P>0.05$)说明糖尿病患者 BGP 的变化女性比男性更明显。糖尿病组 BGP 的变化与血 Ca、P 及 AKP 无明显相关，而仅与 BMD 的变化呈明显相关，说明血清 BGP 的变化与 BMD 一样比血 Ca、P 及 AKP 等更有主要价值。尤其值得注意的是，当 BGP 低于正常时，其血 Ca、P 及 AKP 值仍在正常范围。由此可见，血清 BGP 与 BMD 一样对骨形成和骨转换的变化的反应，较血 Ca、P 及 AKP 更为敏感可靠，现临床已推广应用。

参考文献

1 Price P A, Parthemore J G, Deftos L J, et al. New biochemical marker for bone metabolism. J Clin Invest, 1980, 66: 878.

2 Price P A and Nishimoto S K. Radioimmunoassay for the vitamin K-dependent protein of bone and its discovery in plasma. Proc Natl Acad Sci USA, 1980, 77(4): 2234.

3 Epstein S, Poser J. Difference in serum bone Gla protein with age and sex. Lancet, 1984, 1(11): 307.

4 Delmas P D, Stenner D, Wahner H W, et al. Increase in serum bone r-carboxyglutamic acid protein with aging in women. J clin Invest 1983, 71: 1316.

5 Johansen J S, Thomsen K & Christiansen. Plasma BGP concetrations in health adults, dependence on sex, age, and glomerular filtration. Scand J Clin Lab Invest, 1987, 47: 345.

6 Price P A, Williamson M K, lotbringer J W. Origin of the vitamin K-dependent bone protein found in plasma and its clearance by kidney and bone. J Biol Chem, 19[illegible], 25[illegible]: 12760.