

第五章 方差分析

李美良 副教授/博士

E-mail: liml@sicau.edu.cn

13350567991 (Tel/微信)

方差分析：将总变异分解，从而评定各变异在总变异中相对重要性的统计分析方法。

方差分析是科学试验分析中的一个十分重要的工具。

方差分析与上章t测验一样，是通过将**试验处理的效应与试验误差**进行比较，来进行统计推断，只不过这里是采用方差来度量**处理引起的变异和误差引起的变异**。

第一节 方差分析的基本原理

□ 平方和与自由度的分解

□ F 分布与 F 测验

一、自由度和平方和的分解

总变异 \longrightarrow 处理间变异+误差变异

变异 \longrightarrow 方差 \longrightarrow 平方和除以自由度

方差是平方和除以自由度。要将一个试验资料的总变异分解为各个变异，首先必须将总平方和和总自由度分解为各个变异的相应部分。因此，自由度和平方和的分解是方差分析的第一步。下面用一个例子来说明这一问题。

[例5.1] 以A、B、C、D四种药剂处理水稻种子，每个处理四个重复，各得4个苗高观察值(cm)，试分解其自由度和平方和。

药剂	苗高观察值				总和 T_i	平均数	
A	18	21	20	13	72	18	\bar{y}_i
B	20	24	26	22	92	23	
C	10	15	17	14	56	14	
D	28	27	29	32	116	29	
					$T=336$		

k个处理，每个处理n个重复

1、总变异

把表中的全部观察值作为一个组，根据前面讲过的计算平方和和自由度的公式，可以计算出总变异的平方和和与自由度。

k个处理，每个处理n个重复

药剂	苗高观察值				总和 T_i	平均数
A	18	21	20	13	72	18
B	20	24	26	22	92	23
C	10	15	17	14	56	14
D	28	27	29	32	116	29

总平方和：

$$SS_T = \sum (y - \bar{y})^2 = \sum y^2 - \frac{(\sum y)^2}{nk} = 18^2 + 21^2 + \dots + 32^2 - \frac{336^2}{4 \times 4} = 602$$

其中： $\frac{(\sum y)^2}{nk} = \frac{T^2}{nk}$ 为矫正数，用C表示。

总自由度:

$$DF_T = nk - 1 = 4 \times 4 - 1 = 15$$

2、处理效应

如果没有处理效应，表中各个处理间平均数 \bar{y}_i 从理论上讲均应该相等，因此，可以用 \bar{y}_i 来度量处理效应。

处理间平方和和自由度:

$$SS_t = \frac{\sum T_i^2}{n} - C$$

$$DF_t = k - 1$$

$$SS_t = \frac{\sum T_i^2}{n} - C = \frac{72^2 + 92^2 + 56^2 + 116^2}{4} - 7056 = 504$$

$$DF_t = k - 1 = 4 - 1 = 3$$

3、误差

表中重复间各观察值间，若不存在误差，则各观察值应该相等，由于误差是客观存在的，因而重复间各观察值间必然是有差异的，因此，可以用重复间的差异度量误差：

$$SS_T = SS_t + SS_e,$$

$$DF_T = DF_t + DF_e$$

$$SS_e = SS_T - SS_t = 602 - 504 = 98$$

$$DF_e = (kn - 1) - (k - 1) = 15 - 3 = 12$$

$$s_t^2 = \frac{SS_t}{df_t} = \frac{504}{3} = 168.00$$

$$s_e^2 = \frac{SS_e}{df_e} = \frac{98}{12} = 8.17$$

将上述例子推广到一般，设有 k 个处理，每个处理有 n 次重复，则资料共有 nk 个观察值，其数据分析如下表。

平方和与自由度的分解表

变异来源	DF	SS	S^2
处理间	$k-1$	$\frac{\sum T_i^2}{n} - \frac{T^2}{nk}$	S_t^2
误差	$kn-k$	$SS_T - SS_t$	S_e^2
总变异	$kn-1$	$\sum y^2 - \frac{T^2}{nk}$	S_T^2

二、 F 分布与 F 测验

(一) F 分布

在一个正态总体中随机抽取两个样本，分别求得其方差 s_1^2 与 s_2^2 ，将 s_1^2 和 s_2^2 的比值定义为 F ，那么 F 值的分布就是 F 分布

$$F = \frac{s_1^2}{s_2^2}$$

按上述方法从正态总体中进行一系列抽样，就可得到一系列的 F 值而作成一個 F 分布

F 分布是平均數 $\mu=1$ 、取值區間為 $[0, \infty]$ 的一組曲線，是偏態分布曲線，曲線的形狀決定於 v_1 和 v_2 。

F 分布一定區間的機率可以從已制成的 F 值表中查出（附表4，見441頁）。表中給出了各種 v_1 、 v_2 下右尾機率 $\alpha=0.05$ 、 $\alpha=0.01$ 顯著水平時的臨界 F 值。

(二) F 测验

在方差分析中， F 测验可用于检测某项变异因素的效应是否存在。所以，在计算 F 值时，总是将要测验的那一项变异因素作分子，而以误差变异作分母。

[例5.2]测定东方红3号小麦蛋白质含量10次，得方差为 $S_1^2 = 1.621$ ，测定农大139小麦的蛋白质含量5次，得方差 $S_2^2 = 0.135$ ，两个蛋白质含量的变异差异是否显著？

答：比较两个变异是否显著，用F测验。

1、提出无效假设： $\sigma_1^2 = \sigma_2^2$

2、确定显著水平： $\alpha=0.05$

3、进行显著测验： $F = \frac{s_1^2}{s_2^2} = 1.621/0.139=12.01$

查F值表： $v_1=df_1=10-1=9$ ； $v_2=df_2=5-1=4$ ； $F_{0.05}=6.00$

F大于 $F_{0.05}$ ，差异显著

[例5.3] 在例5.1中算得处理间方差 $s_t^2=168.00$ ，重复间方差 $s_e^2=8.17$ ，具有自由度 $\nu_1=3$ ， $\nu_2=12$ 。试测验处理间变异与重复间变异差异是否显著？

答： $H_0: \sigma_t^2 = \sigma_e^2$ $H_A: \sigma_t^2 \neq \sigma_e^2$

$$\alpha=0.05$$

$$F = \frac{s_t^2}{s_e^2} = \frac{168.00}{8.17} = 20.56$$

查F值表在 $\nu_1=3$, $\nu_2=12$ 时

$$F_{0.05}=3.49$$

实得 $F > F_{0.05}$

差异显著

将例5.1和例5.3的分析结果归纳在一起，列出方差分析表如下：

水稻药剂处理苗高方差分析表

变异来源	DF	SS	S^2	F	$F_{0.05}$
处理间变异	3	504	168.00	20.56	3.49
误差变异	12	98	8.17		
总变异	15	602			

第二节 多重比较

上一节对试验数据通过平方和与自由度的分解，将处理方差与误差方差作比较，由 F 测验推论处理间有显著差异，对有些试验来说方差分析已算告一段落，但对有些试验来说，其目的不仅在于了解一组处理间总体上有无实质性差异，更在于了解哪些处理间存在真实差异，故需进一步做处理平均数间的比较。

一个试验中处理平均数间比较是多重比较。

方法:

- 最小显著差数法
- q 法
- 新复极差法

一、最小显著差数法

最小显著差数法，简称LSD法

$$LSD_{\alpha} = t_{\alpha} s_{\bar{y}_i - \bar{y}_j}$$

$$s_{\bar{y}_i - \bar{y}_j} = \sqrt{\frac{2S_e^2}{n}}$$

[例5.4] 试以LSD法测验各种药剂处理的苗高平均数之间的差异显著性。

$$s_{\bar{y}_i - \bar{y}_j} = \sqrt{\frac{2 \times 8.17}{4}} = 2.02$$

由附表3, $DF_e=12$ 时, $t_{0.05}=2.179$, $t_{0.01}=3.055$

故 $LSD_{0.05}=2.179 \times 2.02=4.40$

$LSD_{0.01}=3.055 \times 2.02=6.17$

不同药剂处理水稻苗高平均数比较(LSD法)

处理	苗高平均数	差异显著性	
		0.05	0.01
D	29	a	A
B	23	b	A B
A	18	c	B C
C	14	c	C

二、 q 法

q 测验法（复极差法）是将 k 个平均数由大到小排列后，根据所比较的两个处理平均数的差数是几个平均数间的极差分别确定最小显著极差 LSR_α 值的。

$$LSR_\alpha = q_\alpha \cdot S_{\bar{y}_i}$$

$$S_{\bar{y}_i} = \sqrt{\frac{S_e^2}{n}}$$

[例5.5] 试以 q 法测验各种药剂处理的苗高平均数之间的差异显著性。

$$s_{\bar{y}_i} = \sqrt{8.17 / 4} = 1.43$$

查附表7，得到当 $DF_e=12$ 时， $p=2,3,4$ 的 q_α 值

LSR_{α} 值

P	$q_{0.05}$	$q_{0.01}$	$LSR_{0.05}$	$LSR_{0.01}$
2	3.08	4.32	4.40	6.18
3	3.77	5.05	5.39	7.22
4	4.20	5.50	6.01	7.87

不同药剂处理水稻苗高平均数比较(q 法)

处理	苗高平均数	差异显著性	
		0.05	0.01
D	29	a	A
B	23	b	AB
A	18	c	BC
C	14	c	C

三、新复极差法

新复极差法（SSR法），与 q 法相似。计算 LSR_{α} 值查的是 SSR_{α} 值（附表8）而不是 q 表。

$$LSR_{\alpha} = S_{\bar{y}_i} \cdot SSR_{\alpha}$$

LSR_{α} 值

P	$SSR_{0.05}$	$SSR_{0.01}$	$LSR_{0.05}$	$LSR_{0.01}$
2	3.08	4.32	4.40	6.18
3	3.23	4.55	4.62	6.51
4	3.33	4.68	4.76	6.69

不同药剂处理水稻苗高平均数比较(SSR法)

处理	苗高平均数	差异显著性	
		0.05	0.01
D	29	a	A
B	23	b	AB
A	18	c	BC
C	14	c	C