第六章方差分析

方差分析的基本功能

对多组处理的样本平均数差异的显著性进行检验

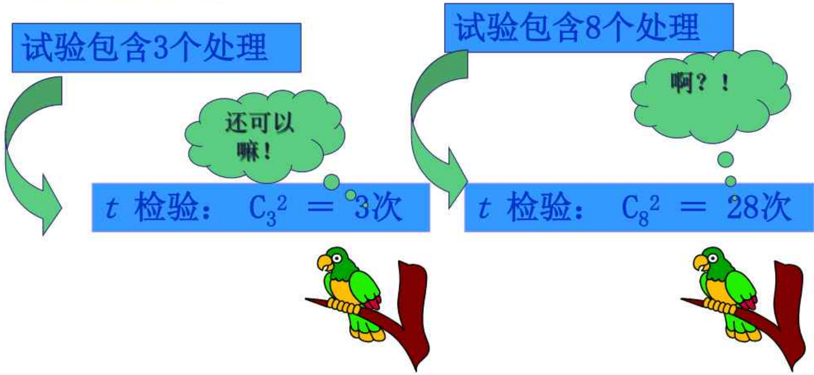
*t*测验可以判断两组数据平均数间的差异显著性，而方差分析既可以判断两组又可以判断多组数据平均数之间的差异显著性。

或许有人会说，我们可以把多组数据化成几个两组数据，用几次t检验来完成这个多组数据差异显著性的判断。那不用方差分析不是也可以吗?

到底这种方法行不行？

对多个处理进行平均数差异显著性检验时，采用t检验法的缺点：

1.检验过程繁琐。



试验包含3个处理

t 检验：　C_3^2 ＝ 3次

试验包含8个处理

t 检验：　C_8^2 ＝ 28次

2.无统一的比较标准。

t检验：C_4^2 ＝6次

需计算 6个标准误 $s_{\bar{y}_i-\bar{y}_j}=\sqrt{\frac{S_i^2}{n_i}+\frac{S_j^2}{n_j}}$

比较时就没有统一的标准

3、犯第一类错误概率增加。

例如我们用t检验的方法检验4个样本平均数之间的差异显著性，

α=0.05

t检验：

C_4^2 ＝6次

6次检验

相互独立

H_0的概率：

1－α ＝0.95

6次都接受H_0的概率$(0.95)^6$＝0.735

犯α错误的概率＝1-0.735＝0.265〉0.05

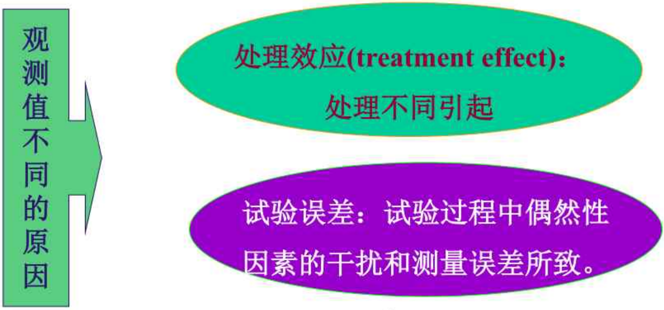
犯α错误的概率明显增加

第一节 方差分析的基本原理

一、方差分析的基本思想、目的和用途

方差：又叫均方，是表示变异程度的量。

在一个多处理试验中，可以得出一系列不同的观测值.



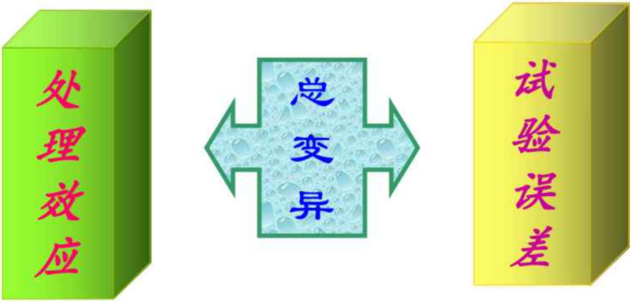
观测值不同的原因

处理效应(treatment effect)：处理不同引起

试验误差：试验过程中偶然性因素的干扰和测量误差所致。

第一节 方差分析的基本原理

方差分析的基本思想



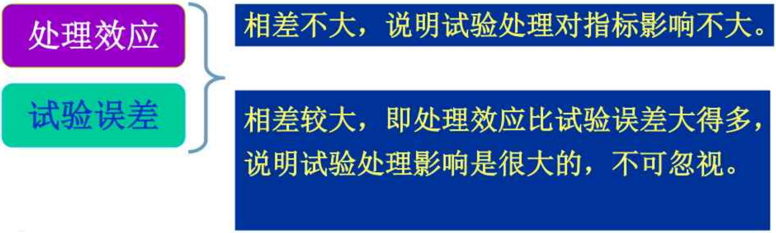
总变异

处理效应

试验误差

方差分析的目的

确定各种原因在总变异中所占的重要程度。



处理效应

试验误差

相差不大，说明试验处理对指标影响不大。

相差较大，即处理效应比试验误差大得多，说明试验处理影响是很大的，不可忽视。

方差分析的用途

1. 判断每个因素水平间的差异显著性

2. 判断各因素间交互作用显著性

3. 用于方差的同质性测验

第一节 方差分析的基本原理

二、方差分析的步骤

1、平方和与自由度的分解

先看下面的例题，这是一个单因素完全随机试验。

样本	A	B	C	总计
	1 2 2 3	3 4 4 5	4 5 6 9	
均数	2	4	6	4

$\bar{y}=4$

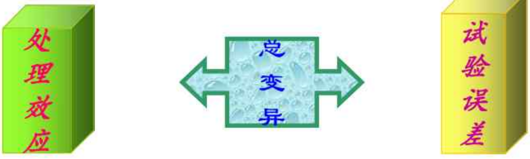
$\bar{y}_A=2$

$\bar{y}_B=4$

$\bar{y}_C=6$

第一节 方差分析的基本原理

平方和的分解



处理效应

总变异

试验误差

$\bar{y}_i-\bar{y}$

$y_{ij}-\bar{y}$

$y_{ij}-\bar{y}_i$

A	B	C
−2	0	2
−2	0	2
−2	0	2
−2	0	2

A	B	C
−3	−1	0
−2	0	1
−2	0	2
−1	1	5

A	B	C
−1	−1	−2
0	0	-1
0	0	0
1	1	3

$SS_t = n\sum(\bar{y}_i-\bar{y})^2 = 32$

$SS_T = \sum(y_{ij}-\bar{y})^2 = 50$

$SS_e = \sum(y_{ij}-\bar{y}_i)^2 = 18$

第一节 方差分析的基本原理

通过前面的平方和的直观分解可以看出：

$$SS_T = SS_t + SS_e$$

当式也可以由公式推导出来：

SS_e

SS_t

$$SS_T = \sum(y_{ij}-\bar{y})^2 = \sum(y_{ij}-\bar{y}_i+\bar{y}_i-\bar{y})^2$$

$$=\sum(y_{ij}-\bar{y}_i)^2+\sum(\bar{y}_i-\bar{y})^2-2\sum(y_{ij}-\bar{y}_i)\sum(\bar{y}_i-\bar{y})$$

因为 $\sum(\bar{y}_i-\bar{y})=0$ 所以 $SS_T = SS_t + SS_e$

第一节 方差分析的基本原理

自由度的分解

总自由度： $df_T = nk-1$

处理项自由度：　$df_t = k - 1$

误差项自由度：　$df_e = df_T - df_t = k(n-1)$

K为处理项数/水平数
n为样本数

第一节 方差分析的基本原理

[例] 以4种药剂处理水稻种子，其中A为对照，每处理各得4个苗高观察值，其结果如下表：试分解其平方和与自由度。

药剂	苗高观察值				总和	平均
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
					T=336	$\bar{y}=21$

$SS_t = n\sum(\bar{y}_i - \bar{y})^2$　　$SS_T = \sum(y_{ij} - \bar{y})^2$　　$SS_e = \sum(y_{ij} - \bar{y}_i)^2$

第一节 方差分析的基本原理

根据矫正数公式进行平方和的分解：

$$C=\frac{T^2}{nk}=\frac{336^2}{4\times4}=7056$$

$$SS_T=\sum y_{ij}^2-C=18^2+21^2+\cdots+32^2-7056=602 \qquad df_T=nk-1=4\times4-1=15$$

$$SS_t=\frac{\sum T_i^2}{n}-C=\frac{72^2+92^2+56^2+116^2}{4}-7056=504 \qquad df_t=k-1=4-1=3$$

$$SS_e=SS_T-SS_t=602-504=98 \qquad df_e=df_T-df_t=k(n-1)=4\times(4-1)=12$$

第一节 方差分析的基本原理

2、求均方,进行F测验,列方差分析表

- 求均方

$$MS_t=\frac{SS_t}{df_t}=\frac{504}{3}=168$$

$$MS_e = \frac{SS_e}{df_e} = \frac{98}{12} = 8.17$$

第一节 方差分析的基本原理

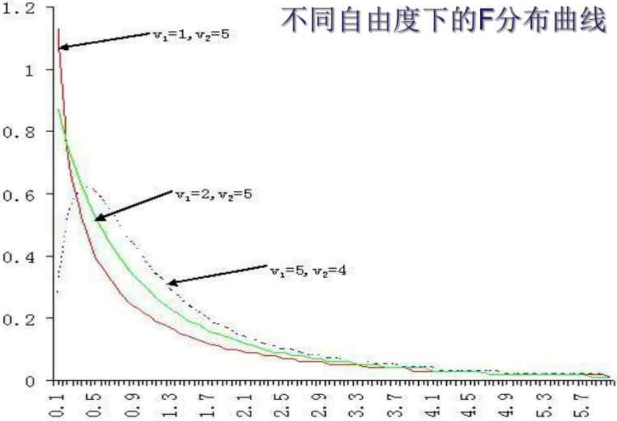
F分布与F测验

从一个正态总体N（μ，σ^2）中，分别随机抽取两个独立样本，分别求得其均方S^2_1和S^2_2，将S^2_1和S^2_2的比值定义为F：

$$F_{(\upsilon_1, \upsilon_2)} = \frac{s_1^2}{s_2^2}$$

第一节 方差分析的基本原理

不同自由度下的F分布曲线



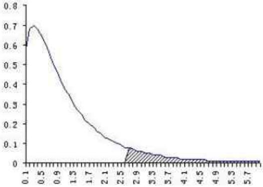
第一节 方差分析的基本原理

F分布的特点：

1、是平均数 $\mu_F=1$ ，取值区间为[0，∞）的一组曲线；

2、在 $\upsilon_1=1$和$\upsilon_2=2$F分布是反向J型，在$\upsilon_1\geq 3$时，曲线转为偏态；

3、F分布下一定区间的概率可以通过书中的附表6查得。



附表6是各种v_1和v_2下右尾概率为0.05和0.01时的临界F值表。

该表时专供测验S_1^2的总体方差是否显著大于S_2^2的总体方差而设计的。

第一节 方差分析的基本原理

对一组处理的重复试验数据经对总平方和与总自由度的分解估计出处理间的均方和处理内均方（误差均方），并通过 $F=MS_t/MS_e$测验处理间所表示出的差异是否真实（比误差大），这一方法即为方差分析法。

这里所测验的统计假设是H_0：$\sigma_t^2 \leq \sigma_e^2$或$\mu_A=\mu_B=\mu_C=\mu_D$对$H_A$：$\sigma_t^2 > \sigma_e^2$或$\mu_A$、$\mu_B$、$\mu_C$和$\mu_D$间存在差异（不一定$\mu_A$、$\mu_B$、$\mu_C$和$\mu_D$间均不等，可能部分不等。）

不同药剂处理水稻苗高的方差分析表

变异来源	df	SS	MS	F	$F_{0.05}$	$F_{0.01}$
药剂处理间	3	504	168	20.56	3.49	5.95
药剂处理内	12	98	8.17			
总变异	15	602				

$20.56 > F_{0.01}$

第二节 多重比较

上节通过F测验可以推论处理间是否有显著差异，但是对于有些试验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，故需进一步来做具体的处理平均数间的比较。

多重比较有多种方法，本节将介绍常用的两种：最小显著差数法（LSD法）和新复极差法（LSR法）。

第二节 多重比较

一、多重比较的原理

（一）、最小显著差数法(LSD法)

最小显著差数（Least Significant Difference，简称LSD法），LSD法实质上是t测验。其基本原理是：在处理间的F测验为显著的前提下，计算出显著水平为a时的最小显著差数LSD$_a$；任何两个平均数的差数（$\bar{y}_i-\bar{y}_j$），如 $\bar{y}_i-\bar{y}_j$ ≥LSD$_a$，即为在a水平上差异显著；反之，则为在a水平上差异不显著，这种方法又称为F测验保护下的最小显著差数法。

$$t=\frac{\bar{y}_i-\bar{y}_j}{s_{\bar{y}_i-\bar{y}_j}}$$

若$|t|\geq t_\alpha$，$\bar{y}_i-\bar{y}_j$在α水平上差异显著。

对于$|t|\geq t_\alpha$，两边同乘以$S_{\bar{y}_1-\bar{y}_2}$，

$|\bar{y}_1-\bar{y}_2|\geq t_\alpha S_{\bar{y}_1-\bar{y}_2}$，令

$\mathrm{LSD}_\alpha = t_\alpha S_{\bar{y}_i-\bar{y}_j}$，当样本容量相等时，

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

第二节 多重比较

（二）新复极差法（LSR法）

LSD法实质上是t测验，但是t测验只适用于两个独立随机样本差异显著性测验，但多重比较中，包括着多个样本，这多个样本中平均数最大的一个与平均数最小的一个比较，实际上已不再是一对独立随机样本的比较，用t测验，必然增大I型错误的概率，容易接受不真实的备择假设。

为此D.B.Duncan提出了新复极差法，又称最小显著极差法（shortest significant ranges,SSR）。

第二节　多重比较

其方法是把多个样本中两个极端平均数的差数当作极差对待，如果极差不显著，则包括在这两个极端处理平均数间的各处理平均数的任何成对比较，其差异也是不显著的。

极差是否显著用极差相当于均数标准差的倍数：SSR=R/$S_{\bar{y}}$ 式中R为极差，SSR为极差相当于均数标准差的倍数 。

在一定自由度下，当平均数个数为2、3、……k时，SSR值已由统计学家求出，见课本附表8。这样只要计算出S，从附表8中查出SSR，就可以计算出LSR：

$$LSR_{\alpha} = S_{\bar{y}} \bullet SSR_{\alpha} = SE \bullet SSR_{\alpha}$$

第二节　多重比较

❖ 多重比较结果的表示方法

● 列梯形表法：将全部平均数从大到小顺次排列，然后算出各平均数间的差数。凡达到a=0.05水平的差数在右上角标一个“*”号，凡达到a=0.01水平的差数在右上角标两个“**”号，

凡未达到a=0.05水平的差数则不予标记。

- 字母标记法：该方法是最常用的多重比较结果的表示方法，在科技论文中一般采用此方法，但是比较过程较复杂。下面重点介绍其标记过程。

第二节 多重比较

❖ 多重比较结果的表示方法

- 列梯形表法：

α=0.01	乙	甲	丙	丁
	32.10	30.58	24.28	20.36

处理	平均数	平均数差异		
		−20.36	− 24.28	− 30.58
乙	32.10	11.74**	7.82**	1.52
甲	30.58	10.22**	6.30**	
丙	24.28	3.92*		
丁	20.36			

第二节 多重比较

- 字母标记法：

全距中平均数个数	$SR_{0.05}$	$SR_{0.01}$	$LSR_{0.05}$	$LSR_{0.01}$
4	3.24	4.42	3.63	4.95
3	3.14	4.31	3.52	4.83
2	3.00	4.13	3.36	4.63

处理	平均数	显著性	
		$\alpha=0.05$	$\alpha=0.01$
乙	32.10	a	A
甲	30.58	a	A
丙	24.28	b	B
丁	20.36	c	B

全距中平均数个数	$SR_{0.05}$	$SR_{0.01}$	$LSR_{0.05}$	$LSR_{0.01}$
6	3.34	4.57	3.74	5.12
5	3.30	4.51	3.70	5.05
4	3.24	4.42	3.63	4.95
3	3.14	4.31	3.52	4.83
2	3.00	4.13	3.36	4.63

处理	平均数	显著性	
		$\alpha=0.05$	$\alpha=0.01$
1	32.10	a	A
2	30.58	a b	A
3	27.28	b	A B
4	23.21	c	B C
5	21.00	c	C
6	20.36	c	C

第二节　多重比较

三、多重比较方法的选择

通过多重比较可以看出，LSD法只用了一个标准，而LSR根据极差的两个极端平均数间的平均数个数多少用了多个标准，LSR法只包括两个处理平均数的极差测验所用的LSR等于LSD，所以，在多重比较中，有时两处理比较时LSD法测验达显著水平，但LSR法测验却不一定达显著水平，即LSR法测验的显著水平高于LSD法。

试验的处理间如果设有对照，各处理与对照的比较或预先安排的个别成对比较相当于两个独立随机样本平均数的比较，一般可选用LSD法；否则应使用LSR法。

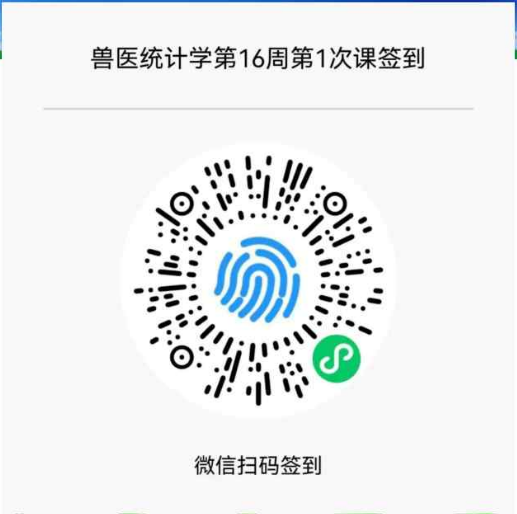
第二节　多重比较

综上所述，方差分析的基本步骤是：

（1）自由度和平方的分解；

（2）求均方，进行F测验，列方差分析表；

（3）若F测验显著，则对各平均数进行必要的多重比较。



兽医统计学第16周第1次课签到

微信扫码签到

第三节 方差分析的线性模型与期望均方

一、方差分析的线性数学模型

方差分析是建立在一定的线性可加模型基础上的。

所谓线性可加模型是指总体每一个变量可以按其变异的原因分解成若干个线性组成部分的数学表达式，它是方差分析的理论

依据。

* * *

第三节 方差分析的线性模型与期望均方

假定有**k**组观测数据，每组有**n**个观测值，则共有**nk**个观测值

处理 重复	1	2	…	i	…	k	
1	y_{11}	y_{21}	…	y_{i1}	…	y_{k1}	
2	y_{12}	y_{22}	…	y_{i2}	…	y_{k2}	
…	…	…	…	…	…	…	
j	y_{1j}	y_{2j}	…	y_{ij}	…	y_{kj}	
…	…	…	…	…	…	…	
n	y_{1n}	y_{2n}	…	y_{in}	…	y_{kn}	
总和	T_1	T_2	…	T_i	…	T_k	$T=\sum y_{ij}$
平均	$\bar{y}_1$	$\bar{y}_2$		$\bar{y}_i$		$\bar{y}_k$	$\bar{y}$

第三节 方差分析的线性模型与期望均方

用线性模型(linear model)来描述每一观测值：

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

- μ —总体平均数　τ_i —处理效应　ε_{ij} —试验误差

y_{ij} —是在第 i 次处理下的第 j 次观测值

第三节　方差分析的线性模型与期望均方

对于由样本估计的线性模型为：

$$y_{ij} = \overline{y} + t_i + e_{ij}$$

$\overline{y}$为样本平均数

t_i为样本处理效应

e_{ij}为试验误差

二、期望均方

根据的 τ_i 不同假定，可将数学模型分为以下三种：**固定模型、随机模型、混合模型**。

(一)固定模型(fixed model)

指各个处理的效应值 τ_i 是固定值，各个的平均效应 $\tau_i = \mu_i - \mu$ 是一个常量，且 $\Sigma\tau_i = 0$。就是说除去随机误差以后每个处理所产生的效应是固定的。实验因素的各水平是根据试验目的事先主观选定的而不是随机选定的。

[例] 以5个水稻品种作大区比较试验，每品种作3次取样，测定其产量所得资料为单向分组资料。本试验需明确各品种的效应，故为固定模型，其方差分析和期望均方的参数估计如下表：

变异来源	DF	SS	MS	期望均方
品种间	4	87.6	21.9	$\sigma^2+n\kappa_\tau^2$
品种内	10	24.0	2.4	σ^2

固定模型的F测验：$F = \frac{S_t^2}{S_e^2} \frac{\hat{\sigma}^2 + n\kappa_\tau^2}{\hat{\sigma}^2}$

若$\tau_i=0$,则F的期望值等于1。所以固定模型假设测验H_0: $\tau_i=0$对H_A: $\tau_i \neq 0$.

第三节 方差分析的线性模型与期望均方

特点

1、在固定模型中，除去随机误差之后的每个处理所产生的效应是固定的，试验重复时会得到相同的结果；

2、方差分析所得到的结论只适合于选定的那几个水平，并不能将其结论扩展到未加考虑的其它水平上。

第三节 方差分析的线性模型与期望均方

(二)随机模型

指各处理的效应值 τ_i 不是固定的数值，而是从平均数为零、方差为σ_τ^2的正态总体中得到的一个随机变量。主要是研究并估计总体变异即方差。

这里 τ_i 是一个随机变量，是从期望均值为 0，方

差为 σ^2 的标准正态总体中得到的随机变量。得出的结论可以推广到多个随机因素的所有水平上。

第三节　方差分析的线性模型与期望均方

如果某些试验条件不能人为控制或通过样本对所属总体做出推断时属于随机模型，例如将从美国引进的玉米在不同纬度生态条件下的情况，来观察该品种对不同地理条件的适应情况，这时各地的气候、水肥、土壤条件是无法人为控制的，就要用随机模型来处理。随机模型得出的结论可以推广到多个随机因素的所有水平上。

第三节　方差分析的线性模型与期望均方

特点

1、在随机模型中，水平确定之后其处理所产生的效应

并不是固定的，试验重复时也很难得到相同的结果；

2、方差分析所得到的结论，可以推广到这个因素的所有水平上。

第三节 方差分析的线性模型与期望均方

固定模型与随机模型的比较：

1. 两者在设计思想和统计推断上有明显不同，固定模型中所得的结论仅在于推断关于特定的处理；而随机模型中的结论将用于推断处理的总体。

2. 二者分析的侧重点也不完全相同，在期望均方和F测验方面也不一样，固定模型主要侧重于效应值的估计，而随机模型则侧重效应方差的估计和测验。

第三节 方差分析的线性模型与期望均方

(三)混合模型

指在多因素试验中既有固定因素又有随机因素时所用的模型。

在试验设计中，固定模型应用最多，随机模型和混合模型相对较少。

第四节　方差分析的基本假定与数据转换

一、方差分析的三个基本假定

1、效应的可加性

即处理效应与环境效应应为线性可加，也即总变异的分解，分解时应按照其线性可加模型进行分解。

2、误差的正态性

即试验误差为独立的随机变数，并作正态分布。

3、误差方差的同质性

即所有试验处理具有共同的误差方差。

第四节　方差分析的基本假定与数据转换

当试验中的试验数据不符合以上三点基本假定时，
要进行数据的处理，使数据符合基本假定才能进行方差分析

二、不符合方差分析基本假定数据的处理方法

1、剔除某些表现“特殊”的观察值、处理或重复

2、将总的试验误差方差分裂为几个较为同质的试验误差方差

3、进行数据转换，用转换后的数据进行方差分析

第四节　方差分析的基本假定与数据转换

常用的数据转换方法有3种：

(1)平方根转换　统计某单位面积上某种昆虫的头数、某种杂草的株数，或统计某种疾病的发病次数等

- 适用对象：稀疏现象的计数资料,其特点是样本的平均数等于方差

- 转换方法：$y \to \sqrt{y}$或者$y \to \sqrt{y+1}$(观察值中有0值时）

(2)对数转换

如抗体的滴度、疾病的潜伏期等

- 适用对象：样本平均数与其标准差或极差存在着相关性或处理效应为乘性或倍加性的资料，这类资料的分布一般呈非正态分布。
- 转换方法：$y \to \lg y$或$y \to \lg(y+1)$(有0值且各数值皆不大于10时）

第四节 方差分析的基本假定与数据转换

(3)反正弦转换

如动物存活率、疾病发生率等百分比数据

- 适用对象：百分数资料，特点是方差与平均数之间存在着函数关系
- 转换方法：$p \to \arcsin(\sqrt{p})$(当$p < 0.3$或$p > 0.7$时转换）

4、采用几个观察值的平均数作方差分析

第五节 完全随机试验结果的方差分析

- 完全随机试验结果的方差分析可分为以下三种情况介绍：

（1）单因素试验，每个处理的重复次数相等；

（2）单因素试验，每个处理的重复次数不相等；

（3）多因素试验，主要介绍两因素试验。

第五节 完全随机试验结果的方差分析

- 单因素试验，每个处理的重复次数不相等
- 计算样本总平方和、类间平方和及误差平方和。

处理	A	B	C	合计
	2	2	3	
	3	3	3	
	4	3	4	
		4	5	
			5	
计数	3	4	5	12
合计	9	12	20	41

先计算出校正项：

$$C=(\sum y)^2/\sum n_i$$

$$=41^2/12=140.08$$

$$SS_T=10.92$$

样本的总平方和为：

$$SS_T=\sum y^2-C=(2^2+3^2+...+5^2)-C$$

$$=151-140.08=109.2$$

第五节 完全随机试验结果的方差分析

- 计算样本总平方和、类间平方和及误差平方和。

处理	A	B	C	合计
	2	2	3	
	3	3	3	
	4	3	4	
		4	5	
			5	
计数	3	4	5	12
合计	9	12	20	41

先计算出校正项：

$$C=(\sum y)^2/\sum n_i$$

$$=41^2/12=140.08$$

$$SS_T=10.92$$

$$SS_t=2.92$$

样本的类间平方和为：$SS_t=\sum_{i=1}^{k}n_i(y_{i.}-y)^2=\sum_{i=1}^{k}(y_{i.}^2/n_i)-C$

$$=9^2/3+12^2/4+20^2/5-140.08=2.92$$

第五节 完全随机试验结果的方差分析

- 计算样本总平方和、类间平方和及误差平方和。

处理	A	B	C	合计
	2	2	3	
	3	3	3	
	4	3	4	
		4	5	
			5	
计数	3	4	5	12
合计	9	12	20	41

先计算出校正项：

$$C=(\sum y)^2/\sum n_i$$

$$=41^2/12=140.08$$

$$SS_T=10.92$$

$$SS_t=2.92$$

$$SS_e=SS_T-SS_t=10.92-2.92=8$$

- 计算各自的自由度。

$$df_T = \sum n_i - 1 = 12 - 1 = 11 \qquad df_t = k - 1 = 3 - 1 = 2$$

$$df_e = df_T - df_t = \sum n_i - k = 11 - 2 = 9$$

第五节 完全随机试验结果的方差分析

- 把计算的各种平方和、自由度、均方和F测验的结果列入方差分析表内。

方差分析表

变异来源	自由度	平方和	均方	F 值	$F_{0.05}$	$F_{0.01}$
处理间	2	2.92	1.46	1.64	4.26	8.02
误差	9	8	0.89			
总变异	11	10.92				

- 该结果表明，应否定H_0，3个处理间差异不显著。

第五节 完全随机试验结果的方差分析

- 多重比较：如果处理间差异达到显著或极显著，也应根据试验的目的采用LSD或LSR法进行多重比较。

但是，由于每一个处理的重复次数可能不相同，会导致在计算标准误时出现困难，可以采用LSD法或计算平均的样本容量，前一种方法较科学。

第五节　完全随机试验结果的方差分析

❖两因素试验

A因素	B因素	观测值	总计	平均
A1	B1	12 13 13	38	12.67
	B2	12 12 11	35	11.67
	B3	13 14 13	40	13.33
A2	B1	13 14 14	41	13.67
	B2	13 13 12	38	12.67
	B3	14 13 13	40	13.33
A3	B1	14 15 15	44	14.67
	B2	14 13 14	41	13.67
	B3	14 15 15	44	14.67
A4	B1	16 16 15	47	15.67
	B2	15 16 15	46	15.33
	B3	16 17 17	50	16.67

可先将每个处理组合看作一个样本，采用多样本资料进行分析。即

$$SS_T=SS_t+SS_e$$

然后，再将SS_t分解为SS_A、SS_B和$SS_{A\times B}$。

第五节　完全随机试验结果的方差分析

❖平方和与自由度的分解

A因素的水平数为a，B因素的水平数为b，重复次数为n。

先计算出校正项：

$$C=(\sum y)^2/(nk)=504^2/3/12=7056,\ k=ab=12$$

$$SS_T=\sum y^2-C=(12^2+12^2+\dots+17^2)-7056=7132-7056=76$$

$$SS_t=\sum T_t^2/n-C=(38^2+35^2+\cdots+50^2)/3-7056=7124-7056=68$$

$$SS_e=SS_T-SS_t=76-68=8$$

$$df_T=nk-1=36-1=35,\ df_t=k-1=12-1=11,$$
$$df_e=k(n-1)=12(3-1)=24=df_T-df_t$$

第五节 完全随机试验结果的方差分析

❖平方和与自由度的分解

对SS_t和df_t进行分解，先将处理组合的总计数列成两向表：

	A1	A2	A3	A4	T_B	平均
B1	**38**	**41**	**44**	**47**	**170**	**14.17**
B2	**35**	**38**	**41**	**46**	**160**	**13.33**
B3	**40**	**40**	**44**	**50**	**174**	**14.50**
T_A	**113**	**119**	**129**	**143**		
平均	**12.56**	**13.22**	**14.33**	**15.89**		

$$SS_A=\sum T_A^2/(bn)-C=(113^2+119^2+129^2+143^2)/9-7056=7113.33-7056=57.33$$
$$SS_B=\sum T_B^2/(an)-C=(170^2+160^2+174^2)/12-7056=7064.67-7056=8.67$$

$$SS_{A\times B}=SS_t-SS_A-SS_B=68-57.33-8.67=2$$

$$df_A=a-1=4-1=3,\ df_B=b-1=3-1=2,$$
$$df_{A\times B}=df_t-df_A-df_B=(a-1)(b-1)=12-3-2=6$$

第五节 完全随机试验结果的方差分析

- 求均方，进行F测验，列方差分析表。

方差分析表

变异来源	自由度	平方和	均方	F 值	$F_{0.05}$	$F_{0.01}$
处理间	11	68	6.18	18.55^{**}	2.22	3.09
A 因素	3	57.33	19.11	57.33^{**}	3.01	4.72
B 因素	2	8.67	4.34	13.01^{**}	3.40	5.61
A×B	6	2	0.33	1	2.51	3.67
误差	24	8	0.33			
总变异	35	76				

- 结果表明，A和B因素间差异达到极显著水平，A×B互作不显著，应进一步进行A和B因素水平间的多重比较。

第五节 完全随机试验结果的方差分析

F测验分母的方差

比较的平均数包含的观测值个数

- 的多重比较
- 间的多重比较。比较标

$$S_{\bar{y}}=\sqrt{Se^2/bn}\ (LSR),\quad S_{\bar{y}_{11}-\bar{y}_2}=\sqrt{2Se^2/bn}\ (LSD)$$

- B因素水平间的多重比较。比较标准误：

$$S_{\bar{y}}=\sqrt{Se^2/an}\ (LSR),\quad S_{\bar{y}_{11}-\bar{y}_2}=\sqrt{2Se^2/an}\ (LSD)$$

- A×B互作间的多重比较。比较标准误：

$$S_{\bar{y}}=\sqrt{Se^2/n}\ (LSR),\quad S_{\bar{y}_{11}-\bar{y}_2}=\sqrt{2Se^2/n}\ (LSD)$$

第五节 完全随机试验结果的方差分析

- 进行必要的多重比较
- A因素水平间的多重比较。比较标准误：

$$SE = \sqrt{Se^2 / bn} = \sqrt{0.33 / 3 / 3} = 0.19$$

全距中平均数个数	$SR_{0.05}$	$SR_{0.01}$	$LSR_{0.05}$	$LSR_{0.01}$
4	3.16	4.24	0.60	0.81
3	3.07	4.13	0.58	0.78
2	2.92	3.96	0.55	0.75

处理	平均数	显著性	
		α=0.05	α=0.01
A4	15.89	a	A
A3	14.33	b	B
A2	13.22	c	C
A1	12.56	d	C

第五节 完全随机试验结果的方差分析

- 进行必要的多重比较
- B因素水平间的多重比较。比较标准误：

$$SE = \sqrt{Se^2 / an} = \sqrt{0.33 / 4 / 3} = 0.15$$

全距中平均数个数	$SR_{0.05}$	$SR_{0.01}$	$LSR_{0.05}$	$LSR_{0.01}$
3	3.07	4.13	0.46	0.62
2	2.92	3.96	0.44	0.59

处理	平均数	显著性	
		$\alpha=0.05$	$\alpha=0.01$
B3	**14.50**	**a**	**A**
B1	**14.17**	**a**	**A**
B2	**13.33**	**b**	**B**

第六节 随机区组试验结果的方差分析

单因素试验

品种	区组			T_i	$\bar{y}_t$
	□	□	□		
A	10.9	9.1	12.2	32.2	10.7
B	10.8	12.3	14.0	37.1	12.4
C	11.1	12.5	10.5	34.1	11.4
D	9.1	10.7	10.1	29.9	10.0
E	11.8	13.9	16.8	42.5	14.2
F	10.1	10.6	11.8	32.5	10.8
G	10.0	11.5	14.1	35.6	11.9
H	9.3	10.4	14.4	34.1	11.4
T_r	83.1	91.0	103.9	T=278.0	$\bar{y}=11.6$
$\bar{y}_i$	10.4	11.4	13.0		

第六节 随机区组试验的方差分析

1、平方和与自由度的分解

$$C=\frac{T^2}{rk}=\frac{278.0^2}{3\times 8}=3220.17$$

$$SS_T=\sum y_{ij}^2-C=10.9^2+9.1^2+\cdots+14.4^2-3220.17=84.61 \qquad df_T=rk-1=3\times 8-1=23$$

$$SS_R=\frac{\sum T_r^2}{k}-C=\frac{83.1^2+91.0^2+103.9^2}{8}-3220.17=27.56 \qquad df_R=r-1=3-1=2$$

$$SS_t=\frac{\sum T_t^2}{r}-C=\frac{32.2^2+37.1^2+\cdots+34.1^2}{3}-3220.17=34.08 \qquad df_t=k-1=8-1=7$$

$$SS_e=SS_T-SS_R-SS_t=84.61-27.56-34.08=22.97 \qquad df_e=df_T-df_R-df_t=23-2-7=14$$

第六节 随机区组试验的方差分析

2、列方差分析表进行F测验

变异来源	DF	SS	MS	F	$F_{0.05}$	$F_{0.01}$
区组间	2	27.56	13.78	8.40**	3.74	6.51
品种间	7	34.08	4.87	2.97*	2.76	4.28
误差	14	22.97	1.64			
总变异	23	84.61				

第六节 随机区组试验的方差分析

3、进行必要的多重比较

$$SE=\sqrt{\frac{MS_e}{r}}=\sqrt{\frac{1.64}{3}}=0.74$$

新复极差测验的最小显著极差

P	2	3	4	5	6	7	8
$SSR_{0.05}$	3.03	3.18	3.27	3.33	3.37	3.39	3.41
$SSR_{0.01}$	4.21	4.42	4.55	4.63	4.70	4.78	4.83
$LSR_{0.05}$	2.24	2.35	2.42	2.46	2.49	2.51	2.52
$LSR_{0.01}$	3.12	3.27	3.37	3.43	3.48	3.54	3.57

第六节 随机区组试验的方差分析

品种	产量	差异显著性	
		5%	1%
E	14.2	a	A
B	12.4	ab	AB
G	11.9	ab	AB
H	11.4	b	AB
C	11.4	b	AB
F	10.8	b	AB
A	10.7	b	AB
D	10.0	b	B

试验结论：E品种与H、C、F、A、D5个品种有5%水平上的差异，E品种与D品种有1%水平上的差异，其余品种之间没有差异显著性。

第六节 随机区组试验的方差分析

两因素试验

A因素	B因素	□ □ □	总计	平均
A1	B1	12 13 13	38	12.67
	B2	12 12 11	35	11.67
	B3	13 14 13	40	13.33
A2	B1	13 14 14	41	13.67
	B2	13 13 12	38	12.67
	B3	14 13 13	40	13.33
A3	B1	14 15 15	44	14.67
	B2	14 13 14	41	13.67
	B3	14 15 15	44	14.67
A4	B1	16 16 15	47	15.67
	B2	15 16 15	46	15.33
	B3	16 17 17	50	16.67
总计		166 171 167		

可先将每个处理组合看作一个样本，采用多样本资料进行分析。即

$$SS_T = SS_t + SSr + SS_e$$

然后，再将SS_t分解为SS_A、SS_B和$SS_{A\times B}$。

第六节 随机区组试验的方差分析

平方和与自由度的分解

A因素的水平数为a，B因素的水平数为b，重复次数为n。

先计算出校正项：

$$C = (\sum y)^2 /(nk) = 504^2 / 3 / 12 = 7056, k = ab = 12$$

$$SS_T = \sum y^2 - C = (12^2 + 12^2 + ... + 17^2) - 7056 = 7132 - 7056 = 76$$

$$SSt = \sum T_t^2 / n - C = (38^2 + 35^2 + \cdots + 50^2) / 3 - 7056 = 7124 - 7056 = 68$$

$$SSr=\sum T_r^{\ 2}/k-C=(166^2+171^2+167^2)/12-7056=7057.167-7056=1.167$$

$$SS_e=SS_T-SSt-SSr=76-68-1.167=6.833$$

$$df_T=nk-1=36-1=35,\ df_t=k-1=12-1=11,$$

$$df_r=3-1=2,\ df_e=df_T-df_t-df_r=35-11-2=22$$

第六节 随机区组试验的方差分析

- 平方和与自由度的分解

对SS_t和df$_t$进行分解，先将处理组合的总计数列成两向表：

	A1	A2	A3	A4	T_B	平均
B1	**38**	**41**	**44**	**47**	**170**	**14.17**
B2	**35**	**38**	**41**	**46**	**160**	**13.33**
B3	**40**	**40**	**44**	**50**	**174**	**14.50**
T_A	**113**	**119**	**129**	**143**		
平均	**12.56**	**13.22**	**14.33**	**15.89**		

$$SS_A=\sum T_A^{\ 2}/(bn)-C=(113^2+119^2+129^2+143^2)/9-7056=7113.33-7056=57.33$$

$$SS_B=\sum T_B^{\ 2}/(an)-C=(170^2+160^2+174^2)/12-7056=7064.67-7056=8.67$$

$$SS_{A\times B}=SS_t-SS_A-SS_B=68-57.33-8.67=2$$

$$df_A=a-1=4-1=3,\ df_B=b-1=3-1=2,$$

$$df_{A\times B}=df_t-df_A-df_B=(a-1)(b-1)=12-3-2=6$$

第六节 随机区组试验的方差分析

- 求均方，进行F测验，列方差分析表。

方差分析表

变异来源	自由度	平方和	均方	F 值	$F_{0.05}$	$F_{0.01}$
区组间	2	1.167	0.58	1.87	3.44	5.72

处理间	11	68	6.18	19.94**	2.26	3.18
A 因素	3	57.33	19.11	61.65**	3.05	4.82
B 因素	2	8.67	4.34	14.00**	3.44	5.72
A×B	6	2	0.33	1.07	2.55	3.76
误差	22	6.83	0.31			
总变异	35	76				

- 结果表明，A和B因素间差异达到极显著水平，A×B互作不显著，应进一步进行A和B因素水平间的多重比较。