基础知识——特征：描述统计—经验分布

其它基础知识：概率、理论分布、抽样分布

统计推断——特征：参数估计（点估计、区间估计）

(估计：平均数、方差、百分率)

差异：假设检验（参数假设检验、非参数假设检验)

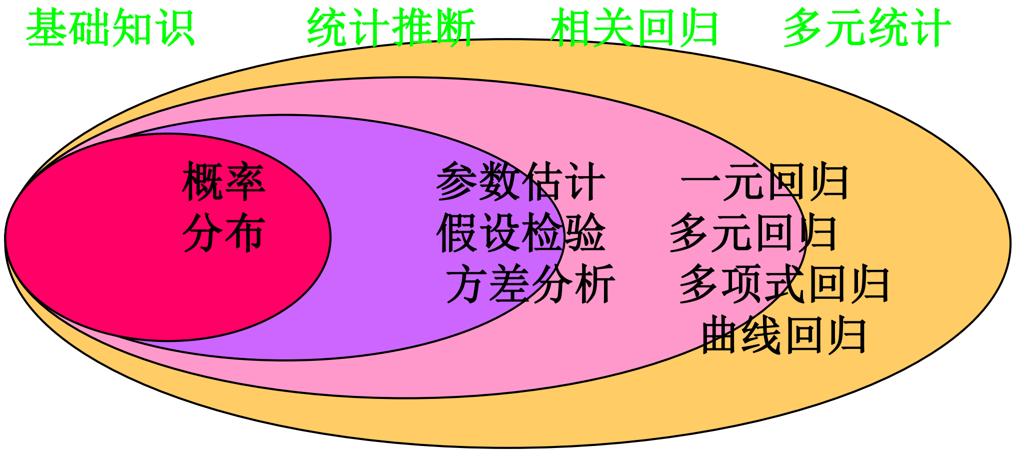
(参数假设检验：平均数、方差、百分率)

(非参数假设检验：分布)

方差分析——检验多个平均数

检验互作

相关回归——关系：



第七章　方差分析

	正态总体		二项总体	未知总体分布
	检验平均数	方差	检验百分率	检验总体分布
一个总体	u, t	χ^2	二项分布, u, t	符号检验，KS检验，适合性检验
两个总体	u, t	F	二项分布, u, t	符号检验，秩和检验，KS检验
多个总体	方差分析	Bartlett		秩和检验

χ^2：独立性检验

变量分析

二维总体分布是否独立

方差分析：7—1 方差分析的基本原理

7—2 单因素方差分析一见上表内容，但下述内容更广

……………………………………………………

7—3 二因素方差分析

7—4 多因素方差分析

7—5 其它：数据转换、缺失数据的估计

李春喜译《生命科学中的数学与统计学》

（**Aulay Mackenzie**）**P118**

- ## 方差分析与 t 检验的来历、关系

1908年William Gosset提出正态分布数据小样本检验。当时他在都伯林的吉尼斯（**Guinness**）啤酒公司工作，该公司拒绝让他发表其检验，但他以“学生 ”**student** 的笔名发表了该检验。这个检验被称为“学生氏 **t** 检验”。该检验只能对两个样本进行检验，不能延伸到三个或更多的样本。

1925年，在罗森斯得农业试验站工作的 **RA Fisher**，将上述的 **t** 检验扩展到了两个样本以上，并进行更复杂试验设计的分析。**Fisher**称其为方差分析。

当方差分析用于两个样本数据时，它给出了与“学生氏 **t** 检验”完全相同的结果。

方差分析的用途

主要用途：

- 多个样本平均数的比较 ----
  单个因素多个水平的平均数的比较

- 多个因素间互作的分析 ---- 交互作用

其它用途：

- 回归方程的假设检验
- 多元相关系数的检验

怎样作方差分析 ----------- Excel
方差分析检验的是什么 -- 期望均方
方差分析之后再做什么 -- 多重比较

7－1 方差分析的基本原理

一、相关术语
二、基本原理
三、数学模型
四、平方和与自由度的分解
五、统计假设的显著性检验
六、多重比较

一、相关术语

试验因素：影响试验指标的原因

非控因素——随机因素。田间试验：温、光、湿、土壤...

可控因素——固定因素。室内试验控制：温、光、湿...

因素水平：各试验因素的不同状态。例：**15℃、20℃、25℃**

试验处理：对受试对象给予的某种外部干预，是试验中实施的因素水平的一个组合。

单因素试验时，试验因素的一个水平就是一个处理；

多因素试验中，实施在试验单位上的具体项目是各因素的某一水平组合。

一个水平组合才是一个处理

试验单位：试验中能接受不同试验处理的独立的试验载体。

书中指"观测总体"，可能不准确，应为"个体"

重复：指在试验中，将一个处理实施在两个(以上)试验单位上

试验误差：由无法控制的因素所引起的差异，常指随机误差

方差分析的误解：并非研究方差
研究平均数的大小转化为研究平均数的方差

二、基本原理

方差分析又叫变量分析，由R.A. Fisher提出。

试验因素
因素水平
试验处理
试验单位
重复
试验误差

教材P91表6－2： 不同饲料喂鱼增重

重复	饲料 A1	A2	A3	A4
1	319	248	221	270
2	279	257	236	308
3	318	268	273	290
4	284	279	249	245
5	359	262	258	286
平均数	311.8	262.8	247.4	279. 8

总平均 275.45

本研究内容是比较不同饲料喂鱼增重是否有差异
统计学本质是比较四个平均数间是否有差异

思路之一：直接对每两个平均数之间进行 t 或 u 检验

问题是：工作不系统且过程繁琐；

t 检验时误差自由度小，估计误差的精确性低，降低检验的灵敏性，容易掩盖差异的显著性；

增大犯α错误的概率：

	不犯错误的概率	犯α错误的概率
单次比较	0.95	1－0.95＝0.05
6 次比较	0.95^6＝0.735	1－0.735＝0.265
	都不犯错	

思路之二：各平均数相同时，彼此间变异较小 -- 方差小

各平均数不同时，彼此间变异较大 -- 方差大

所以，由平均数之间变异(方差)的大小，

可推知平均数之间是否真有差异。

问题是：平均数间方差的大小只是相对的，而不是绝对的，

究竟以什么作为衡量方差大小的标准呢？——误差方差

平均数间方差

样本内方差

——方差分析正是基于此思路：

F＝各处理平均数间方差 / 误差方差

将对平均数的比较转化为对方差的比较 ---- 方差分析

思路之二：各平均数相同时，彼此间变异较小（方差小）
各平均数不同时，彼此间变异较大（方差大）

平均数					平均数间的方差	
	311.8	311.8	311.8	311.8	0	小
	311.8	312.8	313.8	314.8	1.67	中
	311.8	262.8	247.4	279. 8	762.36	大

教材P97表6－2： 不同饲料喂鱼增重

重复	A1	A2	A3	A4
1	319	248	221	270
2	279	257	236	308
3	318	268	273	290
4	284	279	249	245
5	359	262	258	286
平均数	311.8	262.8	247.4	279. 8

平均数间方差 762.36

组间方差 762.36×5＝3811.8

组内方差	1041.7	135.7	399.3	561.2

组内平均方差 534.475

误差方差

平均数间方差

样本内方差

三、数学模型 ---- 线性可加模型

$x_j = \mu + \varepsilon_j$ (不含处理效应)

$x_{ij} = \mu + \tau_i + \varepsilon_{ij}$ (施加处理) (6.1)

x_{ij} 是在第 i 次处理下的第 j 次观察值

μ 是总体平均数

τ_i 是处理效应

ε_{ij}是试验误差，要求相互独立，且服从同一个正态分布$N(0,\sigma^2)$

例：人类 6 胞胎，测量身高或体重

1个同卵6胞胎　3个同卵2胞胎

$x_1 = \mu + \varepsilon_1$	$x_{11} = \mu + G_1 + \varepsilon_{11}$		
$x_2 = \mu + \varepsilon_2$	$x_{12} = \mu + G_1 + \varepsilon_{12}$	σ_1^2	ε_1服从 $N(0,\sigma^2)$
$x_3 = \mu + \varepsilon_3$	$x_{21} = \mu + G_2 + \varepsilon_{21}$		
$x_4 = \mu + \varepsilon_4$	$x_{22} = \mu + G_2 + \varepsilon_{22}$	σ_2^2	ε_2服从 $N(0,\sigma^2)$
$x_5 = \mu + \varepsilon_5$	$x_{31} = \mu + G_3 + \varepsilon_{31}$		
$x_6 = \mu + \varepsilon_6$	$x_{32} = \mu + G_3 + \varepsilon_{32}$	σ_3^2	ε_3服从 $N(0,\sigma^2)$

$x_i = \mu + G + \varepsilon_i$

英国6胞胎姐妹成性感美女惹来追求者无数(图)

1983年11月18日，英国的沃尔顿夫妇在接受不孕症治疗后，一下子生了一组六胞胎女婴，她们成了世界首例、也是唯一一例全部存活的全女婴六胞胎。最新一期时尚杂志《FIRST》披露，23年后，当年的6胞胎女婴个个长成了如花似玉的性感美女，并惹来无数小伙追求。但让沃尔顿夫妇俩“发愁”的是，如果6个女儿在同一年出嫁，那么高昂的嫁妆费非得让他们倾家荡产不可！

打破1/1040亿概率

生下全女婴六胞胎

父母也分不清谁是谁

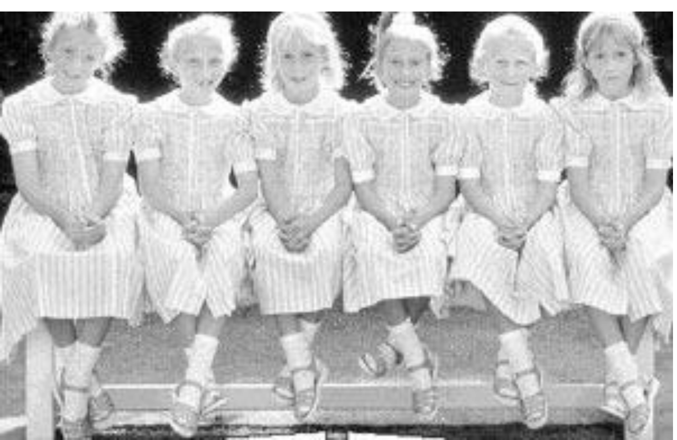
6姐妹广告片约不断

全都长成性感美女

惹来无数小伙追求

5个女儿接连搬回家

老爸最愁她们一起出嫁



世界新纪录：一胞9胎

本报讯 一名马来西亚妇女日前在槟城妇产科医院成功地生下9胞胎，创下了世界记录。但是，在母体里只生活了21周的9个婴儿，在出世后7小时内相继夭折。

这位名叫茱莉娜的产妇今年29岁。在9名婴儿中，最重的两名男婴各为400克，最轻的一名女婴只有310克。9名婴儿总重量为3230克。

茱莉娜说，她曾经到一家私人诊所注射了一种类似“多仔丸”的生育激素。（李士君）

“六胞胎”四个长大成人

河北五胞胎喜迎3周岁

◀这是河北省河间市镇上村五胞胎与母亲王翠英和父亲缴宝存的全家福。3月4日是五胞胎的3岁生日。现在，5个孩子身体健康，聪明可爱。据统计，五胞胎的几率只有六千万分之一，目前全世界报道的五胞胎仅有36例。

新华社发

1：2：3

“超级妈妈”生出“完美组合”

本报综合报道 据英国媒体3月9日报道，英国一位40岁妇女日前一胎生下三个活泼可爱的小宝宝，从而谱写了人类生育史上的一个奇迹：10年前，她生下了一个独生女；2年前，她生下一对双胞胎；现在，居然是一个三胞胎！这个意外的1：2：3的“完美组合”给他们一家带来了意外的惊喜，这位创造奇迹的妇女则被媒体誉为“超级妈妈”。

今年40岁的“超级妈妈”名叫朱莉·沃德，是一位在实验室工作的技术人员，她的丈夫乔·布里蒙特今年56岁，夫妻俩住在英国约克郡，几十年来感情非常和睦。

《楚天都市报》2004年3月11日

最幸运的双胞胎——

一起过100岁生日

英国康沃尔郡特鲁若市双胞胎姐妹贝蒂和鲁妮在今年元旦和家人朋友们一起庆祝了她们的100岁生日。据悉，这对老姐妹共同见证了5位英国君主的更替，她们至今仍爱驾驶汽车，她们好学不倦，在97岁时一起学习西班牙语。而谈起长寿的秘诀，姐妹俩都说：“长寿的秘密就是永远保持微笑。”

据新华社电

线性可加模型：$x_j = \mu + \varepsilon_j$ (不含处理效应)

$$x_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad \text{(施加处理)} \qquad (6.4)$$

样本估计的线性模型：

$$x_{ij} = \bar{x} + t_i + e_{ij} \qquad (6.5)$$

$\bar{x}$ 是样本平均数

t_i 是样本的处理效应＝$\bar{x}_i - \bar{x}$

e_{ij} 是试验误差＝$x_{ij} - \bar{x}_i$

例：主食普通大米、优质大米、黄金大米，用 3个班学生做试验：

t_i 是样本的处理效应： t_1 为普通大米 的效应

t_2 为优质大米 的效应

t_3 为黄金大米 的效应

$\bar{x}$ 是样本平均数：是3个班级的总平均数

转基因水稻

四、平方和与自由度的分解

$$x_{ij} = \bar{x} + t_i + e_{ij} \qquad (6.5)$$

t_i是样本的处理效应$=\bar{x}_i - \bar{x}$

e_{ij}是试验误差$=x_{ij} - \bar{x}_i$

由： $x_{ij} = \bar{x} + t_i + e_{ij}$

$= \bar{x} + (\bar{x}_i - \bar{x}) + (x_{ij} - \bar{x}_i)$

移项： $x_{ij} - \bar{x} = (\bar{x}_i - \bar{x}) + (x_{ij} - \bar{x}_i)$

求平方：$(x_{ij} - \bar{x})^2 = [(\bar{x}_i - \bar{x}) + (x_{ij} - \bar{x}_i)]^2$

$= (\bar{x}_i - \bar{x})^2 + 2(\bar{x}_i - \bar{x})(x_{ij} - \bar{x}_i) + (x_{ij} - \bar{x}_i)^2$

求和：$\sum\sum(x_{ij} - \bar{x})^2 = n\sum(\bar{x}_i - \bar{x})^2 + \sum\sum(x_{ij} - \bar{x}_i)^2$

$$SS_T = SS_t + SS_e$$

求自由度： $(kn-1) = (k-1) + k(n-1)$

$$df_T = df_t + df_e$$

处理均方： $S_t^2 = SS_t / df_t = n\sum(\bar{x}_i - \bar{x})^2 / (k-1)$

误差均方： $S_e^2 = SS_e / df_e = \sum\sum(x_{ij} - \bar{x}_i)^2 / k(n-1)$

$$A^2 = [B + C]^2$$
$$= B^2 + 2BC + C^2$$

$$(x_{ij} - \bar{x})^2 = [(\bar{x}_i - \bar{x}) + (x_{ij} - \bar{x}_i)]^2$$
$$= (\bar{x}_i - \bar{x})^2 + 2(\bar{x}_i - \bar{x})(x_{ij} - \bar{x}_i) + (x_{ij} - \bar{x}_i)^2$$

$$\sum\sum 2(\bar{x}_i - \bar{x})(x_{ij} - \bar{x}_i) = 0$$

利用：离均差之和等于零

离均差乘积之和等于零

例：P91 例6.1： 不同饲料养鱼增重

(1)平方和的计算： $SS_T = SS_t + SS_e$

$SS_T = \sum\sum(x_{ij} - \bar{x})^2 = \sum x^2 - C = 19986.95$

$SS_t = n\sum(\bar{x}_i - \bar{x})^2 = (1/n)\sum T_i^2 - C = 11435.35$

$SS_e = SS_T - SS_t = 8551.60$

(2)自由度的计算：$df_T = df_t + df_e = 3 + 16$

(3)方差的计算

$S_t^2 = 11435.35/3 = 3811.783$

$S_e^2 = 8551.475/16 = 534.475$

五、统计假设的显著性检验——F 检验　P92

检验什么？
如何检验？

F＝处理均方S_t^2 / 误差均方S_e^2 ＝ 3811.8 / 534.5＝7.13

1、期望方差：方差的期望

是指与样本方差对应的总体方差的本质

需要利用H_0假设及方差的同质性：

H_0：　$\mu_1 = \mu_2 = \ldots\ldots = \mu_K$

$\sigma^2{}_1 = \sigma^2{}_2 = \ldots\ldots = \sigma^2{}_K = \sigma^2$　前提

误差均方的期望：$E(Se^2) = \sum\sum(x_{ij} - \overline{x_i})^2 / k(n-1)$

$= \sum [\, \sum(x_{ij} - \overline{x_i})^2 / (n-1) \,] / k$

注：记 $S^2{}_i = \sum(x_{ij} - \overline{x_i})^2 / (n-1) \longrightarrow \sigma^2{}_i$　无偏估计

则：$E(Se^2) = \sum [\, \sigma^2{}_i \,] / k$

$= k\sigma^2 / k$

$= \sigma^2$

处理均方的期望：

$$E(S_t^2) = n\sum(\bar{x}_i - \bar{x})^2/(k-1)$$
$$= \sigma^2 + n[(\sum \tau_i^2)/(k-1)]$$

还可以证明

F 测验：

$$F = E(S_t^2)/E(Se^2) = \frac{\sigma^2 + n[(\sum \tau_i^2)/(k-1)]}{\sigma^2}$$
$$= 1 + ?$$

$$F = S_t^2 / S_e^2 = 3811.8 / 534.5 = 7.13$$

教材P91表6－2： 不同饲料喂鱼增重

重复	A1	A2	A3	A4	
1	319	248	221	270	
2	279	257	236	308	
3	318	268	273	290	
4	284	279	249	245	
5	359	262	258	286	
平均数	311.8	262.8	247.4	279. 8	平均数间的方差 762.36
					组间方差 762.36×5＝3811.8
方差	1041.7	135.7	399.3	561.2	组内平均方差 534.5

$$\sigma^2_1 = \sigma^2_2 = \sigma^2_3 = \sigma^2_4 = \cdots\cdots = \sigma^2_K = \sigma^2$$

2、固定模型与随机模型

F 测验： $F=【\sigma^2+n[(\sum \tau_i^2)/(k-1)]】/\sigma^2=1+?$

F 测验实际上是测验 $n[(\sum \tau_i^2)/(k-1)]$ 是否为0

P88 由对τ_i的不同假定，将数学模型分为固定模型和随机模型

固定模型：记$\kappa^2_\tau=(\sum \tau_i^2)/(k-1)$，并满足$\sum \tau_i=0$

此时，H_0： $\tau_i=0$

H_A： $\tau_i\neq 0$

（或者，H_0： $\mu_1=\mu_2=\cdots\cdots=\mu_K$

H_A： μ_1，μ_2，……，μ_K 不全相等）

随机模型： 记$\sigma^2_\tau=(\sum \tau_i^2)/(k-1)$， 其为$N(0,\sigma^2_\tau)$的方差

此时， H_0： $\sigma^2_\tau=0$

H_A： $\sigma^2_\tau\neq 0$

G.R.Norman《生物统计学基础》
Biostatistics:The Bare Essentials

- 固定模型包括了设计中研究者感兴趣的因素的所有水平。

满足$\sum \tau_i = 0$

- 随机模型只包括了该因素的某些水平，并且意图外推到该因素的所有水平。

固定模型与随机模型的说明

（1）、比较：

固定模型：固定模型的处理效应τ是固定(特定)的——限定这里只有也仅有k个处理τ；研究目的是了解这些特定的处理效应间是否有差异； 若F 测验达显著水平，则进而进行多重比较。

全国1000个品种的猪，选定特定的 4 个品种，研究猪增重

随机模型：随机模型的处理效应τ是随机的——假定这里存在一个关于“处理τ”的总体(实验资料是从这个“处理τ”总体中随机抽取的某些样本，不同的实验，抽取的“处理”就可能不同)；研究的目的是要研究处理τ的变异程度； 若F 测验达显著水平，则进而需要估计σ^2_{τ}的值。

全国1000个品种的猪，随机抽取 4 个品种，研究猪增重的变异

——此模型在数量遗传学中很有用。

（2）、两种模型的不同，只与F 测验的方式和统计推断的内容有关，而与df、SS、S^2的计算无关。

F 检验:

$$F = S_t^2 / S_e^2 = 3811.8 / 534.5 = 7.13$$

方差分析表

表 8-4　单因素固定效应模型方差分析表

变差来源 (source of variation)	平方和 (sum of squares)	自由度 (degree of freedom)	均方 (mean square)	F	均方期望 (expected mean square)
处理间 (between treatments)	SS_A	$a-1$	MS_A	$F=\frac{MS_A}{MS_e}$	$\sigma^2+n\eta_a^2$
误差(error)或处理内 (within treatments)	SS_e	$na-a$	MS_e		σ^2
总和(total)	SS_T	$na-1$			

方差分析

差异源	SS	df	MS	F	P-value	F crit
组间	11435.4	3	3811.78	7.13183	0.00295	3.23887
组内	8551.6	16	534.475			查F值表
总计	19987	19				

例题结果汇总

不同饲料喂鱼增重

重复	A1	A2	A3	A4		
1	319	248	221	270		
2	279	257	236	308		
3	318	268	273	290		
4	284	279	249	245		
5	359	262	258	286		
平均数	311.8	262.8	247.4	279. 8	平均数间的方差	762.36
					组间方差	762.36×5=3811.8
方差	1041.7	135.7	399.3	561.2	组内平均方差	534.5

H_0: $\mu_1 = \mu_2 = \ldots\ldots = \mu_K$

H_A: μ_1, μ_2, ……, μ_K 不全相等

$$SS_T = \sum\sum(x_{ij} - \bar{\bar{x}})^2 = \sum x^2 - C = 19987$$

$$SS_t = n\sum(\bar{x}_i - \bar{\bar{x}})^2 = (1/n)\sum T_i^2 - C = 11435.4$$

$$SS_e = SS_T - SS_t = 8551.6$$

方差分析

差异源	SS	df	MS	F	P-value	F crit
组间	11435.4	3	3811.78	7.13183	0.00295	3.23887
组内	8551.6	16	534.475			查F值表
总计	19987	19				

六、多重比较

1、概述：F 测验是一个整体概念

若是固定模型的实验数据，需进行多重比较

H_0： $\mu_1 = \mu_2 = \ldots\ldots = \mu_K$

H_A： μ_1，μ_2， ……， μ_K 不全相等

2、方法：教材P93－97

(1)、最小显著差数法LSD (Least significant difference)

简单但有缺陷

(2)、最小显著极差法LSR (Least significant range)

新复极差法

q 检验

二者原理相同，只是查表不同

2、方法：教材P93－97

(1)、最小显著差数法 LSD

由假设检验公式： $t=(\bar{x}_1-\bar{x}_2)/s_{(x1-x2)}$

变形： $LSD_\alpha=(\bar{x}_1-\bar{x}_2)=t_\alpha \times s_{(x1-x2)}$

$df=df_e$

其中，平均数差数的标准误 $s_{(x1-x2)}=\sqrt{[(2S^2_e)/n]}$

来历：P 93 公式(6.20)、(6.21)

变动的标准

(2)、最小显著极差法LSR

新复极差检验(SSR)，又称Duncan法

q 检验，又称Newman-keuls 检验

$LSR_\alpha = SSR_\alpha \times s_{\bar{x}}$ 查附表8

$LSR_\alpha = q_\alpha \times s_{\bar{x}}$ 查附表9

平均数标准误 $s_{\bar{x}} = \sqrt{[S^2_e/n]}$

SSRα×平均数标准误 $s_{\bar{x}}$ ＝

10.34

表6-6	4种饲料	喂鱼增重	试验
M	2	3	4
SSR 0.05	3	3.14	3.24
SSR 0.01	4.13	4.31	4.42
LSR 0.05	31.02	32.47	33.5
LSR 0.01	42.7	44.57	45.7

(1)、最小显著差数法**LSD**

(2)、最小显著极差法**LSR**

(3)、用法说明：**P97**

新复极差检验：用于一般的试验

q 检验：用于精度要求高的试验

LSD检验：用于试验中各处理平均数皆与对照相比较

3、结果表示：

(1)、标记字母法：

P94表6－4、P96表6－7

(2)、梯形比较法：

P94表6－5

表6－6　4种饲料　喂鱼增重　试验

M	2	3	4
SSR 0.05	3	3.14	3.24
SSR 0.01	4.13	4.31	4.42
LSR 0.05	31.02	32.47	33.5
LSR 0.01	42.7	44.57	45.7

表6－7　4种饲料　喂鱼增重　差异显著性

饲料	平均数	α＝0.05	α＝0.01	注意方向
A 1	311.8	a	A	A
A 4	279.8	b	AB	AB
A 2	262.8	b	B	B
A 3	247.4	b	B	BC
				BC
				C
				CD

注意方向：

下上下上下上下上……

7－2 单因素方差分析

一、单向分组资料（每组样本容量相等）的方差分析

二、单向分组资料（每组样本容量不相等）的方差分析

一、单向分组资料（每组样本容量相等）的方差分析

P97：表6－10 组内观测次数相等的单因素方差分析

表 6-10　组内观测次数相等的单因素方差分析

变异来源	df	SS	s^2	F
处理间	$k-1$	$SS_t=\frac{\sum T_i^2}{n}-C$	s_t^2	$\frac{s_t^2}{s_e^2}$
处理内	$k(n-1)$	$SS_e=SS_T-SS_t$	s_e^2	
总变异	$kn-1$	$SS_T=\sum x^2-C$		

例：P98例6.2：不同地区黄鼬针毛长度

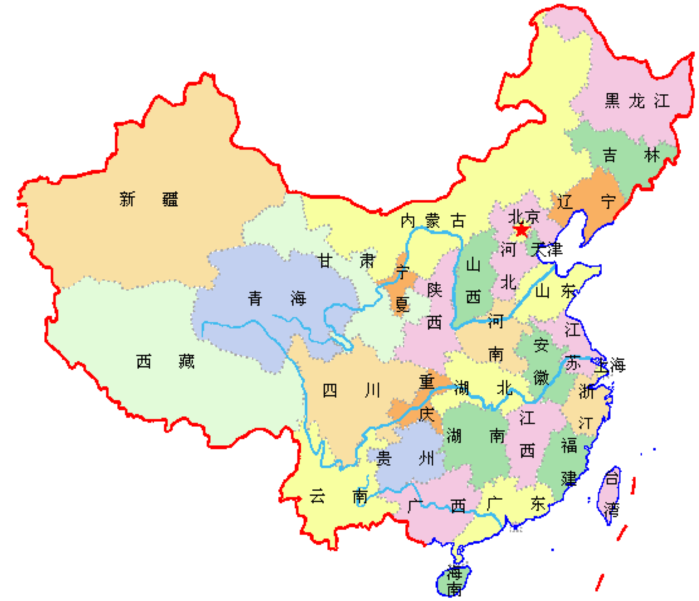
地区	东北	内蒙古	河北	安徽	贵州
1	32	29.2	25.5	23.3	22.3
2	**32.8**	27.4	26.1	25.1	22.5
3	31.2	26.3	25.8	25.1	22.9
4	30.4	26.7	26.7	25.5	23.7
平均	31.6	27.4	26.025	24.75	22.85

差异源	SS	df	MS	F	P-value
组间	173.71	4	43.4275	50.1569	1.7E-08
组内	12.9875	15	0.86583		
总计	186.697	19			

例：P98例6.2：不同地区黄鼬针毛长度

东北最冷　黄鼬针毛最长

贵州最暖　黄鼬针毛最短



黄鼬针毛	长度差异	显著性

		差异显著	性		
地区	平均数	α =0.05	α =0.01		
东北	31.6	a	A		注
内蒙	27.4	b	B		意
河北	26.03	bc	BC		方
安徽	24.75	c	CD		向
贵州	22.85	d	D		

D错

P99表6-13，D错

二、单向分组资料（每组样本容量不相等）的方差分析

P97：表6－10 组内观测次数相等

表 6-10 组内观测次数相等的单因素方差分析

变异来源	df	SS	s^2	F
处理间	$k-1$	$SS_t=\frac{\sum T_i^2}{n}-C$	s_t^2	$\frac{s_t^2}{s_e^2}$
处理内	$k(n-1)$	$SS_e=SS_T-SS_t$	s_e^2	
总变异	$kn-1$	$SS_T=\sum x^2-C$		

P99：表6－14 组内观测次数不相等

表 6-14 组内观测次数不相等的方差分析

变异来源	df	SS	s^2	F
处理间	$k-1$	$SS_t=\sum\frac{T_i^2}{n_i}-C$	s_t^2	$\frac{s_t^2}{s_e^2}$
处理内	$\sum n_i-k$	$SS_e=SS_T-SS_t$	s_e^2	
总变异	$\sum n_i-1$	$SS_t=\sum x^2-C$		

差别：（1）、数学模型中限制条件$\sum \tau_i=0$ 应改为 $\sum n_i\tau_i=0$

（2）、平方和 SS 、 自由度 df

（3）、各 n_i 的平均数 n_0 ——多重比较时使用

例：P100例6.3 小麦种子切胚乳后种植，成熟后测单株粒重

表6－15	小麦试验										
	株号										
处理	1	2	3	4	5	6	7	8	9	10	合计	平均数
整粒 I	21	29	24	22	25	30	27	26			204	25.5
切一半 II	20	25	25	23	29	31	24	26	20	21	244	24.4
切全部 III	24	22	28	25	21	26					146	24.3

方差分析					
差异源	SS	df	MS	F	P-value
组间	6.76667	2	3.38333	0.31757	0.73135
组内	223.733	21	10.654		
总计	230.5	23			

$$n_0 = \frac{(\sum n_i)^2 - \sum n_i^2}{(\sum n_i)(k-1)} = 7.8$$

例：						
徐秉玖编	药物统计	学	P131	表6.5		
表6.5	在5个实验	室中对同	一批药片测得的15	分钟后溶	出百分比	

A	B	C	D	E
68	55	78	75	65
78	62	63	60	60
63	67	78	66	66
56	60	65	69	75
61	67	70	58	75
69	73	74	64	70
			71	
			71	
			65	
			77	
			60	
			63	

$n_0 = 1008/144 = 7$

差异源	SS	df	MS	F	P-value
组间	189.722	4	47.4306	1.1464	0.35326
组内	1282.58	31	41.3737		
总计	1472.31	35			

第七章　方差分析

7－1 方差分析的原理

7－2 单因素方差分析

有重复

一、单向分组资料（每组样本容量　相等）的方差分析

二、单向分组资料（每组样本容量不相等）的方差分析

7－3 二因素方差分析

一、无重复观察值的二因素方差分析

二、有重复观察值的二因素方差分析

7－4 多因素方差分析

一、无重复观察值的三因素方差分析

二、有重复观察值的三因素方差分析

7－5 其它

一、数据转换

二、缺失数据的估计

第七章　方差分析

	正态总体		二项总体	未知总体分布
	检验平均数	方差	检验百分率	检验总体分布
一个总体	u, t	χ^2	二项分布, u, t	符号检验，KS检验，适合性检验
两个总体	u, t	F	二项分布, u, t	符号检验，秩和检验，KS检验
多个总体	方差分析	Bartlett		秩和检验

变量分析

两个总体分布是否独立

χ^2 独立性检验

方差分析：7—1 方差分析的基本原理

7—2 单因素方差分析—见上表内容，但下述内容更广

……………………………………………………

7—3 二因素方差分析

7—4 多因素方差分析

第七章　方差分析

7—1 方差分析的原理

7—2 单因素方差分析

一、单向分组资料（每组样本容量 相等）的方差分析

二、单向分组资料（每组样本容量不相等）的方差分析

-- 有重复 --------------------

7—3 二因素方差分析

一、无重复观察值的二因素方差分析

二、有重复观察值的二因素方差分析

7—4 多因素方差分析

一、无重复观察值的三因素方差分析

二、有重复观察值的三因素方差分析

7—5 其它

一、数据转换

二、缺失数据的估计

单因素

不同的鱼种 ---- 单因素
比较生长速度 /相同的饲料

无重复
不能进行方差分析

红色：1
青色：1
黄色：1

不能求出误差方差

有重复
能够进行方差分析
样本容量不相等

红色：7
青色：2
白色：2

能够求出误差方差

7－3 二因素方差分析

目的：研究两种因素共同影响试验结果
手段：比较两个方向的多个平均数差异

有重复观察值

教材P107

表6－25	某种昆虫	滞育天数	
	温度25℃	30℃	35℃
光照5h/d	143	101	89
	183	100	93
	120	80	101
	107	83	76
10h/d	96	79	80
	103	61	76
	78	83	61
	91	59	67
15h/d	79	60	67
	83	71	58
	96	78	71
	98	64	83

无重复观察值

教材102

表6－19	激素处理	对大豆干	物重影响
	时间H1	H2	H3
浓度M1	13	14	14
M2	12	12	13
M3	3	3	3
M4	10	9	10
M5	2	5	4

方差分析的用途

- 多个样本平均数的比较
- 多个因素间互作的分析 ---- 交互作用

其它用途：

- 回归方程的假设检验
- 多元相关系数的检验

$$A^2 = [B+C]^2 = B^2 + 2BC + C^2$$

$$(x_{ij} - \bar{x})^2 = [(\bar{x}_i - \bar{x}) + (x_{ij} - \bar{x}_i)]^2$$

$$= (\bar{x}_i - \bar{x})^2 + 2(\bar{x}_i - \bar{x})(x_{ij} - \bar{x}_i) + (x_{ij} - \bar{x}_i)^2$$

$$\sum\sum 2(\bar{x}_i - \bar{x})(x_{ij} - \bar{x}_i) = 0$$

利用：离均差之和等于零

离均差乘积之和等于零

$A^2 = [B+C+D+E]^2 =$

	B	C	D	E
B	B^2	2BC	2BD	2BE
C		C^2	2CD	2CE
D			D^2	2DE
E				E^2

一、无重复观察值的二因素方差分析

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

$SS_T = SS_t + SS_e$

1、数学模型：$x_{ij}=\mu+\alpha_i+\beta_j+\varepsilon_{ij}$

直接写出下述公式

平方和分解：$SS_T=SS_A+SS_B+SS_e$

自由度分解：$df_T=df_A+df_B+df_e$

$=(a-1)+(b-1)+(a-1)(b-1)$

计算方差：

$S^2_A=SS_A/df_A$

$S^2_B=SS_B/df_B$

$S^2_e=SS_e/df_e$

F测验：

$F=S^2_A/S^2_e$

$F=S^2_B/S^2_e$

表6－19	激素处理	对大豆干
	时间 H1	H2
浓度 M1	13	14
M2	12	12
M3	3	3

	时间 H1			H2		处理效应
浓度 M1	μ＋α 1+	β 1+ε 11		μ＋α 1+	β 2+ε 12	α 1
M2	μ＋α 2+	β 1+ε 21		μ＋α 2+	β 2+ε 22	α 2
M3	μ＋α 3+	β 1+ε 31		μ＋α 3+	β 2+ε 32	α 3
处理效应		β 1			β 2	

2、教材P102，表6-18：固定模型、随机模型、混合模型

此时 F 检验无差别

规则：F ＝ 1 ＋ ?

表 6-18　无重复观测值的二因素方差分析

变异来源	df	SS	s^2	F	期望方差 $E(s^2)$		
					固定模型	随机模型	混合模型*
A 因素	$a-1$	SS_A	s_A^2	$\frac{s_A^2}{s_e^2}$	$\sigma^2+b\eta_\alpha^2$	$\sigma^2+b\sigma_\alpha^2$	$\sigma^2+b\eta_\alpha^2$
B 因素	$b-1$	SS_B	s_B^2	$\frac{s_B^2}{s_e^2}$	$\sigma^2+a\eta_\beta^2$	$\sigma^2+a\sigma_\beta^2$	$\sigma^2+a\sigma_\beta^2$
误　差	$(a-1)(b-1)$	SS_e	s_e^2		σ^2	σ^2	σ^2
总变异	$ab-1$	SS_T					

单因素方差分析期望方差

变差来源 (source of variation)	平方和 (sum of squares)	自由度 (degree of freedom)	均方 (mean square)	F	均方期望 (expected mean square)
处理间 (between treatments)	SS_A	$a-1$	MS_A	$F=\frac{MS_A}{MS_e}$	$\sigma^2+n\eta_a^2$
误差(error)或处理内 (within treatments)	SS_e	$na-a$	MS_e		σ^2
总和(total)	SS_T	$na-1$			

3、例：教材P102：激素处理对大豆干物重的影响

教材P96			
表6—19	激素处理	对大豆干	物重影响
	时间H1	H2	H3
浓度M1	13	14	14
M2	12	12	13
M3	3	3	3
M4	10	9	10
M5	2	5	4

结果：

处理浓度之间的差异达到显著水平

处理时间之间的差异未达显著水平

方差分析					
差异源	SS	df	MS	F	P-value
行	289.067	4	72.2667	117.189	3.9E-07
列	1.73333	2	0.86667	1.40541	0.29987
误差	4.93333	8	0.61667		
总计	295.733	14			

无重复观察值的二因素方差分析的两个问题:

掩盖试验因素的效果（显著性），增加犯β错误的概率：因为此时估计的误差，是在两因素不存在互作或互作很小的情况下估计的。— —若存在互作，则其误差方差较大

纳伪错误

分母较大

无法估计真正的试验误差：因为每个水平组合只有一个观测值，不能对因素的交互作用进行研究。— —不能剖分出互作与真正的误差方差

两因素或多因素试验，一般应设重复。

4、多重比较：P103

- 用SSR检验
- 平均数标准误

比较A因素各水平间，平均数标准误 $s_{\bar{x}} = \sqrt{[S^2_e / b]}$

比较B因素各水平间，平均数标准误 $s_{\bar{x}} = \sqrt{[S^2_e / a]}$

二、有重复观察值的二因素方差分析

1、数学模型：$x_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$

直接写出

平方和分解：$SS_T = SS_A + SS_B + SS_{AB} + SS_e$

自由度分解：$df_T = df_A + df_B + df_{AB} + df_e$

计算方差：

$$S^2_A = SS_A / df_A$$
$$S^2_B = SS_B / df_B$$
$$S^2_{AB} = SS_{AB} / df_{AB}$$
$$S^2_e = SS_e / df_e$$

F测验：

$$S^2_A / S^2_e$$
$$S^2_B / S^2_e$$
$$S^2_{AB} / S^2_e$$

固定模型

2、教材P106，表6-24：

规则：F ＝ 1 ＋ ?

固定模型：处理所产生的效应值τ_i是固定的，侧重于效应值的估计和比较

随机模型：处理的效应值τ_i不是固定的数值，侧重于效应方差的估计和检验

混合模型：随机抽取 3个省；再从城市和乡村各抽取20个数据。

省份是随机效应型，城乡是固定效应型——混合模型

表 6-24 具重复观测值的二因素分组资料方差分析表

变异来源	SS	df	s^2
因素 A	SS_A	$a-1$	s_A^2
因素 B	SS_B	$b-1$	s_B^2
A × B	SS_{AB}	$(a-1)(b-1)$	s_{AB}^2
误 差	SS_e	$ab(n-1)$	s_e^2
总变异	SS_T	$abn-1$	

变异来源	固定模型		随机模型		混合模型(A 固定,B 随机)	
	F	期望方差	F	期望方差	F	期望方差
因素 A	$\frac{s_A^2}{s_e^2}$	$\sigma^2+bn\eta_\alpha^2$	$\frac{s_A^2}{s_{AB}^2}$	$\sigma^2+n\sigma_{\alpha\beta}^2+bn\sigma_\alpha^2$	$\frac{s_A^2}{s_{AB}^2}$	$\sigma^2+n\sigma_{\alpha\beta}^2+bn\eta_\alpha^2$
因素 B	$\frac{s_B^2}{s_e^2}$	$\sigma^2+an\eta_\beta^2$	$\frac{s_B^2}{s_{AB}^2}$	$\sigma^2+n\sigma_{\alpha\beta}^2+an\sigma_\beta^2$	$\frac{s_A^2}{s_e^2}$	$\sigma^2+an\sigma_\beta^2$
A × B	$\frac{s_{AB}^2}{s_e^2}$	$\sigma^2+n\eta_{\alpha\beta}^2$	$\frac{s_{AB}^2}{s_e^2}$	$\sigma^2+n\sigma_{\alpha\beta}^2$	$\frac{s_{AB}^2}{s_e^2}$	$\sigma^2+n\sigma_{\alpha\beta}^2$
误 差		σ^2		σ^2		σ^2

3、例：教材P107表6-25

采用表6-25的数据计算

表6—25	某种昆虫	滞育天数	
	温度25℃	30℃	35℃
光照 5h/d	143	101	89
	183	100	93
	120	80	101
	107	83	76
10h/d	96	79	80
	103	61	76
	78	83	61
	91	59	67
15h/d	79	60	67
	83	71	58
	96	78	71
	98	64	83

温度和光照是人为控制的固定因素

固定模型

方差分析					
差异源	SS	df	MS	F	P-value
样本	7029.56	2	3514.78	16.3281	2.2E-05
列	7061.06	2	3530.53	16.4013	2.2E-05
交互	1329.94	4	332.486	1.54458	0.2176
内部	5812	27	215.259		
总计	21232.6	35			

计算器的年代
计算机的时代

例：教材P107

用表6-26的数据计算
新数据＝原数据－80

表6－26			
	温度25℃	30℃	35℃
光照5h/d	63	21	9
	58	20	13
	40	0	21
	27	3	-4
10h/d	16	-1	0
	23	-19	-4
	-2	3	-19
	11	-21	-13
15h/d	-1	-20	-13
	3	-9	-22
	16	-2	-9
	18	-16	3

方差分析					
差异源	SS	df	MS	F	P-value
样本	5367.06	2	2683.53	21.9345	2.2E-06
列	5391.06	2	2695.53	22.0326	2.1E-06
交互	464.944	4	116.236	0.95009	0.45053
内部	3303.25	27	122.343		
总计	14526.3	35			

例：教材P108表6-28

啤酒生产中糖化时间的影响因素：

烘烤方式 A（A1、A2） 固定因素

大麦水份 B（B1、B2、B3、B4） 随机因素

大麦水份不均匀又不易控制：随机

2×4＝8 种处理

3 次重复

混合模型

烘烤方式	水份 B1	水份 B2	水份 B3	水份 B4
A1，温度低				
A2，温度高				

4、主效与互作：教材 第四版 P6、P87

主效：各试验因素的相对独立作用为该因素的主效应，简称主效

互作：某一因素在另一因素的不同水平上所产生的效应不同，则二因素间存在交互作用，简称互作。

——两个或多个试验因素的相互作用而产生的效应

共同努力
相互扯皮

例1：小鼠饲料营养实验 （蛋白质α_i 、维生素β_j ）

$$x_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

若不存在互作，$(\alpha\beta)_{ij} = 0$，则主效αi与βj 累加就能决定小鼠的增重，即各因素的最优水平的组合就是最优的处理组合

若存在互作， $(\alpha\beta)_{ij} \neq 0$，则最优处理组合的选定应根据各处理组合的直接表现选定。

广告语：有多少亲朋好友
送多少黄金搭档

最佳组合　最佳搭档

$x_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$

共同努力
相互扯皮

促进作用　x_{ij} = 表现$_i$＋表现$_j$＋(更好)$_{ij}$

无互作用　x_{ij} = 表现$_i$＋表现$_j$＋(0)$_{ij}$

抑制作用　x_{ij} = 表现$_i$＋表现$_j$＋(差)$_{ij}$

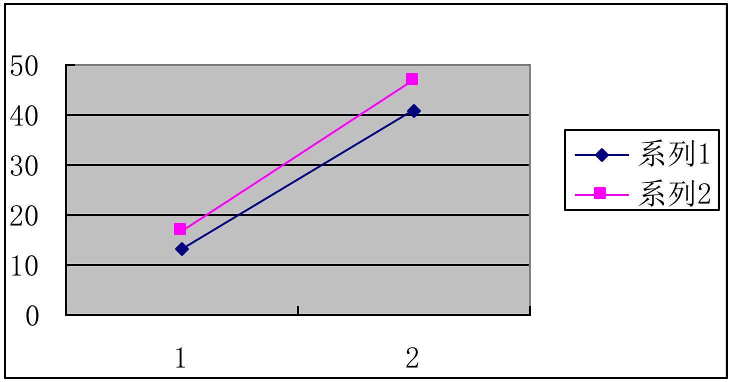
一山不容二虎：x_{ij} = 虎$_i$＋虎$_j$＋(争斗 / 两败俱伤)$_{ij}$

男女搭配，干活不累：

x_{ij} = 男人$_i$＋女人$_j$＋(更好 / 干活不累)$_{ij}$

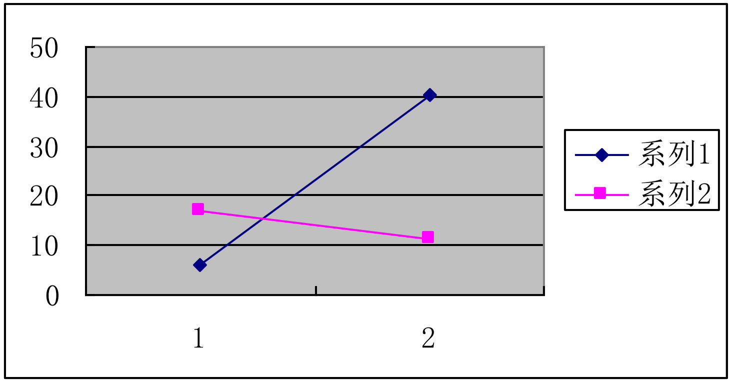
三个臭皮匠合成一个诸葛亮：

x_{ij} =凡人$_i$＋凡人$_j$＋凡人$_k$＋(好)$_{ij}$＋(好)$_{ik}$＋(好)$_{jk}$＋(更好)$_{ijk}$



男—女

狗—蛙



男人女人	100米速度	
	陆地	水中
男人	15	45
	12	36
	13	39
	13	39
	14	42
女人	18	48
	17	55
	16	42
	16	42
	18	48

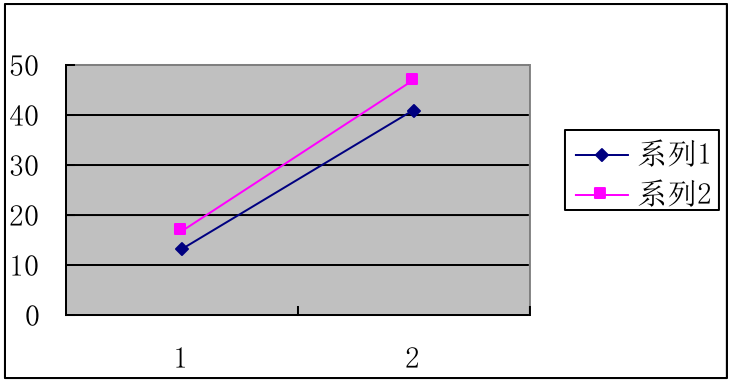
男女：陆、水中跑游速度

狗蛙：陆、水中跑游速度

狗与蛙的	100米速度	
	陆地	水中
狗	7	45
	6	36
	6	39
	5	39
	7	42
蛙	18	12
	17	11
	16	10
	16	10
	18	13

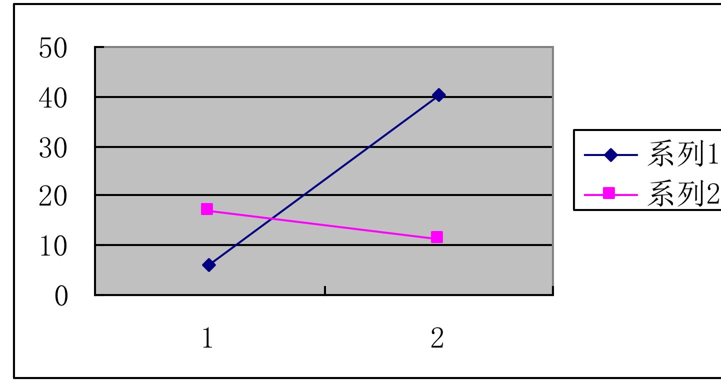
差异源	SS	df	MS	F	P-value
行	135.2	1	135.2	12.5767	0.00269
列	4032.8	1	4032.8	375.144	1.6E-12
交互	12.8	1	12.8	1.1907	0.29135
内部	172	16	10.75		
总计	4352.8	19			

差异源	SS	df	MS	F	P-value
行	414.05	1	414.05	109.682	1.43873E-08
列	994.05	1	994.05	263.325	2.33636E-11
交互	1980.05	1	1980.05	524.517	1.17322E-13
内部	60.4	16	3.775		
总计	3448.55	19			



男—女

狗—蛙



2×2

红高黑低
红高黑低

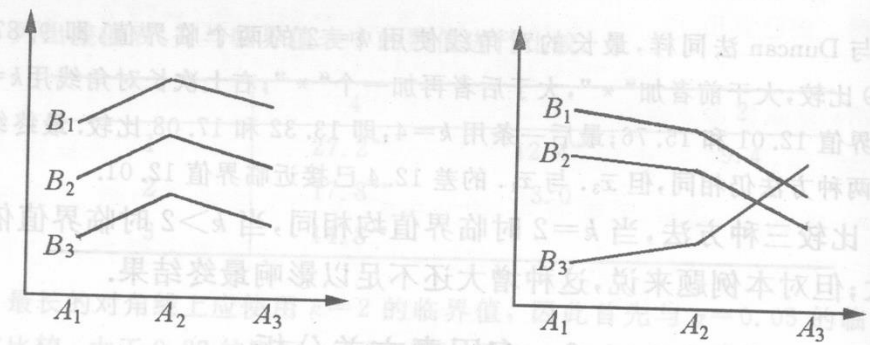
互作：某一因素在另一因素的不同水平上所产生的效应不同，则二因素间存在交互作用，简称互作。

红高黑低
红低黑高

当一个因素的效应明显地依赖于其他因素的水平时，我们称这些因素间有交互效应．例如，由于人的体质不同，药物的疗效也可能会有不同；不同的地施用同样的肥料，增产效果也有不同，等等．交互效应的有无可用一些直观方法粗略估计，例如可用图形来估计（见图4.1）．

3×3

低高低
低高低
低高低



(a) 无交互效应 (b) 有交互效应

图 4.1 交互效应示意图

高中低
高中低
低中高

图中每条曲线代表 B 因素的一个水平．若各曲线平行或近似平行，可认为无交互效应，否则为有交互效应．以上只是一种直观的判断，

5、多重比较：P108

- 用SSR检验
- 平均数标准误

比较A因素各水平间，平均数标准误 $s_{\bar{x}} = \sqrt{[S^2_e / bn]}$

比较B因素各水平间，平均数标准误 $s_{\bar{x}} = \sqrt{[S^2_e / an]}$

7－4 多因素方差分析

一、无重复观察值的三因素方差分析

与下者差别不大，后述

二、有重复观察值的三因素方差分析

例1、教材P112表6-33：

胱氨酸(A_i)：4

蛋氨酸(B_j)：3

蛋白质(C_k)：2

24 种处理

2 次重复

猪的每天增重量

胱氨酸 A	蛋氨酸 B	蛋白质 C	日增重	
0	0	12	1.11	0.97
		14	1.52	1.45
	0.025	12	1.09	0.99
		14	1.27	1.22
	0.05	12	0.85	1.21
		14	1.67	1.24
0.05	0	12	1.3	1
		14	1.55	1.53
	0.025	12	1.03	1.21
		14	1.24	1.34
	0.05	12	1.12	0.96
		14	1.76	1.27
0.1	0	12	1.22	1.13
		14	1.38	1.08
	0.025	12	1.34	1.41
		14	1.4	1.21
	0.05	12	1.34	1.19
		14	1.46	1.39
0.15	0	12	1.19	1.03
		14	0.8	1.29
	0.025	12	1.36	1.16
		14	1.42	1.39
	0.05	12	1.46	1.03
		14	1.62	1.27

例2、生科院98级肖建强同学的生态学实验

鱼数(B_j)： 4
草类(A_i)： 3
时间(C_k)： 3

36 种处理
2 次重复

缺失数据

鱼数	草类	测定时期	重复I	重复II
0	A草	t1	32.7	32.7
		t2	51.4	51.4
		t3	54.5	54.5
	B草	t1	33.9	33.9
		t2	30.7	30.7
		t3	26.9	26.9
	C草	t1	57.8	57.8
		t2	54.1	54.1
		t3	29.4	29.4
1	A草	t1	38.9	39.5
		t2	60.9	43.9
		t3	50.6	57
	B草	t1	26.1	30.7
		t2	40	22.5
		t3	17.1	29
	C草	t1	18.3	95.3
		t2	8.9	9.5
		t3	3.4	3.4
2	A草	t1	43.8	43.8
		t2	59.4	59.4
		t3	43.7	43.7
	B草	t1	46.2	18.5
		t2	19.9	6.3
		t3	19.8	6.3
	C草	t1	24.2	82.6
		t2	6.8	6.8
		t3	2.8	8.1
3	A草	t1	31.7	42.5
		t2	26.1	30.1
		t3	24.9	44.3
	B草	t1	34.2	47.8
		t2	18.3	38
		t3	20.7	9.3
	C草	t1	88.2	38.4
		t2	8.4	6.6
		t3	14.4	2.3

二、有重复观察值的三因素方差分析

1、数学模型：

$$x_{ijkl}=\mu+\alpha_i+\beta_j+\gamma_k+(\alpha\beta)_{ij}+(\alpha\gamma)_{ik}+(\beta\gamma)_{jk}+(\alpha\beta\gamma)_{ijk}+\varepsilon_{ijkl}$$

直接写出下述公式

平方和：

$$SS_T=SS_A+SS_B+SS_C+SS_{AB}+SS_{AC}+SS_{BC}+SS_{ABC}+SS_e$$

自由度：

$$df_T=df_A+df_B+df_C+df_{AB}+df_{AC}+df_{BC}+df_{ABC}+df_e$$

无重复观察值时就无：$(\alpha\beta\gamma)_{ijk}$ SS_{ABC} df_{ABC}

2、内容：

P111

表6-32

固定模型

随机模型

混合模型

ABC 都固定

A、 B 固定，C 随机

A 固定， B、C 随机

ABC 都随机

表 6-32 三因素方差分析表

变异来源	SS	df	s^2	固定模型 F	固定模型 期望方差
A	SS_A	$a-1$	s_A^2	$\frac{s_A^2}{s_e^2}$	$\sigma^2+bcn\eta_\alpha^2$
B	SS_B	$b-1$	s_B^2	$\frac{s_B^2}{s_e^2}$	$\sigma^2+acn\eta_\beta^2$
C	SS_C	$c-1$	s_C^2	$\frac{s_C^2}{s_e^2}$	$\sigma^2+abn\eta_\gamma^2$
$A\times B$	SS_{AB}	$(a-1)(b-1)$	s_{AB}^2	$\frac{s_{AB}^2}{s_e^2}$	$\sigma^2+cn\eta_{\alpha\beta}^2$
$A\times C$	SS_{AC}	$(a-1)(c-1)$	s_{AC}^2	$\frac{s_{AC}^2}{s_e^2}$	$\sigma^2+bn\eta_{\alpha\gamma}^2$
$B\times C$	SS_{BC}	$(b-1)(c-1)$	s_{BC}^2	$\frac{s_{BC}^2}{s_e^2}$	$\sigma^2+an\eta_{\beta\gamma}^2$
$A\times B\times C$	SS_{ABC}	$(a-1)(b-1)(c-1)$	s_{ABC}^2	$\frac{s_{ABC}^2}{s_e^2}$	$\sigma^2+n\eta_{\alpha\beta\gamma}^2$
误　差	SS_e	$abc(n-1)$	s_e^2		σ^2

变异来源	随机模型 F	随机模型 期望方差	混合模型（A、B 固定，C 随机） F	混合模型（A、B 固定，C 随机） 期望方差
A	不能检验	$\sigma^2+bcn\sigma_\alpha^2+cn\sigma_{\alpha\beta}^2+bn\sigma_{\alpha\gamma}^2+n\sigma_{\alpha\beta\gamma}^2$	$\frac{s_A^2}{s_{AC}^2}$	$\sigma^2+bcn\eta_\alpha^2+bn\sigma_{\alpha\gamma}^2$
B		$\sigma^2+acn\sigma_\beta^2+cn\sigma_{\alpha\beta}^2+an\sigma_{\beta\gamma}^2+n\sigma_{\alpha\beta\gamma}^2$	$\frac{s_B^2}{s_{BC}^2}$	$\sigma^2+acn\eta_\beta^2+an\sigma_{\beta\gamma}^2$
C		$\sigma^2+abn\sigma_\gamma^2+bn\sigma_{\alpha\gamma}^2+an\sigma_{\beta\gamma}^2+n\sigma_{\alpha\beta\gamma}^2$	$\frac{s_C^2}{s_e^2}$	$\sigma^2+abn\sigma_\gamma^2$
$A\times B$	$\frac{s_{AB}^2}{s_{ABC}^2}$	$\sigma^2+cn\sigma_{\alpha\beta}^2+n\sigma_{\alpha\beta\gamma}^2$	$\frac{s_{AB}^2}{s_{ABC}^2}$	$\sigma^2+cn\eta_{\alpha\beta}^2+n\sigma_{\alpha\beta\gamma}^2$
$A\times C$	$\frac{s_{AC}^2}{s_{ABC}^2}$	$\sigma^2+bn\sigma_{\alpha\gamma}^2+n\sigma_{\alpha\beta\gamma}^2$	$\frac{s_{AC}^2}{s_e^2}$	$\sigma^2+bn\sigma_{\alpha\gamma}^2$
$B\times C$	$\frac{s_{BC}^2}{s_{ABC}^2}$	$\sigma^2+an\sigma_{\beta\gamma}^2+n\sigma_{\alpha\beta\gamma}^2$	$\frac{s_{BC}^2}{s_e^2}$	$\sigma^2+an\sigma_{\beta\gamma}^2$
$A\times B\times C$	$\frac{s_{ABC}^2}{s_e^2}$	$\sigma^2+n\sigma_{\alpha\beta\gamma}^2$	$\frac{s_{ABC}^2}{s_e^2}$	$\sigma^2+n\sigma_{\alpha\beta\gamma}^2$
误　差		σ^2		σ^2

3、例1：P112表6－33

胱氨酸(A_i)： 4

蛋氨酸(B_j)： 3

蛋白质(C_k)： 2

24 种处理

2 次重复

固定模型

胱氨酸 A	蛋氨酸 B	蛋白质 C	日增重	
0	0	12	1.11	0.97
		14	1.52	1.45
	0.025	12	1.09	0.99
		14	1.27	1.22
	0.05	12	0.85	1.21
		14	1.67	1.24
0.05	0	12	1.3	1
		14	1.55	1.53
	0.025	12	1.03	1.21
		14	1.24	1.34
	0.05	12	1.12	0.96
		14	1.76	1.27
0.1	0	12	1.22	1.13
		14	1.38	1.08
	0.025	12	1.34	1.41
		14	1.4	1.21
	0.05	12	1.34	1.19
		14	1.46	1.39
0.15	0	12	1.19	1.03
		14	0.8	1.29
	0.025	12	1.36	1.16
		14	1.42	1.39
	0.05	12	1.46	1.03
		14	1.62	1.27

方差分析						
变异来源	df	SS	MS	F	F0.05	F0.01
胱氨酸 A	3	0.0427	0.0142	0.445	3.01	4.72
蛋氨酸 B	2	0.0526	0.0263	0.824	3.4	5.61
蛋白质 C	1	0.5355	0.5355	16.787**	4.26	7.82
A□	6	0.2543	0.0424	1.329	2.51	3.67
A□	3	0.2399	0.08	2.508	3.01	4.72
B□	2	0.0821	0.041	1.285	3.4	5.61
A□□	6	0.0685	0.0114	0.357	2.51	3.67
误差	24	0.7653	0.0319			
总变异	47	2.0409				

例2：生科院98级肖建强同学的实验

固定模型

变因	df	SS	MS	F	$F_{0.05}$	$F_{0.01}$
A草类间	2	4328.93	2164.46	6.29**	3.44	5.72
B鱼数间	3	1622.58	540.86	1.57	3.05	4.82
C时间间	2	3833.47	1916.73	5.57*	3.44	5.72
AB	6	2282.96	380.49	1.11	2.55	3.76
AC	4	7226.05	1806.51	5.25**	2.82	4.31
BC	6	1600.60	266.77	＜1		
ABC	12	1317.06	109.76	＜ 1		
机误	36-14	7571.78	344.17			
总	72-1-14	29783.42				

7－5 缺失数据的估计

缺区估计的数据并不能提供任何新的信息

教材P115

原则：使补上缺失的数据后，误差平方和最小。

方法：

(1)、试验有重复，且每种搭配条件下试验数据至少有一个数据没有丢失：则丢失的数据用同一种搭配条件下没有丢失的数据的平均数来代替它；方差分析中误差平方和的自由度等于原有的自由度减去丢失数据的个数。

例：肖建强同学的生态学实验

(2)、有些搭配条件下没有数据：教材P115-116

误差平方和——令其最小

有些搭配条件下没有数据：教材P115

缺失一个数据的估计：$x-(305+x)/8-(112+x)/4+(1216+x)/32=0$

缺失两个数据的估计：$x-(305+x)/8-(112+x)/4+(1183+x+y)/32=0$

$y-(245+y)/8-(111+y)/4+(1183+x+y)/32=0$

	B1	B2	B3	B4	B5	B6	B7	B8	合计
A1	30	39	41	42	42	39	38	38	309
A2	37	46	x	43	51	44	35	49	305+x
A3	27	37	36	24	37	41	33	43	278
A4	30	42	35	40	46	47	38	46	324
合计	124	164	112+x	149	176	171	144	176	1216+x

	B1	B2	B3	B4	B5	B6	B7	B8	合计
A1	30	39	41	42	42	39	38	38	309
A2	37	46	x	43	51	44	35	49	305+x
A3	27	37	36	24	37	41	y	43	245+y
A4	30	42	35	40	46	47	38	46	324
合计	124	164	112+x	149	176	171	111+y	176	1183+x+y

7－6 方差分析的基本假定和数据转换

教材P116

1、方差分析的基本假定：

正态性：试验误差是服从正态分布$N(0,\sigma^2)$的独立的随机变量

可加性：处理效应与误差效应是可加的

$$x_{ij}=\mu+\alpha_i+\beta_j+\varepsilon_{ij}$$

方差同质性：所有试验的误差方差具备同质性，即方差齐性

$$\sigma^2_1=\sigma^2_2=\sigma^2_3 \;\ldots\ldots = \sigma^2_K=\sigma^2$$

方差非同质的一种常见情形：处理平均数与方差有一定关系

方差非同质时：数据转换常可同时改进误差方差的异质性、以及非正态性或非可加性。

等重复试验的精度高于不等重复试验，应尽量采用等重复试验

2、数据转换的方式

a、平方根转换：

用于数据为Poisson分布、或样本平均数与方差有比例关系

一般情况：$\sqrt{(x)}$

数据较小时：$\sqrt{(x+1)}$

b、对数转换：

用于总体为对数正态分布的资料，或者资料中的效应成比例而不是可加的，或者平均数与标准差成比例时

一般情况：lg(x)、 ln(x)

原始数据包括 0 时：lg(x+1)

c、反正弦转换：转换后的数值是以度为单位的角度，角度转换

用于数据是比例数或以百分率表示的，其分布趋于二项分布

$$\theta = \sin^{-1}\sqrt{(P)}$$

多重比较时需用转换后的数据进行计算

解释分析最终结果时应还原为原来的数值

注意：

方差分析的步骤：

1、看是否需要进行缺区估计、数据转换

2、方差同质性检验

3、方差分析表：自由度、平方和、方差、F检验

4、多重比较——固定模型时

不看课本与笔记本，能够直接写出下面的答案吗？

1、一阶中心矩：

$m_k = \sum (x - \overline{x}) / n =$

2、对立事件：

$P(A) + P(\overline{A}) =$

3、完全事件系：

$P(A_1 + A_2 + \cdots + A_n) = P(A_1) + P(A_2) + \cdots + P(A_n) =$

4、独立事件的乘法定理：

$P(A_1 \cdot A_2 \cdot \cdots \cdot A_n) =$

5、正态分布$N(\mu, \sigma^2)$的数字特征：

$E(x) =$ $V(x) =$

$\gamma_1 =$ $\gamma_2 =$

6、计算概率：$P(|u| > 1.96) =$

$P(|u| > 2.58) =$

作业：

第二版： **6.1 6.2 6.3 6.6 6.10 6.11**

第三版： **6.1 6.2 6.3 6.5 6.7 6.8**

第四版： **6.1 6.2 6.3 6.5 6.6 6.7**