

基础数理统计

(研究生公共课)

肖柳青 主讲





第五章 方差分析与正交设计 (ANOVA--The Analysis of Variance)

主要内容

- 5.1 单因素方差分析
- 5.2 双因素方差分析
- 5.3 正交试验设计



§ 5.1 单因素方差分析

问题的提出



引例1 在饲料养鸡增肥的研究中,某饲料研究所提出三

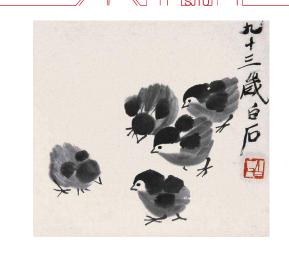
种饲料配方:



 A_1 是以鱼粉为主的饲料,

 A_2 是以槐树粉为主的饲料,

 A_3 是以苜蓿粉为主的饲料。



为比较三种饲料的效果,特选 24 只相似的雏鸡随机均分为三组,每组各喂一种饲料,60天后观察它们的重量。试验结果如下表所示:



表5.1 鸡饲料试验数据

1 1

饲料A	鸡重(克)							
A_1	1073	1009	1060	1001	1002	1012	1009	1028
A_2	1107	1092	990	1109	1090	1074	1122	1001
A_3	1093	1029	1080	1021	1022	1032	1029	1048



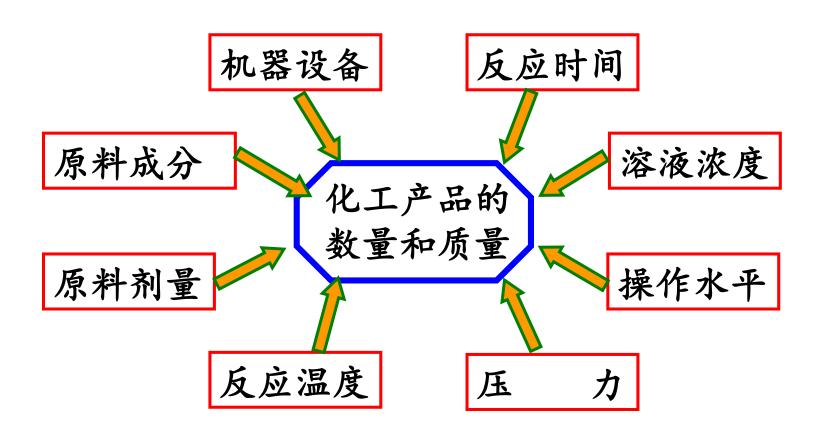
本引例中,我们要比较的是三种饲料对鸡的增肥作用是否相同。为此,我们把饲料称为因子,记为A,而三种不同的配方称为因子A的三个水平,记为 A_1 , A_2 , A_3 , 使用配方 A_i 下第j 只鸡60天后的重量用 y_{ii} 表示,i=1, 2, 3, j=1, 2,..., 8。

我们的目的是比较三种饲料配方下鸡的平均重量是否相等,为此,需要做一些基本假定,把所研究的问题归结为一个统计问题,然后用方差分析的方法进行解决。



主要概念

常常通过试验了解。。。

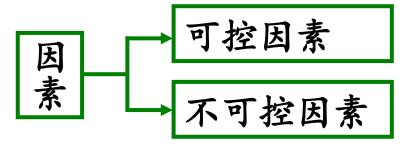




主要概念:

试验指标——试验中要考察的指标.

因素——影响试验指标的条件.



水 平——因素所处的状态.

单因素试验——在一项试验中只有一个因素改变.

多因素试验——在一项试验中有多个因素在改变.



例2 设有三台机器,用来生产规格相同的铝合金薄板.取样,测量薄板的厚度精确至千分之一厘米.得结果如下表所示.

表5.2 铝合金板的厚度

		· V / / > C
机器I	机器工	机器皿
0.236	0.257	0.258
0.238	0.253	0.264
0.248	0.255	0.259
0.245	0.254	0.267
0.243	0.261	0.262

试验指标:薄板的厚度

因素: 机器

水平: 不同的三台机器是因素的三个不同的水平 假定除机器这一因素外, 其他条件相同, 属于单因素 试验.



例3 一火箭用四种燃料,三种推进器作射程试验. 每种燃料与每种推进器的组合各发射火箭两次,得 射程如下(以海里计).

表5.3 火箭的射程

推进器(是	推进器(<i>B</i>)		B_2	B_3
	A_1	58.2	56.2	65.3
	7 1	52.6	41.2	60.8
	A_2	49.1	54.1	51.6
燃料(A)	\mathbf{A}_2	42.8	50.5	48.4
/3m/4-1 (2 ·)	Λ	60.1	70.9	39.2
	A_3	58.3	73.2	40.7
	A_4	75.8	58.2	48.7
	~ 4	71.5	51.0	41.4



表5.2 火箭的射程

推进器(<i>B</i>)		<i>B</i> ₁	B_2	B_3
	A ₁	58. 2 52. 6	56. 2 41. 2	65. 3 60. 8
	A_2	49. 1	54. 1	51. 6
燃料(A)	_	42. 8 60. 1	50. 5 70. 9	48. 4 39. 2
	A ₃	58. 3	73. 2	40. 7
	A ₄	75. 8 71. 5	58. 2 51. 0	48. 7 41. 4

试验指标:射程

因素:推进器和燃料

水平: 推进器有3个, 燃料有4个.

两因素试验





例2



表5.2 铝合金板的厚度

机器 I	机器工	机器皿
0.236	0.257	0.258
0.238	0.253	0.264
0.248	0.255	0.259
0.245	0.254	0.267
0.243	0.261	0.262

试验目的:考察各台机器所生产的薄板的厚度 有无显著的差异.即考察机器这一因素对厚度有无 显著的影响.



例2

表5.2 铝合金板的厚度

机器	机器	机器
0. 236	0. 257	0. 258
0. 238	0. 253	0. 264
0. 248	0. 255	0. 259
0. 245	0. 254	0. 267
0. 243	0. 261	0. 262

问题分析: 在每一个水平下进行独立试验, 结果是一个随机变量, 将数据看成是来自三个总体的样本值. 设总体均值分别为 μ_1, μ_2, μ_3 .

检验假设 $H_0: \mu_1 = \mu_2 = \mu_3$,

 $H_1: \mu_1, \mu_2, \mu_3$ 不全相等.



检验假设 $H_0: \mu_1 = \mu_2 = \mu_3$,

 $H_1: \mu_1, \mu_2, \mu_3$ 不全相等.

进一步假设各总体均为正态总体,且各总体的方差相等,但参数均未知.

问 题——检验具有方差齐性的多个正态总体均值是否相等.

解决方法——方差分析法,一种统计方法.



一.数学模型

设因素A有s个水平 A_1, A_2, \cdots, A_s ,在水平 A_j ($j = 1,2,\cdots,s$)下,进行 n_j ($n_j \geq 2$)次独立试验,得到如下表的结果. $\frac{1}{8}$ 5.4

观察结果 水平	A_1	A_2	• • •	A_s
	X_{11}	X_{12}	• • •	X_{1s}
	X_{21}	X_{22}	• • •	X_{2s}
	•	•		:
	$X_{n_1^{-1}}$	$X_{n_2^2}$	• • •	$X_{n_s s}$
样本总和	$T_{ullet 1}$	$T_{ullet 2}$	• • •	T_{ullet_S}
样本均值	$\overline{X}_{ullet 1}$	$\overline{X}_{ullet 2}$	• • •	$\overline{X}_{ullet_{\mathbf{x}}}$
总体均值	μ_1	μ_2	• • •	μ_{s}



假设

(1)各个水平 A_j ($j = 1, 2, \dots, s$)下的样本 X_{1j}, X_{2j} ,

 $\dots, X_{n_j j}$ 来自具有相同方差 σ^2 ,均值分别为 $\mu_j (j=1,$

 $2, \dots, s$)的正态总体 $N(\mu_j, \sigma^2), \mu_j$ 与 σ^2 均未知;

(2)不同水平 A_j 下的样本之间相互独立.

$$X_{ij} \sim N(\mu_j, \sigma^2), \qquad X_{ij} - \mu_j \sim N(0, \sigma^2).$$

记 $X_{ij} - \mu_j = \varepsilon_{ij}$,表示随机误差,那么 X_{ij} 可写成



单因素试验方差分析的数学模型:

$$X_{ij} = \mu_j + \varepsilon_{ij}$$
,
 $\varepsilon_{ij} \sim N(0, \sigma^2)$, 各 ε_{ij} 独立,
 $i = 1, 2, \dots, n_j$, $j = 1, 2, \dots, s$,
 $\mu_j = \sigma^2$ 均未知

二. 方差分析

需要解决的问题:

(1) 检验假设

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_s$$
,

$$H_1: \mu_1, \mu_2, \dots, \mu_s$$
不全相等.

(2)估计未知参数 $\mu_1, \mu_2, \dots, \mu_s, \sigma^2$.



1、数学模型的等价形式

记
$$n = \sum_{j=1}^{s} n_j, \mu = \frac{1}{n} \sum_{j=1}^{s} n_j \mu_j.$$

总平均

水平 A_j 的效 应,表示水平 A_j 下的总体 平均值与总 平均的差异,

$$\delta_j = \mu_j - \mu_j j = 1,2,\dots,s.$$

$$n_1\delta_1 + n_2\delta_2 + \cdots + n_s\delta_s = 0$$



原数学模型

$$X_{ij} = \mu_j + \varepsilon_{ij}$$
,
 $\varepsilon_{ij} \sim N(0, \sigma^2)$, 各 ε_{ij} 独立,
 $i = 1, 2, \dots, n_j, j = 1, 2, \dots, s$,
 $\mu_j = 1, 2, \dots, s$

$$X_{ij} = \mu + \delta_j + \varepsilon_{ij},$$
 $\varepsilon_{ij} \sim N(0, \sigma^2),$ 各 ε_{ij} 独立,
 $i = 1, 2, \dots, n_j, j = 1, 2, \dots, s,$

$$\sum_{j=1}^{s} n_j \delta_j = 0$$

改写为





检验假设

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_s,$$

$$H_1: \mu_1, \mu_2, \dots, \mu_s$$
不全相等.

等价于

检验假设
$$H_0: \delta_1 = \delta_2 = \cdots = \delta_s = 0,$$
 $H_1: \delta_1, \delta_2, \cdots, \delta_s$ 不全为零.



2、平方和的分解



$$S_T = \sum_{j=1}^{s} \sum_{i=1}^{n_j} (X_{ij} - \overline{X})^2 - \text{总偏差平方和(总变差)}$$

$$\overline{X}_{\bullet j} = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{ij} - \text{水} + A_j \text{ Time of } A_$$



$$S_T = \sum_{j=1}^{s} \sum_{i=1}^{n_j} (X_{ij} - \overline{X})^2$$

$$= \sum_{i=1}^{s} \sum_{j=1}^{n_j} [(X_{ij} - \overline{X}_{\bullet j}) + (\overline{X}_{\bullet j} - \overline{X})]^2$$

$$= \sum_{j=1}^{s} \sum_{i=1}^{n_j} (X_{ij} - \overline{X}_{\bullet j})^2 + \sum_{j=1}^{s} \sum_{i=1}^{n_j} (\overline{X}_{\bullet j} - \overline{X})^2$$

$$+2\sum_{j=1}^{s}\sum_{i=1}^{n_{j}}(X_{ij}-\overline{X}_{\bullet j})(\overline{X}_{\bullet j}-\overline{X})$$

= 0



$$S_T = \sum_{j=1}^{s} \sum_{i=1}^{n_j} (X_{ij} - \overline{X}_{\bullet j})^2 + \sum_{j=1}^{s} \sum_{i=1}^{n_j} (\overline{X}_{\bullet j} - \overline{X})^2$$

$$=S_E+S_A$$

$$S_E = \sum_{i=1}^{s} \sum_{i=1}^{n_j} (X_{ij} - \overline{X}_{\bullet j})^2$$
 —误差(组内)平方和

$$S_A = \sum_{j=1}^{s} \sum_{i=1}^{n_j} (\overline{X}_{\bullet j} - \overline{X})^2 = \sum_{j=1}^{s} n_j (\overline{X}_{\bullet j} - \overline{X})^2$$

$$=\sum_{j=1}^{3}n_{j}\overline{X}_{\bullet j}^{2}-n\overline{X}^{2}$$
 —因素A的效应(组间)
平方和



3、 S_E , S_A 的统计性质

$$egin{align*} S_E &= \sum_{j=1}^s \sum_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2 \ &= \sum_{i=1}^{n_1} (X_{i1} - ar{X}_{ullet 1})^2 + \dots + \sum_{i=1}^{n_s} (X_{is} - ar{X}_{ullet s})^2 \ &\sum_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ \sum_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=1}^{n_j} (X_{ij} - ar{X}_{ullet j})^2
ot length \ & \int_{i=$$



又由于各 X_{ij} 独立,所以由 χ^2 分布的可加性知

$$S_E/\sigma^2 \sim \chi^2(\sum_{j=1}^s (n_j-1))$$

即

$$S_E/\sigma^2 \sim \chi^2(n-s)$$
,其中 $n = \sum_{j=1}^s n_j$.

根据 χ^2 分布的性质可以得到

$$S_E$$
的自由度为 $n-s$;

$$E(S_E) = (n-s)\sigma^2.$$





在
$$H_0: \mu_1 = \mu_2 = \dots = \mu_s$$
成立时,

$$S_T \sim ?$$

$$\frac{S_T}{\sigma^2} \sim ?$$



$$S_{A} = \sum_{j=1}^{s} n_{j} \overline{X}_{\bullet j}^{2} - n \overline{X}^{2}$$

$$E(S_{A}) = E[\sum_{j=1}^{s} n_{j} \overline{X}_{\bullet j}^{2} - n \overline{X}^{2}]$$

$$= \sum_{j=1}^{s} n_{j} E(\overline{X}_{\bullet j}^{2}) - n E(\overline{X}^{2})$$

$$= \sum_{j=1}^{s} n_{j} [\frac{\sigma^{2}}{n_{j}} + (\mu + \delta_{j})^{2}] - n [\frac{\sigma^{2}}{n} + \mu^{2}]$$

$$= (s - 1)\sigma^{2} + 2\mu \sum_{j=1}^{s} n_{j} \delta_{j} + n\mu^{2} + \sum_{j=1}^{s} n_{j} \delta_{j}^{2} - n\mu^{2}$$

$$= (s - 1)\sigma^{2} + \sum_{j=1}^{s} n_{j} \delta_{j}^{2}$$

 H_0 为真时, S_A 与 S_E 独立, $S_A/\sigma^2 \sim \chi^2(s-1)$

(柯赫伦分解定理)



4、假设检验问题的拒绝域

$$\boldsymbol{H}_0: \boldsymbol{\delta}_1 = \boldsymbol{\delta}_2 = \cdots = \boldsymbol{\delta}_s = \boldsymbol{0},$$

检验假设 $H_0: \delta_1 = \delta_2 = \cdots = \delta_s = 0,$ $H_1: \delta_1, \delta_2, \cdots, \delta_s$ 不全为零.

$$H_0$$
为真时, $S_A/\sigma^2 \sim \chi^2(s-1)$

$$E(\frac{S_A}{s-1}) = \sigma^2$$
,即 $\frac{S_A}{s-1}$ 是 σ^2 的无偏估计.

$$H_1$$
为真时, $\sum_{j=1}^{3} n_j \delta_j^2 > 0$,

$$E(\frac{S_A}{s-1}) = \sigma^2 + \frac{1}{s-1} \sum_{j=1}^{s} n_j \delta_j^2 > \sigma^2$$



由
$$E(S_E) = (n-s)\sigma^2$$
 故 $E(\frac{S_E}{n-s}) = \sigma^2$

即不管 H_0 是否为真, $S_E/(n-s)$ 都是 σ^2 的无偏估计.

$$F = \frac{S_A/(s-1)}{S_E/(n-s)}$$

- (1)分子和分母相互独立;
- (2) 分母 S_E 的数学期望始终是 σ^2 ;
- $(3)H_0$ 为真时,分子的期望为 σ^2, H_0 不真时,分子取值有偏大的趋势.

拒绝域形如
$$F = \frac{S_A/(s-1)}{S_E/(n-s)} \ge k$$



因为 H_0 为真时, $S_A/\sigma^2 \sim \chi^2(s-1)$, $S_E/\sigma^2 \sim \chi^2(n-s)$

从而 H_0 为真时,

$$\frac{S_A/(s-1)}{S_E/(n-s)} = \frac{S_A/\sigma^2}{(s-1)} / \frac{S_E/\sigma^2}{(n-s)} \sim F(s-1,n-s).$$

检验假设

$$\boldsymbol{H}_0: \boldsymbol{\delta}_1 = \boldsymbol{\delta}_2 = \cdots = \boldsymbol{\delta}_s = \boldsymbol{0},$$

$$H_1: \delta_1, \delta_2, \dots, \delta_s$$
不全为零.

拒绝域为

$$F = \frac{S_A/(s-1)}{S_E/(n-s)} \ge F_{1-\alpha}(s-1, n-s)$$



单因素试验方差分析表5.5

方差	e 来源	平方和	自由度	均方	FE
因	素A	S_A	s – 1	$\overline{S}_A = \frac{S_A}{s-1}$	$F = \overline{S}_A / \overline{S}_E$
误	差	S_E	n-s	$\overline{S}_E = \frac{S_E}{n-s}$	
总	和	S_T	n-1		

记
$$T_{\bullet j} = \sum_{i=1}^{n_j} X_{ij}, j = 1, \dots, s, T_{\bullet \bullet} = \sum_{j=1}^{s} \sum_{i=1}^{n_j} X_{ij},$$

$$S_T = \sum_{j=1}^{s} \sum_{i=1}^{n_j} X_{ij}^2 - \frac{T_{\bullet \bullet}^2}{n}, \quad S_A = \sum_{j=1}^{s} \frac{T_{\bullet j}^2}{n_j} - \frac{T_{\bullet \bullet}^2}{n},$$

$$S_E = S_T - S_A.$$



例1. 采用例1的数据,将原始数据减去1000,列表给出计算过程:

例1的计算表

水平		数据(原始数据-1000)						T_i	T_i^2	$\sum_{j=1}^m y_{ij}^2$	
A_1	73	9	60	1	2	12	9	28	194	37636	10024
A_2	107	92	-10	109	90	74	122	1	585	342225	60355
A_3	93	29	80	21	22	32	29	48	354	125316	20984
									1133	505177	91363





利用**单因素试验方差分析表5.5**,可算得各偏差平方和为:

$$S_T = 91363 - \frac{1133^2}{24} = 37876.0417,$$
 $f_T = 24 - 1 = 23$
 $S_A = \frac{505177}{8} - \frac{1133^2}{24} = 9660.0833,$ $f_A = 3 - 1 - 2$
 $S_e = S_T - S_A = 37876.0417 - 9660.0833 = 28215.9584,$ $f_e = 3(8 - 1) = 21$

把上述诸平方和及其自由度填入方差分析表



例1的方差分析表5.6

来源	平方和	自由度	均方和	F比
因子	9660.0833	2	4830.0417	3.5948
误差	28215.9584	21	1343.6171	
总和	37876.0417	23		

若取 α =0.05,则 $F_{0.05}$ (2,21)=3.47,由于F=3.5948>3.47,故认为因子A(饲料)是显著的,即三种饲料对鸡的增肥作用有明显的差别。



续例2 设有三台机器,用来生产规格相同的铝合金薄板.取样,测量薄板的厚度精确至千分之一厘米.得结果如下表所示.

表5.7 铝合金板的厚度

		<u> </u>
机器	机器	机器
0. 236	0. 257	0. 258
0. 238	0. 253	0. 264
0. 248	0. 255	0. 259
0. 245	0. 254	0. 267
0. 243	0. 261	0. 262

取 $\alpha = 0.05$, 检验假设

$$H_0: \mu_1 = \mu_2 = \mu_3, H_1: \mu_1, \mu_2, \mu_3$$
不全相等.

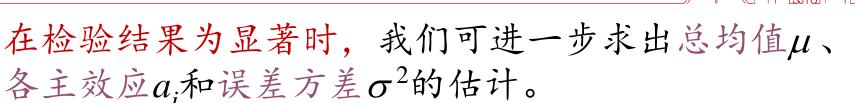


方差来源		平方和	自由度	均方	F 比
因	素A	0.00105333	2	0.00052667	32.92
误	差	0.000192	12	0.000016	
总	和	0.00124533	14		

 $F = 32.92 > F_{0.95}(2,12) = 3.89$. 在水平0.05下拒绝 H_0 . 各机器生产的薄板厚度有显著差异. 即机器作为一个因素对试验结果的影响是显著的。



5、未知参数的估计



由模型(5.1.1)知诸 X_{ij} 相互独立,且 X_{ij} ~ $N(\mu + a_i, \sigma^2)$,因此,可使用极大似然方法求出一般平均 μ 、各主效应 a_i 和误差方差 σ^2 的估计,由极大似然估计的不变性,各水平均值 μ_i 的极大似然估计。

由于 S_E 不是 σ^2 的无偏估计,可修偏 $\hat{\sigma}^2 = S_E/(n-s)$



未知参数的点估计

$$E(S_E/(n-s)) = \sigma^2$$
 $\hat{\sigma}^2 = S_E/(n-s)$ $E(\overline{X}) = \mu$ $\hat{\mu} = \overline{X}$ 无偏估计 $E(\overline{X}_{\bullet j}) = \mu_j, j = 1,2,\cdots,s$ $\hat{\mu}_j = \overline{X}_{\bullet j}$ $\delta_j = \mu_j - \mu, j = 1,2,\cdots,s$ $\hat{\delta}_j = \overline{X}_{\bullet j} - \overline{X}$ 若拒绝 H_0 ,需对两总体 $N(\mu_j,\sigma^2),N(\mu_k,\sigma^2)$ 的均值差 $\mu_j - \mu_k = \delta_j - \delta_k$ 作出区间估计. $E(\overline{X}_{\bullet j} - \overline{X}_{\bullet k}) = \mu_j - \mu_k$, $D(\overline{X}_{\bullet j} - \overline{X}_{\bullet k}) = \sigma^2(\frac{1}{n_j} + \frac{1}{n_k})$



$$\overline{X}_{\bullet j} - \overline{X}_{\bullet k}$$
与 $\hat{\sigma}^2 = S_E/(n-s)$ 独立,

数
$$\frac{(\overline{X}_{\bullet j} - \overline{X}_{\bullet k}) - (\mu_j - \mu_k)}{\sqrt{\overline{S}_E(\frac{1}{n_j} + \frac{1}{n_k})}}$$

$$=\frac{(\overline{X}_{\bullet j}-\overline{X}_{\bullet k})-(\mu_{j}-\mu_{k})}{\sigma\sqrt{1/n_{j}+1/n_{k}}}\bigg/\sqrt{\frac{\overline{S}_{E}}{\sigma^{2}}}\bigg/(n-s) \sim t(n-s).$$

均值差 $\mu_j - \mu_k = \delta_j - \delta_k$ 的置信水平为 $1 - \alpha$ 的置信

区间为
$$\left(\overline{X}_{\bullet j} - \overline{X}_{\bullet k} \pm t_{1-\frac{\alpha}{2}}(n-s)\sqrt{\overline{S}_{E}(\frac{1}{n_{j}} + \frac{1}{n_{k}})}\right)$$



例4 求例1中的未知参数 σ^2 , μ_j , δ_j (j=1,2,3)的点估计及均值差的置信水平为0.95的置信区间。

解
$$\hat{\sigma}^2 = S_E/(n-s) = 0.000016$$
,
 $\hat{\mu}_1 = \overline{x}_{\bullet 1} = 0.242$, $\hat{\mu}_2 = \overline{x}_{\bullet 2} = 0.256$, $\hat{\mu}_3 = \overline{x}_{\bullet 3} = 0.262$.
 $\hat{\mu} = \overline{x} = 0.253$, $\hat{\delta}_1 = \overline{x}_{\bullet 1} - \overline{x} = -0.011$,
 $\hat{\delta}_2 = \overline{x}_{\bullet 2} - \overline{x} = 0.003$, $\hat{\delta}_3 = \overline{x}_{\bullet 3} - \overline{x} = 0.009$.

$$t_{0.975}(n-s) = t_{0.975}(21) = 2.1788,$$

$$t_{0.975}(21)\sqrt{\bar{S}_E(\frac{1}{n_j}+\frac{1}{n_k})}=0.006,$$





$$(0.242 - 0.256 \pm 0.006) = (-0.020, -0.008)$$

 $\mu_1 - \mu_3$ 的置信水平为0.95的置信区间为

$$(0.242 - 0.262 \pm 0.006) = (-0.026, -0.014)$$

 $\mu_2 - \mu_3$ 的置信水平为0.95的置信区间为

$$(0.256 - 0.262 \pm 0.006) = (-0.012,0)$$





例1. 继续例1, 此处我们给出诸水平均值的估计。

因子A的三个水平均值的估计分别为

$$\widehat{\mu}_1 = 1000 + \frac{194}{8} = 1024.25,$$

$$\widehat{\mu}_2 = 1000 + \frac{585}{8} = 1073.125,$$

$$\widehat{\mu}_3 = 1000 + \frac{354}{8} = 1044.25,$$

从点估计来看,水平2(以槐树粉为主的饲料) 是最优的。



误差方差的无偏估计为

$$\hat{\sigma}^2 = MS_E = 1343.6171$$

利用(5.1.23)可以给出诸水平均值的置信区间。此处, $\hat{\sigma} = \sqrt{1343.6171} = 36.6554$,若取 $\alpha = 0.05$,则

 $t_{\alpha/2}(n-s) = t_{0.05}(21) = 2.0796, \hat{\sigma}t_{0.025}(21)/\sqrt{8} = 26.9509$,于是三个水平均值的0.95置信区间分别为

 μ_1 :1024.25 ± 26.9509 = [997.2891, 1051.2109],

 μ_2 : $1073.125 \pm 26.9509 = [1046.1741, 1100.0759],$

 $\mu_3 : 1044.\overline{25 \pm 26.9509} = [1017.2891, 1071.2109].$





在单因子试验的数据分析中可得到如下三个结果:

- > 因子是否显著;
- \rightarrow 试验的误差方差 σ^2 的估计;
- \rightarrow 诸水平均值 μ_i 的点估计与区间估计。

在因子A显著时,通常只需对较优的水平均值作参数估计,在因子A不显著场合,参数估计无需分不同水平进行。





某食品公司对一种食品设计了四种新包装。 为考察哪种包装最受顾客欢迎,选了10个地段繁 华程度相似、规模相近的商店做试验, 其中二种 包装各指定两个商店销售, 另二个包装各指定三 个商店销售。在试验期内各店货架排放的位置、 空间都相同, 营业员的促销方法也基本相同, 经 过一段时间,记录其销售量数据,列于表7左半 边. 其相应的计算结果列于右侧。



表5.7 销售量数据及计算表

包装类型	销售量			m_{i}	T_{i}	T_i^2/m_i	$\sum_{j=1}^{m_i} \boldsymbol{\mathcal{Y}}_{ij}^2$
A_1	12	18		2	30	450	468
A_2	14	12	13	3	39	507	509
A_3	19	17	21	3	57	1083	1091
A_4	24	30		2	54	1458	1476
和				n=10	<i>T</i> =180		



由此可求得各类偏差平方和如下

$$\left(\frac{T^2}{n} = \frac{180^2}{10} = 3240\right)$$

$$S_T = 3544 - 3240 = 304, \quad f_T = 10 - 1 = 9$$

 $S_A = 3498 - 3240 = 258, \quad f_A = 4 - 1 = 3$
 $S_e = 304 - 258 = 46, \quad f_e = 10 - 4 = 6$

方差分析表如表5.8所示。

若取 α =0.01, 查表得 $F_{0.01}$ (3,6)=9.78, 由于F=11.22>9.78, 故我们可认为各水平间有显著差异。



表5.8 例5的方差分析表

来源	平方和	自由度	均方和	F比
因子A	258	3	86	11.22
误差e	46	6	7.67	
总和T	304	9		



由于因子显著,我们还可以给出诸水平均值的估计。因子A的四个水平均值的估计分别为

$$\hat{\mu}_1 = 30/2 = 15, \quad \hat{\mu}_2 = 39/3 = 13,$$

 $\hat{\mu}_3 = 57/3 = 19, \quad \hat{\mu}_4 = 54/2 = 27,$

由此可见, 第四种包装方式效果最好。 误差方差的无偏估计为

$$\widehat{\sigma^2} = MS_E = 7.67$$





进一步,也可以给出诸水平均值的置信区间。 此处, $\hat{\sigma} = \sqrt{7.67} = 2.7695$,若取 $\alpha = 0.05$,则 $t_{\alpha/2}(n-s) = t_{0.05}(6) = 2.4469$ $\hat{\sigma}t_{0.975}(6) = 6.7767$, 于是效果较好的第三和第四个水平均值的 0.95置信区间分别为

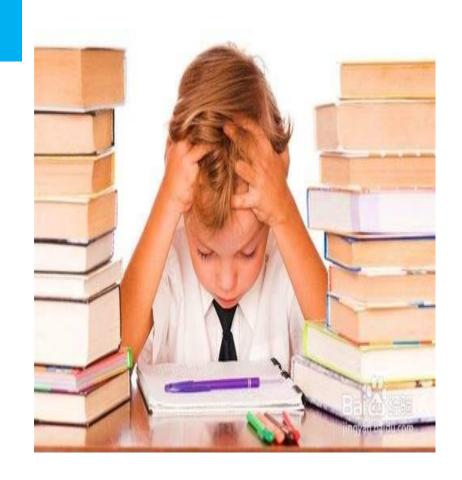
 μ_3 : 19 ± 6.7767/ $\sqrt{3}$ = [15.0875, 22.9125], μ_4 : 27 ± 6.7767/ $\sqrt{2}$ = [22.2081, 31.7919].





第12次作业:

- 孙 p.145 习题五
- **1**, 4, 10.



谢谢!

