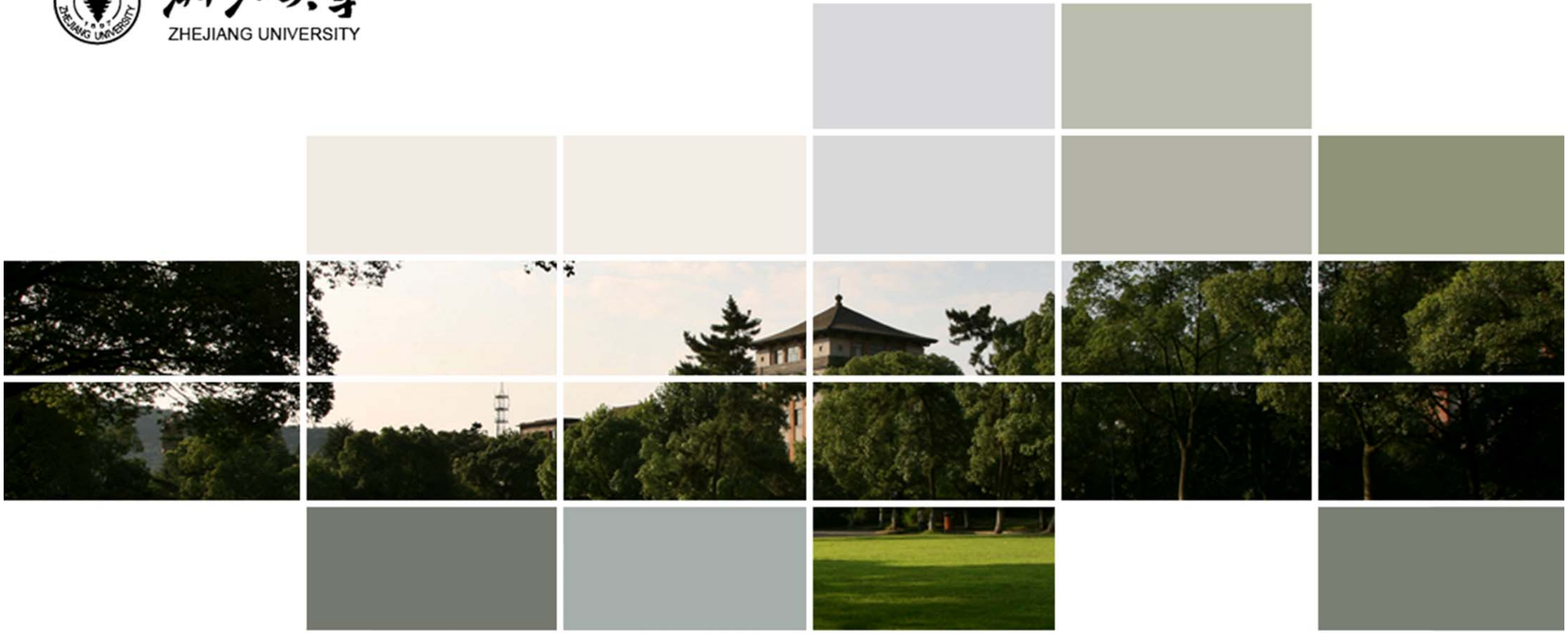
浙江大學 ZHEJIANG UNIVERSITY



第62讲　单因素方差分析
（参数估计及均值的多重比较）

在第61讲建立了单因素方差分析模型：

$$
\begin{cases}
X_{ij} = \mu_i + \varepsilon_{ij}, \\
\varepsilon_{ij} \sim N(0, \sigma^2), \text{各} \varepsilon_{ij} \text{独立}, \\
j = 1, 2, \cdots, n_i, \quad i = 1, 2, \cdots, r.
\end{cases}
$$

解决了问题一：检验假设

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

除此之外，还有两个问题需要解决.

问题二：未知参数的估计

单因素试验方差分析模型中的未知参数有：

$\sigma^2, \mu_i, i=1,\ldots,r.$其无偏估计为：

（1） σ^2 的估计 $\hat{\sigma}^2 = \dfrac{S_E}{n-r} = MS_E;$

（2） μ_i 的估计 $\hat{\mu}_i = \bar{X}_{i\bullet}, i=1,2,\ldots,r.$

问题三：在方差分析中，

如果拒绝原假设H_0，只能说明均值不全相等。

那么，它们中有没有部分是相等的？

比较$N(\mu_i,\sigma^2)$和$N(\mu_j,\sigma^2)$的差异的两个方法：

（1）作 $\mu_i-\mu_j(i\neq j)$的区间估计；

（2）作 $H_0:\mu_i=\mu_j, H_1:\mu_i\neq\mu_j$ 的假设检验。

（1）置信区间

因为$E(\bar{X}_{i\bullet}-\bar{X}_{j\bullet})=\mu_i-\mu_j$, $D(\bar{X}_{i\bullet}-\bar{X}_{j\bullet})=\sigma^2\left(\dfrac{1}{n_i}+\dfrac{1}{n_j}\right)$,

且$\bar{X}_{i\bullet}-\bar{X}_{j\bullet}$与$\hat{\sigma}^2=MS_E$相互独立。

$$故\ \frac{(\bar{X}_{i\bullet}-\bar{X}_{j\bullet})-(\mu_i-\mu_j)}{\sqrt{MS_E(1/n_i+1/n_j)}}=\frac{(\bar{X}_{i\bullet}-\bar{X}_{j\bullet})-(\mu_i-\mu_j)}{\sigma\sqrt{(1/n_i+1/n_j)}}\Bigg/\sqrt{\frac{S_E}{\sigma^2(n-r)}}$$

$$\sim t(n-r)$$

$$\left(\bar{X}_{i\bullet}-\bar{X}_{j\bullet}\pm t_{\alpha/2}(n-r)\sqrt{MS_E(1/n_i+1/n_j)}\right)$$

得$(\mu_i-\mu_j)$的水平为$1-\alpha$的置信区间

（2）假设检验

$$H_0:\mu_i=\mu_j, H_1:\mu_i\neq\mu_j, i\neq j$$

检验统计量 $t_{ij}=\dfrac{\bar{X}_{i\bullet}-\bar{X}_{j\bullet}}{\sqrt{MS_E(1/n_i+1/n_j)}}$,

当H_0成立时，$t_{ij}\sim t(n-r)$,

拒绝域$W=\{|t_{ij}|\geq t_{\alpha/2}(n-r)\}$.

例1. 为了比较三种不同类型日光灯管的寿命(小时)，现从每种类型日光灯管中抽取 8个，总共 24 个日光灯管进行老化试验，下页数据是经老化试验后测算得出的各个日光灯管的寿命(小时)。假设每种类型日光灯管寿命均服从正态分布$N(\mu_i,\sigma^2),(i=1,2,3)$.

类型	寿命(小时)
类型I	5290 6210 5740 5000 5930 6120 6080 5310
类型II	5840 5500 5980 6250 6470 5990 5470 5840
类型III	7130 6660 6340 6470 7580 6560 7290 6730

(1) 试判断三种不同类型日光灯管的平均寿命是否存在差异. (α=0. 05)

(2) 估计未知参数 $\sigma^2, \mu_i (i=1,2,3)$.

(3) 若存在差异，请进行多重检验. (α=0. 05)

这里因素是日光灯管类型，共有3个水平，这是一个单因素方差分析问题，要检验的假设是“不同类型日光灯管平均寿命是否存在显著差别”。

本例的Excel计算见实验26.

解：(1) 检验假设 $H_0: \mu_1 = \mu_2 = \mu_3$,

$H_1: \mu_1, \mu_2, \mu_3$不全相等.

单因素方差分析

SUMMARY				
组	观测数	求和	平均	方差
行 1	8	45680	5710.0	206400.0
行 2	8	47340	5917.5	115935.7
行 3	8	54760	6845.0	191971.4

方差分析表

差异源	SS	df	MS	F	P-value	F crit
组间	5844100	2	2922050	17.04458	0.00004	3.4668
组内	3600150	21	171435.7			
总计	9444250	23				

>3.4668，拒绝H_0

<0.05，拒绝H_0

(2) σ^2的估计$\hat{\sigma}^2 = MS_E = 171435.7;$

μ_i的估计分别为： 5710, 5917.5, 6845.

(3) 检验假设$H_{01}: \mu_1 = \mu_2, H_{11}: \mu_1 \neq \mu_2,$

$$检验统计量 t_{12} = \frac{\bar{X}_{1\bullet} - \bar{X}_{2\bullet}}{\sqrt{MS_E (1/n_1 + 1/n_2)}}$$

计算得$t_{12} = -1.002,$查表得$t_{0.025}(21) = 2.0796,$

$|t_{12}| < t_{0.025}(21),$不拒绝$H_{01}.$

(3) 检验假设 $H_{02}: \mu_1 = \mu_3, H_{12}: \mu_1 \neq \mu_3,$

检验统计量 $t_{13} = \dfrac{\bar{X}_{1\bullet} - \bar{X}_{3\bullet}}{\sqrt{MS_E(1/n_1 + 1/n_3)}}$

计算得 $t_{13} = -5.482$, 查表得 $t_{0.025}(21) = 2.0796$,

$|t_{13}| > t_{0.025}(21)$, 拒绝 H_{02}.

(3) 检验假设$H_{03}:\mu_2=\mu_3, H_{13}:\mu_2\neq\mu_3,$

$$\text{检验统计量}\ t_{23}=\frac{\bar{X}_{2\bullet}-\bar{X}_{3\bullet}}{\sqrt{MS_E(1/n_2+1/n_3)}}$$

计算得$t_{23}=-4.480,\quad t_{0.025}(21)=2.0796,$

$|t_{23}|>t_{0.025}(21),\quad$ 拒绝$H_{03}.$

方差分析的前提

进行方差分析必须具备三个基本的条件:

（1） 独立性. 数据是来自 r 个独立总体的简单随机样本.

（2） 正态性. r 个独立总体均为正态总体.

（3） 方差齐性. r 个正态总体的方差是相同的， 即满足假设

$$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_r^2.$$

- 方差分析和其它统计推断一样，样本的独立性对方差分析是非常重要的，
- 在实际应用中会经常遇到非随机样本的情况，这时使用方差分析得出的结论不可靠.
- 因此，在安排试验或采集数据的过程中，一定要注意样本的独立性问题.

- 在实际中，几乎没有一个总体真正服从正态分布，而方差分析却依赖于正态性假设．不过根据经验，方差分析F检验对正态性的假设并不是非常敏感，
- 即，实际所得到的数据，若没有异常值和偏性，或者说，数据显示的分布比较对称的话，即使样本容量比较小(如每个水平下的样本容量仅为5左右)，方差分析的结果仍是值得依赖的.

- 方差齐性对于方差分析是非常重要的，方差齐性的检验可采用如下的经验准则：
- 当最大样本标准差不超过最小样本标准差的两倍时，方差分析F检验结果近似正确.

例2　判断例1的方差齐性是否成立。

解：由例1的计算结果，最大标准差为454.31，最小标准差为340.49，所以可以认为方差相等。