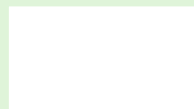


第四章 多个样本均数比较的方差分析

analysis of variance
(ANOVA)

 琤 教授

中南大学公共卫生学院流行病与卫生统计学系

例4-2 某医生为了研究一种降血脂新药的临床疗效，按统一纳入标准选择**120**名高血脂患者，采用完全随机设计方法将患者等分为**4**组，进行双盲试验。6 测得**低密度脂蛋白**作为试验结果，见表4-3。问**4**个处理组患者的低密度脂蛋白含量总体均数有无差别？

讲述内容

- 第一节 方差分析的基本思想及其应用条件
- 第二节 完全随机设计资料的方差分析
- 第三节 随机区组设计资料的方差分析
- 第四节 拉丁方设计资料的方差分析
- 第五节 两阶段交叉设计资料的方差分析
- 第六节 多个样本均数间的多重比较
- 第七节 多样本方差比较的Bartlett检验
和Levene检验

第一节

方差分析的基本思想及 其应用条件

目的：推断多个总体均数是否有差别。

↓
也可用于两个

方法：方差分析，即多个样本均数比较的 F 检验。(由英国统计学家Fisher首创)

基本思想：根据试验设计的类型，将全部测量值总的离均差平方和及其自由度分解为两个或多个部分，除随机误差作用外，每个部分的变异可由某个因素的作用(或某几个因素的交互作用)加以解释。(P48)

应用条件(P49):

总体——正态且方差相等

$$N_1(\mu_1, \sigma^2), N_2(\mu_2, \sigma^2)$$

样本——独立、随机

设计类型:

完全随机设计资料的方差分析

随机区组设计资料的方差分析

拉丁方设计资料的方差分析

两阶段交叉设计资料的方差分析

完全随机设计资料的方差分析的基本思想

表 4-1 g 个处理组的试验结果

处理分组	测量值						统计量		
1 水平	X_{11}	X_{12}	$\cdots$	X_{1j}	$\cdots$	X_{1n_1}	n_1	$\bar{X}_1$	S_1
2 水平	X_{21}	X_{22}	$\cdots$	X_{2j}	$\cdots$	X_{2n_2}	n_2	$\bar{X}_2$	S_2
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
g 水平	X_{g1}	X_{g2}	$\cdots$	X_{gj}	$\cdots$	X_{gn_g}	n_g	$\bar{X}_g$	S_g
合计	X						N	$\bar{X}$	S

第*i*个处理组第*j*个观察结果

方差分析的**目的**就是在 $H_0: \mu_1 = \mu_2 = \dots = \mu_g$ 成立的条件下，通过分析各处理组均数($\bar{x}_i$)之间的差别大小，推断 g 个总体均数间有无差别，从而说明处理因素的效果是否存在。

变异（离均差平方和）分解：

➤ 离均差平方和 (sum of squares of deviations from mean, SS) :

对各组和求和

$$SS = \sum_{i=1}^g \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2 = \sum_{i,j}^N (X - \bar{X})^2$$

第i组内求和

$$= (N - 1)S^2$$

结合表4-1

1.总变异: 全部观测值大小不同, 这种变异称为总变异。反映所有测量值之间总的变异程度, 计算公式为

$$SS_{\text{总}} = \sum_{i=1}^g \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2 = \sum_{i=1}^g \sum_{j=1}^{n_i}$$

$$= \sum_{i,j}^N X_{ij}^2 - \quad \quad \quad \nu_{\text{总}} = N - 1$$

校正系数:

$$C = \frac{(\sum_{i=1}^g \sum_{j=1}^{n_i} X_{ij})^2}{N} = \frac{(\sum_{i,j}^N X_{ij})^2}{N}$$

2. 组间变异：各处理组由于接受处理的水平不同，各组的样本均数也大小不等，这种变异称为组间变异。**表示**为各组均数与总均数的离均差平方和，**反映**处理因素的作用和随机误差的影响，计算公式为：

如果**无**—— g 个总体均数**无**差别
如果**有**—— g 个总体均数**有**差别

$$SS_{\text{组间}} = \sum_{i=1}^g n_i (\bar{X}_i - \bar{X})^2 = \sum_{i=1}^g \frac{(\sum_{j=1}^{n_i} X_{ij})^2}{n_i} - C$$

$$\nu_{\text{组间}} = g - 1$$

3. 组内变异： 各组内各测量值 X_{ij} 与其所在组的均数的差值的平方和，**表示**随机误差的影响。计算公式为：

$$SS_{\text{组内}} = \sum_{i=1}^g \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$$

$$v_{\text{组内}} = N - g$$

三种变异的关系：

$$SS_{\text{总}} = SS_{\text{组间}} + SS_{\text{组内}}$$

$$V_{\text{总}} = V_{\text{组间}} + V_{\text{组内}}$$

均方差，均方 (mean square, MS)。

$$MS_{\text{组间}} = \frac{SS_{\text{组间}}}{v_{\text{组间}}}$$

$$MS_{\text{组内}} = \frac{SS_{\text{组内}}}{v_{\text{组内}}}$$

检验统计量:

$$F = \frac{MS_{\text{组间}}}{MS_{\text{组内}}}, \quad v_1 = v_{\text{组间}}, \quad v_2 = v_{\text{组内}}$$

如果 $\mu_1 = \mu_2 = \dots$ 则 μ_g 都为随机误差 的估计, F 值应接近于 1。

如果 $\mu_1, \mu_2, \dots$ 不全相等, F 值将明显大于 1。

用 F 界值 (单侧界值) 确定 P 值。



$$F = \frac{S_1^2}{S_2^2}, \quad \nu_1 = n_1 - 1, \quad \nu_2 = n_2 - 1$$

F 分布曲线的形状由两个参数 $\nu_1 = n_1 - 1$

$\nu_2 = n_2 - 1$ 决定, F 的取值范围为 0

$\sim$ 为 $\nu_1 = 10$ 、 $\nu_2 = \infty$ 和 $\nu =$

图形。

统计学家为应用的方便编制了附表 3 的 F 界值表, 表中单尾的界值用 $F_{\alpha, (\nu_1, \nu_2)}$ 表示, 双尾的界值则用 $F_{\alpha/2, (\nu_1, \nu_2)}$ 表示。

求得F值后，查F界值表得P值(F值愈大，P值愈小)，然后按所取的 α 水准作出推断结论。

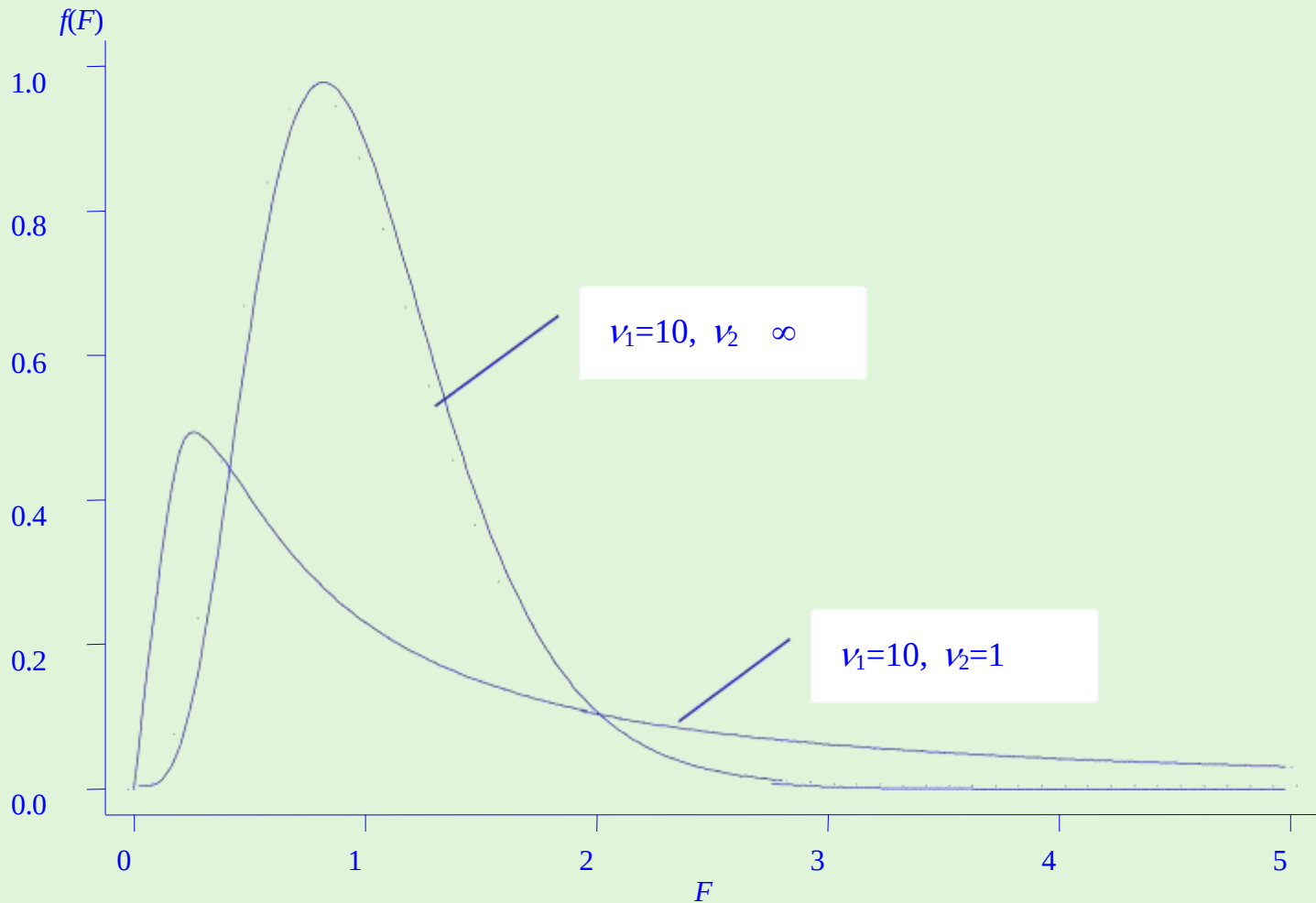


图3-10 不同自由度时F分布的图形



方差分析的**基本思想**:

- 根据试验设计的类型，将全部观测值总的离均差平方和及其自由度分解为两个或多个部分，除随机误差作用外，每个部分的变异可由某个因素的作用(或某几个因素的交互作用)加以解释。
- 通过比较不同变异来源的均方，借助 F 分布做出统计推断，从而推论各种研究因素对试验结果有无影响。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/545001300020011114>