

# 《生物实验设计》

## 第六章 方差分析

王超

广东药科大学

Email: wangchao@gdpu.edu.cn

2022-09-29



廣東藥科學  
GUANGDONG PHARMACEUTICAL UNIVERSITY

# 第六章 方差分析

- 样本平均数的假设检验适用于样本与总体或者两个样本之间的差异显著性检验
- 实际研究中，常需要 3 个及 3 个以上样本平均数进行比较
  - 如果两两相互比较，随着样本平均数个数增加而剧增
  - $n$  个样本平均数需要比较的次数为  $C_n^2$
- 导致：
  - 检验过程繁琐
  - 无统一的试验误差，误差估计的精确性和检验的灵敏性低
  - 推断的可靠性降低

# 方差分析

- 方差分析 ANOVA

- 将所有处理的观测值作为一个整体，一次比较就对所有各组间样本平均数是否有差异做出判断
- 差异不显著，则认为他们是相同的
- 差异显著，进一步比较是哪一组数据与其他数据不同

- 方差分析的用途

- 多个样本平均数的比较
- 分析多个因素间的交互作用
- 回归方程的假设检验
- 方差的同质性检验

# 第一节 假设检验的原理与方法

## 一、假设检验的概念

- 试验数据往往存在一定的差异，这种差异可能
  - 由于随机误差产生
  - 由于试验处理所引起
- 试验处理的效应往往和随机误差混淆，不容易分开
- 通过概率的计算和假设检验作出正确判断

### 假设检验

- 根据总体的理论分布和小概率原理，对未知或不完全知道的总体提出两种彼此对立的假设，然后由样本的实际结果，经过一定的计算，作出在一定概率意义上应该接受的那种假设的推断

如果：

- 抽样结果使小概率事件发生
  - 则拒绝假设
- 抽样结果没有使小概率事件发生
  - 则接受假设

小概率事件：概率  $\leq 0.05$  或  $\leq 0.01$  的事件为小概率事件

- ① 提出假设
- ② 确定显著水平
- ③ 计算统计数与相应的概率
- ④ 推断是否接受假设

### (一) 提出假设

- 对总体提出假设，一般是两个彼此对立的假设
  - 无效假设或零假设  $H_0$ :
    - 处理的效应跟总体参数之间没有真实的差异，试验结果中的差异是误差所致，即处理“无效”
  - 备择假设  $H_A$ :
    - 处理结果中的差异是由于总体参数不同所引起的，即处理“有效”
  - 无效假设与备择假设是对立事件：接受  $H_0$  则否定  $H_A$ ，接受  $H_A$  则否定  $H_0$
- $H_0$  随研究内容的不同而不同：
  - $H_0$  必须有意义
  - 根据  $H_0$  可以算出因抽样误差而获得样本结果的概率



## (一) 提出假设

以样本平均数的假设为例：

- 对一个样本平均数的假设（样本与总体）
  - 假设平均数为  $\bar{x}$  的样本来自于一组具有  $\mu$  的总体，提出：
    - $H_0 : \mu = \mu_0$
    - $H_A : \mu \neq \mu_0$
- 对两个样本平均数相比较的假设（样本与样本）
  - 假设两个样本平均数  $\bar{x}_1$  和  $\bar{x}_2$  分别来自具有平均数  $\mu_1$  和  $\mu_2$  的两个总体，提出：
    - $H_0 : \mu_1 = \mu_2$
    - $H_A : \mu_1 \neq \mu_2$
- 可从假设的总体中推论其平均数的随机抽样分布，从而可以算出其一个样本平均数指定值出现的概率，这样就可以根据样本与总体的关系，作为假设检验的理论依据

## (一) 提出假设

- 克矽平能否治疗矽肺病？

矽肺病患者血红蛋白含量的平均数  $\mu_0 = 126(mg/L)$ ,  $\sigma^2 = 240(mg/L)^2$  的正态分布  $N(126, 240)$

克矽平对 6 名患者进行治疗, 治疗后测得平均血红蛋白含量  $\bar{x} = 136(mg/L)$

- $\bar{x}$  和  $\mu_0$  之间的差值是由抽样误差还是药物治疗造成的？

## (二) 确定显著水平

- 确定一个否定  $H_0$  的概率标准, 显著水平  $\alpha$
- 人为规定的小概率界限
- 常用  $\alpha = 0.05$  和  $\alpha = 0.01$
- 根据研究需要调整

```
qnorm(0.025, mean = 0, sd = 1)
```

```
## [1] -1.959964
```

```
qnorm(0.005, mean = 0, sd = 1)
```

```
## [1] -2.575829
```

## (三) 计算统计数与相应的概率

在  $H_0: \mu = \mu_0$  的前提下,

$$u = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{136 - 126}{\sqrt{40}} = 1.58$$

- 在  $N(126, 40)$  的总体中, 以  $n = 6$  进行随机抽样, 得到平均值  $\bar{x} = 136$  与 126 相差 10 以上的概率是  
 $P(|u| > 1.58) = 2 * 0.05705 = 0.1141$
- 假设检验所计算的是超过实得差异得概率
- 概率的大小是推断  $H_0$  是否正确的依据

```
pnorm(-1.58, 0, 1)
```

```
## [1] 0.05705343
```

## (四) 推断是否接受假设

- 小概率原理：小概率事件在单次抽样试验中几乎是不可能发生的
- 如果概率大于显著水平则不认为是小概率事件，应该接受  $H_0$
- 差异显著水平 (0.05 或 0.01)
- 差异显著水平的标记方法 (\* 或 \*\*)
- 概率值为 0.1141，大于 0.05 的显著水平，所以接受  $H_0$
- 所以在治疗前后血红蛋白含量没有显著差异，差值应归于误差导致的

果蝇翅膀大小、细胞大小和细胞数量的倍数变化以及成年翼的大小比例差异 (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ )

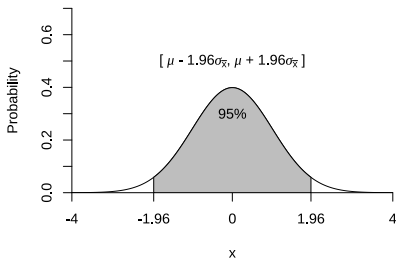
*Matsushita, R., Nishimura, T. Trehalose metabolism confers developmental robustness and stability in Drosophila by regulating glucose homeostasis. Commun Biol 3, 170 (2020).*

假设检验的步骤概括为：

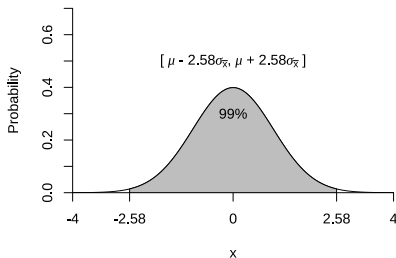
- ① 对样本所属总体提出无效假设  $H_0$  和备择假设  $H_A$
- ② 确定检验的显著水平  $\alpha$
- ③ 在  $H_0$  正确的前提下，计算抽样分布的统计数或相应的概率值
- ④ 根据小概率原理，进行差异是否显著的判断并得出结论

在标准正态分布下，样本平均数的抽样分布

$$N(\mu = 0, \sigma^2 = 1), \alpha = 0.05$$

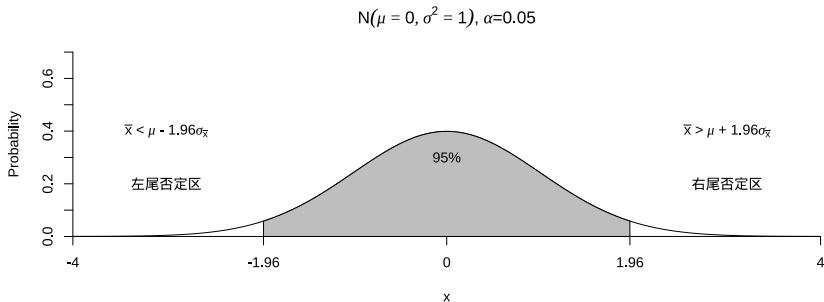


$$N(\mu = 0, \sigma^2 = 1), \alpha = 0.01$$



- 区间  $[\mu - u_{\alpha}\sigma_{\bar{x}}, \mu + u_{\alpha}\sigma_{\bar{x}}]$ ，其中  $u_{\alpha}$  根据  $u$  分布查表或者计算获得
- 对于一定的  $\alpha$ ，落在区间的  $\bar{x}$  有  $1 - \alpha$ ，落在区间外的是  $\alpha$
- $1 - \alpha$  相当于接受  $H_0$  的区域-接受区
- $\alpha$  相当于否定  $H_0$  的区域-否定区

否定区被接受区隔开，分为左尾和右尾两个：



- 临界点是  $\mu \pm u_{\alpha}\sigma_{\bar{x}}$



- 具有两个否定区的检验称为双尾检验
  - 这时候备择假设有两种可能,  $\mu > \mu_0$   $\mu < \mu_0$ , 落入左尾或者右尾否定区
- 某些情况下, 双尾检验不符合实际
  - 已知处理后产生的效应并提出无效假设  $H_0: \mu \leq \mu_0$ , 备择假设  $H_A: \mu > \mu_0$

# 第一节 概率基础知识

## 四、假设检验中的两类错误