# 生物统计与试验设计

## Biostatistics and Experimental Design


主讲：杨泽峰

扬州大学农学院

# 第二章 描述性统计数

- 2.1 变量与次数分布
- 2.2 次数分布表
- 2.3 次数分布图
- **2.4 集中趋势的统计数**
- 2.5 离散趋势的统计数

# 2.4 集中趋势的统计数

反映变量集中性的特征数是平均数（mean），平均数是数量资料的代表值，表示整个资料内变数的中心位置，并且可以作为一组资料的代表，与另一组资料进行比较。

平均数主要包括有算术平均数、中位数、众数、几何平均数及调和平均数。其中最常用的是算术平均数。

# 2.4.1 算术平均数

- **算术平均数**（arithmetic mean）是指总体或样本资料中各观察值的总和除以观察值个数所得到的商，简称平均数或均数。
- 总体的算术平均数用 $\mu$ 表示；
- 样本的算术平均数用 $\bar{y}$ 表示。

# 2.4.1 算术平均数

## （1）样本的算术平均数

**设某一资料包含 $n$ 个观察值：** $y_1$， $y_2$， …， $y_n$， **则样本平均数：**

$$\bar{y} = \frac{\bar{y}_1 + \bar{y}_2 + \cdots \bar{y}_n}{n} = \frac{\sum_{i=1}^{n} y_i}{n}$$

**其中：** $\sum$ **为求和符号；** $\sum_{i=1}^{n} y_i$ **表示从一个观察值** $y_1$ **累加到** $n$ **个观察值** $y_n$ **。**

# 2.4.1 算术平均数

## （1）样本的算术平均数

【例 2-6】：测得 20 株某品种玉米的叶片数为 13、14、14、15、14、16、15、16、15、17、16、14、15、16、15、16、17、15、16、15，试求该品种玉米平均每株的叶片数是多少：

计算：

$$\sum y=13+14+\cdots+15=304$$

$$\bar{y}=\frac{\sum y}{n}=\frac{13+14+\cdots 15}{20}=\frac{304}{20}=15.2$$ （片）

# 2.4.1 算数平均数

## (1) 样本的算术平均数

**平均数的基本性质:**

① 样本各观察值与平均数之差的和为零，即离均差之和等于零。

$$\sum(y-\bar{y})=0$$

② 样本各观察值与平均数之差的平方和为最小，即离均差平方和为最小。

$$\sum(y-\bar{y})^2=\sum(y-a)^2$$，其中常数 $a \neq \bar{y}$

# 2.4.1 算术平均数

## （2）总体的算术平均数

对于总体而言，通常用 $\mu$ 表示总体平均数，具有 $N$ 个观察值的有限总体的平均数成为总体平均数。总体数的计算公式为：

$$\mu=\frac{y_1+y_2+\cdots y_N}{N}=\frac{\sum_{i=1}^{N} y_i}{N}$$ 或简写为 $$\mu=\frac{\sum y_i}{N}$$

# 2.4.2 几何平均数

计算平均增长率，需要用几何平均数（geometric mean），记为 $G$。$G$ 的定义为：

$$G=\left(y_1\cdot y_2\cdots y_n\right)^{\frac{1}{n}}$$

对上式取对数得：

$$\lg G=\frac{\lg y_1+\lg y_2+\cdots\lg y_n}{n}=\frac{\sum(\lg y)}{n}$$

因而可知，几何平均数是观察值对数的算术平均数的反对数。

# 2.4.3 调和平均数



计算平均速率，需要用到调和平均数（harmonic mean），记为 $H$。

$$H=\frac{1}{\frac{1}{n}\left(\frac{1}{y_1}+\frac{1}{y_2}+\cdots+\frac{1}{y_n}\right)}=\frac{1}{\frac{1}{n}\sum\frac{1}{y}}$$

由上可知，调和平均数是观察值倒数的算术平均数的反倒数。

就同一资料而言，往往具有 $\bar{y}>G>H$ 的关系。

$H$ 具有最能减少极端大观察值的作用。

# 2.4.4 众数

**众数**（mode）是以出现频率最大定义的，计作 $M_0$。

如 100 个麦穗小穗数中，18 个小穗出现的频率最大，故 $M_0$=18。

对于连续性变量，则只能通过次数分布表求解：

$$M_0 = L + \frac{f_2}{f_1 + f_2} i$$

其中：$L$ 为次数最多组的低限， $f_1$ 为次数最多组上方组的次数，

$f_2$ 为次数最多组下方组的次数，$i$ 为组距。

表 2-6　100 个麦穗每穗小穗数的次数分布

| 编号 | 每穗小穗数 | 次数 |
|---|---|---|
| 1 | 15 | 7 |
| 2 | 16 | 17 |
| 3 | 17 | 27 |
| 4 | 18 | 29 |
| 5 | 19 | 14 |
| 6 | 20 | 6 |

# 2.4.5 中位数

**中位数**（median），又称中数，记作 $M_d$。

中位数以比它大和比它小的观察值各占 50%而定义的。将观察值按大小依次排列，当观察值数目为奇数时，最中间的观察值就是 $M_d$；当观察值数目为偶数时，最中间的两个观察值的算术平均数为 $M_d$。

由次数分布表计算中位数：$$M_d = L_{M_d} + \frac{\frac{n}{2} - A}{f_{M_d}} i$$

其中：$L_{M_d}$ 为中位数所在组的低限，$f_{M_d}$ 为中位数所在组的次数，$n$ 为样本容量，$A$ 为中位数所在组上方各组的累积次数，$i$ 为组距。

# 2.4.5 中位数



**【例2-7】 根据次数分布表求取众数和中位数。**

表 2-3　102 株水稻株高（cm）的次数分布表

| 编号 | 组区间 | 组中值 | 次数 |
|---|---|---|---|
| 1 | 101-102 | 101.5 | 2 |
| 2 | 102-103 | 102.5 | 3 |
| 3 | 103-104 | 103.5 | 7 |
| 4 | 104-105 | 104.5 | 13 |
| 5 | 105-106 | 105.5 | 16 |
| 6 | 106-107 | 106.5 | 18 |
| 7 | 107-108 | 107.5 | 15 |
| 8 | 108-109 | 108.5 | 14 |
| 9 | 109-110 | 109.5 | 8 |
| 10 | 110-111 | 110.5 | 5 |
| 11 | 111-112 | 111.5 | 1 |

**（1）众数：**

$$M_0 = L + \frac{f_2}{f_1 + f_2} i = 106 + \frac{15\times1}{16+15} = 106.4839$$

**（2）中位数**

$$M_d = L_{M_d} + \frac{\frac{n}{2} - A}{f_{M_d}} i = 106 + \frac{\frac{102}{2} - 41}{18} \times 1 = 106.5556$$