

# 《生物实验设计》

## 第二章 资料整理与特征数计算

王超

广东药科大学

Email: wangchao@gdpu.edu.cn

2022-09-03

## 第二章 资料整理与特征数计算

- 在试验及调查中能够获得大量的原始数据，这是在一定条件下对某种具体事物或现象观察的结果，称之为资料。
- 统计分析就是对资料的整理分析，列出统计表，绘出统计图，计算特征数。

对资料的分类整理，必须坚持同质的原则。

只有同质的试验数据，才能根据科学原理来分类，使资料能正确反映事物的本质和规律。

- 数量性状资料（定量）
- 质量性状资料（定性）

### (一) 数量性状资料

数量性状资料一般由计数和测量或度量得到的。

- 计数资料：由计数法得到的数据（非连续变量）。
- 计量资料：由测量或度量所得的数据（连续变量）
  - 依试验的要求和测量仪器或工具的精度而定。

### (二) 质量性状资料

对某种现象只能观察而不能测量的资料。

一般需要先把质量性状资料数量化，可以采用：

#### ① 统计次数法

- 根据某一质量性状的类别统计其次数或频数，以次数和频数作为该质量性状的数据。

#### ② 评分法

- 评分是用数字级别表示某现象在表现程度上的差别。
- 根据评分将质量性状资料进行量化，然后参照计数资料的处理方法进行。

### (一) 调查

#### ① 全面调查/普查

- 对研究对象的每一个个体逐一进行调查
- 范围广、时间长、工作量大
- 极少数情况

#### ② 抽样调查

- 根据一定的原则对研究对象抽取一部分个体进行测量或度量
- 把得到的数据资料作为样本进行统计处理，然后利用样本特征数对总体参数进行推断
- 要想无偏差估计总体，重要的是采用科学的抽样方法

### 抽样调查

#### ① 随机抽样

在试验过程中对试验单位的抽样、分组等都必须遵守随机原则，避免人为主观因素的影响

#### ② 顺序抽样

- 按照既定顺序从总体中抽取一定数量的个体构成样本

#### ③ 典型抽样

- 根据初步资料或经验判断，有意识、有目的地选择一个典型群体作为代表进行调查，以估计整个总体

可以混合地才采用以上集中抽样方法。



### 随机抽样

随机抽样必须满足两个条件：

- 总体中每个个体被抽中的机会是均等的
- 总体中任意一个个体是否被抽中是相互独立的（不受其他个体影响）

随机抽样的方法：

- ① 简单随机抽样
- ② 分层随机抽样
- ③ 整体抽样
- ④ 双重抽样

### (二) 试验

通过一定数量有代表性的试验单位，在一定的条件下进行的有探索性的研究工作。

试验处理原则：

- ① 随机
- ② 重复
- ③ 局部控制

试验设计方法在第九章介绍。

- 原始资料的检查与核对
- 频数分布表
- 频数分布图

# 频数分布表

100 只鸡每月产蛋数（用 `rnorm` 随机生成这样一组数据）

```
set.seed(2022)
```

```
egg <- round(rnorm(100, mean = 14, sd = 1.5))
```

```
egg
```

```
##      [1] 15 12 13 12 14 10 12 14 15 14 16 14 13 14 14 14 13 13 1
##     [26] 13 15 14 13 14 14 15 12 14 13 16 16 14 14 14 14 15 12 1
##     [51] 17 15 14 16 14 15 14 16 13 15 12 12 14 14 13 14 12 16 1
##     [76] 14 16 16 16 15 12 17 16 12 11 15 16 16 18 14 15 14 15 1
```

```
summary(egg)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      10.00   13.00   14.00   14.25   15.00   18.00
```

利用 `summary` 可以大致了解数据的分布情况。

```
fdt_eg <- table(egg) # 次数统计
addmargins(fdt_eg)
```

```
## egg
##  10  11  12  13  14  15  16  17  18 Sum
##   1   2  12  13  29  21  17   3   2 100
```

```
prop.table(fdt_eg) # 频率统计
```

```
## egg
##   10   11   12   13   14   15   16   17   18
## 0.01 0.02 0.12 0.13 0.29 0.21 0.17 0.03 0.02
```

```
addmargins(prop.table(fdt_eg))
```

```
## egg
##   10   11   12   13   14   15   16   17   18 Sum
## 0.01 0.02 0.12 0.13 0.29 0.21 0.17 0.03 0.02 1.00
```

## 300 个麦穗的每穗穗粒数

```
set.seed(2022)
wheat <- round(rnorm(300, mean = 40, sd = 7))
wheat[1:100]

##      [1] 46 32 34 30 38 20 33 42 45 42 47 39 33 41 40 39 35 33 47 46 43 43 48 48 38
##     [26] 34 45 42 36 38 41 46 29 38 34 48 48 41 41 39 38 46 32 30 40 34 42 38 31 43
##     [51] 52 47 41 49 42 44 40 48 37 43 31 31 38 41 34 40 32 48 56 43 39 49 43 40 49
##     [76] 39 47 50 50 46 33 53 48 32 25 44 49 50 59 40 45 42 43 37 60 36 27 45 42 43

summary(wheat)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      20.00   34.00   40.00   39.69   45.00   60.00
```

# R demo

```
fdt_wt <- table(cut(wheat, breaks = seq(15, 60, 5), include.lowest = TRUE), addmargins(fdt_wt))
```

```
##
```

```
## [15,20] (20,25] (25,30] (30,35] (35,40] (40,45] (45,50] (50,55]
```

##	1	3	27	57	68	77	53
----	---	---	----	----	----	----	----

```
prop.table(fdt_wt) # 频率统计
```

```
##
```

```
## [15,20] (20,25] (25,30] (30,35] (35,40]
```

##	0.003333333	0.010000000	0.090000000	0.190000000	0.226666667	0.176666667	0.033333333	0.013333333
----	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------

```
## (45,50] (50,55] (55,60]
```

##	0.176666667	0.033333333	0.013333333	1.000000000
----	-------------	-------------	-------------	-------------

```
addmargins(prop.table(fdt_wt))
```

```
##
```

```
## [15,20] (20,25] (25,30] (30,35] (35,40]
```

##	0.003333333	0.010000000	0.090000000	0.190000000	0.226666667	0.176666667	0.033333333	0.013333333	1.000000000
----	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------

```
## (45,50] (50,55] (55,60] Sum
```

##	0.176666667	0.033333333	0.013333333	1.000000000
----	-------------	-------------	-------------	-------------

# 计量资料的整理

```
set.seed(2022)
fish <- round(rnorm(150, mean = 55, sd = 9))
fish[1:100]

##      [1] 63 44 47 42 52 29 45 58 62 57 64 53 46 56 55 54 49 46 64 63 58 58 65 66 52
##     [26] 47 61 58 50 53 56 62 41 53 48 65 65 56 56 53 53 62 45 42 56 48 58 53 43 58
##     [51] 70 64 57 66 58 61 55 66 51 59 43 43 52 56 47 55 45 65 76 59 54 67 59 56 67
##     [76] 53 64 67 68 63 46 72 65 45 35 60 67 67 80 55 62 57 59 52 81 50 38 61 57 59

summary(fish)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      29.00   50.00   56.50   55.69   63.00   81.00
```



# R demo

```
fdt_wt <- table(cut(wheat, breaks = seq(15, 60, 5), include.lowest = TRUE), addmargins(fdt_wt))
```

```
##
```

```
## [15,20] (20,25] (25,30] (30,35] (35,40] (40,45] (45,50] (50,55]
```

##	1	3	27	57	68	77	53
----	---	---	----	----	----	----	----

```
prop.table(fdt_wt) # 频率统计
```

```
##
```

```
## [15,20] (20,25] (25,30] (30,35] (35,40]
```

##	0.003333333	0.010000000	0.090000000	0.190000000	0.226666667	0.176666667	0.033333333	0.013333333
----	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------

```
## (45,50] (50,55] (55,60]
```

##	0.176666667	0.033333333	0.013333333					
----	-------------	-------------	-------------	--	--	--	--	--

```
addmargins(prop.table(fdt_wt))
```

```
##
```

```
## [15,20] (20,25] (25,30] (30,35] (35,40]
```

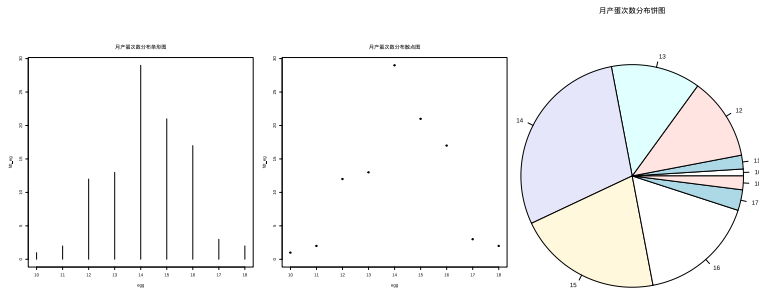
##	0.003333333	0.010000000	0.090000000	0.190000000	0.226666667	0.176666667	0.033333333	0.013333333	1.000000000
----	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------

```
## (45,50] (50,55] (55,60] Sum
```

##	0.176666667	0.033333333	0.013333333	1.000000000					
----	-------------	-------------	-------------	-------------	--	--	--	--	--

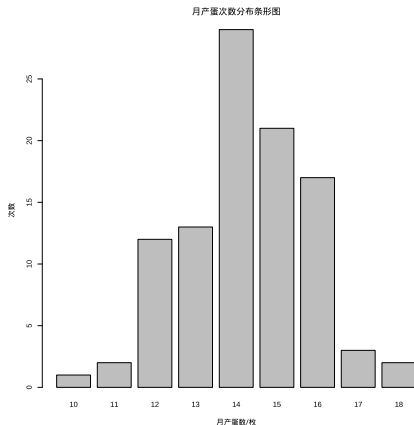
# 频数分布图

```
plot(fdt_eg, type = "h", main = " 月产蛋次数分布条形图")  
plot(fdt_eg, type = "p", main = " 月产蛋次数分布散点图", pch = 20)  
pie(prop.table(fdt_eg), main = " 月产蛋次数分布饼图")
```



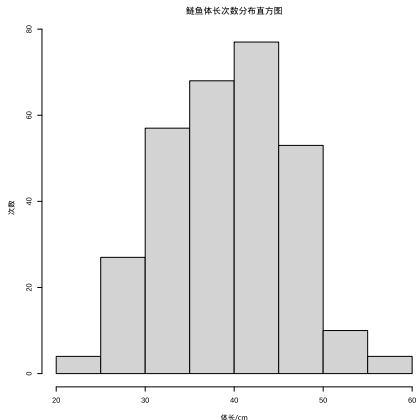
# plot 之外的方法

```
barplot(fdt_eg, main = " 月产蛋次数分布条形图",  
        ylab = " 次数", xlab = " 月产蛋数/枚")
```



# 直方图

```
hist(wheat, main=" 鲢鱼体长次数分布直方图", ylab=" 次数", xlab=" 体长/cm")
```



## 第二节 资料特征数的计算

# 一、平均数

- 算数平均数:

$$\mu = \frac{x_1 + x_2 + x_3 + \cdots + x_N}{N} = \frac{1}{N} \sum_{i=1}^N x_i$$

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_N}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

- 中位数:

$$M_d = x_{\frac{n+1}{2}}$$

或者

$$M_d = (x_{\frac{n}{2}} + x_{\frac{n}{2}+1})/2$$

- 众数:  $M_o$
- 几何平均数:

$$G = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdots x_n} = \sqrt[n]{\prod_{i=1}^n x_i}$$

# 算数平均数计算方法

1. 直接及算法
2. 加减常数法