

### 目录

1

▶ **第一节：频数分布**

2

▶ **第二节：集中趋势的描述**

3

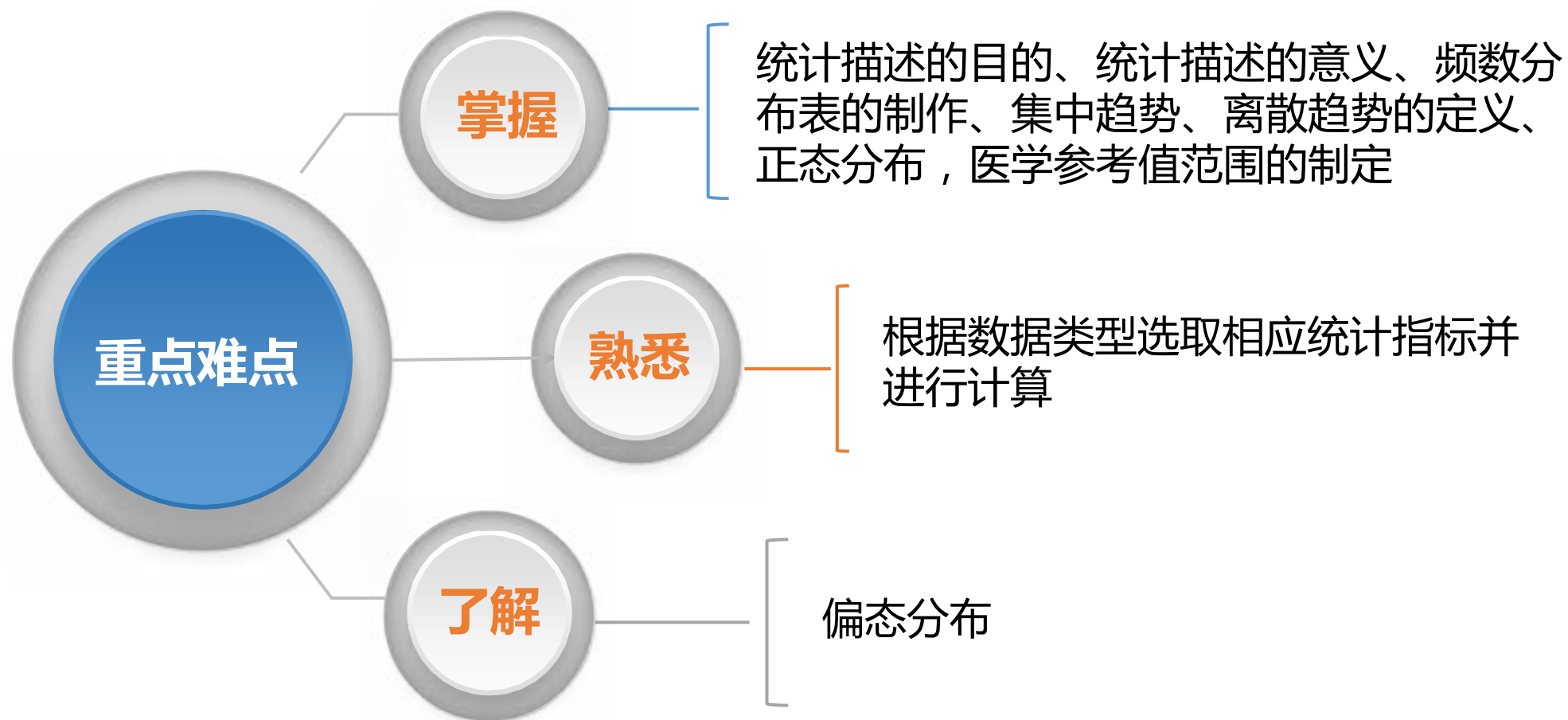
▶ **第三节：离散趋势的描述**

4

▶ **第四节：正态分布**

5

▶ **第五节：医学参考值范围的制定**



# 第一节

# 频数分布

# 一、频数分布

---

(一) 频数分布表

(二) 频数分布图

(三) 频数分布表和频数分布图的用途

### (一) 频数分布表

将原始计量数据整理、编制成更好地展示样本数据分布特征的频数分布表（简称频数表，frequency table）来了解其分布规律。

频数分布（frequency distribution）通常是针对样本而言。对于连续变量，频数分布为 $n$ 个变量值在各变量值区间内个数的分配[见表2-1第（1）栏和第（2）栏]。对于离散变量，频数分布为 $n$ 个变量值在各（或各几个）变量值处的个数分配[见表2-2第（1）栏和第（2）栏]。表2-2为某医院1 123名产后出血孕妇的人流次数分布表，为离散型计量资料。

例2-1 某医院用随机抽样方法检查了138名正常成年女子的红细胞数，其测量结果如下，试编制频数分布表。

3.96	4.23	4.42	3.59	5.12	4.02	4.32	3.72	4.76	4.16	4.61	4.26
3.77	4.20	4.36	<u>3.07</u>	4.89	3.97	4.28	3.64	4.66	4.04	4.55	4.25
4.63	3.91	4.41	3.52	5.03	4.01	4.30	4.19	4.75	4.14	4.57	4.26
4.56	3.79	3.89	4.21	4.95	3.98	4.29	3.67	4.69	4.12	4.56	4.26
4.66	4.28	3.83	4.20	5.24	4.02	4.33	3.76	4.81	4.17	3.96	3.27
4.61	4.26	3.96	4.23	3.76	4.01	4.29	3.67	3.39	4.12	4.27	3.61
4.98	4.24	3.83	4.20	3.71	4.03	4.34	4.69	3.62	4.18	4.26	4.36
5.28	4.21	4.42	4.36	3.66	4.02	4.31	4.83	3.59	3.97	3.96	4.49
5.11	4.20	4.36	4.54	3.72	3.97	4.28	4.76	3.21	4.04	4.56	4.25
4.92	4.23	4.47	3.60	5.23	4.02	4.32	4.68	4.76	3.69	4.61	4.26
3.89	4.21	4.36	3.42	5.01	4.01	4.29	3.68	4.71	4.13	4.57	4.26
4.03	<u>5.46</u>	4.16	3.64	4.16	3.76						

### 1. 求极差（全距）

极差也称全距，即最大值和最小值之差，记作 $R$ 。

$$\text{本例 } R = 5.46 - 3.07 = 2.39(\times 10^{12}/L)$$

### 2. 确定组数和组距

根据极差选定适当“组段”数。组段数通常取 10~15 组。本例组距

$$i = 2.39/12 = 0.199 \approx 0.20(\times 10^{12}/L)$$

**3. 确定组距** 每个组段的下限为、上限为，变量值的归组统一为  $L \leq X < U$   
最后组段写出上限。

### 4. 分组划记并统计频数

**表2-1 138名正常成年女子的红细胞数 (  $\times 10^{12}/L$  ) 频数分布**

组段	频数 $f$	组中值 $X$	$fX$	$fX^2$
(1)	(2)	(3)	(4) = (2) $\times$ (3)	(5) = (3) $\times$ (4)
3.07~	2	3.17	6.34	20.10
3.27~	3	3.37	10.11	34.07
3.47~	9	3.57	32.13	114.70
3.67~	14	3.77	52.78	198.98
3.87~	22	3.97	87.34	346.74
4.07~	30	4.17	125.10	521.67
4.27~	21	4.37	91.77	401.03
4.47~	15	4.57	68.55	313.27
4.67~	10	4.77	47.70	227.53
4.87~	6	4.97	29.82	148.21
5.07~	4	5.17	20.68	106.92
5.27~5.47	2	5.37	10.74	57.67
合计	138	—	583.06	2 490.89

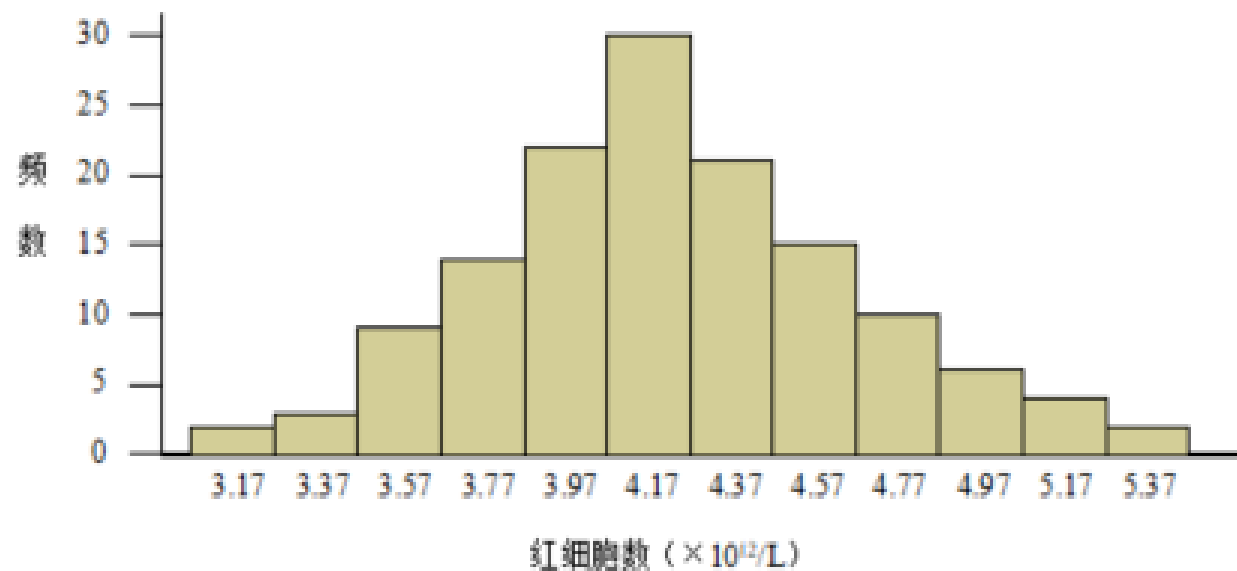


表2-1中138名正常成年女子的红细胞数为连续型计量资料；而表2-2中1123产后出血孕妇的人流次数为离散型计量资料。

**表2-2 某医院1 123名产后出血孕妇的人流次数分布**

人流次数	人数	累积频数	累积频率 (%)
(1)	(2)	(3)	(4)
0	402	402	35.80
1	330	732	65.18
2	232	964	85.84
3	118	1 082	96.35
4	27	1 109	98.75
5	11	1 120	99.73
6	3	1 123	100.00
合计	1 123	—	—

### (二) 频数分布图



- 138名正常成年女子的红细胞数 ( $\times 10^{12}/L$ ) 频数分布

### (三) 频数表与频数分布图的用途

#### 1. 揭示计量资料的分布特征

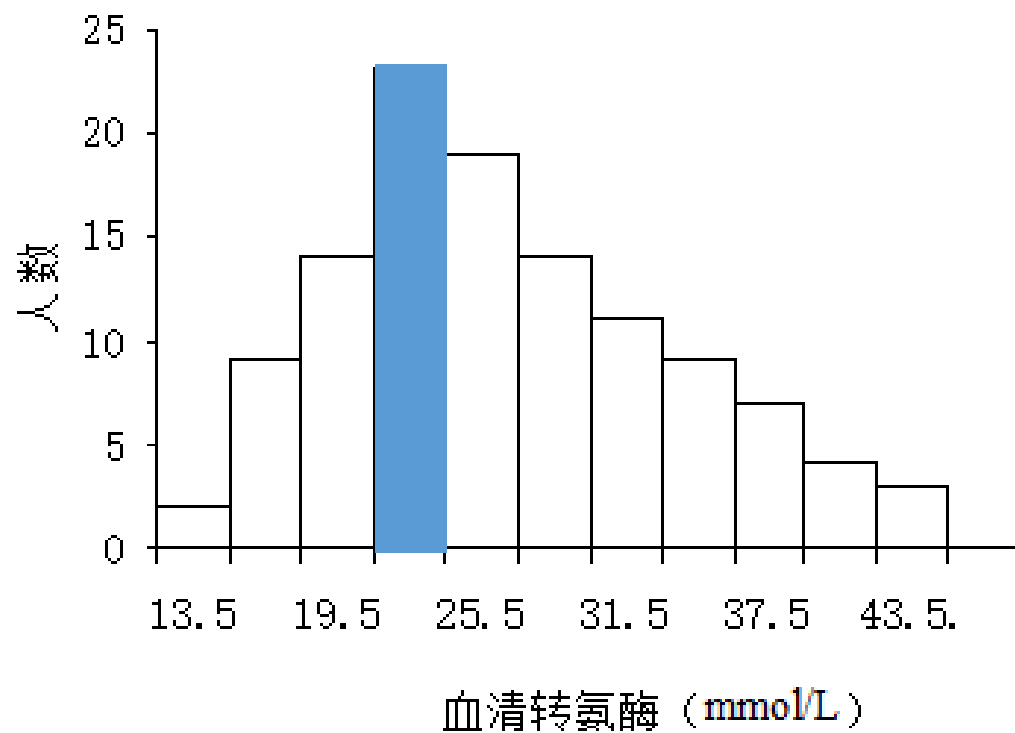
**(1) 对称分布**：若各组段频数的分布以频数最多的组段为中心左右两侧大体对称（总体则完全对称），就认为该资料是对称分布

**(2) 偏态分布**：正偏态分布  
负偏态分布

### 正偏态分布（右偏态分布）

表2-3 115名正常成年女子的血清转氨酶（mmol/L）含量分布

转氨酶含量	人数
12~	2
15~	9
18~	14
21~	23
24~	19
27~	14
30~	11
33~	9
36~	7
39~	4
42~45	3

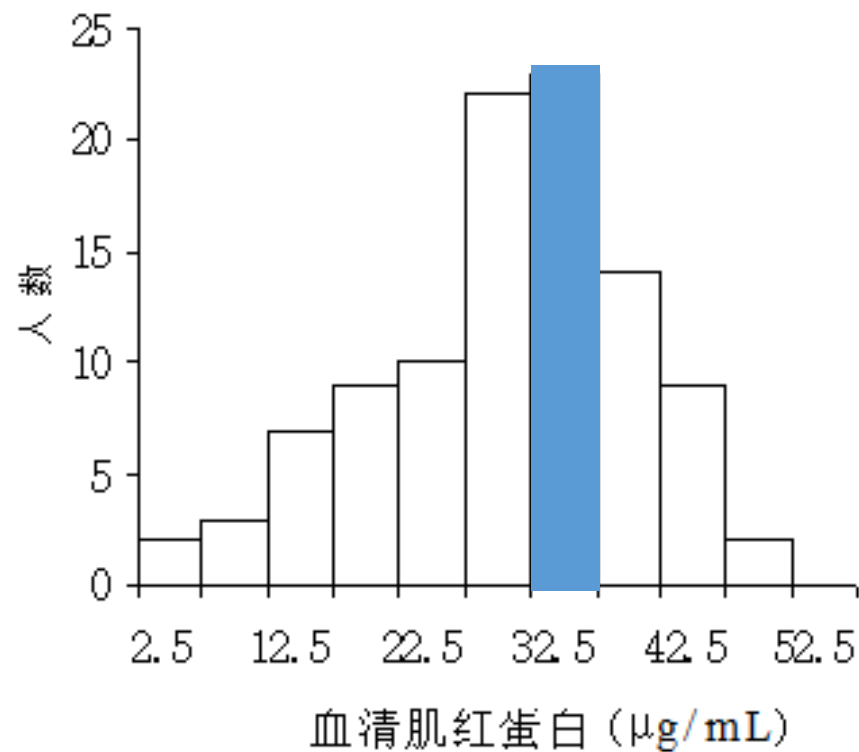


- 115名正常成年女子血清转氨酶的频数分布

### 负偏态分布（左偏态分布）

表2-4 101名正常人的血清肌红蛋白( $\mu\text{g/mL}$ )含量分布

肌红蛋白含量	人数
0~	2
5~	3
10~	7
15~	9
20~	10
25~	22
30~	23
35~	14
40~	9
45~50	2



- 101名正常人血清肌红蛋白的频数分布

### 2. 描述频数分布特征

- 变异的范围在 $3.07 \sim 5.46(\times 10^{12}/L)$
- 有明显的统计分布规律，数据主要集中在 $3.47 \sim 4.87(\times 10^{12}/L)$ 之间，尤以 $4.07 \sim 4.27(\times 10^{12}/L)$ 组段的人数最多，且上下组段数的频数分布基本对称。

### 3. 便于发现一些特大或特小的离群值

### 4. 进一步做统计分析处理



# 第二节

## 集中趋势的描述

# 二、集中趋势的描述

---

**(一) 算术平均数**

**(二) 几何均数**

**(三) 中位数**

**(四) 百分位数**

### (一) 算术均数 (mean)

- 简称均数，用于反映一组呈对称性分布的变量值在数量上的平均水平或者说是集中位置的特征值。
- 应用：对称性分布，尤其是正态分布。
- 符号：总体用  $\mu$  表示，样本用  $\bar{X}$  表示。

### 1. 直接计算法 适用于 $n$ 较小时 (例如 $n < 30$ )

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{\sum X}{n}$$

例2-2：用直接法计算例2-1某医院随机抽查的138名正常成年女子的红细胞数的均数。

$$\bar{X} = \frac{3.96 + 4.23 + \dots + 3.76}{138} = 4.23(\times 10^{12}/L)$$

### 2. 加权法（频数表法，大样本资料）

$$\bar{X} = \frac{f_1X_1 + f_2X_2 + f_3X_3 + \dots + f_kX_k}{f_1 + f_2 + f_3 + \dots + f_k} = \frac{\sum fX}{\sum f}$$

$$\left\{ \begin{array}{l} X = \frac{\text{该组上限} + \text{该组下限}}{2} \text{ (组中值)} \\ f \text{ 各组段的频数} \\ k \text{ 频数表的组段个数} \end{array} \right.$$

**表2-1 138名正常成年女子的红细胞数 ( ×10<sup>12</sup>/L ) 频数分布**

组段	频数 <i>f</i>	组中值 <i>X</i>	<i>fX</i>	<i>fX</i> <sup>2</sup>
(1)	(2)	(3)	(4) = (2) × (3)	(5) = (3) × (4)
3.07~	2	3.17	6.34	20.10
3.27~	3	3.37	10.11	34.07
3.47~	9	3.57	32.13	114.70
3.67~	14	3.77	52.78	198.98
3.87~	22	3.97	87.34	346.74
4.07~	30	4.17	125.10	521.67
4.27~	21	4.37	91.77	401.03
4.47~	15	4.57	68.55	313.27
4.67~	10	4.77	47.70	227.53
4.87~	6	4.97	29.82	148.21
5.07~	4	5.17	20.68	106.92
5.27~5.47	2	5.37	10.74	57.67
合计	138	—	583.06	2 490.89

$$\bar{X} = \frac{2 \times 3.17 + 3 \times 3.37 + \dots + 2 \times 5.37}{2 + 3 + \dots + 2} = \frac{583.06}{138} = 4.23(\times 10^{12}/L)$$

例2-3 利用表2-1  
计算某医院随机  
抽查的138名正  
常成年女子的红  
细胞数的均数。

### (二)几何均数 ( geometric mean )

- 意义：可用于反映一组经对数转换后呈对称分布的变量值在数量上的平均水平。
- 应用：观察值间按倍数关系变化，经对数转换后呈对称分布的资料，常见于描述抗体滴度、细菌计数、血清凝集效价等免疫学和微生物学资料。
- 符号:用  $G$  表示。

### 1. 计算公式

$$G = \sqrt[n]{X_1 X_2 \dots X_n} \text{ 或 } G = \lg^{-1} \left( \frac{\sum \lg X}{n} \right)$$

例2-4 某地5例微丝蚴患者治疗7年后用间接荧光抗体试验测得其抗体滴度倒数分别为10, 20, 40, 40, 160, 试计算其几何均数。

或

$$G = \sqrt[5]{10 \times 20 \times 40 \times 40 \times 160} = 34.8$$

$$G = \lg^{-1} \frac{\lg 10 + \lg 20 + \lg 40 + \lg 40 + \lg 160}{5} = 34.8$$



### 2. 加权法

适用于相同观察值较多或频数表资料，计算公式为：

$$G = \lg^{-1}\left(\frac{\sum f \lg X}{\sum f}\right)$$

例2-5 69例类风湿关节炎 ( RA ) 患者血清EBV-VCA-IgG抗体滴度的分布见表2-5第 ( 1 ) 和 ( 2 ) 栏, 求其平均抗体滴度。

**表2-5 69例RA患者血清EBV-VCA-IgG抗体测定结果**

抗体滴度	人数 $f$	滴度倒数 $X$	$\lg X$	$f \cdot \lg X$
(1)	(2)	(3)	(4)	(5)
1: 10	4	10	1.0000	4.0000
1: 20	3	20	1.3010	3.9030
1: 40	10	40	1.6021	16.0210
1: 80	10	80	1.9031	19.0310
1: 160	11	160	2.2041	24.2451
1: 320	15	320	2.5051	37.5765
1: 640	14	640	2.8062	39.2868
1: 1280	2	1280	3.1072	6.2144
合 计	69	—	—	150.2778

$$G = \lg^{-1}\left(\frac{150.2778}{69}\right) = \lg^{-1}(2.1779) = 150.6$$

即抗体的平均滴度为 1:150.6。

### (三)中位数 ( median )

- 意义：是将 $n$ 个变量值从小到大排列，位置居于中间的那个变量值。
- 应用：偏态分布资料；一端或两端无确切数值的资料（开口资料）；分布不明的资料。
- 符号:样本用 $M$ 表示。

### 1. 直接计算法 (小样本时)

$$n \text{ 为奇数时} \quad M = X_{\left(\frac{n+1}{2}\right)}$$

$$n \text{ 为偶数时} \quad M = \frac{1}{2} \left( X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)} \right)$$

例2-6 7名病人患某病的潜伏期分别为1, 3, 5, 5, 8, 10, 17天, 求其中位数。

本例 $n=7$ ,为奇数

$$M = X_{\left(\frac{7+1}{2}\right)} = X_{(4)} = 5(\text{天})$$

例2-7 14名病人食物中毒的发病时间分别为1, 2, 3, 3, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4天, 求其中位数。

本例n=14,为偶数

$$M = \frac{1}{2} \left( X_{\left(\frac{14}{2}\right)} + X_{\left(\frac{14}{2}+1\right)} \right) = \frac{1}{2} (X_7 + X_8) = \frac{1}{2} (4 + 4) = 4(\text{天})$$

例2-8 试计算表2-2某医院1 123名产后出血孕妇人工流产次数的中位数。

本例为离散型计量资料。因 $n=1123$ ，故中位数是从小到大排序后居于 $(n+1)/2=(1123+1)/2=562$ 位的观察值。据表，排在第1~402位的观察值均为“0”，其累积频率为35.80%，排在403~732位的观察值均为“1”，其累积频率为65.18%，余类推。第562位数属于第二个变量值，即人流次数为“1”，故某医院1 123名产后出血孕妇人工流产次数的中位数  $X_{(562)}=1$ 。

### (四) 百分位数 ( percentile )

百分位数是一种位置指标，用 $P_x$ 来表示。一个百分位数将全部变量值分为两部分，在不包含 $X\%$ 的全部变量值中有 $P_x$ 的变量值比它小，有 $(100-X)\%$ 的变量值比它大。



1. 直接计算法 将 $n$ 个变量值从小到大排列, 设 $(n+1)X\% = j + g$   $j$  为整数部分  $g$  为小数部分, 则

$$\text{当 } g = 0 \text{ 时 } P_x = X_{(j)}$$

$$\text{当 } g \neq 0 \text{ 时 } P_x = (1-g)X_{(j)} + gX_{(j+1)}$$

例2-9 调查某医院细菌性痢疾治愈者的住院天数，119名患者的住院天数从小到大的排列如表，试求第5百分位数和第99百分位数。

**表2-6 119名细菌性痢疾治愈者的住院天数**

患者编号	住院天数
1	1
2	1
3	2
4	2
5	2
6	3
7	4
8	4
9	5
⋮	⋮
116	39
117	40
118	40
119	42

$(119+1) \times 5\% = 6, j = 6$  按公式计算, 得

$$P_5 = X_{(6)} = 3(\text{天})$$

$(119+1) \times 99\% = 118.8, j = 118, g = 0.8$  按公式 (2-8) 计算, 得

$$P_{99} = 0.2X_{(118)} + 0.8X_{(119)} = 0.2 \times 40 + 0.8 \times 42 = 41.6(\text{天})$$

$P_5$  的意义是该医院有5%的细菌性痢疾治愈者的住院天数少于3天，或者说有95%的细菌性痢疾治愈者的住院天数多于3天； $P_{99}$  的意义是绝大多数（99%）细菌性痢疾治愈者的住院天数少于41.6天。

### 2. 频数表法

$$P_X = L_X + \frac{i_X}{f_X} (nX\% - \sum f_L)$$

式中 $L_X$ 、 $i_X$ 和 $f_X$ 分别为第 $X$ 百分位数所在组段的下限、组距和频数， $\sum f_L$ 为小于 $L_X$ 各组段的累计频数， $n$ 为总例数。

当 $X\% = 50\% = \frac{1}{2}$ 时，中位数的计算公式为

$$M = P_{50} = L_{50} + \frac{i_{50}}{f_{50}} \left( \frac{n}{2} - \sum f_L \right)$$

例 2-10 某地118名链球菌咽喉炎患者的潜伏期频数表见第(1)、(2)栏，求中位数及第25、第75百分位数

**表2-7 118名链球菌咽喉炎患者的潜伏期**

天 数	人 数	累计频数	累计频率 (%)
(1)	(2)	(3)	(4)
12~	4	4	3.4
24~	17	21	17.8
36~	32	53	44.9
48~	24	77	65.3
60~	18	95	80.5
72~	12	107	90.7
84~	5	112	94.9
96~	4	116	98.3
108~120	2	118	100.0

将相应数据代入公式得：

$$M = P_{50} = 48 + \frac{12}{24} \left( \frac{118}{2} - 53 \right) = 51 \text{ (天)}$$

同理

$$P_{25} = 36 + \frac{12}{32} (118 \times 25\% - 21) = 39.2 \text{ (天)}$$

$$P_{75} = 60 + \frac{12}{18} (118 \times 75\% - 77) = 67.7 \text{ (天)}$$

### 2. 百分位数的应用

- 用于描述样本或总体观察值序列某百分位置的水平。
- 用于确定偏态分布资料的参考值范围。
- 用于描述偏态资料的变异程度。



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/738020041015006100>