

# 离散趋势的描述



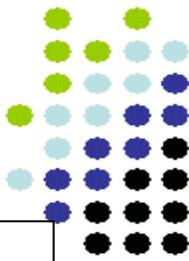
分组	身高					均数
甲组	90	95	100	105	110	100
乙组	96	98	100	102	104	100
丙组	96	100	99	101	104	100

$$\text{甲组: } \frac{90 + 95 + 100 + 105 + 110}{5} = 100 \text{ cm}$$

$$\text{乙组: } \frac{96 + 98 + 100 + 102 + 104}{5} = 100 \text{ cm}$$

$$\text{丙组: } \frac{96 + 100 + 99 + 101 + 104}{5} = 100 \text{ cm}$$

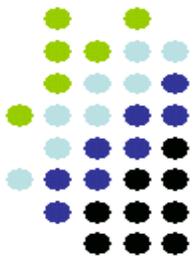
# 离散趋势的描述



分组	身高					均数	极差
甲组	90	95	100	105	110	100	110- 90=20 (cm)
乙组	96	98	100	102	104	100	104- 96=8 (cm)
丙组	96	100	99	101	104	100	104- 96=8 (cm)

# 离散趋势的描述

---



反映数据的离散度（ **Dispersion** ）。即个体观察值的变异程度。常用的指标有：

1. 极差(**Range**) (全距)
2. 百分位数与四分位数间距  
**Percentile and Quartile range**
3. 方差 **Variance**与标准差**Standard Deviation**
4. 变异系数 **Coefficient of Variation**

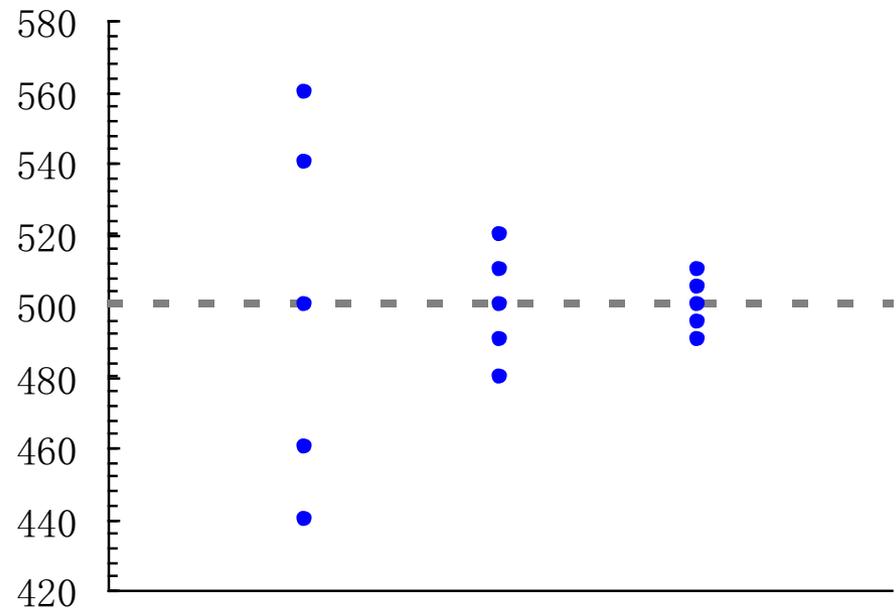
# 极差(range)



$$R = X_{\max} - X_{\min}$$

优点：简便

缺点：1. 只利用了两个  
极端值  
2.  $n$ 大， $R$ 也会大  
3. 不稳定

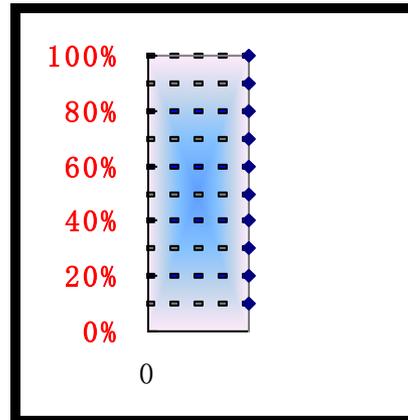


# 四分位数间距



四分位间距:

$$QR = P_{75} - P_{25}$$



$P_{100}$  (max)

$P_{75}$

$P_{50}$  (中位数)

$P_{25}$

$P_0$  (min)



某地118名链球菌咽喉炎患者的潜伏期频数表  
如下，计算其四分位数间距

天数 (1)	人数, $f$ (2)	累计频数 $\Sigma f$ (3)	累计频率 (%)
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~	2	118	100.0



---

$$P_x = L + i \times \frac{(n \times x\% - \sum f_L)}{f_x}$$

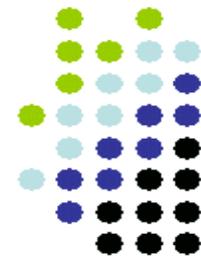
$$Q_3 = P_{75} = 60 + 12 \times \frac{(118 \times 75\% - 77)}{18} = 67.7(\text{天})$$

$$Q_1 = P_{25} = 36 + 12 \times \frac{(118 \times 25\% - 21)}{32} = 39.2(\text{天})$$

$$QR = Q_3 - Q_1 = 67.7 - 39.2 = 28.5(\text{天})$$

# 方差与标准差

---



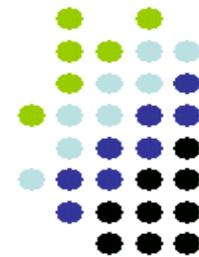
方差：反映一组数据的平均离散水平

$$\text{总体方差 } \sigma^2 = \frac{\sum (X - \mu)^2}{N}$$

$$\text{样本方差 } S^2 = \frac{\sum (X - \bar{X})^2}{n-1} = \frac{\sum X^2 - (\sum X)^2 / n}{n-1}$$

# 标准差

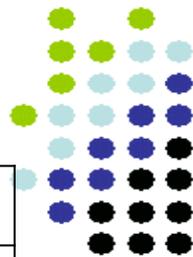
---



**标准差**（**standard deviation**）即方差的正平方根；其单位与原变量 $X$ 的单位相同。

$$\text{样本标准差 } S = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}} = \sqrt{\frac{\sum X^2 - (\sum X)^2 / n}{n-1}}$$

$$\text{频数表样本标准差 } S = \sqrt{\frac{\sum fX^2 - (\sum fX)^2 / \sum f}{\sum f - 1}}$$



分组	身高					均数	$\Sigma X$	$\Sigma X^2$
甲组	90	95	100	105	110	100	500	50250
乙组	96	98	100	102	104	100	500	50040
丙组	96	100	99	101	104	100	500	50034

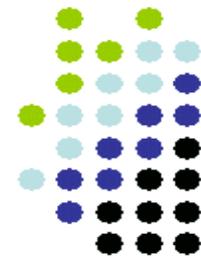
$$\text{甲组: } S = \sqrt{\frac{50250 - \frac{500^2}{5}}{5-1}} = 7.9 (cm)$$

$$\text{乙组: } S = \sqrt{\frac{50040 - \frac{500^2}{5}}{5-1}} = 3.2 (cm)$$

$$\text{丙组: } S = \sqrt{\frac{50034 - \frac{500^2}{5}}{5-1}} = 2.9 (cm)$$

# 变异系数

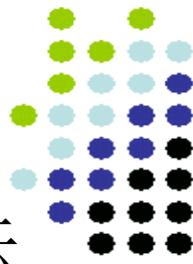
---



变异系数(coefficient of variation, CV)

$$CV = \frac{S}{\bar{X}} \times 100\%$$

- 适用条件:
- ①观察指标单位不同, 如身高、体重
  - ②同单位资料, 但均数相差悬殊



- 
- 如某地7岁男孩身高的均数为123.10 cm，标准差为4.71 cm；体重均数为22.91 kg,标准差为2.26 kg。那个指标的变异度大。

$$\text{身高: } CV = \frac{S}{\bar{X}} \times 100\% = \frac{4.71}{123.10} \times 100\% = 3.83\%$$

$$\text{体重: } CV = \frac{S}{\bar{X}} \times 100\% = \frac{2.26}{22.29} \times 100\% = 10.14\%$$