第二章 描述性统计

2－1 资料的整理

一、资料的类型：（一）质量性状资料

（二）数量性状资料

二、资料的搜集：（一）调查

（二）试验

三、资料的整理方法：（一）次数分布表

（二）次数分布图与频率分布图

（三）研究频数分布的意义

2－2 基本特征数

一、平均数

二、变异数

三、偏斜度与峭度 －－ －－ －－教材中没有

四、试验数据中异常值的分析－－教材中没有

2－1 资料的整理

一、资料的类型

男女特征

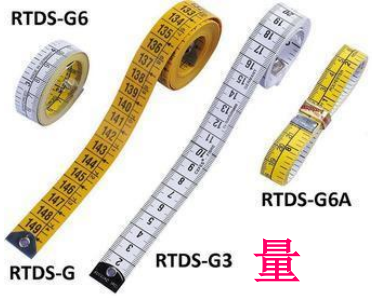
描述：特征－－指标

定性

定量

	体重	身高	年龄	容貌	吃饭	花钱	生活费	零花钱
男生	重	高	大	帅	多	多	多	多
女生	轻	矮	小	靓	少	少	少	少

称



量

测

Y		x1		x2	
平均	63.8571	平均	172.417	平均	20.8643
标准误差	1.39689	标准误差	0.81648	标准误差	0.11194
中位数	60.5	中位数	172	中位数	21
众数	60	众数	177	众数	21
标准差	9.05288	标准差	5.29141	标准差	0.72543
方差	81.9547	方差	27.999	方差	0.52625
峰度	0.8388	峰度	-0.9797	峰度	-0.7632
偏度	1.01457	偏度	0.16925	偏度	0.07571
区域	40	区域	19	区域	2.5
最小值	50	最小值	163	最小值	19.5
最大值	90	最大值	182	最大值	22
求和	2682	求和	7241.5	求和	876.3
观测数	42	观测数	42	观测数	42

体重		身高		年龄	
平均	49.0526	平均	160.816	平均	20.7421
标准误差	0.77381	标准误差	0.87956	标准误差	0.1461
中位数	50	中位数	160	中位数	21
众数	50	众数	160	众数	21
标准差	4.77011	标准差	5.42195	标准差	0.90064
方差	22.7539	方差	29.3976	方差	0.81115
峰度	0.49155	峰度	-0.1281	峰度	-0.4794
偏度	0.32831	偏度	0.33512	偏度	-0.3952
区域	22	区域	24	区域	3
最小值	38	最小值	150	最小值	19
最大值	60	最大值	174	最大值	22
求和	1864	求和	6111	求和	788.2
观测数	38	观测数	38	观测数	38

Totel		月生活费		月零花钱	
平均	458.421	平均	291.579	平均	166.842
标准误差	32.5007	标准误差	27.6704	标准误差	20.1319
中位数	400	中位数	240	中位数	200
众数	400	众数	200	众数	100
标准差	141.667	标准差	120.612	标准差	87.753
方差	20069.6	方差	14547.4	方差	7700.58
峰度	-0.3116	峰度	-0.5373	峰度	1.0923
偏度	0.86857	偏度	1.0945	偏度	0.89476
区域	450	区域	320	区域	350
最小值	300	最小值	180	最小值	50
最大值	750	最大值	500	最大值	400
求和	8710	求和	5540	求和	3170
观测数	19	观测数	19	观测数	19

Totel		月生活费		月零花钱	
平均	361.111	平均	244.444	平均	116.667
标准误差	16.4474	标准误差	16.1128	标准误差	10.6948
中位数	400	中位数	200	中位数	100
众数	400	众数	200	众数	100
标准差	69.7802	标准差	68.3608	标准差	45.3743
方差	4869.28	方差	4673.2	方差	2058.82
峰度	-0.6873	峰度	-0.2506	峰度	0.09541
偏度	-0.0066	偏度	0.84458	偏度	0.82646
区域	250	区域	250	区域	150
最小值	250	最小值	150	最小值	50
最大值	500	最大值	400	最大值	200
求和	6500	求和	4400	求和	2100
观测数	18	观测数	18	观测数	18

指标量化方法：　　称、量、测、计数

估——容貌

品德

我讲课如何?

赋值——性别（男＝1，女＝0）

是党员吗（是＝1，否＝0）

	体重	身高	年龄	容貌	吃饭	花钱	生活费	零花钱
男生	63.9	172.4	20.9			458	292	167
女生	49.1	160.8	20.7			361	244	117

- 女：有香港 18
- 男：无香港

- 湖北人：<平均

2－1 资料的整理

一、资料的类型

（一）质量性状资料：　　特征

统计次数法：（男女，血型，肤色，花色）每种类型的次数

评分法：病害程度（0级、1级、2级、……）

（二）数量性状资料：

计数－－间断性变量：班级人数，种子粒数

计量－－连续性变量：身高，体重，植株高度，种子粒重

注意：由于观测方法的不同，质量性状可以转化为数量性状。例如人的肤色、花色等。

肉眼看 --- 仪器测

在可见光中，颜色越深的光线其波长越短

二、资料的搜集

（一）调查：

普查＝全面调查：生物学研究中应用极少

抽样调查：随机抽样必须满足的条件：机会均等；

相互独立。

全省究竟有多少人

第六次人口普查启动

本报讯(记者彭肇一 通讯员彭宪明)昨日，省政府第六次人口普查领导小组召开会议，标志着我省第六次人口普查工作全面启动。领导小组组长、常务副省长李宪生在会上发表重要讲话。

据介绍，全省第六次人口普查标准时间点为2010年11月1日零时。此次人口普查的主要任务是查准人口总量，查清人口结构，查实人口迁移流动状况。

届时，全省约40万名专业普查员将进入普查区逐户、逐人进行询问，询问内容主要包括姓名、性别、年龄、受教育程度等多项内容。

根据国务院有关规定，我国人口普查每十年进行一次。我省第五次人口普查于2000年进行，普查结果为：截至2000年11月1日零时，我省总人口为6027.82万人。

人口普查，十年一次

民调

（二）试验：

在研究中，对无限总体，一般通过试验获取样本资料。

电视收视率

- 电视收视率是指某一时段内收看某电视频道(或某电视节目)的人数(或家户数)占电视观众总人数(或家户数)的百分比。
- 虽然收视率本身只是一个简单的数字，但是在看似简单的数字背后却是一系列科学的基础研究、抽样和建立固定样组、测量、统计和数据处理的复杂过程。

国民党 -- 民进党

民调与政权

- 2010年3月3日及3月3日晚间调查
- 成功访问 871 位设籍台北县的成年民众，另 320 人拒访；
- 在 95% 的信心水平下，抽样误差在正负 3.3% 以内。
- 调查以台北县的住宅电话为母体作尾数两位随机抽样，并依据台北县民众之性别和年龄结构进行加权。

1082.2375万票

投票
唱票

三、资料的整理方法

（一）次数分布表

选班长

评先进

姓名	得票			得票数
赵一	正正丅	正正丅	丅	16
钱二	正正丅	正一		12
张三	正正丅	丅		9
李四	正正一			6
王五	正一			4
刘六	丅			2
田七	一			1
			总数	50

三、资料的整理方法

（一）次数分布表

1、计数资料：

例P11

从表2-1得到表2-2

见右边表、图

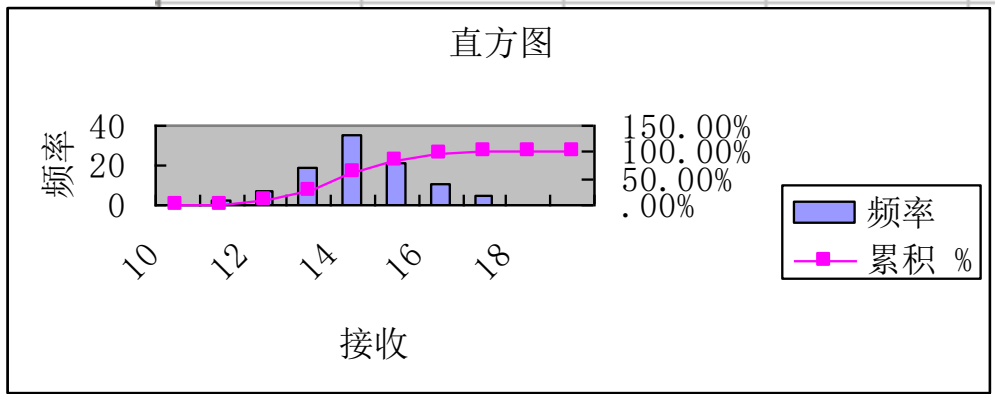
次数

频率＝次数/总次数

累积频率

例P11：100只鸡每月产蛋数

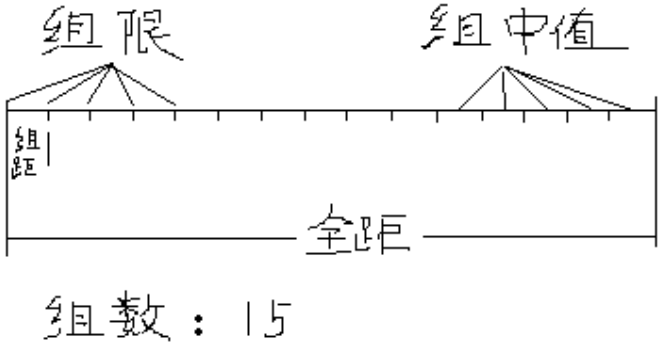
产蛋数量	次数	频率	累积频率
11	2	0.02	0.02
12	7	0.07	0.09
13	19	0.19	0.28
14	35	0.35	0.63
15	21	0.21	0.84
16	11	0.11	0.95
17	5	0.05	1



例P12：300个麦穗穗粒数

2、计量资料：表2－4

表2－6



全距＝最大值－最小值

组数：表2－5　样本容量与分组数的关系

组距＝全距／组数

组限：每个组变量值的起止界限（下限、上限）

组中值＝　（下限＋上限）／2

例P12：

150尾鲢鱼的体长

表 2-6　150 尾鲢鱼体长(cm)的次数分布表

组限/cm	组中值/cm	次　数	频　率	累积频率
35～	37.5	3	0.0200	0.0200
40～	42.5	4	0.0267	0.0467
45～	47.5	17	0.1133	0.1600
50～	52.5	28	0.1867	0.3467
55～	57.5	40	0.2666	0.6133
60～	62.5	25	0.1667	0.7800
65～	67.5	17	0.1133	0.8933
70～	72.5	6	0.0400	0.9333
75～	77.5	7	0.0467	0.9800
80～	82.5	2	0.0133	0.9933
85～	87.5	1	0.0067	1.0000

样本容量与分组数

向书坚、张学毅：《统计学》

(1) 确定组数

组数取决于原始数据中数据的多少和变量值极差的大小，极差等于数据中最大变量值与最小变量值之差，亦称全距。一般而言，数据越多，极差越大，分组数目就应该越多一些。但根据惯例，人们很少使用少于 6 个或大于 16 个的分组数目。

美国学者 Sturges 提出，当总体的分布接近正态分布时，可以根据数据个数 n 来近似地确定分组数目 k，经验公式为：

$$k = 1 + \frac{\lg n}{\lg 2} \qquad (2.1)$$

根据式(2.1)，可得到如下的分组数目查对表(表 2-8)。

表 2-8 经验分组数目查对表

n	15—24	25—44	45—89	90—179	180—359	360—719	720—1 439
k	5	6	7	8	9	10	11

式(2.1)只能作为确定组数的参考，不能生搬硬套，因为原始数据的分布并不都是接近正态分布的，需要结合实际情况进行分析后决定是否采用。

李春喜：表2-5

样本容量	分组数
30-50	5-8
60-100	7-10
100-200	9-12
200-500	10-18
500以上	15-30

三、资料的整理方法

（一）次数分布表

（二）次数分布图与频率分布图

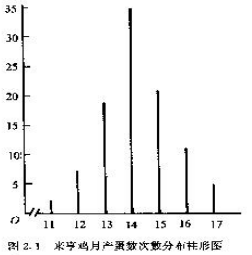


图2-1　条形图

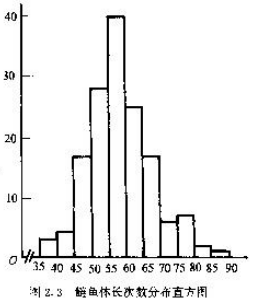


图2-3　直方图

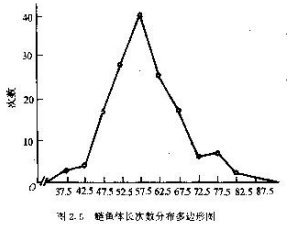


图2-4　多边形图

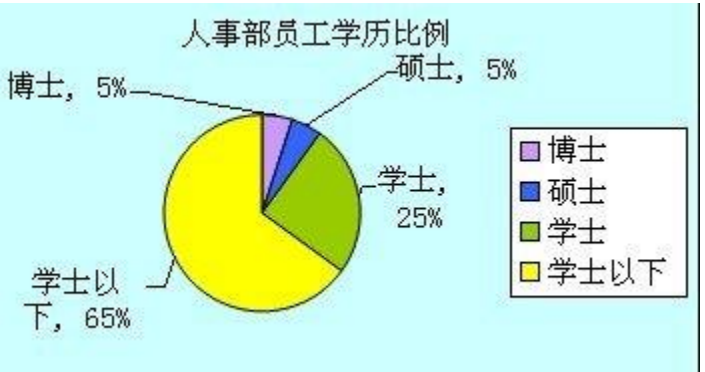


图2-2　饼图

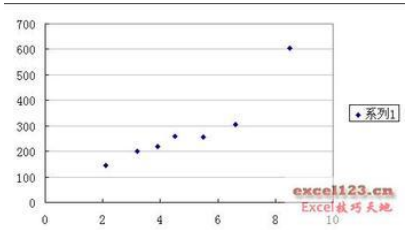
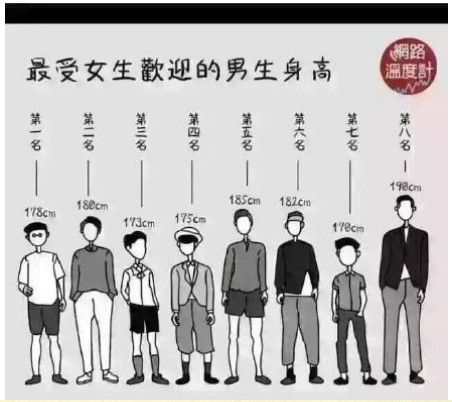


图2-5　散点图



一个连的士兵的身高的分布

1.58-1.60 1.62-1.64 1.66-1.68 1.70-1.72 1.74-1.76 1.78-1.80 1.82-1.84 1.86-

The distribution of height in a company of soldiers (from Stern, 1973).

遗传学分离比例

三、资料的整理方法

1 : 3

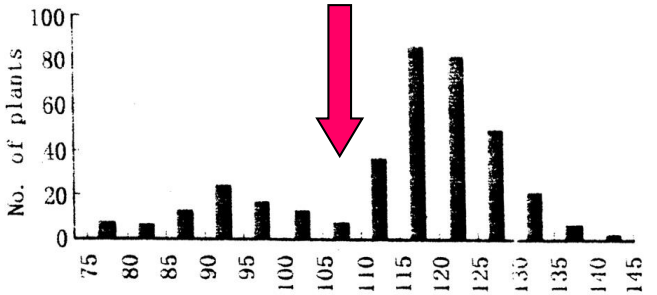
（三）研究频数分布的意义

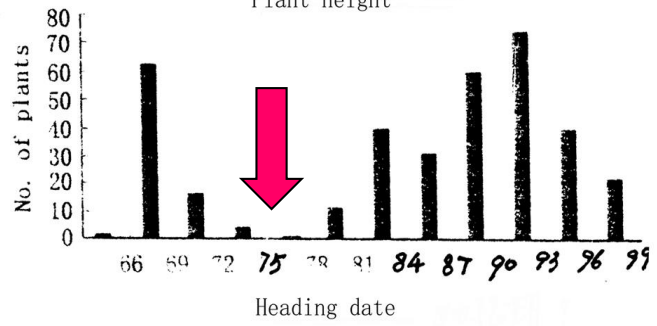
统计学角度：集中程度

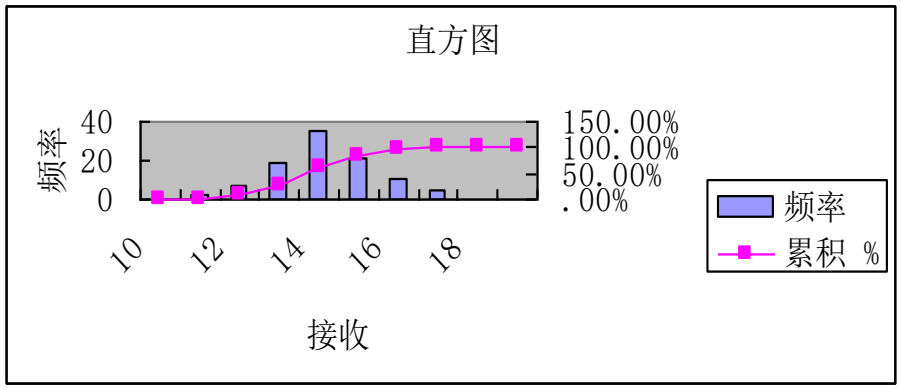
分散程度

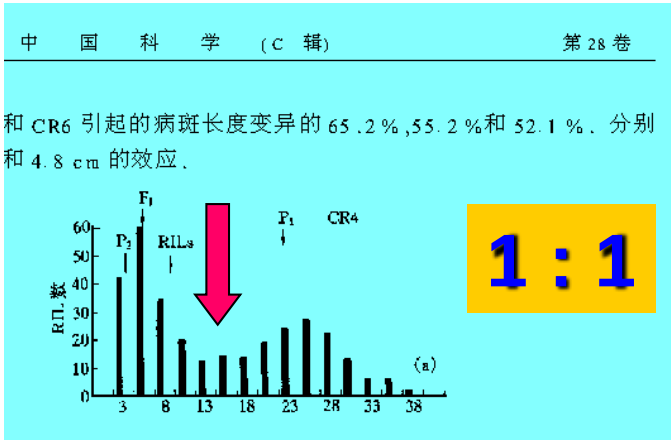
图的形状

专业知识角度：









2－2 基本特征数

一、平均数　　集中程度

二、变异数　　分散程度

三、偏斜度与峭度－－教材中没有　　图的形状

2－2 基本特征数

一、 平均数

（一）平均数的种类

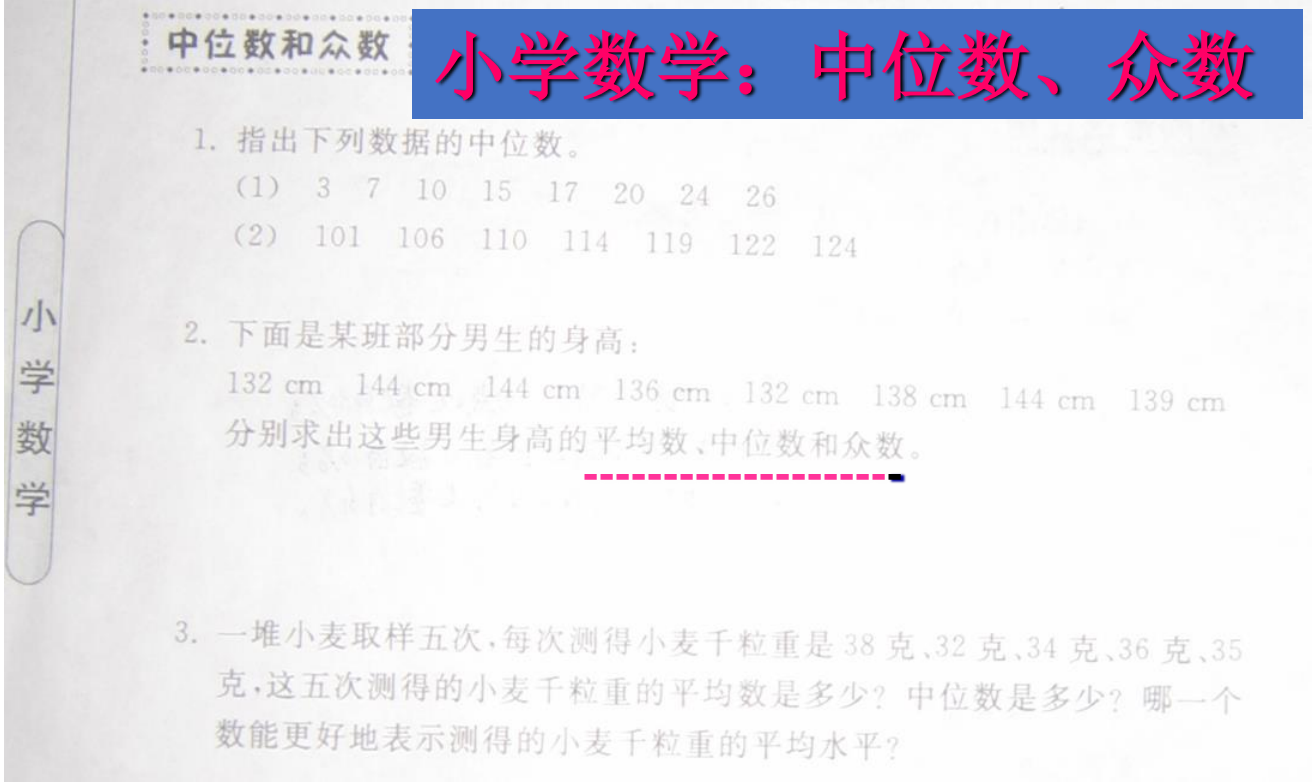
（二）算术平均数的计算方法

（三）算术平均数的特性

（四）算术平均数的作用

（五）附：求和符号的三个运算法则

小学数学：中位数、众数



一、 平均数

（一）平均数的种类：

算术平均数：

几何平均数：P17

调和平均数

中位数： {2，5，8} {3，6，9} {1，4，7，10}

众数： 聚众闹事

直接平均法
加权平均法

（二）算术平均数的计算方法：

P17：未分组资料： $\bar{x} = (\sum x_i) / n$

分组资料： $\bar{x} = (\sum f_i x_i) / n$ $\quad n = \sum f_i$

例： {2，5，8} $\quad \bar{x} = (\sum x_i) / n = 15 / 3 = 5$

{3，6，9} $\quad \bar{x} = (\sum x_i) / n = 18 / 3 = 6$

{1，4，7，10} $\quad \bar{x} = (\sum x_i) / n = 22 / 4 = 5.5$

例： {6，6，6，8，8，8}

$\bar{x} = (\sum f_i x_i) / n = (3 \times 6 + 3 \times 8) / 6 = 7$

（三）算术平均数的特性：

P18：$\sum (x_i - \bar{x}) = 0$

$\sum (x_i - \bar{x})^2 < \sum (x_i - a)^2$

（四）算术平均数的作用：

P19：1、指出数据资料的中心位置，标志着资料所代表性状的数量水平和质量水平；

2、作为样本或资料的代表数与其它资料进行比较。

（五）附：求和符号的三个运算法则

$$\sum c = nc$$

$$\sum c\,x_i = c\sum x_i$$

$$\sum (x_i \pm y_i) = \sum x_i \pm \sum y_i$$

例：$\{8，8，8\}$　　$\sum c = nc = 3\times 8 = 24$

例：$\{2c，5c，8c\}$　$\sum c\,x_i = c\sum x_i = c\times (2+5+8) = c\times 15$

例：$\{2，5，8\}$

$\{3，6，9\}$　　$\sum (x_i \pm y_i) = \sum x_i \pm \sum y_i$

$= (2+5+8) \pm (3+6+9)$

$= 15 \pm 18$

第二章 描述性统计

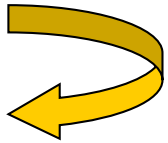
2－2 基本特征数

二、 变异数

（一）极差：

（二）方差：

（三）标准差：



平方
开方

（四）变异系数：

（一）极差： 两极之差

$R = \max \{x_1, x_2, \cdots, x_n\} - \min \{x_1, x_2, \cdots, x_n\}$

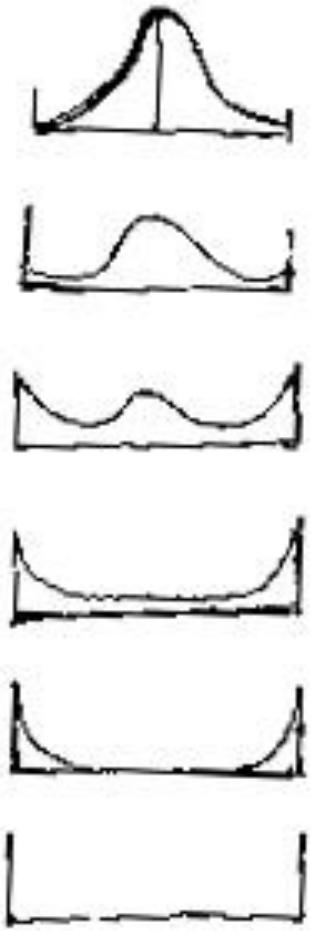
它只反映两极的信息，未反映中间的信息，所以它不是最佳指标。

例： {2，5，8} R＝6

{3，6，9} R＝6

{1，4，7，10} R＝9

{6，6，6，8，8，8} R＝2



均方

（二）方差： 样本 方差 $S^2 = \sum (x_i - \bar{x})^2 / (n-1)$

总体 方差 $\sigma^2 = \sum (x_i - \mu)^2 / N$

（三）标准差 ： 样本 标准差 S

总体 标准差 σ

1、种类：样本方差，样本标准差

总体方差，总体标准差

2、来历: $\sum (x_i - \bar{x}) = 0$，

$\sum (x_i - \bar{x})^2 = ?$

自由度：n－1

3、 作用： P22

4、 计算： P20 未分组资料 P21 分组资料

P20 未分组资料

例：{2，5，8}

$$S^2 = \sum (x_i - \bar{x})^2 / (n-1)$$
$$= [(2-5)^2 + (5-5)^2 + (8-5)^2] / (3-1) = 9$$

例：{3，6，9}

$$S^2 = \sum (x_i - \bar{x})^2 / (n-1)$$
$$= [(3-6)^2 + (6-6)^2 + (9-6)^2] / (3-1) = 9$$

例：{1，4，7，10}

$$S^2 = \sum (x_i - \bar{x})^2 / (n-1) = 15$$

P21 分组资料

例：{6，6，6，8，8，8}

$$S^2 = \sum f_i (x_i - \bar{x})^2 / (n-1)$$
$$= [3 \times (6-7)^2 + 3 \times (8-7)^2] / (6-1)$$
$$= 1.2$$

（四）变异系数：用于比较两个或多个样本的变异程度

$$CV = (s / \bar{x}) \times 100\%$$

例：

{2，5，8} S＝3 $\bar{x}$＝5 $CV = (s / \bar{x}) \times 100\% = 60\%$

{3，6，9} S＝3 $\bar{x}$＝6 $CV = (s / \bar{x}) \times 100\% = 50\%$

{1，4，7，10} S＝3.9 $\bar{x}$＝5.5 $CV = (s / \bar{x}) \times 100\% = 70\%$

例：三个人的钱：

{20，50，80} S＝30 $\bar{x}$＝50 CV ＝60%

{10020，10050，10080} S＝30 $\bar{x}$＝10050 CV ＜0.3%

极差 **R = 60**

第二章 描述性统计

2-2

楚天都市报 电话:86795929 13886158070 E-mail:xuhanxiong@163.net 三镇街头 责任编辑 徐汉雄 版面编辑 柳 逸 2004.3.25 星期四 23

怀揣20元 异乡过一周

体验生存,河南大学生江城"出招"

本报记者易敏 刘中灿 实习生李黎

昨日中午，记者经过武昌水果湖路时，看到一名女孩站在人行道上，摆出一种奔跑的姿势，却又一动不动，仿如一尊雕塑。她的身边放着一个纸牌，上写"行为艺术：给钱就动……"纸牌旁的一个帽子里，装有一些零钱。

女孩名叫王娟，20岁，河南省漯河职业技术学院大三学生。她说，来汉是参加学校组织的一个"就业生存挑战赛"活动，以了解社会。3月21日晚，学校用车将20名选手(14男6女)送到武汉，22日上午解散。每个选手身上都只有20元钱，要求通过各种方式在汉找工作、生活一个星期。另有4名老师坐镇武汉，并给每名选手配了张30元的手机卡，以便联系。每个选手还有一个"锦囊"，遇到困境时可以打开，若实在走投无路，还可向老师求救。

记者看到，王娟的记事本上写着：2元钱买了张地图；2元钱坐公交车；1元钱买了包子，充当早饭、晚饭……她说，还没有找到满意的工作，已花去10多元钱，所以想出在街头表演的办法为生。昨日上午，1个多小时的表演，赚来10.5元。

为了"落脚"，她跑了一所高校的许多女生寝室，最后终于说动一间寝室，以做清洁、打开水交换晚上免费留住。

记者与其所在的学校取得联系，得知该校就业指导处吕爱华处长就在武汉。

吕老师证实了此事，并说，从目前掌握的情况看，20人大多找到了工作，或有了落脚处，活动成功有望。

俗话说，急中生智，看了王娟同学的生存"奇招"，不由感叹：的确如此。

面临困境时，常人难免悲观消沉，或是急火攻心、不知所措；甚至因此"剑走偏锋"，做出些错事、蠢事也是有的。而王娟等同学，却成功地应对了20元异乡讨生活的挑战。

一种有益的尝试

余人月

这种生存体验，对挑战自己的智力与观念，无疑是极有益的尝试。对我等常人，也是一种启示：面临类似考验时，能否放低身段，开启心智，设法做些力所能及的事，绝处求生？而不是抱着"等、靠、要"的思想，怨天尤人。

人的潜力无穷尽，倘遇波折，积极面对，努力向上，必定能战胜自己，迎来"柳暗花明"。

例：三个人的钱：

{20，50，80}　　S=30　$\bar{x}$=50　CV =60%

{10020，10050，10080}　　S=30　$\bar{x}$=10050　CV <0.3%

前者会有生与死的差别
后者仅有饱与饥的差别

CV的优点与缺点：

《Biostatistics:The Bare Essentials》
《生物统计学基础》，凌莉主译，P25

- **优点：**
- **缺点：不足之处**

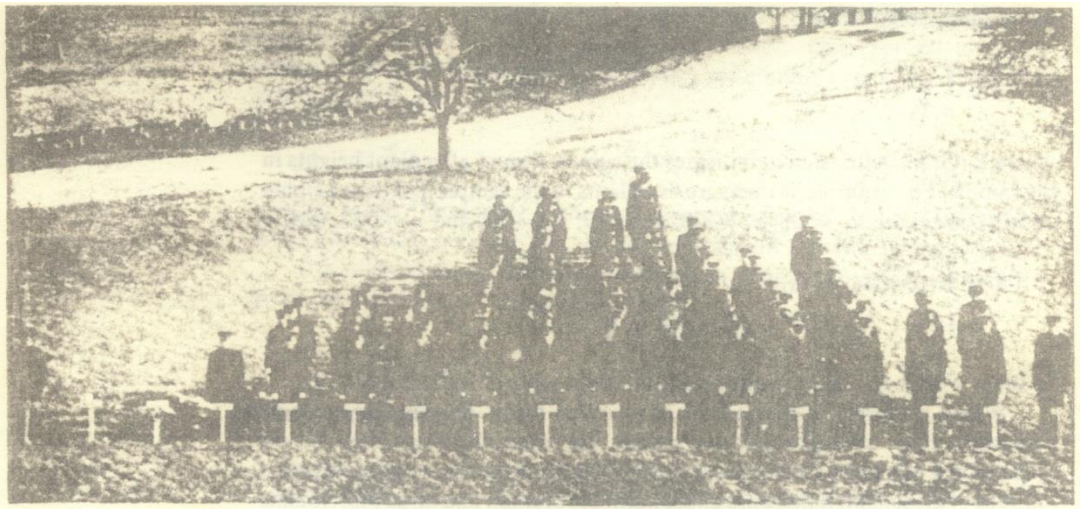
第一，每个统计检验几乎都会包含SD，但不能将CV以某种方式引入到任何一种检验中--------无法进行检验。

第二，更重要的是，CV对测量标准很敏感。

每个数字乘以一个常数，CV不变，很好

每个数字加上一个常数，均值增加导致CV下降

总结：



The distribution of height in a company of soldiers (from Stern, 1973).

计算公式

平均数＝ $\bar{x} = (\sum x_i) / n$

方差＝ $S^2 = \sum (x_i - \bar{x})^2 / (n-1)$

标准差＝S

$$\bar{x} \pm S = 172.4 \pm 5.3$$

相当于 68.26% 的观察值落在此区间

Y		x1		x2	
平均	63.8571	平均	172.417	平均	20.8643
标准误差	1.39689	标准误差	0.81648	标准误差	0.11194
中位数	60.5	中位数	172	中位数	21
众数	60	众数	177	众数	21
标准差	9.05288	标准差	5.29141	标准差	0.72543
方差	81.9547	方差	27.999	方差	0.52625
峰度	0.8388	峰度	-0.9797	峰度	-0.7632
偏度	1.01457	偏度	0.16925	偏度	0.07571
区域	40	区域	19	区域	2.5
最小值	50	最小值	163	最小值	19.5
最大值	90	最大值	182	最大值	22
求和	2682	求和	7241.5	求和	876.3
观测数	42	观测数	42	观测数	42
体重		身高		年龄	
平均	49.0526	平均	160.816	平均	20.7421
标准误差	0.77381	标准误差	0.87956	标准误差	0.1461
中位数	50	中位数	160	中位数	21
众数	50	众数	160	众数	21
标准差	4.77011	标准差	5.42195	标准差	0.90064
方差	22.7539	方差	29.3976	方差	0.81115
峰度	0.49155	峰度	-0.1281	峰度	-0.4794
偏度	0.32831	偏度	0.33512	偏度	-0.3952
区域	22	区域	24	区域	3
最小值	38	最小值	150	最小值	19
最大值	60	最大值	174	最大值	22
求和	1864	求和	6111	求和	788.2
观测数	38	观测数	38	观测数	38

总结

吕应堂教授实验室宣传栏

代表性成果之一：

拟南芥迟开花突变体的获得

	叶片数	开花时间
突变型	12.5 ±0.6	45.1 ±1.6
野生型	8.7 ±1.5	31.6 ±1.1

2－2 基本特征数

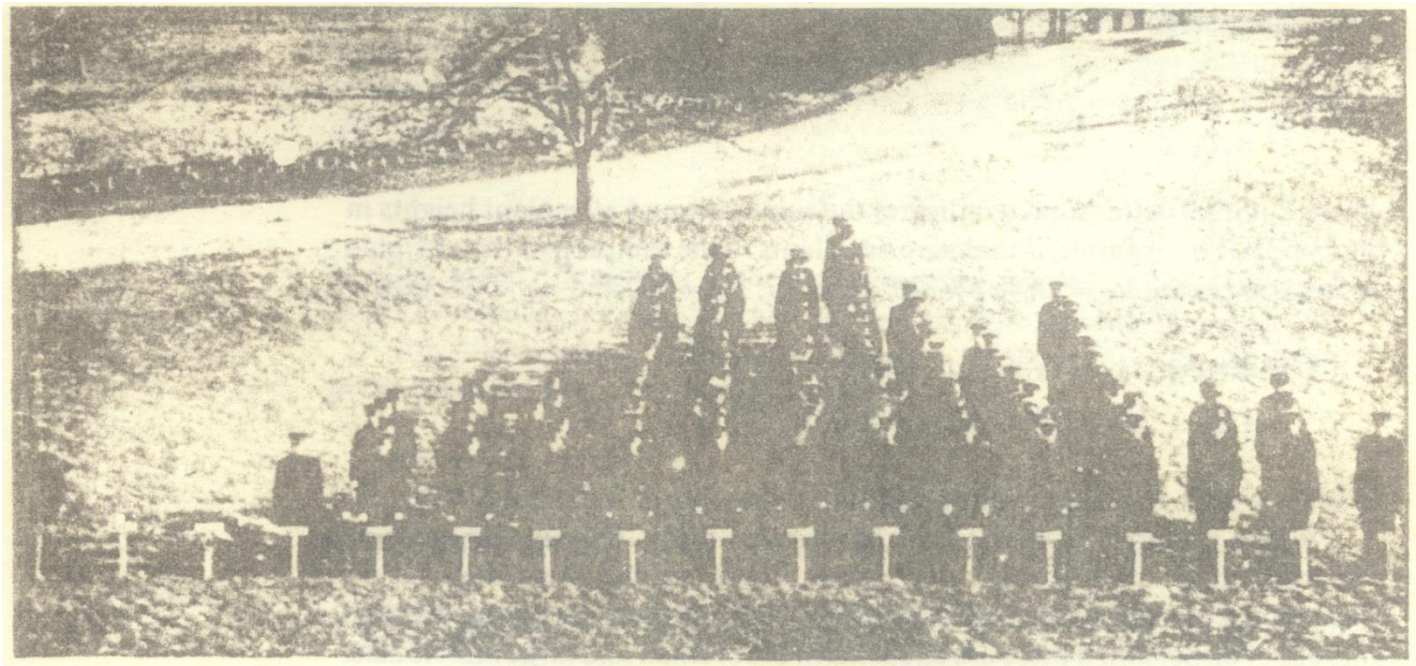
三、偏斜度与峭度－－教材中没有

（一）中心矩与原点矩

（二）偏斜度

（三）峭度

（四）应用



The distribution of height in a company of soldiers (from Stern, 1973).

三、偏斜度与峭度

（一）中心矩与原点矩

	中心矩	原点矩
K阶	$m_k = \Sigma (x - \bar{x})^k / n$	$m'_k = \Sigma x^k / n$
1阶		$m'_1 = \Sigma x / n = \bar{x}$
2阶	$m_2 = \Sigma (x - \bar{x})^2 / n = \sigma^2$	$m'_2 = \Sigma x^2 / n$
3阶	$m_3 = \Sigma (x - \bar{x})^3 / n$	$m'_3 = \Sigma x^3 / n$
4阶	$m_4 = \Sigma (x - \bar{x})^4 / n$	$m'_4 = \Sigma x^4 / n$

K = 1,2,3,4,5,6,7，，，，，，

最有用的是什么？

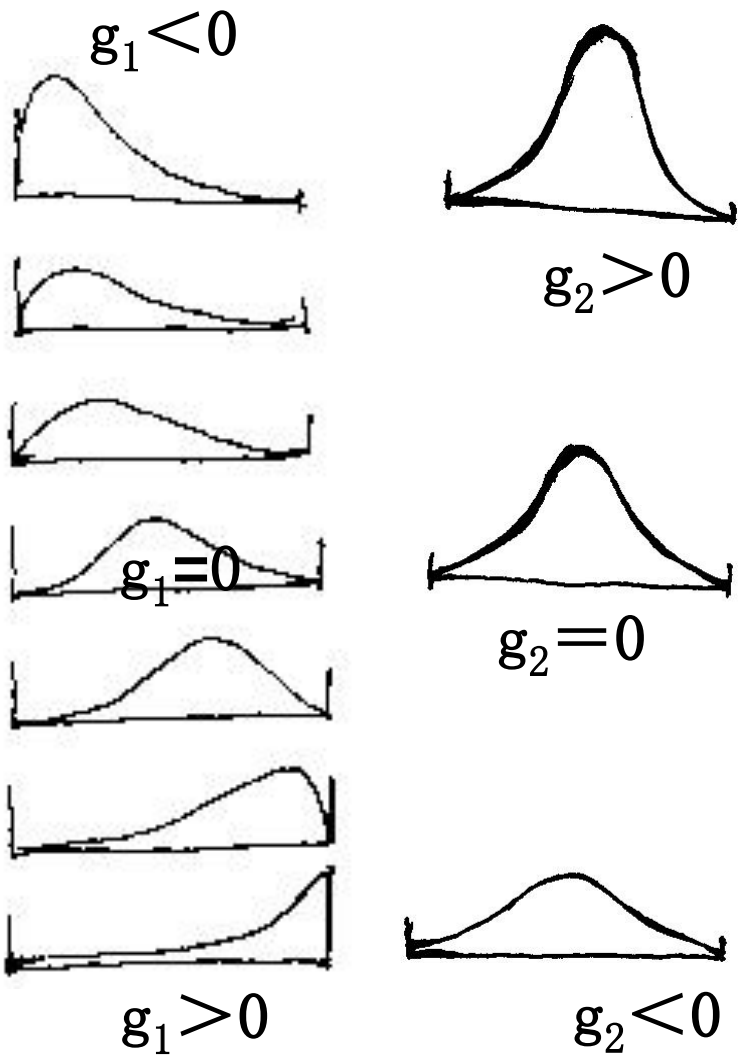
以标准正态分布为标准

（二）偏斜度 ： g_1

$$g_1 = m_3 / \sigma^3 = m_3 / m_2^{3/2}$$

（三）峭度： g_2

$$g_2 = m_4 / \sigma^4 - 3 = m_4 / m_2^2 - 3$$



当 n 较小时

（二）偏斜度：g_1　　三阶

$$g_1 = \frac{n}{(n-1)(n-2)}\sum\left(\frac{x_i - \bar{x}}{s}\right)^3 \qquad SEg_1 = \left[\frac{6n(n-1)}{(n-2)(n+1)(n+3)}\right]^{1/2} \to \sqrt{\frac{6}{n}}$$

（三）峭度：g_2　　四阶

$$g_2 = \left[\frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum\left(\frac{x_i - \bar{x}}{s}\right)^4\right] - \frac{3(n-1)^2}{(n-2)(n-3)} \qquad SEg_2 = \left[\frac{24n(n-1)}{(n-3)(n-2)(n+3)(n+5)}\right]^{1/2} \to \sqrt{\frac{24}{n}}$$

向书坚、张学毅主编《统计学》，北京大学出版社，2009

计算偏态系数的公式有很多，但常用的方法主要有皮尔逊测度法和中心矩法两种。

1. 皮尔逊测度法

皮尔逊测度法就是利用算术平均数与众数的关系来测度数据分布偏斜程度的一种方法。我们知道，对于完全对称分布，算术平均数、众数和中位数三者必然相等，若不相等，则分布呈偏态。在偏态分布情况下，三者之间存在差距。分布越偏斜，算术平均数与众数或中位数的距离越远。因此，通常用算术平均数与众数的差距除以标准差来测定偏态的偏斜程度。计算公式为：

$$SK_p = \frac{\bar{x} - M_o}{\sigma} \tag{3.29}$$

经验证明，在适度偏态的情况下，$-3 \leqslant SK_p \leqslant 3$。

当 $\bar{x} = M_o$，$SK_p = 0$ 时，数据分布呈对称分布；

当 $\bar{x} > M_o$，$SK_p > 0$ 时，数据分布呈右（正）偏分布；

当 $\bar{x} < M_o$，$SK_p < 0$ 时，数据分布呈左（负）偏分布。

2. 中心矩法

（四）应用：

数量遗传学：
DH群体 Doubled-haploid population

无基因互作时：
$g_1=0$　$g_2\leqslant 0$

有基因互作时：
$g_1 \neq 0$　$g_2>0$
$g_1 >0$　$g_2>0$　互补作用
$g_1 <0$　$g_2>0$　重叠作用

武汉植物学研究 2000, 18(5): 431－434
Journal of Wuhan Botanical Research

水稻苗期根系性状的数量遗传分析

陈明明　董　伟　胡中立*　章志宏
（武汉大学生命科学学院　武汉　430072）

QUANTITATIVE GENETIC ANALYSIS OF SEEDLING ROOT CHARACTERS IN RICE

Chen Mingming　Dong Wei　Hu Zhongli*　Zhang Zhihong
(College of Life Sciences, Wuhan University　Wuhan　430072)

关键词　水稻，加倍单倍体群体，根系性状，遗传力，基因互作

表 1　亲本、DH 群体的性状表现及各性状遗传参数的估计

Table 1　Performance of the characters studied for the parents and the DH population, Estimates of heritability(h^2), number of genetic factors(k) involved in the characters, and the coefficients of skewness(g_1) and kurtosis(g_2) in the distribution of the DH population

试验 Test	性状 Characters	亲本 Parents		DH 群体 DH population		遗传力 h^2	基因数目 k	偏度系数 g_1	峰度系数 g_2
		圣630 Gui630	02428	均值 Means	变幅 Range				
T	新发根率 Percentage of newly-developed roots	9.35	11.41	6.83	1.71－18.54		11.4	1.3408**	2.0367**
	新发根干重 Dry weight of newly-developed roots	2.22	2.52	1.86	0.22－5.26		11.6	0.9736**	0.2587
	地上部干重 Dry weight of plant part above the ground	23.73	22.08	25.71	11.57－46.15		7.7	0.6062**	0.5645
D	新发根率 Percentage of newly-developed roots	2.37	3.09	2.22	0.32－6.18		12.5	1.1678**	1.9440**
	新发根干重 Dry weight of newly-developed roots	4.70	5.20	3.60	0.77－12.07		13.5	1.5039**	2.4305**
	地上部干重 Dry weight of plant part above the ground	198.46	168.31	160.81	43.07－285.04		6.4	0.0476	0.0353
	新发根数 No. of newly-developed roots	25.3	23.4	14.41	6－39	0.66	8.4	0.8975**	0.2836
	苗高 Plant height	71.1	82.9	73.96	54.2－98.6	0.33	10.8	－0.0692	0.2071
	单株分蘖数 No. of tillers per plant	2.60	1.00	1.13	0－3	0.45	6.7	0.2902	－0.5625

（四）应用：

金融数学：

Tobin ：不要把鸡蛋放在一个篮子里

分散风险

1981年诺贝尔经济学奖　《经济学论文集:消费和经济计量学》(1975年)

Markowitz ：证券组合模型

投资者希望证券组合具有：最大的期望收益

最小的方差

最大的偏斜度

研究涉及金融微观分析及数学、计算机在金融经济学方面的应用。
1990年诺贝尔经济学奖

Matlab在马柯维茨均值-方差模型的简单应用

踩 0　浏览 671　收藏 4　评论　+ 加入豆单　举报　手机观看

Markowitz(1952)发展了一个在不确定条件下严格陈述的可操作的资产组合选择理论：均值-方差方法 *Mean-Variance methodology.*

马科维茨(H. Markowitz, 1927～)
《证券组合选择理论》

模型理论

经典马柯维茨均值-方差模型为：

$$\begin{cases} \min \sigma_p^2 = X^T \Sigma X \\ \max E(r_p) = X^T R \\ s.t. \sum_{i=1}^{n} x_i = 1 \end{cases}$$

其中，$R = (R_1, R_2, \ldots, R_n)^T$；$R_i = E(r_i)$ 是第 i 种资产的预期收益率；$X = (x_1, x_2, \ldots, x_n)^T$ 是投资组合的权重向量；$\sum = (\sigma_{ij})_{n \times n}$ 是 n 种资产间的协方差矩阵；$R_p = E(r_p)$ 和 σ_p^2 分别是投资组合的期望回报率和回报率的方差。

四、试验数据中异常值的分析：检验异常值的方法

1、3S法

当样本容量 n＞10时，如果每个观察值与其测量结果的算术平均数之差大于3倍标准差 s 时，该测量数据应舍弃。

因为，在多次试验中，测量值落在 $\bar{x}$ ± 3s 范围内的概率为0.9973

正态分布理论

2、狄克松法（Dixon检验法）

适用于样本容量为3-30的小样本

3、格拉布斯法（Grubbs检验法）

样本容量≥3，样本容量在50以上更好

顾志峰等《实用生物统计学》，科学出版社，2012

第二章 描述性统计

练习题：P19（第四版：P22-23）

2.1

2.5

2.6

2.7

2.8

注：用活页纸做作业，不用本子。

EXCEL统计软件应用