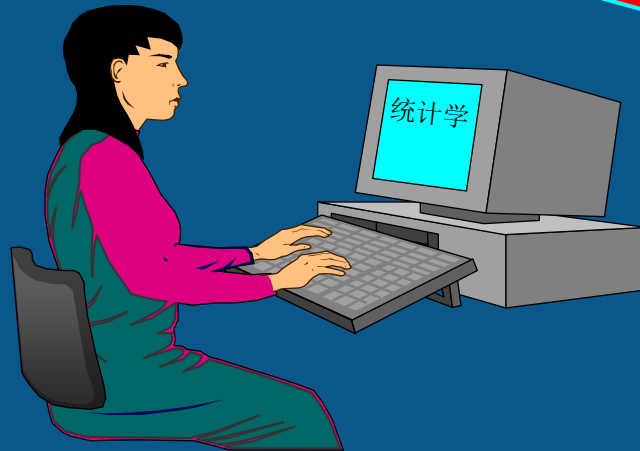


第 3 章 数据的图表展示

PowerPoint



第 3 章 数据的图表展示

3.1 数据的预处理

3.2 品质数据的整理与显示

3.3 数值型数据的整理与显示

3.4 合理使用图表

学习目标

1. 了解数据预处理的内容和目的
2. 掌握分类和顺序数据的整理与显示方法
3. 掌握数值型数据的整理与显示方法
4. 用**Excel**作频数分布表和图形
5. 合理使用图表

3.1 数据的预处理

- 3.1.1 数据审核
- 3.1.2 数据筛选
- 3.1.3 数据排序
- 3.1.4 数据透视表

数据的预处理



1. 数据审核
 - 检查数据中的错误
2. 数据筛选
 - 找出符合条件的数据
3. 数据排序
 - 升序和降序
 - 寻找数据的基本特征
4. 数据透视
 - 按需要汇总

数据审核

数据审核—原始数据 (raw data)

1. 完整性审核

- 应调查的单位或个体是否有遗漏
- 所有的调查项目或变量是否填写齐全

2. 准确性审核

- 数据是否真实反映实际情况，内容是否符合实际
- 数据是否有错误，计算是否正确等

数据的审核 — 二手数据 (second hand data)

1. 适用性审核

- 弄清楚数据的来源、数据的口径以及有关的背景材料
- 确定数据是否符合自己分析研究的需要

2. 时效性审核

- 尽可能使用最新的数据

3. 确认是否有必要做进一步的加工整理

数据筛选与排序

数据筛选

(data filter)

1. 当数据中的错误不能予以纠正，或者有些数据不符合调查的要求而又无法弥补时，需要对数据进行筛选
2. 数据筛选的内容
 - 将某些不符合要求的数据或有明显错误的数
据予以剔除
 - 将符合某种特定条件的数据筛选出来，而不
符合特定条件的数据予以剔除

数据筛选 (data filter)

用Excel进行数据筛选

8名学生的考试成绩数据

	A	B	C	D	E
1	姓名	统计学成绩	数学成绩	英语成绩	经济学成绩
2	张松	69	68	84	86
3	王翔	91	75	95	94
4	田雨	54	88	67	78
5	李华	81	60	86	64
6	赵颖	75	96	81	83
7	宋媛	83	72	66	71
8	袁方	75	58	76	90
9	陈凤	87	76	92	77

数据排序

(data rank)

1. 按一定顺序将数据排列，以发现一些明显的特征或趋势，找到解决问题的线索
2. 排序有助于对数据检查纠错，以及为重新归类或分组等提供依据
3. 在某些场合，排序本身就是分析的目的之一
4. 排序可借助于计算机完成

数据排序 (方法)

1. 分类数据的排序

- 字母型数据，排序有升序降序之分，但习惯上用升序
- 汉字型数据，可按汉字的首位拼音字母排列，也可按笔画排序，其中也有笔画多少的升序降序之分

2. 数值型数据的排序

- 递增排序：设一组数据为 x_1, x_2, \dots, x_n ，递增排序后可表示为： $x_{(1)} < x_{(2)} < \dots < x_{(n)}$
- 递减排序：可表示为： $x_{(1)} > x_{(2)} > \dots > x_{(n)}$

数据透视表

数据透视表 (pivot table)

1. 可以从复杂的数据中提取有用的信息
2. 可以对数据表的重要信息按使用者的习惯或分析要求进行汇总和作图
3. 形成一个符合需要的交叉表(列联表)
4. 在利用数据透视表时，数据源表中的首行必须有列标题

	A	B	C	D	E	F
1	编号	性别	家庭所在地区	平均月生活费(元)	月平均衣物支出(元)	买衣物首选因素
2	1	男	大型城市	800	200	价格
3	2	女	中小城市	600	180	款式
4	3	男	大型城市	1000	300	品牌
5	4	男	中小城市	400	40	价格

数据透视表

(用Excel创建数据透视表)

第1步：选择【插入】→【数据透视表】

第2步：在【表/区域】框内选定数据区域（在操作前将光标放在任意数据单元格内，系统会自动选定数据区域）。选择放置数据透视表的位置。系统默认是新工作表，如果要将透视表放在现有工作表中，选择【现有工作表】，并在【位置】框内点击工作表的任意单元格（不要覆盖数据）。点击【确定】

第3步：用鼠标右键单击数据透视表，选择【数据透视表选项】，在弹出的对话框中点击【显示】，并选中【经典数据透视表布局】，然后【确定】

第4步：将数据透视的一个字段拖至“行”位置，将“另一个字段”拖至“列”的位置（行列可以互换），再将要计数的变量拖至“值字段”位置，即可生成需要的频数分布表

3.2 品质数据的整理与展示

3.2.1 分类数据的整理与图示

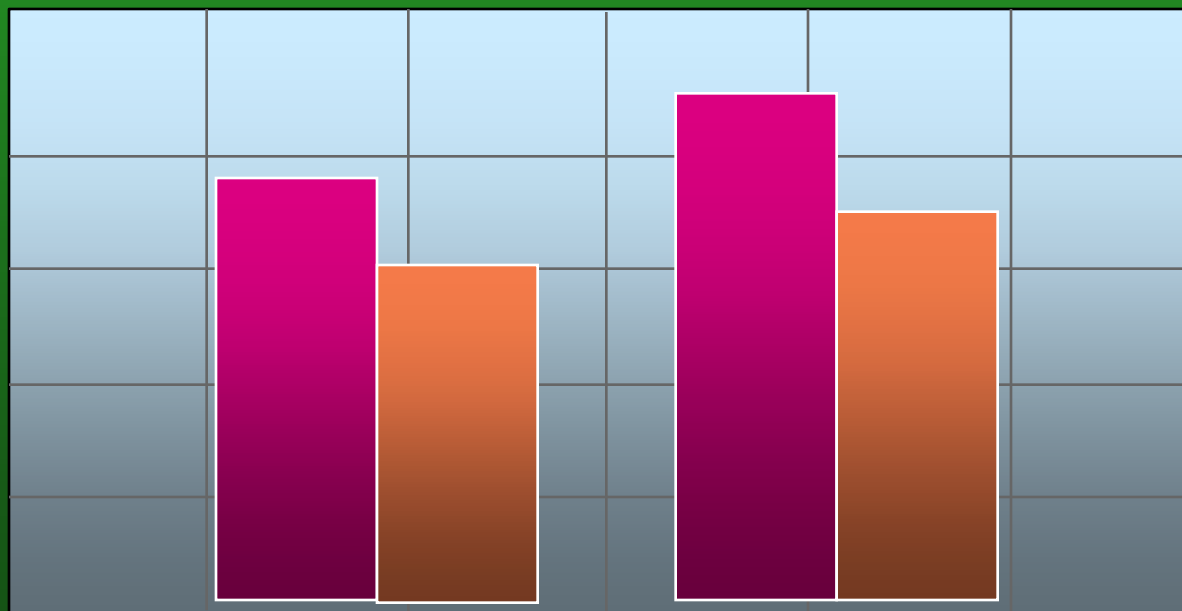
3.2.2 顺序数据的整理与图示

数据的整理与显示

(基本问题)

1. 要弄清所面对的数据类型
 - 不同类型的数据，采取不同的处理方式和方法
2. 对分类数据和顺序数据主要是作分类整理
3. 对数值型数据则主要是作分组整理
4. 适合于低层次数据的整理和显示方法也适合于高层次的数据；但适合于高层次数据的整理和显示方法并不适合于低层次的数据

分类数据的整理与图示

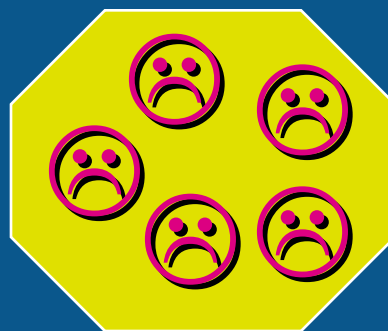


分类数据的整理 (基本过程)

1. 列出各类别
2. 计算各类别的频数
3. 制作频数分布表
4. 用图形显示数据

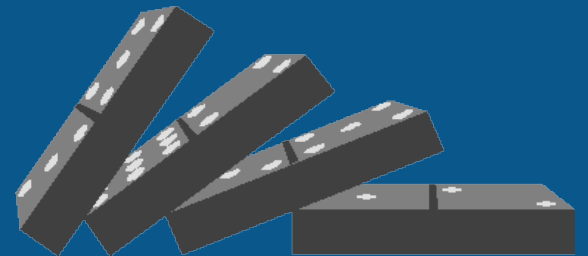


分类	频数	比例	百分比	比率
A				
B				
C				
D				
E				



分类数据的整理 (可计算的统计量)

1. 频数(**frequency**)：落在各类别中的数据个数
2. 比例(**proportion**)：某一类别数据个数占全部数据个数的比值
3. 百分比(**percentage**)：将对比的基数作为100而计算的比值
4. 比率(**ratio**)：不同类别数值个数的比值



分类数据整理—频数分布表

(例题分析)

【例】一家市场调查公司为研究不同品牌饮料的市场占有率，对随机抽取的一家超市进行了调查。调查员在某天对50名顾客购买饮料的品牌进行了记录，如果一个顾客购买某一品牌的饮料，就将这一饮料的品牌名字记录一次。右边就是记录的原始数据

	A	B	C	D	E	F
1	顾客性别	饮料类型	顾客性别	饮料类型	顾客性别	饮料类型
2	女	碳酸饮料	女	碳酸饮料	女	其他
3	男	绿茶	男	绿茶	女	碳酸饮料
4	男	矿泉水	男	其他	女	其他
5	女	矿泉水	女	碳酸饮料	女	果汁
6	男	碳酸饮料	男	绿茶	男	绿茶
7	男	矿泉水	男	绿茶	女	果汁
8	女	碳酸饮料	女	碳酸饮料	女	碳酸饮料
9	女	绿茶	男	碳酸饮料	女	果汁
10	男	果汁	女	绿茶	男	矿泉水
11	男	碳酸饮料	男	矿泉水	女	碳酸饮料
12	女	矿泉水	女	绿茶	女	绿茶
13	女	其他	女	碳酸饮料	女	其他
14	男	碳酸饮料	女	矿泉水	女	果汁
15	男	绿茶	男	其他	男	绿茶
16	男	碳酸饮料	男	碳酸饮料	女	其他
17	女	其他	女	果汁	女	矿泉水
18	男	矿泉水	男	矿泉水		



 用Excel制作频数分布表

分类数据整理—频数分布表

(例题分析)

表 3—5 不同类型饮料的频数分布

饮料类型

	频率	百分比	有效百分比	累积百分比
有效 果汁	6	12.0	12.0	12.0
矿泉水	10	20.0	20.0	32.0
绿茶	11	22.0	22.0	54.0
其他	8	16.0	16.0	70.0
碳酸饮料	15	30.0	30.0	100.0
合计	50	100.0	100.0	

表 3—6 不同性别顾客的频数分布

顾客性别

	频率	百分比	有效百分比	累积百分比
有效 男	22	44.0	44.0	44.0
女	28	56.0	56.0	100.0
合计	50	100.0	100.0	

表 3—7 饮料类型和顾客性别的交叉频数分布

饮料类型*顾客性别 交叉制表

计数

		顾客性别		合计
		男	女	
饮料类型	果汁	1	5	6
	矿泉水	6	4	10
	绿茶	7	4	11
	其他	2	6	8
	碳酸饮料	6	9	15
合计		22	28	50

分类数据整理—频数分布表

(例题分析)

表 3—8 顾客性别和饮料类型的列联表及其分析

饮料类型* 顾客性别 交叉制表

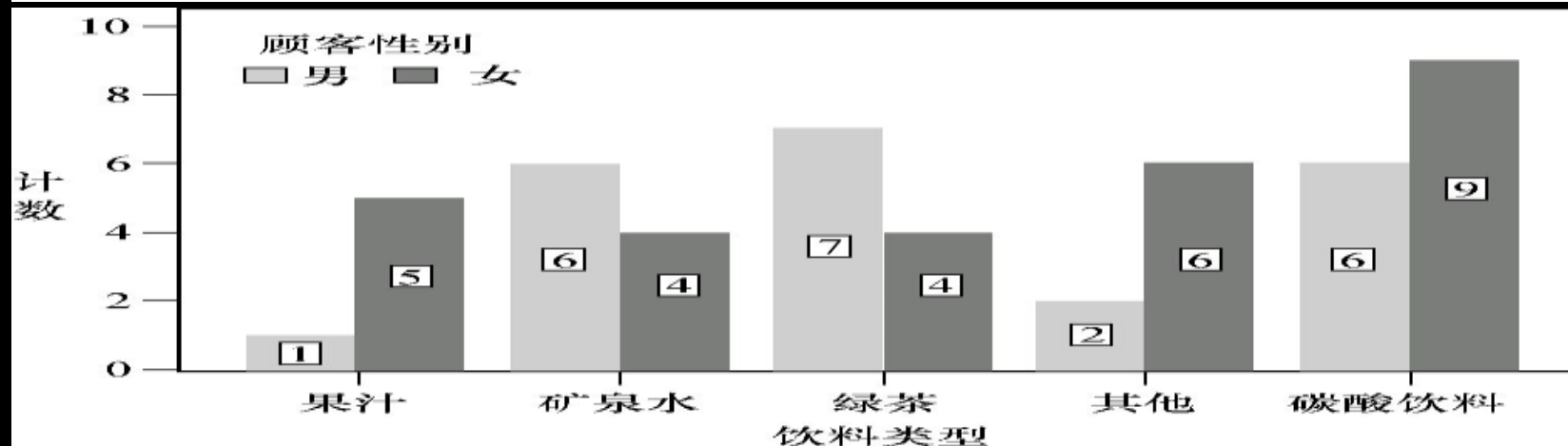
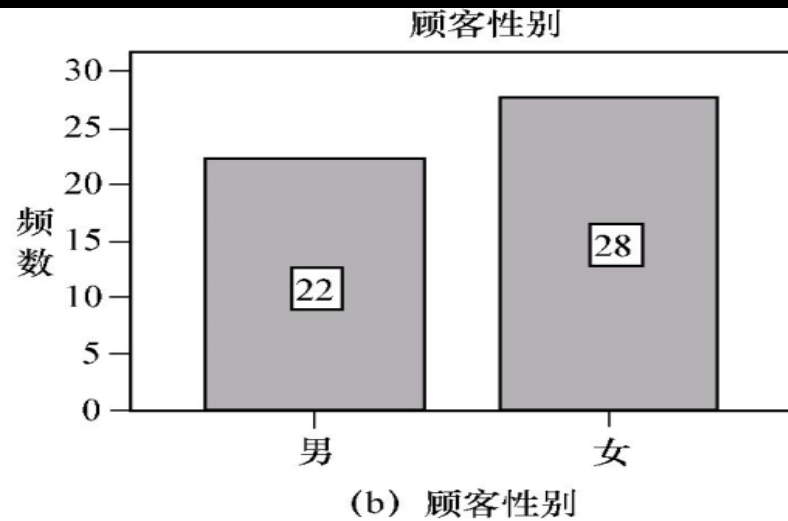
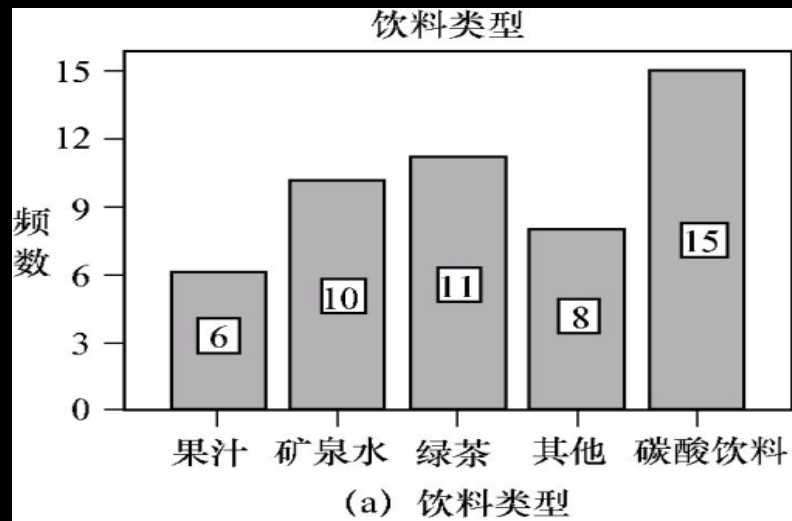
			顾客性别		合计
			男	女	
饮料类型	果汁	计数	1	5	6
		饮料类型 中的 %	16.7%	83.3%	100.0%
		顾客性别 中的 %	4.5%	17.9%	12.0%
		总数的 %	2.0%	10.0%	12.0%
	矿泉水	计数	6	4	10
		饮料类型 中的 %	60.0%	40.0%	100.0%
		顾客性别 中的 %	27.3%	14.3%	20.0%
		总数的 %	12.0%	8.0%	20.0%
	绿茶	计数	7	4	11
		饮料类型 中的 %	63.6%	36.4%	100.0%
		顾客性别 中的 %	31.8%	14.3%	22.0%
		总数的 %	14.0%	8.0%	22.0%
	其他	计数	2	6	8
		饮料类型 中的 %	25.0%	75.0%	100.0%
		顾客性别 中的 %	9.1%	21.4%	16.0%
		总数的 %	4.0%	12.0%	16.0%
	碳酸饮料	计数	6	9	15
		饮料类型 中的 %	40.0%	60.0%	100.0%
		顾客性别 中的 %	27.3%	32.1%	30.0%
		总数的 %	12.0%	18.0%	30.0%
合计		计数	22	28	50
		饮料类型 中的 %	44.0%	56.0%	100.0%
		顾客性别 中的 %	100.0%	100.0%	100.0%
		总数的 %	44.0%	56.0%	100.0%

分类数据的图示—条形图 (bar Chart)

1. 用宽度相同的条形的高度或长短来表示各类别数据的图形
2. 有单式条形图、复式条形图等形式
3. 主要用于反映分类数据的频数分布
4. 绘制时，各类别可以放在纵轴，称为**条形图**，也可以放在横轴，称为**柱形图 (column chart)**

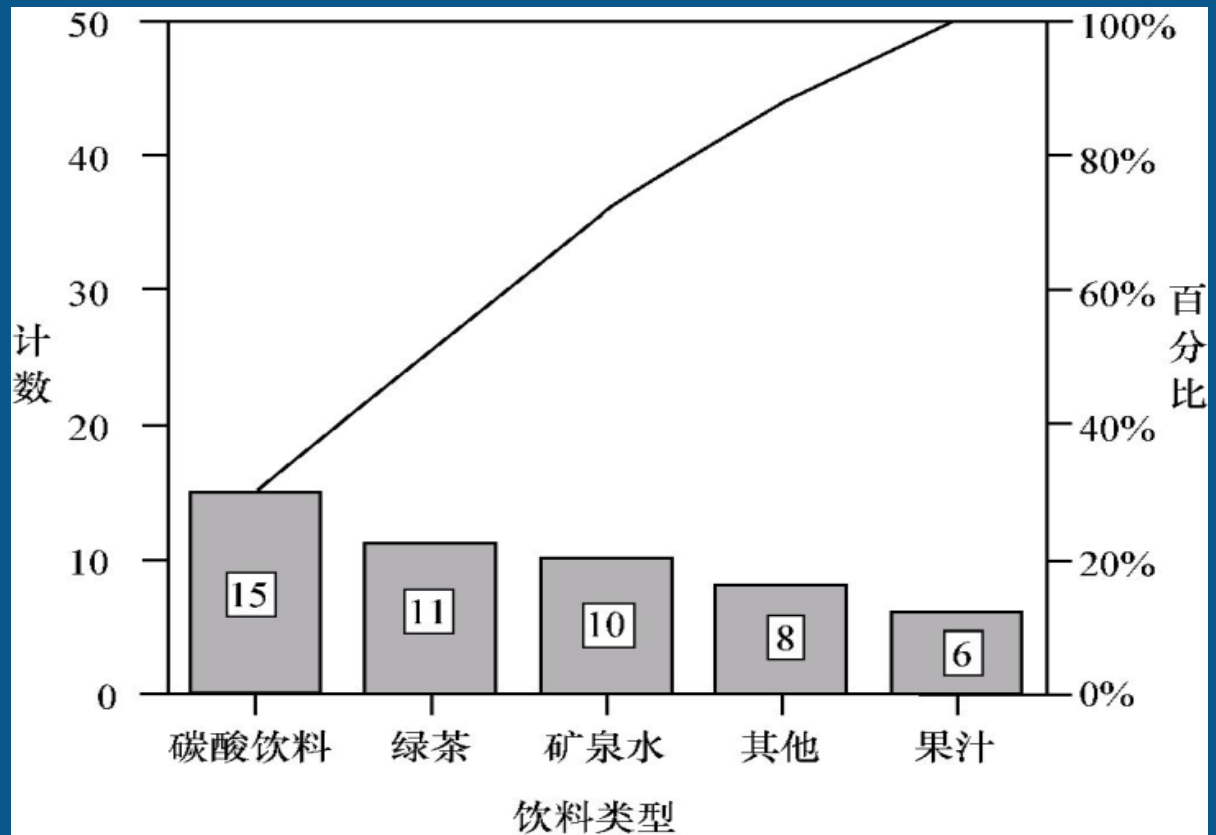
分类数据的图示—条形图

(例题分析)



分类数据的图示－帕累托图 (pareto chart)

1. 按各类别数据出现的频数多少排序后绘制的柱形图
2. 主要用于展示分类数据的分布



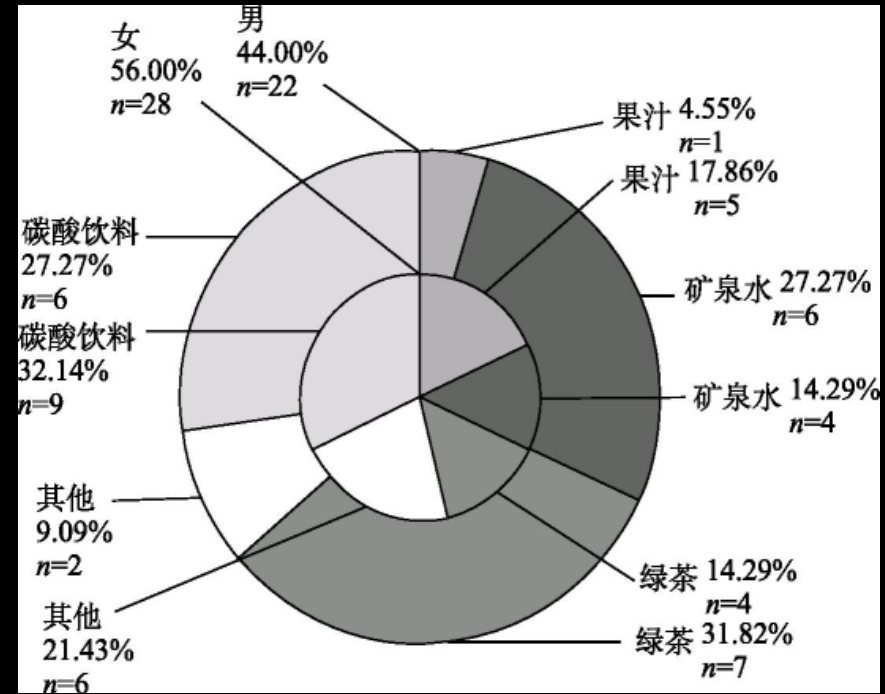
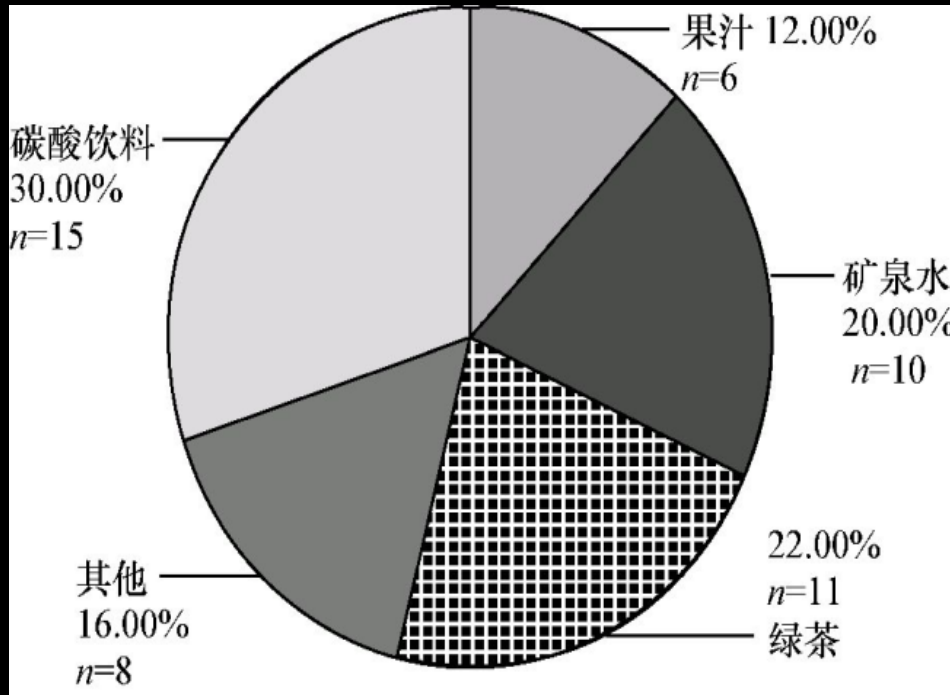
分类数据的图示—饼图

(pie Chart)

1. 也称圆形图，是用圆形及圆内扇形的角度来表示数值大小的图形
2. 主要用于表示样本或总体中各组成部分所占的比例，用于研究结构性问题
3. 绘制圆形图时，样本或总体中各部分所占的百分比用圆内的各个扇形角度表示，这些扇形的中心角度，按各部分数据百分比乘以 360° 确定

分类数据的图示—饼图

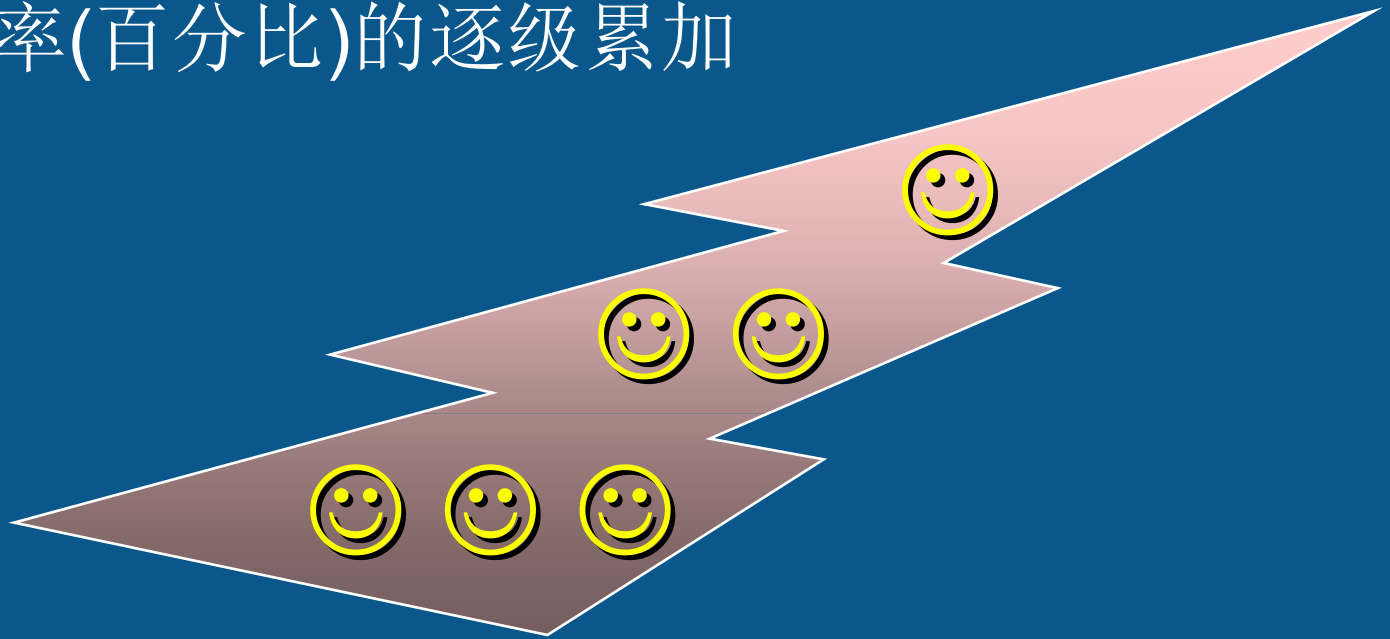
(例题分析)



顺序数据的整理与图示

顺序数据的整理 (可计算的统计量)

1. 累积频数(**cumulative frequencies**): 各类别频数的逐级累加
2. 累积频率(**cumulative percentages**): 各类别频率(百分比)的逐级累加



顺序数据的频数分布表 (例题分析)

【例】在一项城市住房问题的研究中，研究人员在甲乙两个城市各抽样调查300户，其中的一个问题是：“您对您家庭目前的住房状况是否满意？”

1. 非常不满意；2. 不满意；3. 一般；4. 满意；5. 非常满意。

甲城市家庭对住房状况评价的频数分布

回答类别	甲城市					
	户数 (户)	百分比 (%)	向上累积		向下累积	
			户数 (户)	百分比 (%)	户数 (户)	百分比 (%)
非常不满意	24	8	24	8.0	300	100.0
不满意	108	36	132	44.0	276	92
一般	93	31	225	75.0	168	56
满意	45	15	270	90.0	75	25
非常满意	30	10	300	100.0	30	10
合计	300	100.0	—	—	—	—

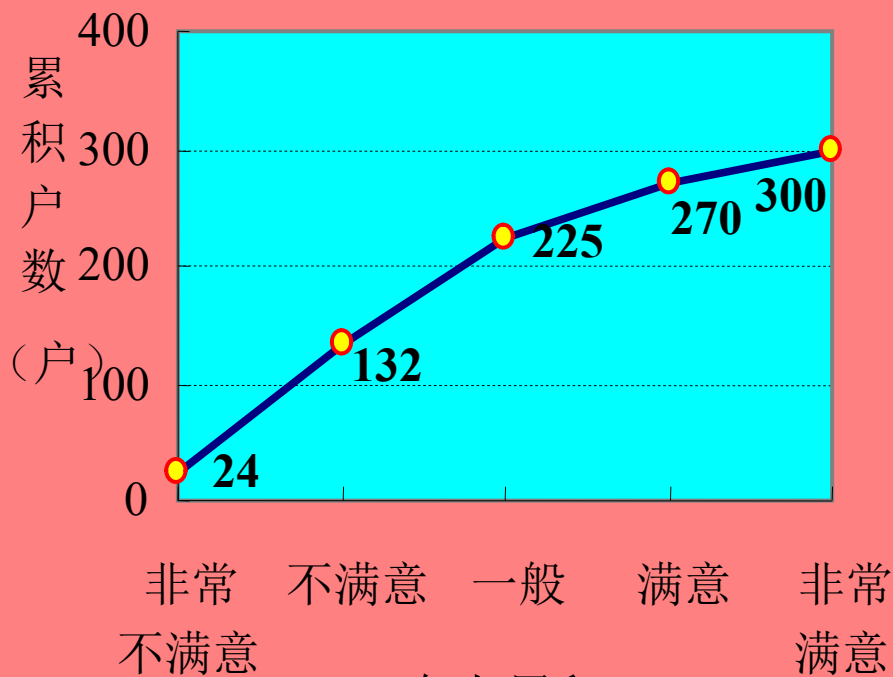
顺序数据的频数分布表 (例题分析)

乙城市家庭对住房状况评价的频数分布

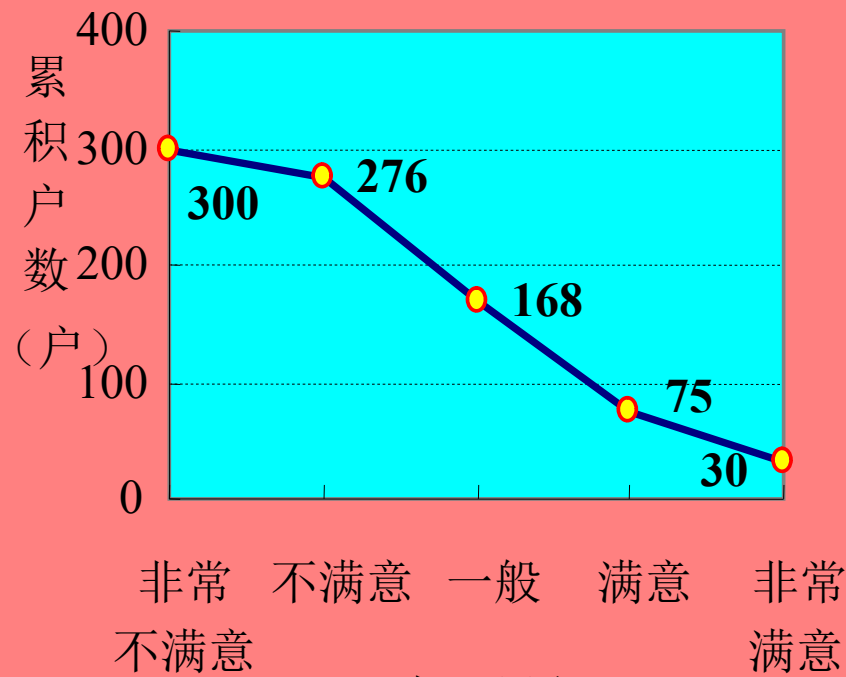
回答类别	乙城市					
	户数 (户)	百分比 (%)	向上累积		向下累积	
			户数 (户)	百分比 (%)	户数 (户)	百分比 (%)
非常不满意	21	7.0	21	7.0	300	100.0
不满意	99	33.0	120	40.0	279	93.0
一般	78	26.0	198	66.0	180	60.0
满意	64	21.3	262	87.3	102	34.0
非常满意	38	12.7	300	100.0	38	12.7
合计	300	100.0	—	—	—	—



顺序数据的图示—累计频数分布图 (例题分析)



(a) 向上累积



(b) 向下累积

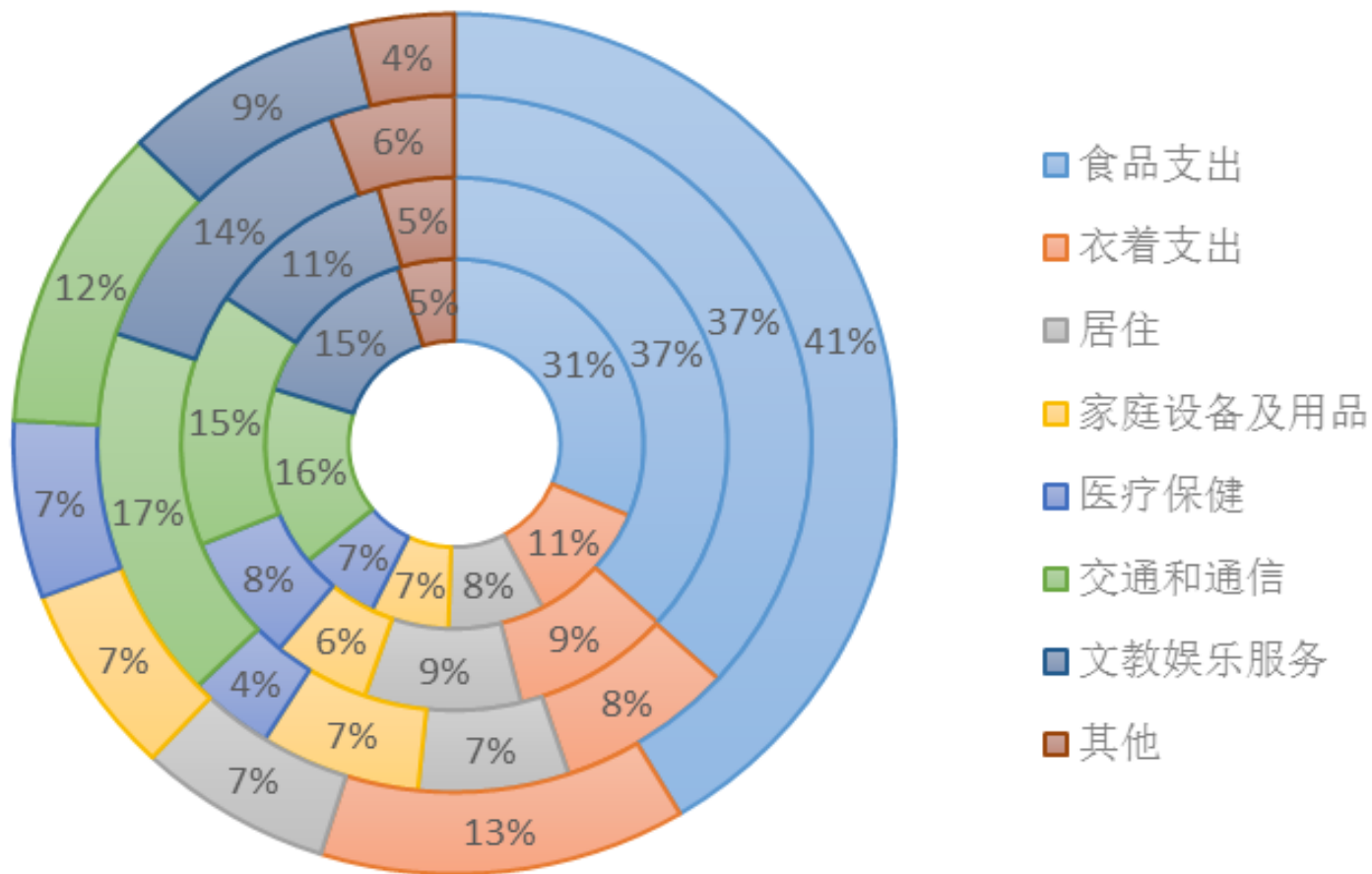
甲城市家庭对住房状况评价的累积频数分布

环形图

(doughnut chart)

1. 环形图中间有一个“空洞”，样本或总体中的每一部分数据用环中的一段表示
2. 与饼图类似，但又有区别
 - 饼图只能显示一个总体各部分所占的比例
 - 环形图则可以同时绘制多个样本或总体的数据系列，每一个样本或总体的数据系列为一个环
3. 用于结构比较研究
4. 用于展示分类和顺序数据

环形图 (例题分析)



3.3 数值型数据的整理与展示

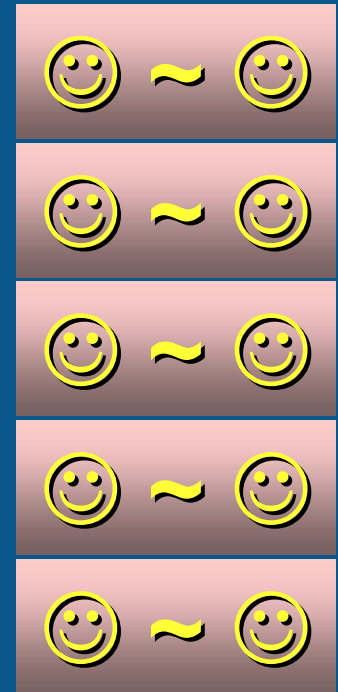
3.3.1 数据分组

3.3.2 数值型数据的图示

数据分组

组距分组 (要点)

1. 将变量值的一个区间作为一组
2. 适合于连续变量
3. 适合于变量值较多的情况
4. 需要遵循“不重不漏”的原则
5. 可采用等距分组，也可采用不等距分组



组距分组 (步骤)

1. 确定组数：组数的确定应以能够显示数据的分布特征和规律为目的。在实际分组时，组数一般为 $5 \leq K \leq 15$
2. 确定组距：组距(Class Width)是一个组的上限与下限之差，可根据全部数据的最大值和最小值及所分的组数来确定，即
组距 = (最大值 - 最小值) ÷ 组数
3. 统计出各组的频数并整理成频数分布表

组距分组 (几个概念)

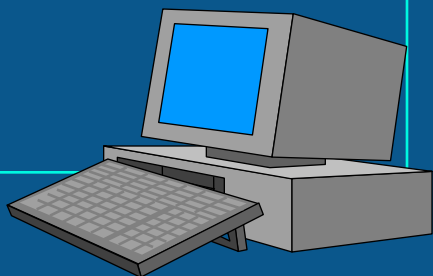
1. 下限(**lower limit**) : 一个组的最小值
2. 上限(**upper limit**) : 一个组的最大值
3. 组距(**class width**) : 上限与下限之差
4. 组中值(**class midpoint**) : 下限与上限之间的中点值

$$\text{组中值} = \frac{\text{下限值} + \text{上限值}}{2}$$

频数分布表的编制

(例题分析)

【例】某电脑公司连续一个月各天的销售量数据(单位:台)。试对数据进行分组



	A	B	C	D	E	F	G	H	I	J
1	某电脑公司2002年前四个月各天的销售量数据(单位:台)									
2	234	159	187	155	172	183	182	177	163	158
3	143	198	141	167	194	225	177	189	196	203
4	187	160	214	168	173	178	184	209	176	188
5	161	152	149	211	196	234	185	189	196	206
6	150	161	178	168	174	153	186	190	160	171
7	228	162	223	170	165	179	186	175	197	208
8	153	163	218	180	175	144	178	191	197	192
9	166	196	179	171	233	179	187	173	174	210
10	154	164	215	233	175	188	237	194	198	168
11	174	226	180	172	190	172	187	189	200	211
12	156	165	175	210	207	181	205	195	201	172
13	203	165	196	172	176	182	188	195	202	213

等距分组表

(上下组限重叠)

	A	B	C
1	按销售量分组(台)	频数 (天)	频率 (%)
2	140~150	4	3.33
3	150~160	9	7.50
4	160~170	16	13.33
5	170~180	27	22.50
6	180~190	20	16.67
7	190~200	17	14.17
8	200~210	10	8.33
9	210~220	8	6.67
10	220~230	4	3.33
11	230~240	5	4.17
12	合计	120	100

等距分组表 (上下组限间断)

	A	B	C
1	按销售量分组(台)	频数 (天)	频率 (%)
2	140~149	4	3.33
3	150~159	9	7.50
4	160~169	16	13.33
5	170~179	27	22.50
6	180~189	20	16.67
7	190~199	17	14.17
8	200~209	10	8.33
9	210~219	8	6.67
10	220~229	4	3.33
11	230~239	5	4.17
12	合计	120	100

等距分组表 (使用开口组)

	A	B	C
1	按销售量分组(台)	频数 (天)	频率 (%)
2	150以下	4	3.33
3	150~159	9	7.50
4	160~169	16	13.33
5	170~179	27	22.50
6	180~189	20	16.67
7	190~199	17	14.17
8	200~209	10	8.33
9	210~219	8	6.67
10	220~229	4	3.33
11	230以上	5	4.17
12	合计	120	100

数值型数据的图示

➡ 分组数据 — 直方图和折线图

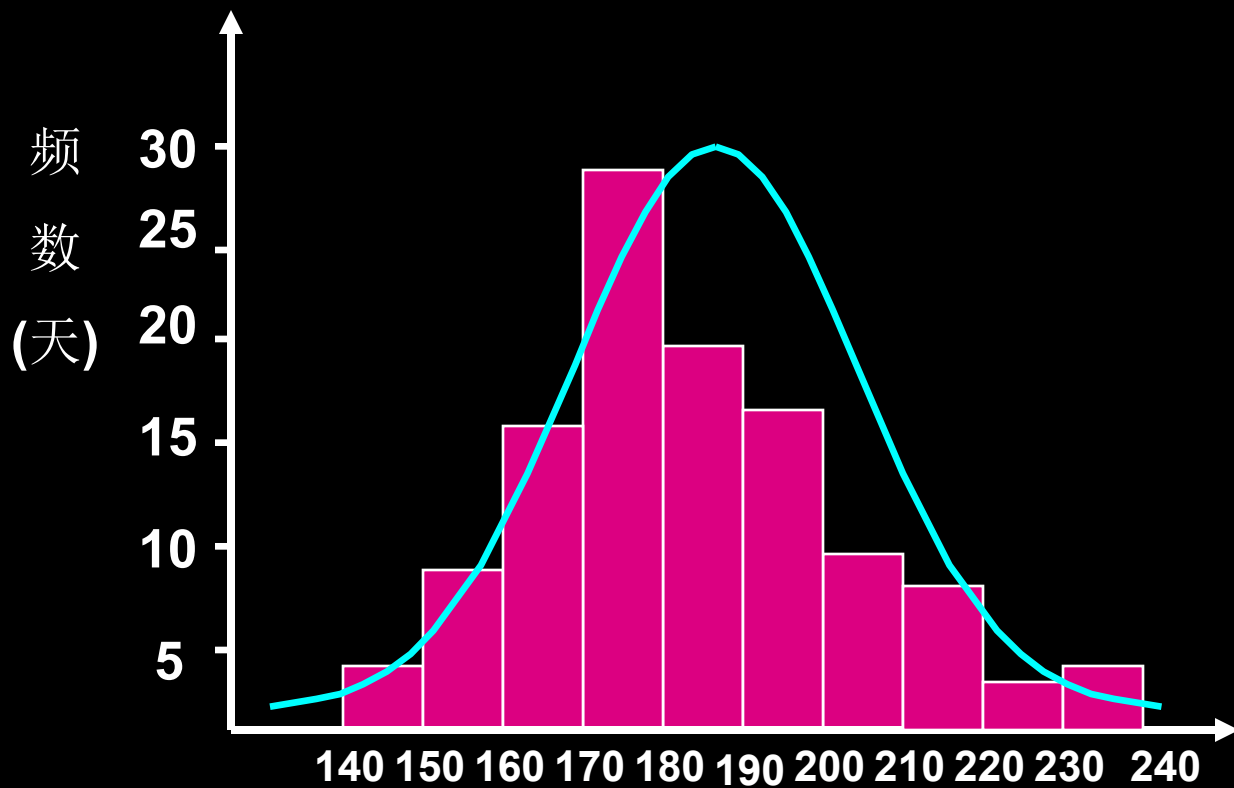


分组数据—直方图 (histogram)

1. 用于展示分组数据分布的一种图形
2. 用矩形的宽度和高度来表示频数分布
 - 本质上是用矩形的 *面积* 来表示频数分布
3. 在直角坐标中，用横轴表示数据分组，纵轴表示频数或频率，各组与相应的频数就形成了一个矩形，即直方图

分组数据的图示 (直方图的绘制)

我一眼就看出
来了，销售量
在170~180
之间的天数最
多!



某电脑公司销售量分布的直方图

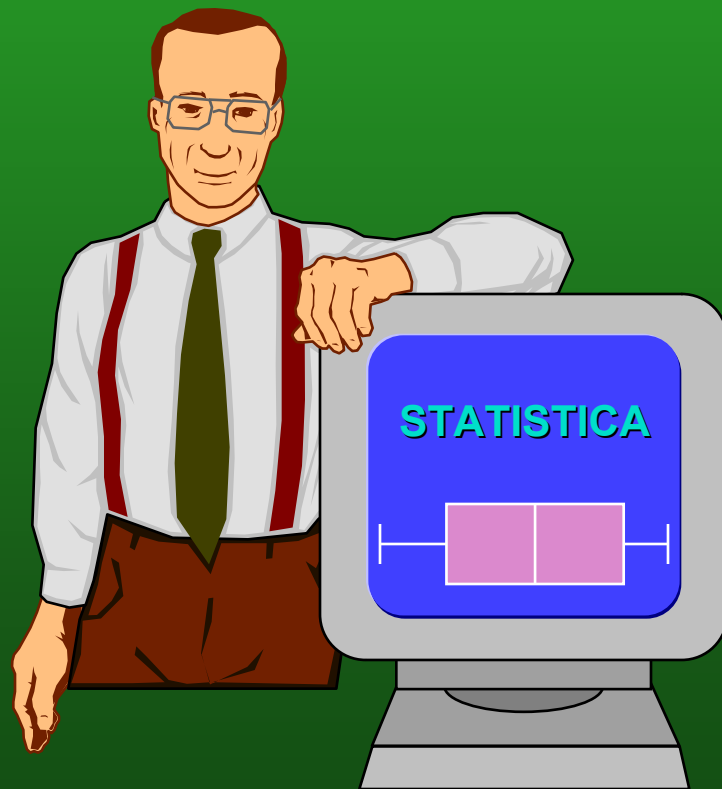
分组数据—直方图

(直方图与条形图的区别)

1. 条形图是用条形的长度(横置时)表示各类别频数的多少, 其宽度(表示类别)则是固定的
2. 直方图是用面积表示各组频数的多少, 矩形的高度表示每一组的频数或百分比, 宽度则表示各组的组距, 其高度与宽度均有意义
3. 直方图的各矩形通常是连续排列, 条形图则是分开排列
4. 条形图主要用于展示分类数据, 直方图则主要用于展示数值型数据

数值型数据的图示

未分组数据 — 茎叶图和箱线图



未分组数据—茎叶图

(stem-and-leaf display)

1. 用于显示未分组的原始数据的分布
2. 由“茎”和“叶”两部分构成，其图形是由数字组成的
3. 以该组数据的高位数值作树茎，低位数字作树叶
4. 树叶上只保留最后一位数字
6. 茎叶图类似于横置的直方图，但也有区别
 - 直方图可观察一组数据的分布状况，但没有给出具体的数值
 - 茎叶图既能给出数据的分布状况，又能给出每一个原始数值，保留了原始数据的信息
 - 直方图适用于大批量数据，茎叶图适用于小批量数据

未分组数据—茎叶图

(例题分析)

电脑销售额 Stem-and-Leaf Plot

Frequency	Stem &	Leaf
3.00	14 .	134
1.00	14 .	9
5.00	15 .	02334
4.00	15 .	5689
8.00	16 .	00112334
8.00	16 .	55567888
13.00	17 .	0112222233444
14.00	17 .	555566778888999
7.00	18 .	0012234
13.00	18 .	5667777888999
6.00	19 .	001244
11.00	19 .	55666667788
5.00	20 .	01233
5.00	20 .	56789
6.00	21 .	001134
2.00	21 .	58
1.00	22 .	3
3.00	22 .	568
4.00	23 .	3344
1.00	Extremes	(>=237)

Stem width: 10
Each leaf: 1 case(s)

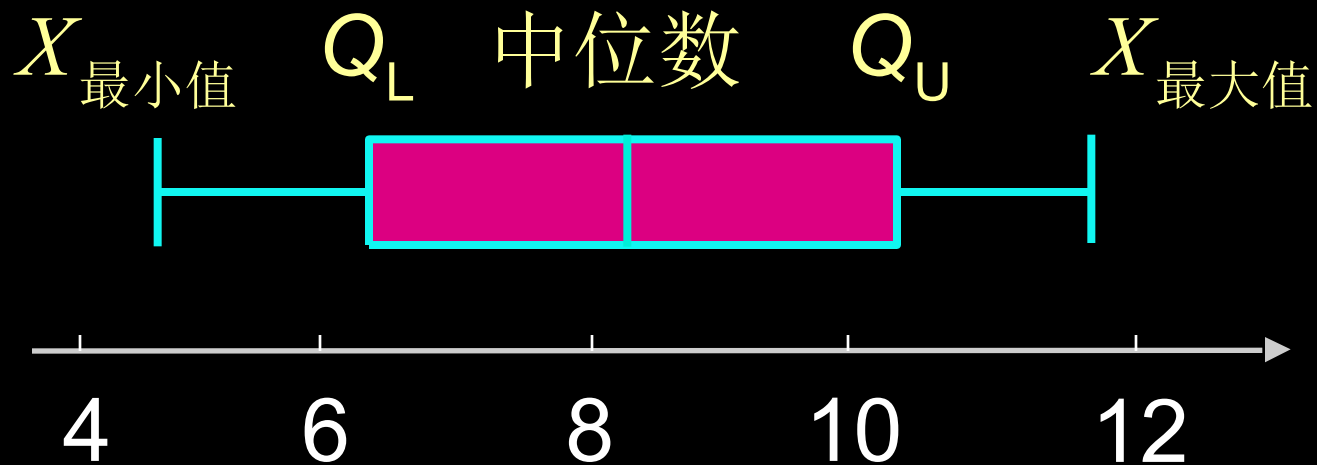
某电脑公司销售量分布的茎叶图

未分组数据 — 箱线图

(box plot)

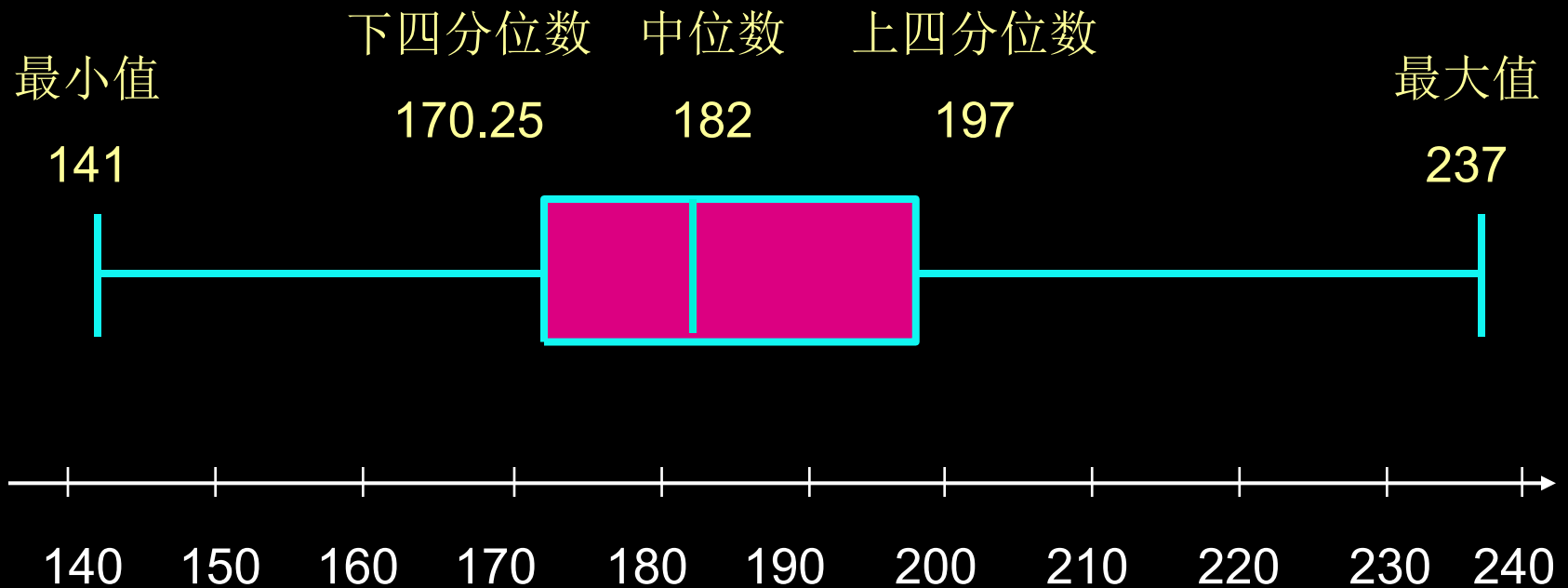
1. 用于显示未分组的原始数据的分布
2. 由一组数据的5个特征值绘制而成，它由一个箱子和两条线段组成
3. 绘制方法
 - 首先找出一组数据的5个特征值，即最大值、最小值、中位数 M_e 和两个四分位数(下四分位数 Q_L 和上四分位数 Q_U)
 - 连接两个四分位数画出箱子，再将两个极值点与箱子相连接
 - 该箱线图也称为**Median/Quart./Range**箱线图

未分组数据—单批数据箱线图 (箱线图的构成)



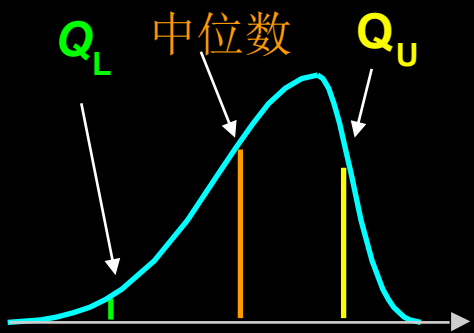
Median/Quart./Range箱线图

未分组数据—单批数据箱线图 (例题分析)



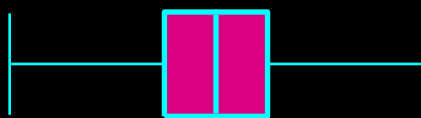
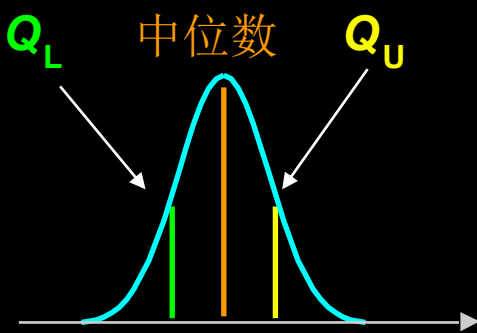
某电脑公司销售量数据的
Median/Quart./Rang箱线图

分布的形状与箱线图



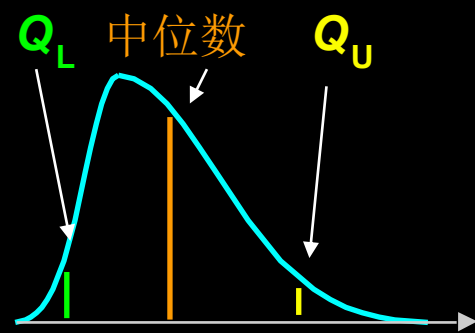
左偏分布

Left-skewed distribution



对称分布

Bell-shaped distribution



右偏分布

Right-skewed distribution

不同分布的箱线图

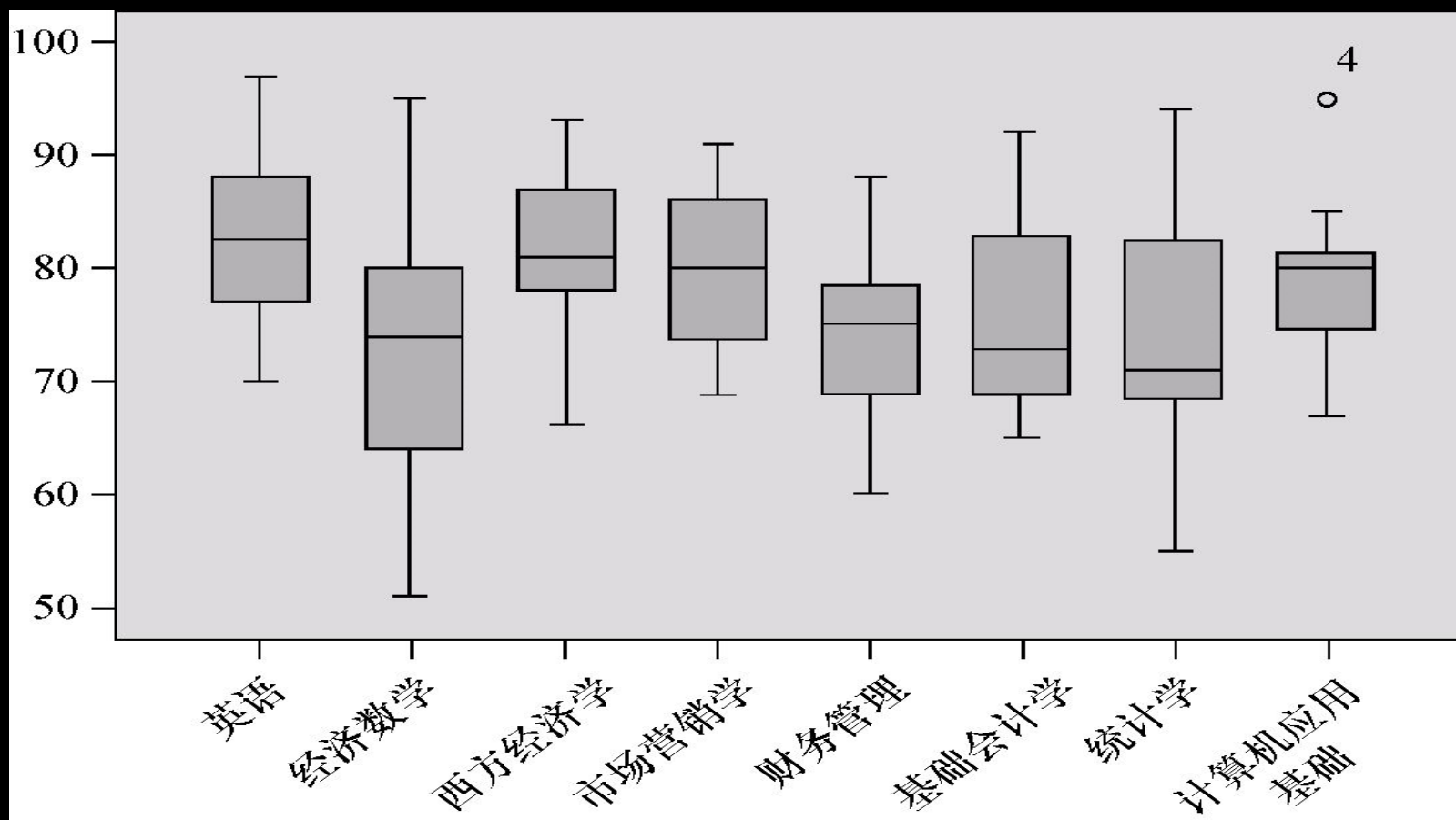
未分组数据—多批数据箱线图 (例题分析)

【例】从某大学经济管理专业二年级学生中随机抽取11人，对8门主要课程的考试成绩进行调查，所得结果如表。试绘制各科考试成绩的批比较箱线图，并分析各科考试成绩的分布特征

11名学生各科的考试成绩数据

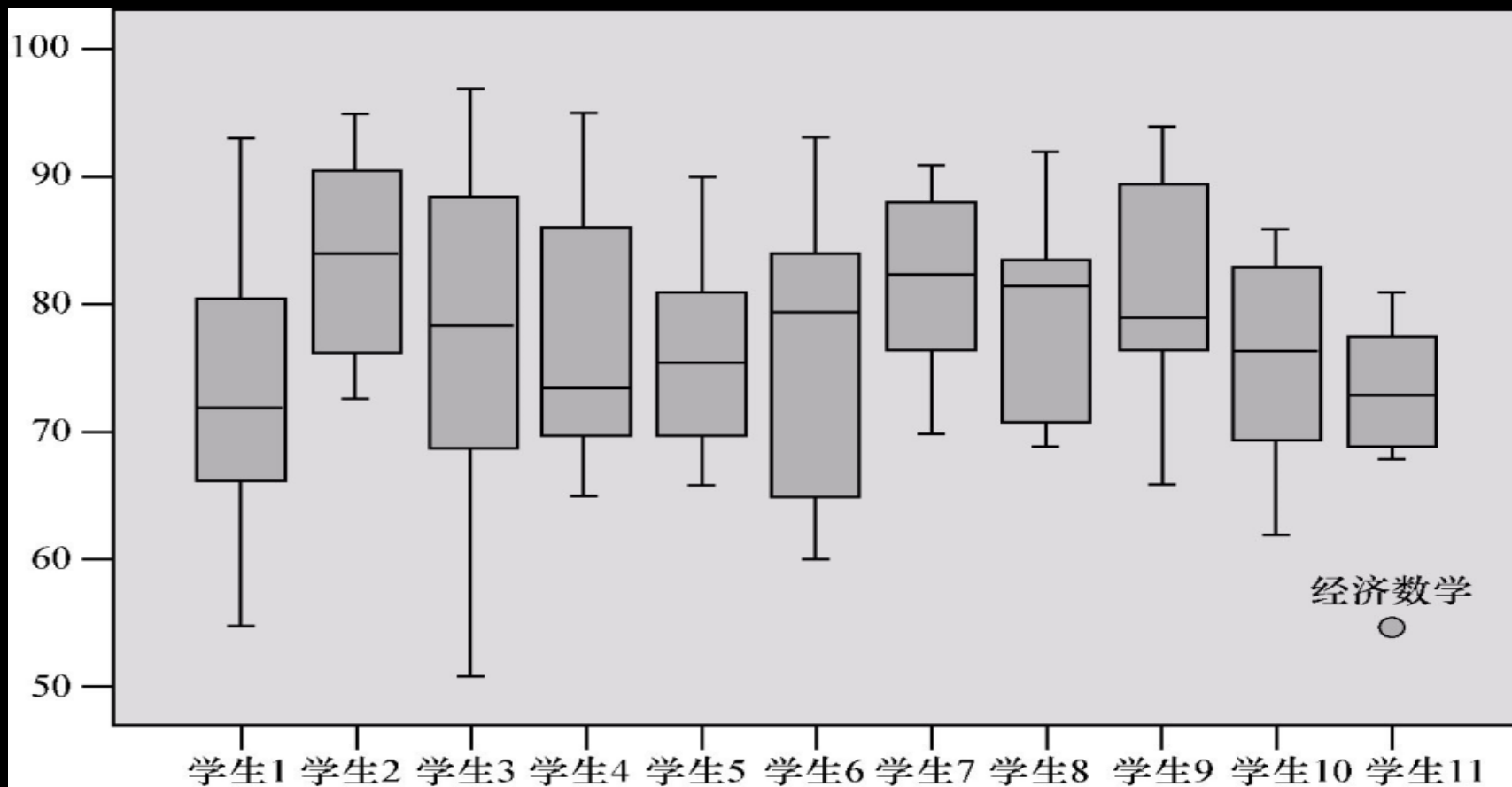
课程名称	学生编号										
	1	2	3	4	5	6	7	8	9	10	11
英语	76	90	97	71	70	93	86	83	78	85	81
经济数学	65	95	51	74	78	63	91	82	75	71	55
西方经济学	93	81	76	88	66	79	83	92	78	86	78
市场营销学	74	87	85	69	90	80	77	84	91	74	70
财务管理	68	75	70	84	73	60	76	81	88	68	75
基础会计学	70	73	92	65	78	87	90	70	66	79	68
统计学	55	91	68	73	84	81	70	69	94	62	71
计算机应用基础	85	78	81	95	70	67	82	72	80	81	77

未分组数据—多批数据箱线图 (例题分析—Median/Quart./Range)



8门课程考试成绩的Median/Quart./Range箱线图

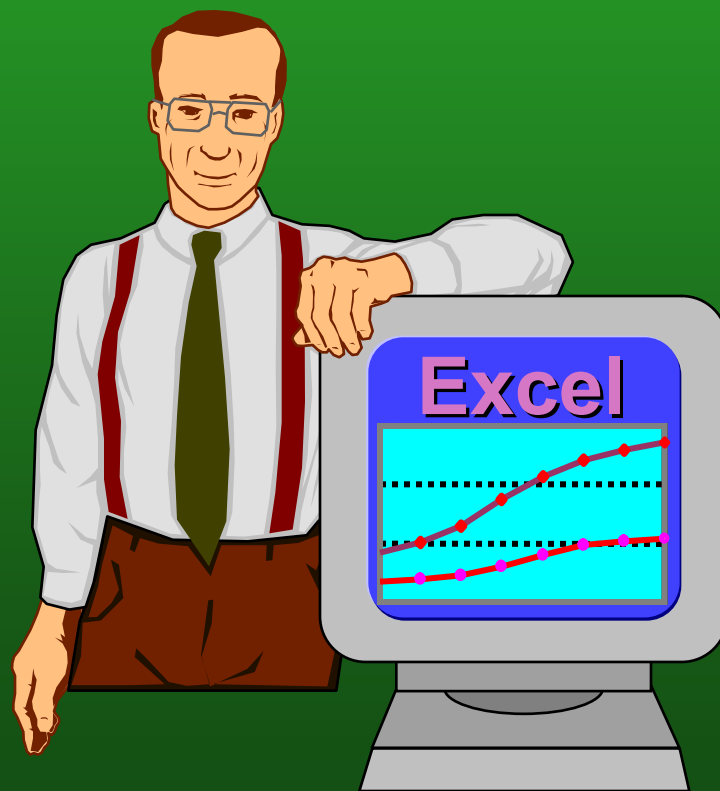
未分组数据—多批数据箱线图 (例题分析—Median/Quart./Range)



11名学生8门课程考试成绩的Median/Quart./Range箱线图

数值型数据的图示

时间序列数据—线图



时间序列数据—线图

(line plot)

1. 表示时间序列数据趋势的图形
2. 时间一般绘在横轴，数据绘在纵轴
3. 图形的长宽比例大致为**10 : 7**
3. 一般情况下，纵轴数据下端应从“**0**”开始，以便于比较。数据与“**0**”之间的间距过大时，可以采取折断的符号将纵轴折断

时间序列数据—线图

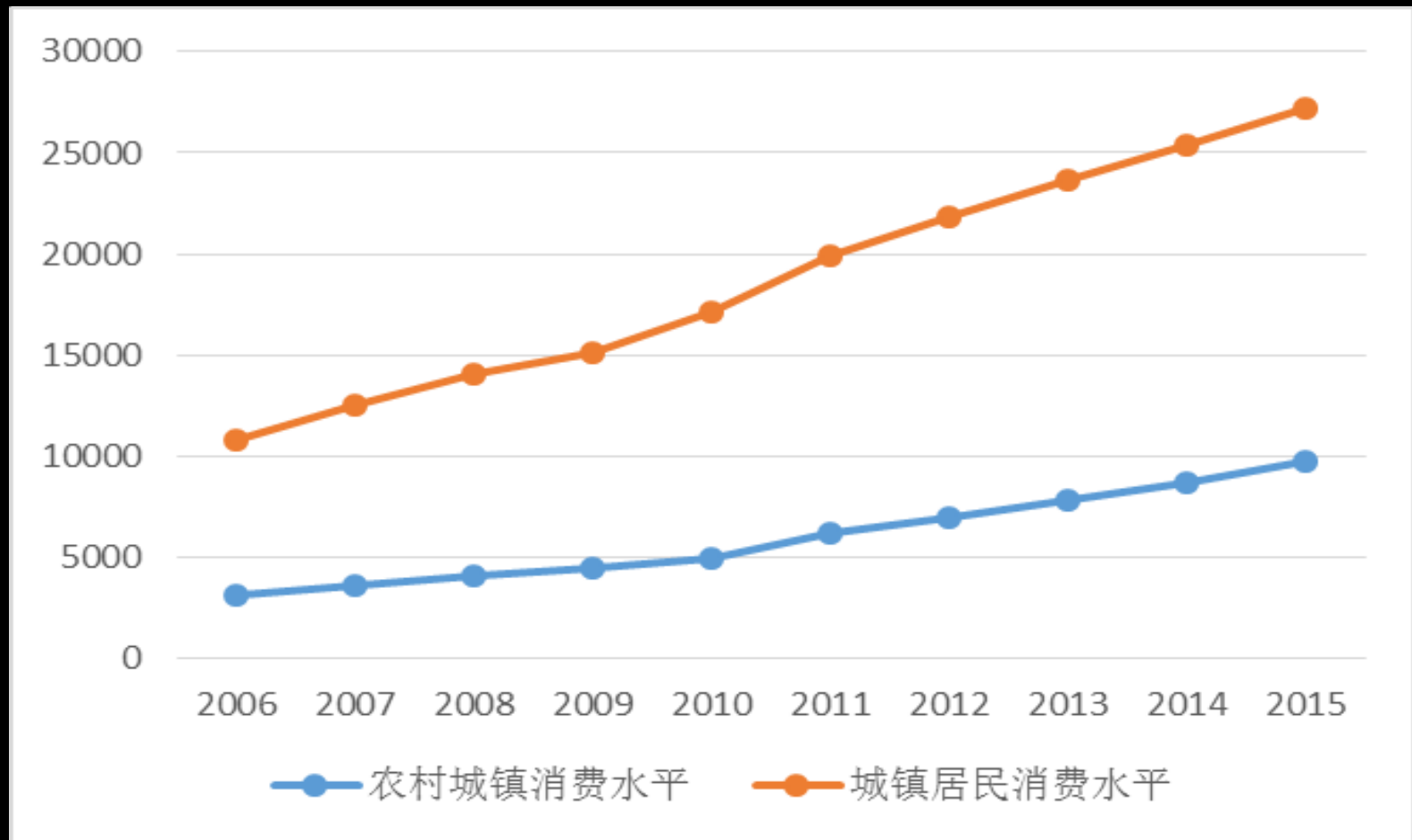
(例题分析)

【例】2006年—2015年我国城乡居民人均消费水平数据（单位：元）如表3—17所示。绘制时间序列图，分析城乡居民消费水平的变化趋势和特点

	A	B	C
1	年份	农村城镇消费水平	城镇居民消费水平
2	2006	3066	10739
3	2007	3538	12480
4	2008	4065	14061
5	2009	4402	15127
6	2010	4941	17104
7	2011	6187	19912
8	2012	6964	21861
9	2013	7773	23609
10	2014	8711	25424
11	2015	9679	27210

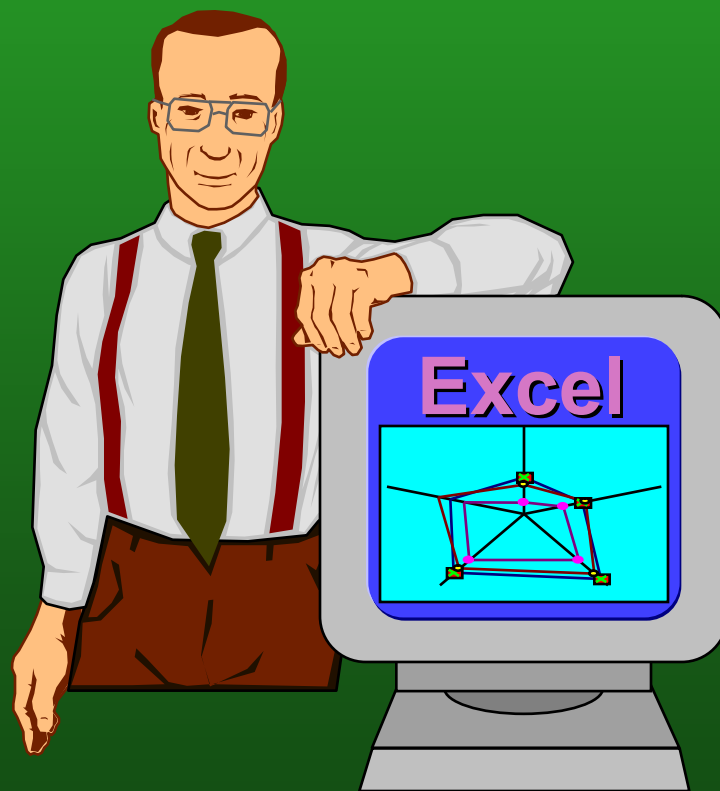
时间序列数据—线图

(例题分析)



数值型数据的图示

多变量数据的图示

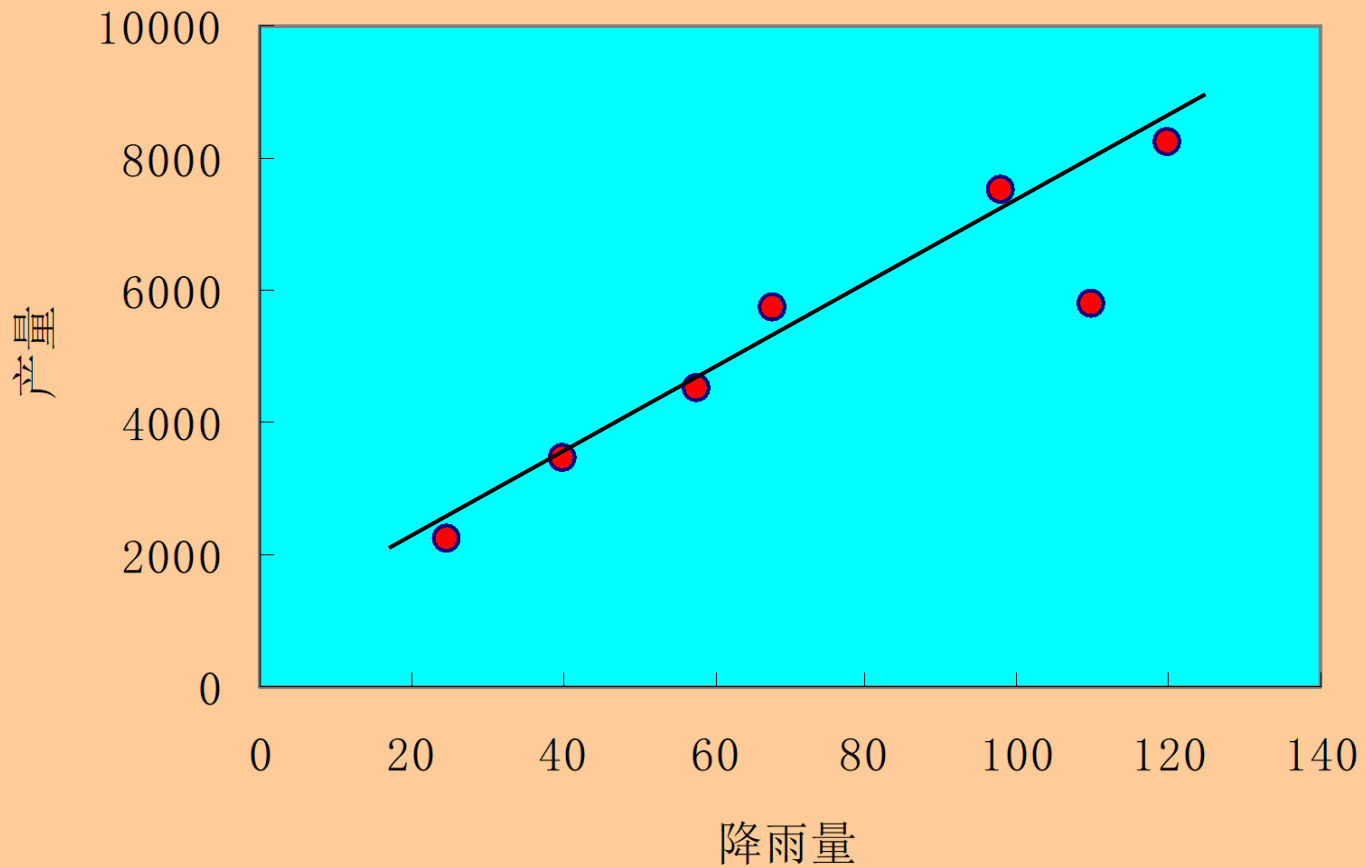


两个变量间的关系 — 二维散点图 (2D Scatterplots)

1. 展示两个变量之间的关系
2. 用横轴代表变量 x ，纵轴代表变量 y ，每组数据 (x_i, y_i) 在坐标系中用一个点表示， n 组数据在坐标系中形成的 n 个点称为散点，由坐标及其散点形成的二维数据图

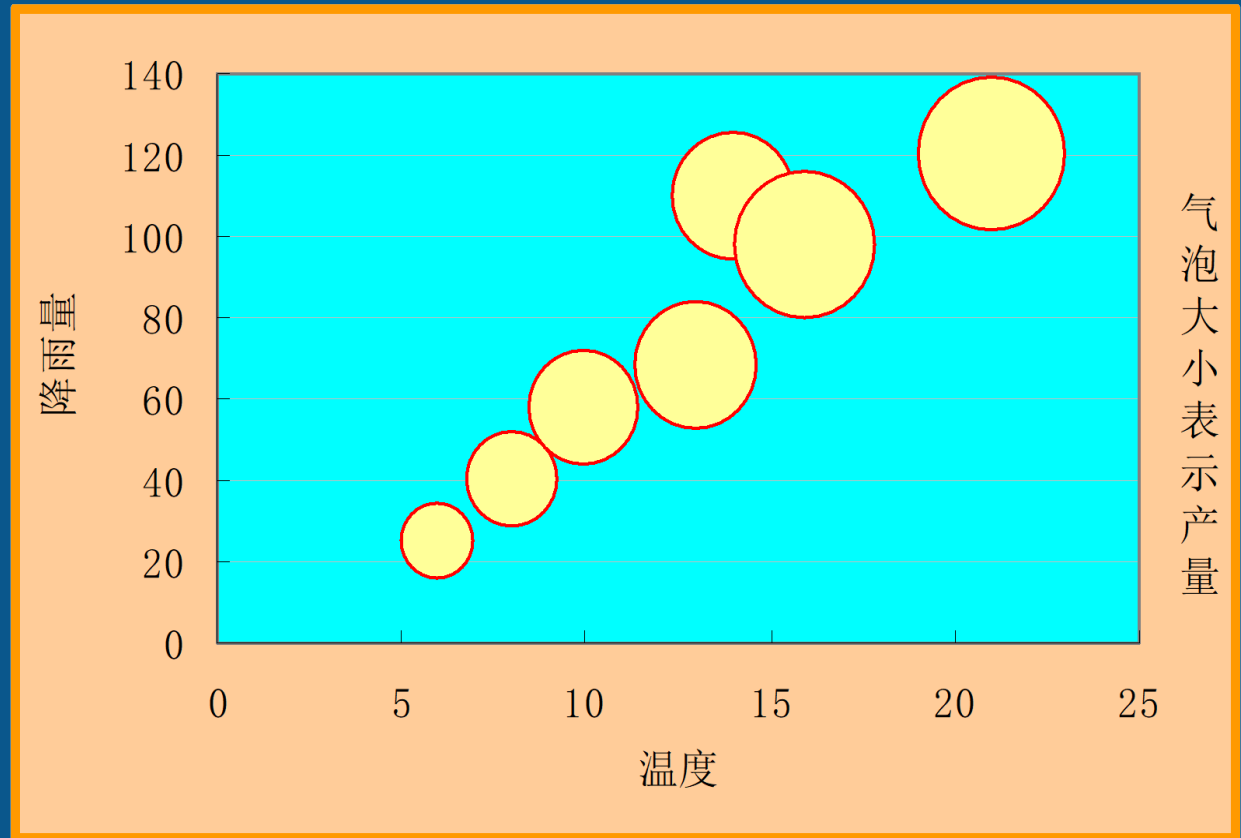
温度 / °C	降雨量/mm	产量/kg/hm ²
6	25	2250
8	40	3450
10	58	4500
13	68	5750
14	110	5800
16	98	7500
21	120	8250

两个变量间的关系—二维散点图 (2D Scatterplots)



三个变量间的关系—气泡图 (bubble chart)

1. 显示三个变量之间的关系
2. 图中数据点的大小依赖于第三个变量



多变量数据—雷达图 (radar chart)

1. 也称为蜘蛛图(spider chart)
2. 显示多个变量的图示方法
3. 在显示或对比各变量的数值总和时十分有用
3. 假定各变量的取值具有相同的正负号，总的绝对值与图形所围成的区域成正比
4. 可用于研究多个样本之间的相似程度

多变量数据 — 雷达图

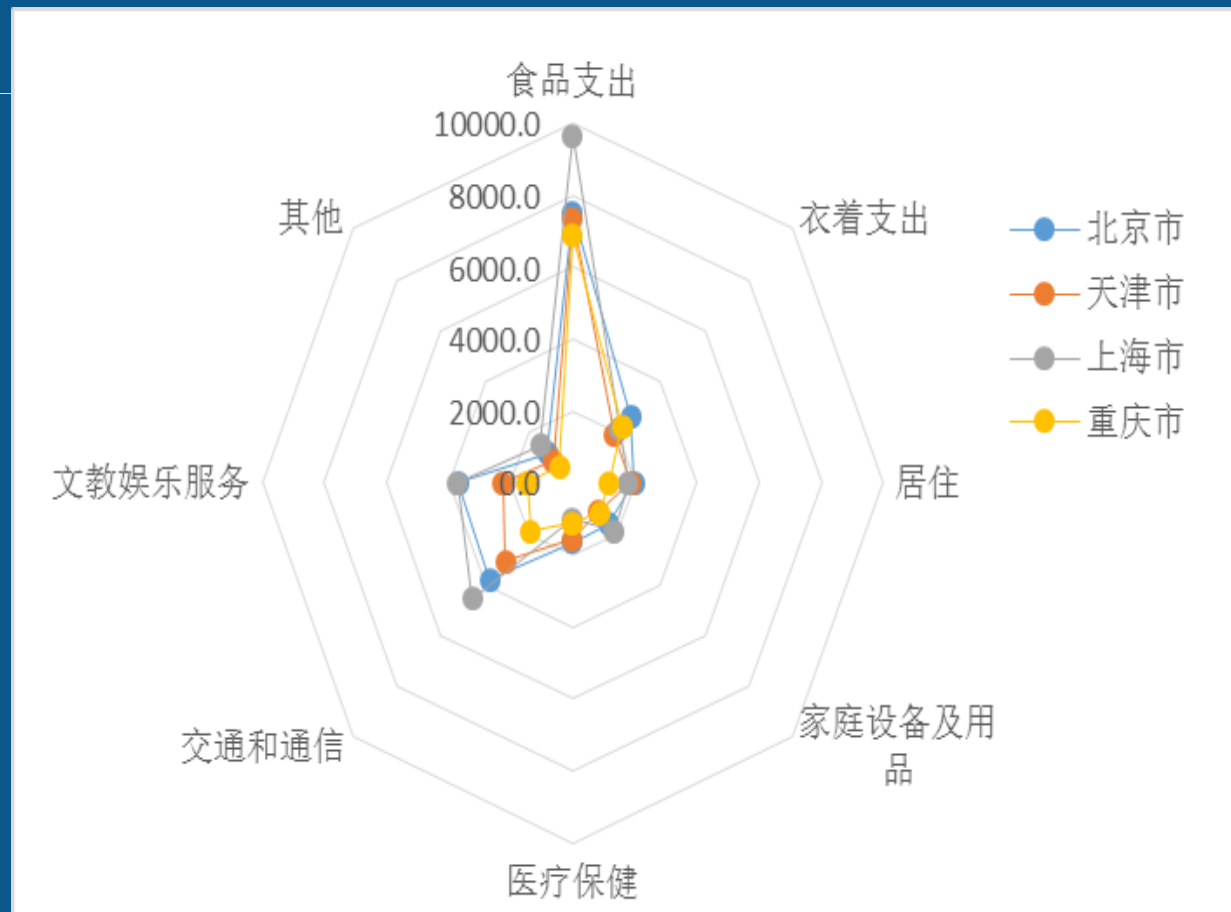
(雷达图的制作)

- 设有 n 组样本 S_1, S_2, \dots, S_n , 每个样本测得 P 个变量 X_1, X_2, \dots, X_P , 要绘制这 P 个变量的雷达图, 其具体做法是
- 先做一个圆, 然后将圆 P 等分, 得到 P 个点, 令这 P 个点分别对应 P 个变量, 在将这 P 个点与圆心连线, 得到 P 个幅射状的半径, 这 P 个半径分别作为 P 个变量的坐标轴, 每个变量值的大小由半径上的点到圆心的距离表示
 - 将同一样本的值在 P 个坐标上的点连线。这样, n 个样本形成的 n 个多边形就是一个雷达图

多变量数据—雷达图

(例题分析)

【例】沿用例3.4。绘制雷达图，比较北京、天津、上海和重庆的家庭消费支出的特点和相似性



3.4 合理使用图表

3.4.1 鉴别图形优劣的准则

3.4.2 统计表的设计

鉴别图表优劣的准则

1. 一张好的图表应包括以下基本特征

- 显示数据
- 让读者把注意力集中在图表的内容上，而不是制作图表的程序上
- 避免歪曲
- 强调数据之间的比较
- 服务于一个明确的目的
- 有对图表的统计描述和文字说明

2. 5种鉴别图表优劣的准则：一张好的图表应当

- 精心设计、有助于洞察问题的实质
- 使复杂的观点得到简明、确切、高效的阐述
- 能在最短的时间内以最少的笔墨给读者提供最大量的信息
- 是多维的
- 表述数据的真实情况

统计表的结构

2002~2003年城镇居民家庭抽样调查资料

表头

		2002年	2003年
项目	单位		
调查户数	户	45317	48028
平均每户家庭人口	人	3.04	3.01
平均每户就业人口	人	1.58	1.58
平均每户就业面	%	51.97	52.49
平均一名就业者负担人数	人	1.92	1.91
平均每人全部年收入	元	8177.40	9061.22
#可支配收入	元	7702.80	8472.20
平均每人消费性支出	元	6029.88	6510.94

列标题

数字资料

行标题

资料来源：《中国统计年鉴2004》，中国统计出版社，2004，第359页。
注：本表为城市和县城的城镇居民家庭抽样调查资料。

附加

统计表的设计

1. 合理安排统计表的结构
2. 总标题内容应满足**3W** 要求
3. 数据计量单位相同时，可放在表的右上角标明，不同时应放在每个变量后或单列出一列标明
4. 表中的上下两条横线一般用粗线，其他线用细线
5. 通常情况下，统计表的左右两边不封口
6. 表中的数据一般是右对齐，有小数点时应以小数点对齐，而且小数点的位数应统一
7. 对于没有数字的表格单元，一般用“—”表示
8. 必要时可在表的下方加上注释

本章小结

1. 数据预处理的内容和目的
2. 分类和顺序数据的整理与显示方法
3. 数值型数据的整理与显示方法
4. 合理使用图表
5. 用**Excel**作频数分布表和图形

结 束



THANKS