

统计表与统计图

- 1.图表是语言的一种重要形式，如果运用得当的话，比起表格来能更明快清晰地进行沟通；
- 2.决定图表形式的不是数据，也不是尺寸，而是你想说明的主题，你想指出的内容要点；
- 3.图表贵精不贵多，只有当图表能够帮助你表达主题时才使用；
- 4.图表是直观教具，它不能取代书写和讲述，只有在帮助你传达主题时，它才起作用。

麦肯锡《用图表说话》

从著名的统计学家 南丁格尔 说起

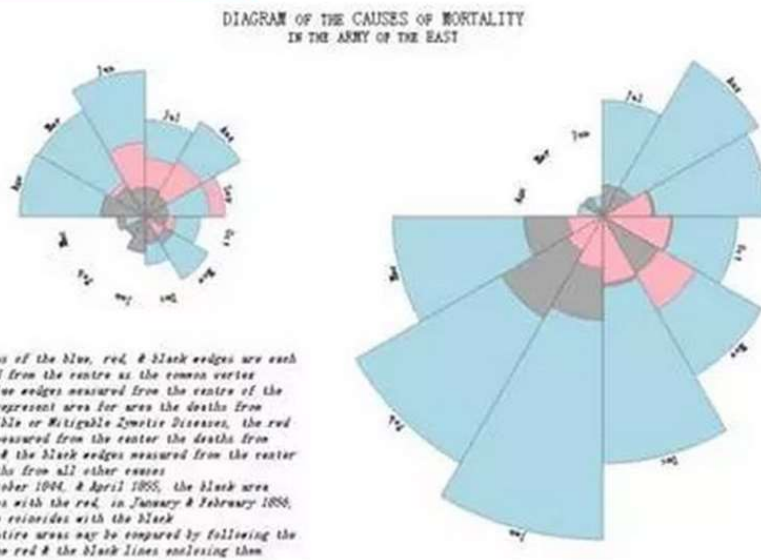


克里米亚战争（1853-1856）
被称为近代科技战争的开端，
电报、铁路、战地护理等在战争中被广泛应用

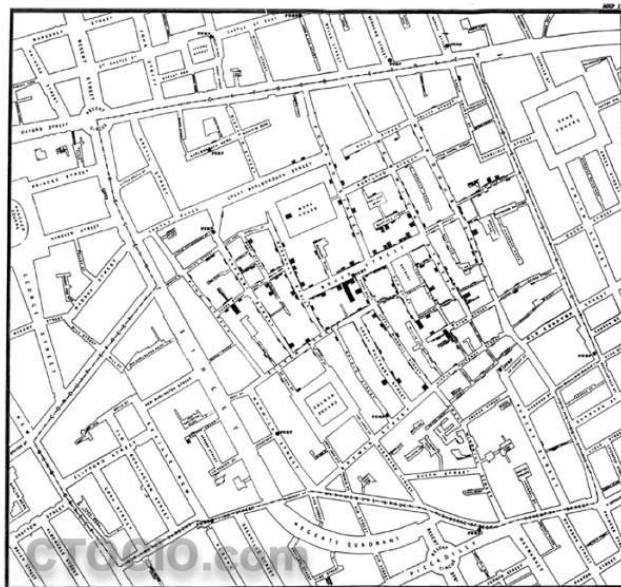


南丁格尔
“克里米亚的天使”
“提灯女神”
英国护士和统计学家

南丁格尔玫瑰图：人类历史上最有影响力的五个数据可视化信息图第二位



伦敦霍乱地图：人类历史上最有影响力的五个数据可视化信息图第五位



英国流行病专家John Snow

1

统计表

投票 最多可选5项

你觉得统计表制作时应遵循什么原则？

- ☐ A 重点突出
- ☐ B 结构简单明了
- ☐ C 主次分明
- ☐ D 层次清楚
- ☐ E 符合逻辑

表1 第三组病人各年存活及死亡情况

年份 (1)	病例数 (2)	存活数 (3)	住院期死 亡总例数 (4)=(2)-(3)	急性期 死亡数 (5)	住院期总病 死率(%) (6)=(4)/(2)	急性期病 死率(%) (7)=(5)/(2)
1964	17	9	8	7	47.1	41.2
1965	13	8	5	4	38.5	30.8
1966	15	8	7	6	46.7	40.0
1967	15	9	6	6	40.0	40.0
1968	12	8	4	4	33.3	33.3
合计	72	42	30	27	41.7	37.5

表1. 1964~1968年急性心肌梗塞患者的病死率 (修改表)

年份	病例数	死亡例数		病死率(%)	
		住院期	急性期	住院期	急性期
1964	17	8	7	47.1	41.2
1965	13	5	4	38.5	30.8
1966	15	7	6	46.7	40.0
1967	15	6	6	40.0	40.0
1968	12	4	4	33.3	33.3
合计	72	30	27	41.7	37.5

表1 第三组病人各年存活及死亡情况 (原表)

年份 (1)	病例数 (2)	存活数 (3)	住院期死 亡总例数 (4)=(2)-(3)	急性期 死亡数 (5)	住院期总病 死率(%) (6)=(4)/(2)	急性期病 死率(%) (7)=(5)/(2)
1964	17	9	8	7	47.1	41.2
1965	13	8	5	4	38.5	30.8
1966	15	8	7	6	46.7	40.0
1967	15	9	6	6	40.0	40.0
1968	12	8	4	4	33.3	33.3
合计	72	42	30	27	41.7	37.5

统计表的结构

表号 标题 (可包括时间、地点、事件)				
(主语)		(谓语)		
横标目的总	纵标目的总标目		纵标目的总标目	
标目	纵标目	纵标目	纵标目	纵标目
横标目	×××	×××	××.××	××.××
:	××	×××	×.××	×.××
:	:	:	:	:
:	:	:	:	:
合计	×××	××××	××.××	××.××
(备注)				

标题：简明确切，概括地说明表的内容。

标目：
1. 简明确切。
2. 标目的排列应有时间顺序、自然顺序。
3. 标目的层次。

内容（数字）：
1. 准确、清晰。
2. 性质相近的数字排在一起。
3. 同一指标数字精度应一致。

线条：一般无纵线，横线不宜过多

关于备注

表2 某省某工厂 1994、1998年四项检测指标异常检出率

检测指标	1994年			1998年		
	受检人数	异常人数	检出率 (%)	受检人数	异常人数	检出率 (%)
血压	519	55	10.60	582	38	6.52
心率	519	44	8.48	582	39	6.70
TTT ^Δ	519	36	6.94	582	23	3.95
GPT [#]	519	20	3.85	582	16	2.75

Δ: TTT (麝香草酚浊度试验), #: GPT (谷丙转氨酶)。

(丁建生等. 中国卫生统计 1999; 16(3):166)

不同溶液冲洗伤口效果比较

	试验组			对照组
	庆大霉素	新洁尔灭	四环素	生理盐水
总例数	30	30	30	30
感染例数	1	3	5	8
百分比	3.3%	10%	16.7%	26.7%

表3. 不同溶液冲洗伤口的效果

组别	总例数	感染例数	感染率 (%)
试验组			
庆大霉素	30	1	3.3
新洁尔灭	30	3	10.0
四环素	30	5	16.7
对照组 (生理盐水)	30	8	26.7

大家来找茬

表4. 不同心理分值的冠心病危险因素水平比较

危险因素	心理得分值								P 值
	1 (252 人)		2 (253 人)		3 (252 人)		4 (253 人)		
	$\bar{X} \pm S$	%	$\bar{X} \pm S$	%	$\bar{X} \pm S$	%	$\bar{X} \pm S$	%	
年龄 (岁)	35.2±6.5		37.0±6.3		36.5±6.8		37.8±6.5		<0.05
收 缩 压 (mmHg)	120.7±13.4		121.2±13.2		121.1±13.2		120.4±12.8		<0.5
舒 张 压 (mmHg)	78.8±10.2		77.9±10.5		78.2±11.0		78.4±10.6		<0.39
体力活动	2.1±0.2		2.1±0.1		2.1±0.2		2.3±0.3		<0.08
体重指数	23.1±3.2		24.0±3.5		24.8±3.1		25.8±3.1		<0.01
吸烟率 (%)		70.8		69.4		70.7		71.1	<0.41
吸烟量 (支/天)	8.0±1.0		10.0±2.0		15.0±2.0		15.0±2.0		<0.001
饮酒率 (%)		52.3		55.5		53.1		52.8	<0.13
饮酒量 (克/天)	60.1±7.5		78.2±8.5		79.3±6.8		106.8±10.2		<0.001
受教育程度	4.5±0.8		4.2±0.9		3.5±0.8		3.4±0.8		<0.05
社会支持	8.7±1.2		7.5±1.1		7.0±1.2		7.0±1.2		<0.05
慢性疾病数构成 (%)									
0		81.6		79.3		77.5		73.9	<0.18
1		15.1		16.2		16.5		15.0	<0.43
≥2		3.3		4.5		6.0		11.1	<0.05

表5. 某年某地居民不同心理分值的冠心病危险因素水平比较($\bar{X} \pm S$)

心理 分值	例数	年龄 (岁)	收缩压 (mmHg)	舒张压 (mmHg)	体力 活动	体重指数	吸烟量 (支/天)	饮酒量 (克/天)	受教育 程度	社会支持
1	252	35.2±6.5	120.7±13.4	78.8±10.2	2.1±0.2	23.1±3.2	8±1	60.1±7.5	4.5±0.8	8.7±1.2
2	253	37.0±6.3	121.2±13.2	77.9±10.5	2.1±0.1	24.0±3.5	10±2	78.2±8.5	4.2±0.9	7.5±1.1
3	252	36.5±6.8	121.1±13.2	78.2±11.0	2.1±0.2	24.8±3.1	15±2	79.3±6.8	3.5±0.8	7.0±1.2
4	253	37.8±6.5	120.4±12.8	78.4±10.6	2.3±0.3	25.8±3.1	15±2	106.8±10.2	3.4±0.8	7.0±1.2
P 值	-	<0.05	0.5	0.39	0.08	<0.01	<0.001	<0.001	<0.05	<0.05

表6. 某年某地居民不同心理分值的冠心病危险因素水平比较

心理分值	例数	吸烟率 (%)	饮酒率 (%)	慢性疾病数构成 (%)		
				0	1	≥2
1	252	70.8	52.3	81.6	15.1	3.3
2	253	69.4	55.5	79.3	16.2	4.5
3	252	70.7	53.1	77.5	16.5	6.0
4	253	71.1	52.8	73.9	15.0	11.1

百度文库>教育专区>高等教育>理学>统计学统计图表 (乱马小黑上传)

统计表 小结

内容一个中心
主谓一定分清
标目简明合理
线条数字要精

2

统计图

投票 最多可选6项

你绘制过什么类型的统计图？

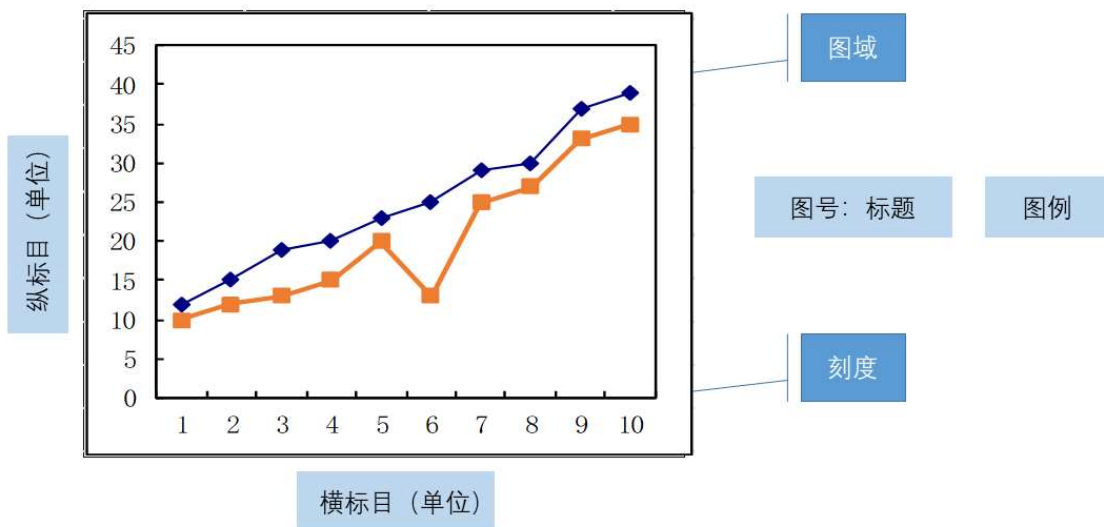
- A 条形图 (bar graph)
- B 线图 (line graph)
- C 散点图 (scatter plot)
- D 饼图 (pie chart)
- E 直方图 (histogram)
- F 箱图 (boxplot)

投票 最多可选6项

以前你是遵循什么原则为你的数据挑选统计图类型的？

- A 随便哪个类型，能画出统计图就好
- B 优先选择好看的
- C 根据想要说明的主题
- D 尝试不同的统计图，挑选看得顺眼的
- E 数据类型
- F 其他

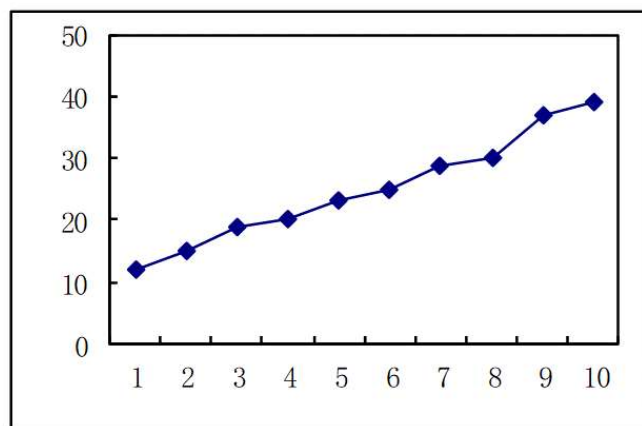
统计图的结构



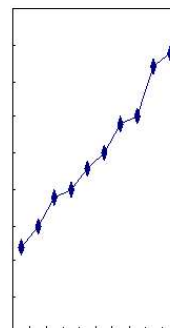
单选题 1分

图域的纵：横的长度比例一般为多少？

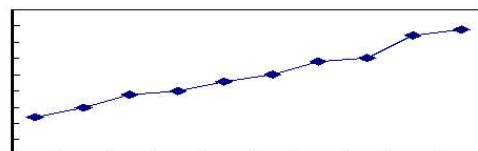
- ☒ A 2:3
- ☐ B 3:5
- ☐ C 5:7
- ☐ D 7:9



5:7



3:1



1:3

常用的统计图1：条形图

种类

简单条形图

集群条形图

堆积条形图



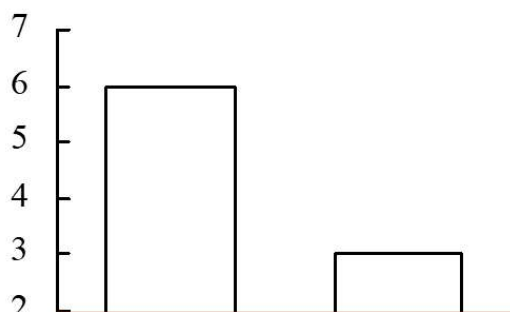
表2 某省某工厂 1994、1998年四项检测指标异常检出率

检测指标	1994年			1998年		
	受检人数	异常人数	检出率 (%)	受检人数	异常人数	检出率 (%)
血压	519	55	10.60	582	38	6.52
心率	519	44	8.48	582	39	6.70
TTT ^Δ	519	36	6.94	582	23	3.95
GPT [#]	519	20	3.85	582	16	2.75

Δ: TTT (麝香草酚浊度试验), #: GPT (谷丙转氨酶)。

(丁建生等. 中国卫生统计 1999; 16(3):166)

1. 1994年四项检测指标异常人数统计图
2. 1994、1998年四项检测指标异常人数统计图
3. 1994、1998年四项检测指标异常检出率统计图



甲

乙

纵轴尺度必须起点为零示意

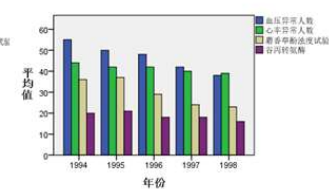
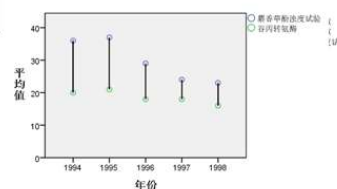
常用的统计图2：线图

种类

简单线图

多线线图

垂直线图



1. 1994~1998年血压指标异常人数统计图
2. 1994~1998年四项检测指标异常人数统计图
3. 1994~1998年肝功能异常指标TTT与GPT检出率差异

表7. 1994~1998四项指标异常检出率

检测指标	1994年			1995年			1996年			1997年			1998年		
	受检人数	异常人数	检出率 (%)	受检人数	异常人数	检出率 (%)	受检人数	异常人数	检出率 (%)	受检人数	异常人数	检出率 (%)	受检人数	异常人数	检出率 (%)
血压	519	55	10.60	539	50	9.28	550	48	8.73	575	42	8.73	582	38	6.52
心率	519	44	8.48	539	42	7.79	550	42	7.64	575	40	7.64	582	39	6.70
TTT ^Δ	519	36	6.94	539	37	6.86	550	29	5.27	575	24	5.27	582	23	3.95
GPT [®]	519	20	3.85	539	21	3.90	550	18	3.27	575	18	3.27	582	16	2.75

常用的统计图之比较：线图与条形图

异：

主题：线图更关注变化趋势，条图凸显个体绝对数值

横轴：线图为连续变量（关联），反应变量的发展变化情况

条形图为分类变量（相互独立），反应变量的相对大小

同：

纵轴：可以计数资料，也可为计量资料

单选题 1分

请问，1994年四项检测指标异常人数统计图，应该使用

☒ A 条形图

☐ B 线形图

常用的统计图2：半对数线图

表8. 绝对差与相对比的比较

A→B	绝对差	相对比 (A/B)	对数差 (lgA-lgB)
1000→100	1000-100=900	1000/100=10	lg1000-lg100=3-2=1
100→10	100-10=90	100/10=10	lg100-lg10=2-1=1
10→1	10-1=9	10/1=10	lg10-lg1=1-0=1

应用：
比较事物的发展
速度(相对比)

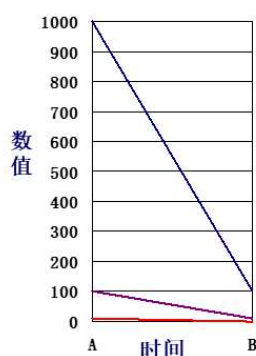


图1. 绝对差示意图 (算术格纸)

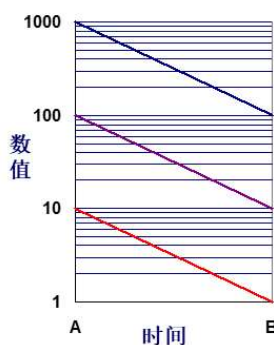


图2. 相对比示意图 (半对数格纸)

表9. 某市1949~1957年15岁以下儿童结核病和白喉死亡率 (1/10万)

年份	结核病死亡率	白喉死亡率
1949	150.2	20.1
1950	148.0	16.6
1951	141.0	14.0
1952	130.0	11.8
1953	110.4	10.7
1954	98.2	6.5
1955	72.6	3.9
1956	68.0	2.4
1957	54.8	1.3

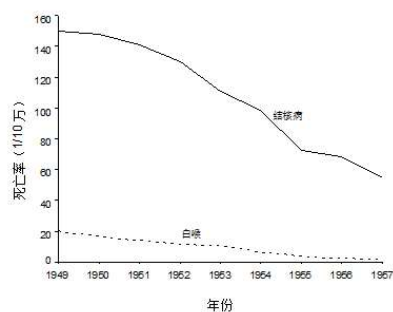


图3. 普通线图

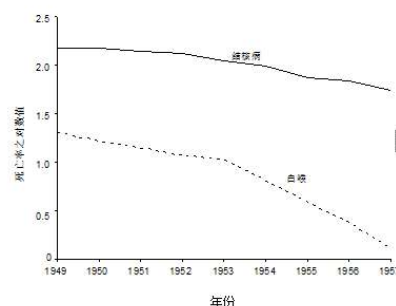
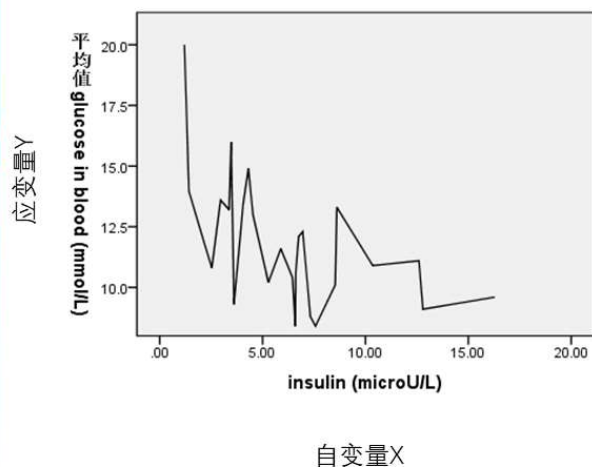


图4. 半对数图

图 2-14 某市 1949~1957 年 15 岁以下儿童结核病和白喉死亡率

常用的统计图3：散点图



线图局限性：

横坐标需有序均匀分布
样本量比较有限

点图：点的密集程度
点的趋势

相关关系

常用的统计图4：饼图或圆图

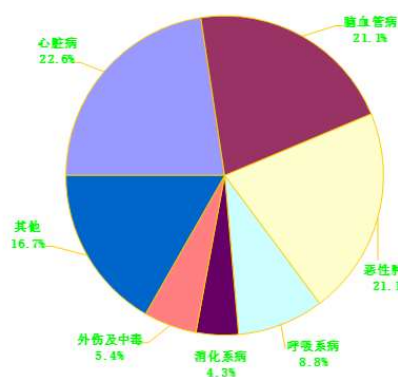


图5 1984年我国部分城市的死因构成

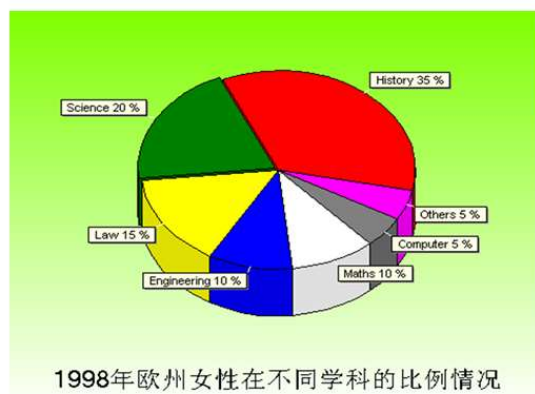
各个扇形面积（ 3.6° ）的大小反映了各组成部分百分比（1%）的大小。

注意事项：

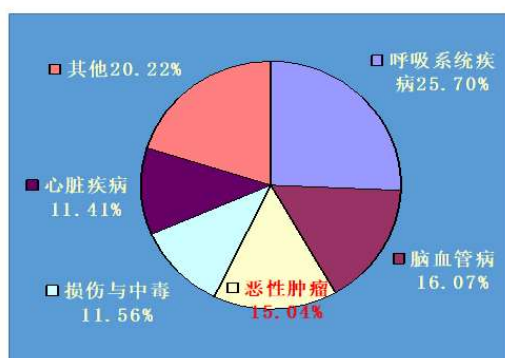
- ①各扇形按大小或逻辑顺序顺时针排列。
- ②从时针的9时或12时开始作图。
- ③扇形内要注明简要文字和百分比的数值。

表10. 1998欧洲各专业女性学生比例

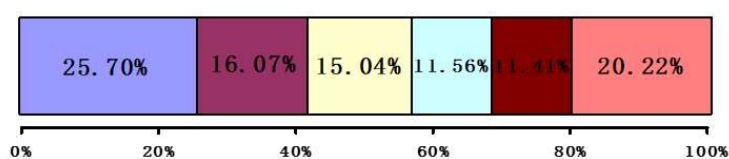
专业	百分比 (%)
历史	35
科学	20
法律	15
工程	10
数学	10
计算机	5
其他	5
	100



常用的统计图5：百分条图



用条图的外形
表达饼图的内涵



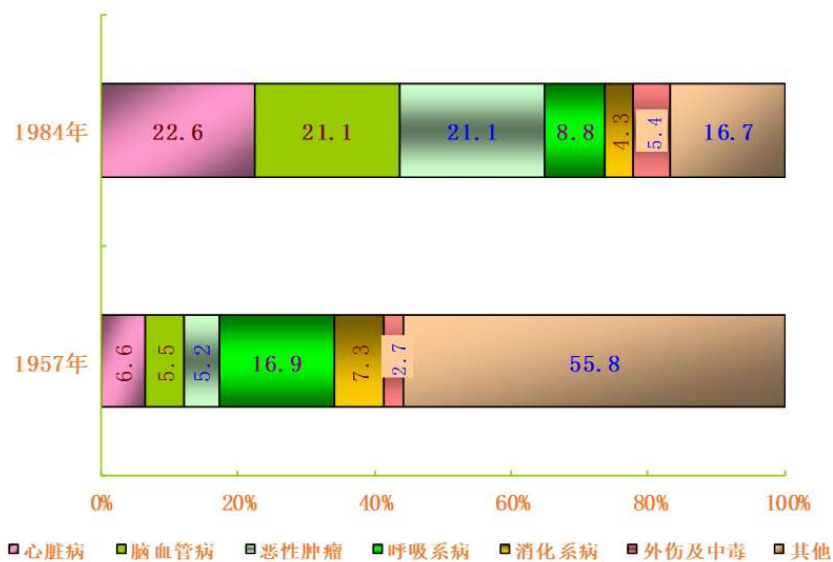


图7. 我国部分城市两个年度的死因构成比

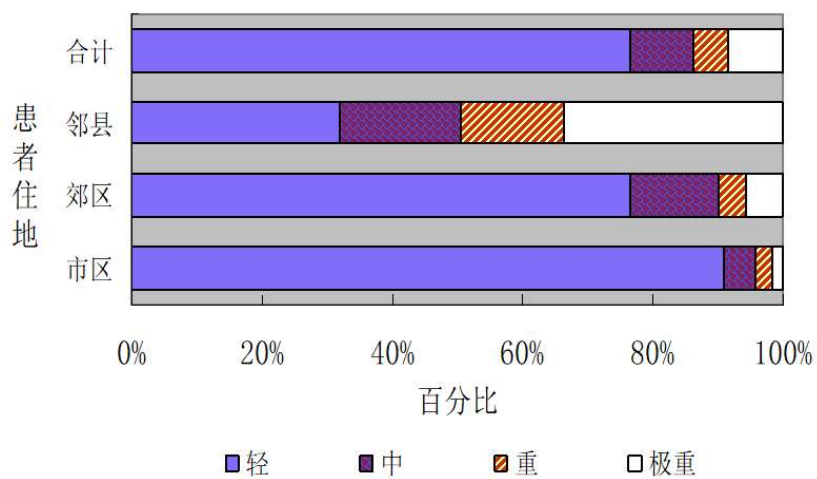
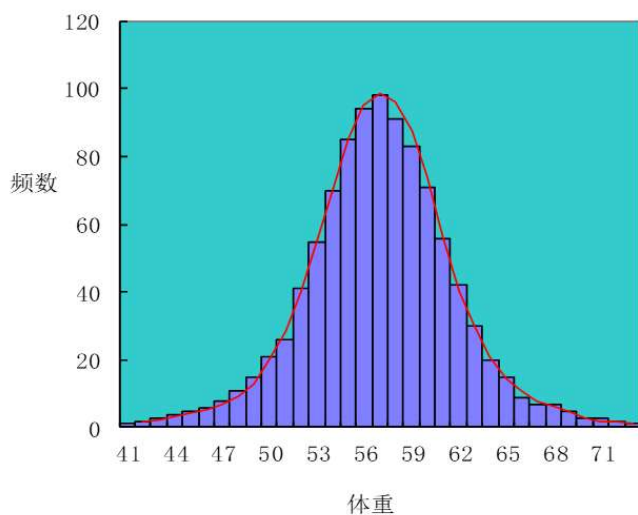


图8. 某市医院收治不同住地乙脑患者中不同病情者百分比

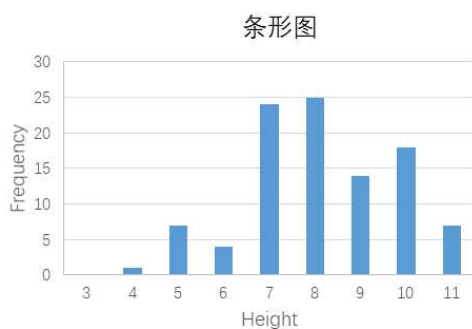
常用的统计图6：直方图



离散化的密度曲线

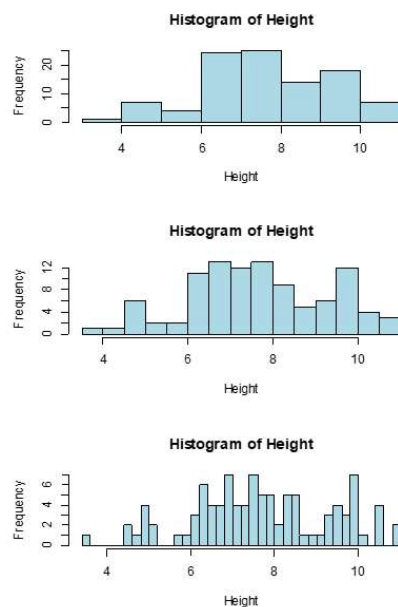
图9. 体重数据的分布

直方图与条形图的异同

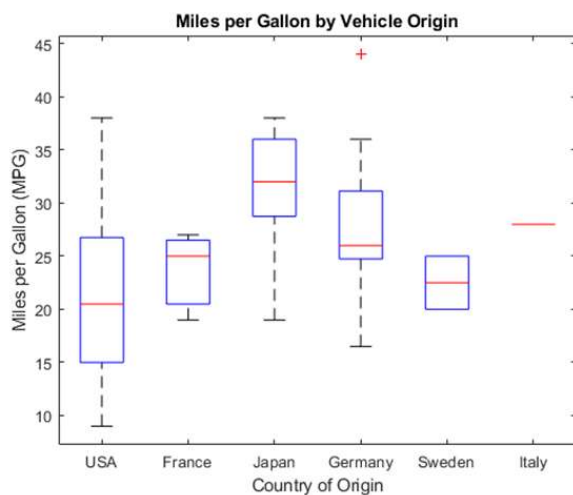


1. 横轴与纵轴的标目
2. bar间距与宽度
3. 数据类型
4. 图表制作方法

直方图

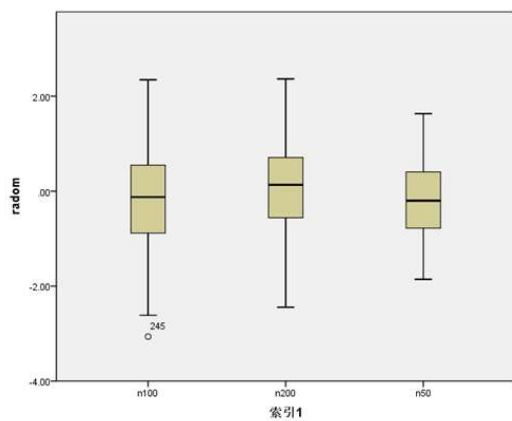


常用的统计图7：箱图 (boxplot)

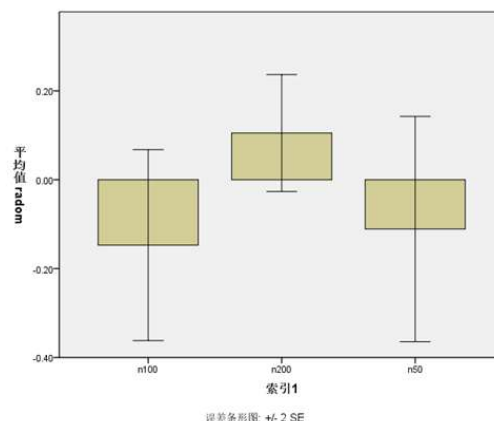


outlier
99.5%
3rd Quantile
Median 中位数
1st Quantile
0.5%
outlier

箱图 vs 误差条图



箱图



误差条图

资料与选图

图形	资料性质	分析目的
直条图	相互独立	直条长短表达数值大小
圆图	构成比	圆的扇行面积表达内部构成比
百分条图	构成比	长条各段的长度（面积）表达内部构成比
普通线图	连续性资料	线段的升降表达事物的动态（差值）变化
半对数线图	连续性资料	线段的升降表达事物的发展速度
直方图	数值变量的频数表资料	直方面积表达各组段的频数或频率
散点图	双变量资料	用点的密集程度或趋势表达两变量间相关关系
箱图	统计资料	表征数据的集中与离散程度

单选题 1分

观察肝癌死亡率和乙肝病毒感染率之间的关系

- A 百分条图
- B 直方图
- C 散点图**
- D 普通线图
- E 直条图
- F 半对数线图
- G 箱图

多选题 1分

观察脑血管意外病人的年龄分布

- ☒ A 百分条图
- ☒ B 直方图
- ☐ C 散点图
- ☐ D 普通线图
- ☒ E 直条图
- ☐ F 半对数线图
- ☐ G 箱图

单选题 1分

观察传染病发病率的变化趋势

- ☐ A 百分条图
- ☐ B 直方图
- ☐ C 散点图
- ☒ D 普通线图
- ☐ E 直条图
- ☐ F 半对数线图
- ☐ G 箱图

单选题 1分

观察传染病死亡在全死因中所占比重大小

- A 百分条图
- B 直方图
- C 散点图
- D 普通线图
- E 直条图**
- F 半对数线图
- G 箱图

单选题 1分

观察传染病死亡率随时间变化的速度

- A 百分条图
- B 直方图
- C 散点图
- D 普通线图
- E 直条图
- F 半对数线图**
- G 箱图

单选题 1分

观察不同地区的脑瘤死亡率

- A 百分条图
- B 直方图
- C 散点图
- D 普通线图
- E 直条图**
- F 半对数线图
- G 箱图

单选题 1分

请问南丁格尔的玫瑰图本质上是一种

- A 百分条图**
- B 直方图
- C 散点图
- D 普通线图
- E 直条图
- F 半对数线图
- G 箱图

填空题 1分

你觉得绘制统计图需要遵循什么原则？挑选你觉得正确的并请按重要性进行排序 [填空1]

- A 主题清晰
- B 视觉美观
- C 形式简洁
- D 条目清晰
- E 启发思考
- E 其他 [填空2]

观点分享1

- 1.图表是语言的一种重要形式，如果运用得当的话，比起表格来能更明快清晰地进行沟通；
- 2.决定图表形式的不是数据，也不是尺寸，而是你想说明的主题，你想指出的内容要点；
- 3.图表贵精不贵多，只有当图表能够帮助你表达主题时才使用；
- 4.图表是直观教具，它不能取代书写和讲述，只有在帮助你传达主题时，它才起作用。

麦肯锡《用图表说话》

观点分享2

怎样正确地使用图表？

1. 确定主题——你想表达什么观点？
2. 比较图表还是文字更有效？
3. 哪种图表最适合你的目的（所要表达的观点）？
4. 检查是否真实、有效地展示了数据？
5. 检查是否启发读者思考数据的本质和规律？
6. 检查是否表达了你的观点？

好的图表应该具备以下五点要求：

1. 巧妙地展示有意义的数据；
2. 防止对数据的曲解；
3. 鼓励读者比较不同的数据；
4. 吸引读者注意数据的实质，而不是其他；
5. 与数据统计和文字描述有机地整合。

摘自新浪博客：小蚊子数据分析的博客

观点分享3

统计图的要素

形式要切合主题：概括性陈述（启发性思考）
重点要清晰明了：可视化使得结论能够更加直观的展示
结构要简单易懂：图形的层次要符合逻辑便于理解

视觉美的享受！！

