




## 心理统计 第三讲：次数分布

严超赣  
Chao-Gan Yan, Ph.D.  
yancg@psych.ac.cn  
http://rfmri.org/yan

Institute of Psychology, Chinese Academy of Sciences

1

## 第三讲 次数分布

- 1 综述
- 2 次数分布表
- 3 次数分布图
- 4 茎和叶图
- 5 次数分布的形状
- 6 百分位数, 百分位等级, 插值法

2

### 描述统计的目标:

- 使得一组数据易于理解  
→ 图和表

3

### 一个例子

- 一位研究者调查了151位大学生, 问他们在过去一年中的应激程度, 用的是0到10点量表。151个大学生的数据如下:
- 4, 7, 7, 7, 8, 8, 7, 8, 9, 4, 7, 3, 6, 9, 10, 5, 7, 10, 6, 8, 7, 8, 7, 8, 7, 4, 5, 10, 10, 0, 9, 8, 3, 7, 9, 7, 9, 5, 8, 5, 0, 4, 6, 6, 7, 5, 3, 2, 8, 5, 10, 9, 10, 6, 4, 8, 8, 8, 4, 8, 7, 3, 8, 8, 8, 8, 7, 9, 7, 5, 6, 3, 4, 8, 7, 5, 7, 3, 3, 6, 5, 7, 5, 7, 8, 8, 7, 10, 5, 4, 3, 7, 6, 3, 9, 7, 8, 5, 7, 9, 9, 3, 1, 8, 6, 6, 4, 8, 5, 10, 4, 8, 10, 5, 5, 4, 9, 4, 7, 7, 7, 6, 6, 4, 4, 4, 9, 7, 10, 4, 7, 5, 10, 7, 9, 2, 7, 5, 9, 10, 3, 7, 2, 5, 9, 8, 10, 10, 6, 8, 3

4

### 用次数分布表整理

在应激量表每一点上的学生数目

应激分数	学生数目	百分比
10	14	9.3
9	15	9.9
8	26	17.2
7	31	20.5
6	13	8.6
5	18	11.9
4	16	10.6
3	12	7.9
2	3	2.0
1	1	0.7
0	2	1.3

5

### 不满意百分率最高的项目

- A2 Policies applied fairly 41%
- F1 Training to do job effectively 37%
- D2 Receive all information 34%
- F2 Encouraged to improve skills 31%
- F3 Feedback on performance 30%

6

## 不满意百分率最低的项目

- A1 Contribution to Company's success 2%
- G1 Good company to work for 7%
- X1 Familiar with objectives and values 9%
- X2 Local consistent with company 11%
- X3 Practices values 13%
- B1 Deliver to customers 13%
- B2 Commitment to customer satisfaction 13%
- B3 Meeting customers needs 14%

7

## 本章学习目标

1. 学会作次数分布, 累计次数分布, 和分组次数分布表
2. 学会用茎和叶图, 直方图, 棒图, 线图表达次数分布
3. 了解正偏态分布和负偏态分布
4. 掌握百分位数和插值法

8

## 次数分布 – 最简单的描述统计

- 描述统计的目的: 简化和整理数据的表达。
- 次数分布 (Frequency Distribution): 是指一批数据在某一量度的每一个类目所出现的次数情况
- 组织此类数据的第一种方法是: 建立次数分布表
- 次数分布表和次数分布图就是表达一组数据是如何在某一度量上分布的

9

## 又一个例子

- 例1: 某个班的26个学生在一次测验中的分数如下 (10分为满分):  
9, 2, 3, 8, 10, 9, 9, 2, 1, 2, 9, 8, 2, 5, 2, 9, 9, 3, 2,  
5, 7, 2, 10, 1, 2, 9  
将这些分数作成一个个次数分布表

X	f
10	2
9	7
8	2
7	1
6	0
5	2
4	0
3	2
2	8
1	2

10

## 简单次数分布表

- 次数分布表是将位于量数的每一个类目的个体的个数组织成表格的形式
  - 变量的值? -填充x列  
从大到小, 每个可能的值
  - 每个值出现多少次 (发生次数)? -填充f列
  - 观察的总数? 将次数行求和, 将得到  $S f = N$
  - 变量的总值? 最简单的方法就是求(X)和 (f)的乘积列, 然后将结果求和  $S(Xf)$

11

## 简单次数分布表

- 例1: 对于下面的次数分布表:
  - a) 此分布中共有几个分数 ( $N = ?$ )
  - b) 对这些分数求和  $\sum X$

x	f	xf
4	2	8
3	4	12
2	5	10
1	3	3
$\Sigma$	14	33

$$N = \sum f = 14$$

$$\sum X = \sum xf = 33$$

12

## 每个分数的比率和百分率

- Proportion= $p=f/n$

x	f	$p=f/n$	$\%=p(100)$
10	2	$2/20=0.10$	10%
9	5	$5/20=0.25$	25%
8	7	$7/20=0.35$	35%
7	3	$3/20=0.15$	15%
6	2	$2/20=0.10$	10%
5	0	$0/20=0$	0%
4	1	$1/20=0.05$	5%

13

## 分组次数分布表 (Grouped Frequency Tables)

- 当变量的值分布在一个较广的范围时，需要分组次数分布表简洁地描述数据
- 常常以区间或类别的形式出现，而不是某一特定值。例如学生成绩：(A = 90-99, B = 80-89, ...)
- 编制分组次数分布表的步骤
  1. 把所有数据按大小顺序排列
  2. 求全距
  3. 定组数
  4. 定组距
  5. 写出区间上下限
  6. 统计每个区间的次数

14

## 分组的“惯常法则” (rules of thumbs)

- 1. 分组次数分布表应该有大约10 (5-15) 个区间，目的是使这组数据易于直观感受和理解
- 2. 所有区间的宽度应该相等
- 3. 组距应该是个比较简单的数字，如2, 5, 10, 20
- 4. 每个区间开始的分数应该是组距的倍数

15

## 又一个例子

- 例3: 以下是22个同学的测验分数，编制分组次数分布表  
94,92,82,87,81,74,75,78,76,76,63,60,62,54,51,56,59,43,32,36,25,18

16

## 又一个例子

- 例3: 以下是22个同学的测验分数，编制分组次数分布表  
94,92,82,87,81,74,75,78,76,76,63,60,62,54,51,56,59,43,32,36,25,18

分数的分布范围是18-94

- 2为组距，分38组
- 5为组距，分16组
- 10为组距，分9组

x	f
90-99	2
80-89	3
70-79	5
60-69	3
50-59	4
40-49	1
30-39	2
20-29	1
10-19	1

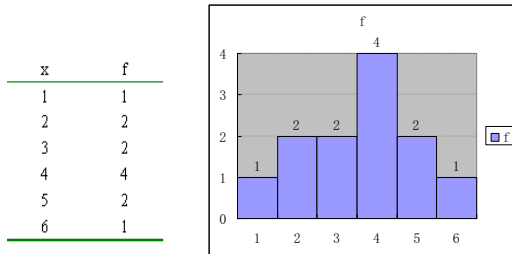
17

## 精确界限 (Real Limits)

- 对于一个连续型变量，每个分数实际对应量尺上的一段区间。分割这些区间的界限叫做精确界限 (real limits)。分割两个邻近分数的精确界限位于两个分数的中间。
- 每个分数有两个精确界限，一个在区间的顶端，称为精确上限 (upper real limit)，另一个在区间的底端，称为精确下限 (lower real limit)。
- 注意一个区间的精确上限也是高一个区间的精确下限

18

## 次数分布图



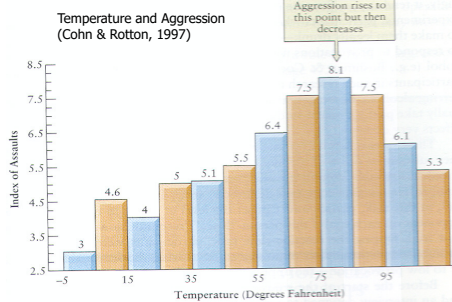
19

## 直方图 (histogram)

- 用一些垂直条画在每个分数之上
  - 垂直条的高度代表次数
  - 垂直条的宽度代表分数的精确区间.
  - 临近的垂直条连接在一起
  - 只有数据是等距或等比量度 (连续变量) 时, 才能用直方图

20

## 直方图



21

## 绘制一个直方图来表达例2的分布

- 水平轴 - X 轴 (abscissa) - X 的值
- 垂直轴 - Y 轴 (ordinate) - 次数
- 绘制直方图的一般规则
  - X 轴与Y 轴的交叉点应该是0
  - Y 轴高度大约是X 轴的2/3 到3/4
- 绘制直方图的方法
  - 用坐标纸绘制
  - 用Excel, SPSS, R, Python, Matlab, JASP等电脑软件绘制

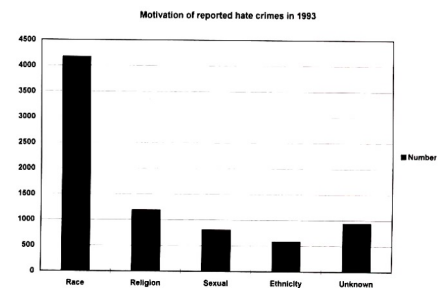
22

## 棒图 (条形图; bar graph)

- 表达质性研究的数据(qualitative data).
- 用一些垂直条画在每个分数(或类别)之上
  - 垂直条的高度代表次数
  - 每个垂直条之间有一段空间.
  - 只有数据是命名或顺序量度(离散型变量)时, 才能用棒图

23

## Bar Graph



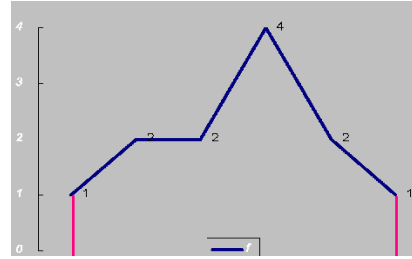
24

## 折线图 (line graph)

- 又称次数分布多边形 (frequency distribution polygon)
  - 在每一分数上面画点
  - 点的高度代表次数
  - 一条连续折线将这些点连接在一起
  - 表达数据的变化趋势

25

## 折线图 (line graph)



26

## 茎和叶图 (stem and leaf display)

- 例4: 考察下列茎和叶图

				茎	叶
88	80	75	97	9	7 6 5
57	69	74	96	9	4 4 3 2 0
86	79	68	56	8	8 6 5 5
95	85	79	74	8	3 2 1 1 0
54	68	94	67	7	9 9 8 7 7 6 5 5 5
73	85	78	53	7	4 4 3 1 1 0
83	65	77	48	6	9 8 8 7 5
94	71	93	82	5	7 6
77	71	64	81	5	4 3
76	70	92	64	4	8
89	81	75	63	4	
75					

表(1)某校 98 级概率统计成绩      图(1)成绩茎叶图表

27

## 如何准确地用图表达数据

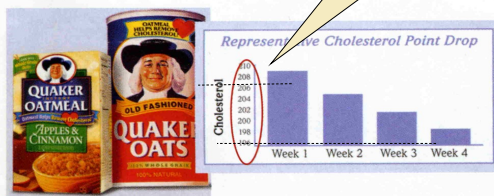
- 把复杂的观点传递得清晰, 准确, 高效.
- 在最短的时间里用最少的笔墨和空间传递最多的信息.
- 图形的实际差别与数据的差别成正比.
- 标注清楚, 详细, 完全. 测度单位要一致.

28

## 误导读者的图

### GLOATING OATS

At first glance, the graph from a Quaker... cholesterol levels plummet when you eat... close look at the left side shows that the num... age from 196 to 210, with levels falling from 209 to 199. A dp... e, but no nosedive.



Submit items for Selling It to our Editorial Department. See p

29

## 次数分布的形状

- 用3个特征可以完整地描述一个分布:
  - 形状 (shape)
  - 集中趋势 (central tendency)
  - 变异性 (variability)

30

### 对称分布 (symmetrical distribution)

- 可以画一条垂直线穿过分布的中央，使得分布的一边恰是另一边的镜像。



31

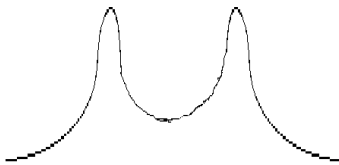
### 偏态分布 (skewed distribution)

- 分数堆积在分布的一端，而另一端成为比较尖细的尾端 (tail)
  - 偏态分布尾端向右的称为正偏态 (positively skewed) (因为其尾端指向正数)
  - 偏态分布尾端向左的称为负偏态 (negatively skewed) .



32

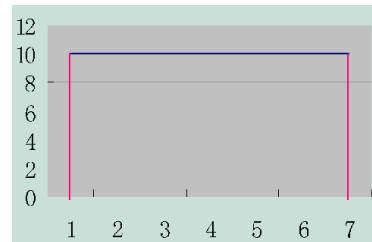
### 双峰分布 (bi-modal distribution)



33

### 矩形分布 (rectangle distribution)

- 一个分布中所有的值都是相同频次



34

### 百分位数 (percentile)

- 以上是描述观察的整体，而我们也可用次数分布来描述某一个别点在一个集中的位置
- 原始分数本身并不提供该点在一个集中的位置的信息
- 一个分数的等级 (rank) 或百分位数等级 (percentile rank) : 某一分布中分数在某一值之下或等于该值的个体所占的百分比。当一个分数被其百分位数等级，该分数称为百分位数 (percentile) .

35

### 累计次数分布

- $cf$  = 累计次数 (cumulative frequency)
- $c\%$  =  $cf/N \times 100\%$  累积百分比 (cumulative percentage)

例 4: 此表是一次词汇测验的分数:

X	f	p	%	cf	c%
5	2	.05	5	40	100
4	10	.25	25	38	95
3	16	.40	40	28	70
2	8	.20	20	12	30
1	4	.10	10	4	10

36

## 如何确定百分位数

此表是一次词汇测验的分数：

X	f	p	%	cf	c%
5	2	.05	5	40	100
4	10	.25	25	38	95
3	16	.40	40	28	70
2	8	.20	20	12	30
1	4	.10	10	4	10

cf = 累积次数 (cumulative frequency)

c% = 累积百分比 (cumulative percentage)

- 95 百分位数等级的所对应的测验分数是多少？
- 如果你在测验中得到 4 分，你的百分位数等级是多少？

37

## 如何确定百分位数

- 对于分数 4，其对应的累积百分比是 95%。但注意：分数 4 意味着一个人得分在 3.5 和 4.5 之间。累积百分比表明组距的精确上限。因此，95 的百分位数是与 4.5 相对应（而不是 4.0）。
- 找出分布中 4 分的精确上限和精确下限的累积次数。
  - 对于分数 4.5，其对应的累积百分比是 95
  - 对于分数 3.5，其对应的累积百分比是 70
  - 对于分数 4.0，其对应的累积百分比是多少呢？

38

## 插值法 (Interpolation)

- 有时你所感兴趣的值并未出现在表内。此时你需要做基于经验的猜测。
- 插值法的假设是在所求解点的附近 1 个组距单位区间之间的分数和对应的百分比的变化是线性的。
  - 关键在于要找到距求解点最近的两个区间，对于较远的区间，则不满足分数和对应的百分比线性变化的假设。

39

X	f	p	%	cf	c%
5	2	.05	5	40	100
4	10	.25	25	38	95
3	16	.40	40	28	70
2	8	.20	20	12	30
1	4	.10	10	4	10

- 对于分数 4.5，其对应的累积百分比是 95

▪ 对于分数 3.5，其对应的累积百分比是 70

对于分数 4.0，其对应的累积百分比是 X？

40

4.5 --- 95%

4.0 --- ?

3.5 --- 70%

$$\frac{4.0-3.5}{4.5-3.5} = \frac{X-70}{95-70}$$

$$4.5-3.5 \quad 95-70$$

$$X=82.5$$

41

## 练习

x	f	cf	c%
20-24	10	50	100%
15-19	10	40	80%
10-14	15	30	60%
5-9	10	15	30%
0-4	5	5	10%

42

## 解答

$$\begin{aligned} 14.5 & \text{-----} 60 \\ X & \text{-----} 50 \\ 9.5 & \text{-----} 30 \\ (X-9.5) / (14.5-9.5) & = (50-30) / (60-30) \\ X & = 12.83 \end{aligned}$$

43

## 在研究中应用次数分布表和次数分布图举例

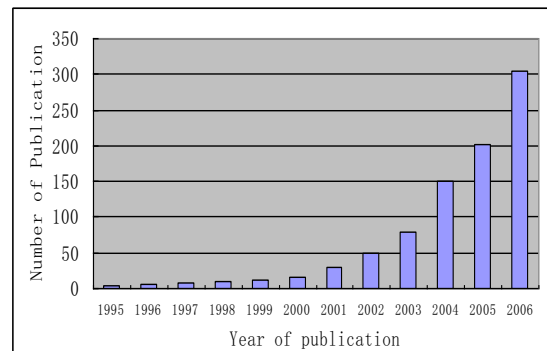
1. 被试的人口学资料常常用到次数分布表
2. 一篇综述中,为描述工作倦怠文献逐年增加的情况,我们用棒图表示

44

Table 1. Participant's Demographic Information (n=158)

	Frequency	Percent
Age group		
25 or below	49	31.0
26-30	72	45.6
31-35	20	12.7
36-40	12	7.6
40-45	5	3.2
Education		
High school	10	6.3
college	54	34.2
bachelor	80	50.6
master or higher	14	8.8
Industry		
telecommunication	15	9.5
Internet	25	15.8
advertising	69	43.7
trade and busyness	27	17.1
IT	22	13.9
Work Experience		
5 years or below	98	62.0
6-10 years	37	23.4
11-15 years	14	8.9
16 years or above	9	5.7

45



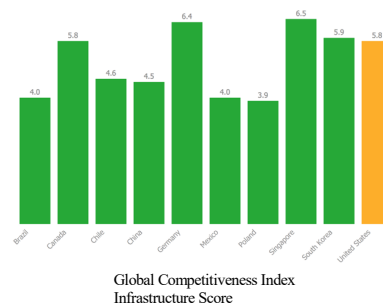
46

## Bar graph (条形图)

- The pictorial representation of grouped data, in the form of vertical or horizontal rectangular bars
- the lengths of the bars are equivalent to the measure of data, are known as bar graphs or bar charts.
- The bars drawn are of uniform width, and the variable quantity is represented on one of the axes. Also, the measure of the variable is depicted on the other axes.

47

## Bar graph (条形图)

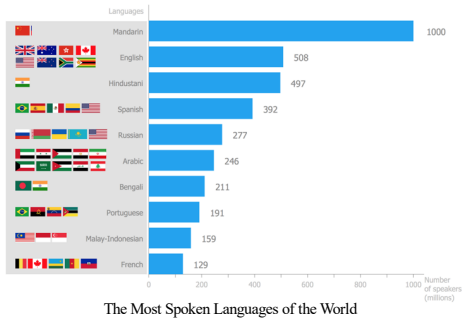


<https://www.conceptdraw.com/solution-park/graphs-charts-basic-bar-graphs>

48



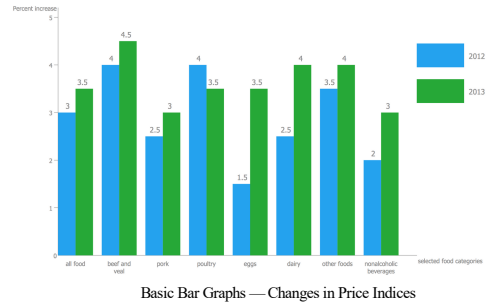
## Bar graph (条形图)



<https://www.conceptdraw.com/solution-park/graphs-charts-basic-bar-graphs>

49

## Bar graph (条形图)



<https://www.conceptdraw.com/solution-park/graphs-charts-basic-bar-graphs>

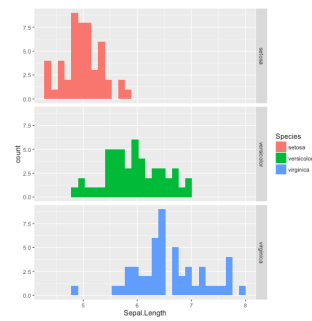
50

## Histogram (直方图)

- a graphical representation of a grouped frequency distribution with continuous classes.
- It is an area diagram and can be defined as a set of rectangles with bases along with the intervals between class boundaries and with areas proportional to frequencies in the corresponding classes.

51

## Histogram (直方图)



<http://book.biologistsguide2computing.com/en/stable/data-visualisation.html>

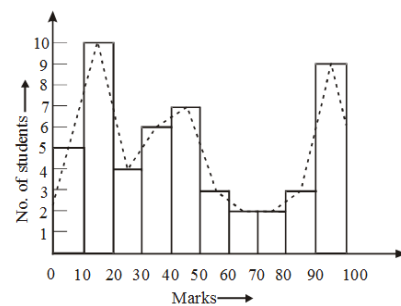
52

## Frequency polygon (频数折线图)

- A frequency polygon is almost identical to a histogram.
- Statistics deals with the collection of data and information for a particular purpose.
- Frequency polygons are a visually substantial method of representing quantitative data and its frequencies.

53

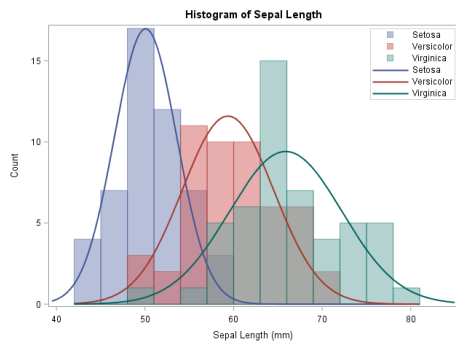
## Frequency polygon (频数折线图)



<https://www.aplustopper.com/frequency-polygon/>

54

## Frequency polygon (频数折线图)



<https://sasnr.com/sas-histogram-example/>

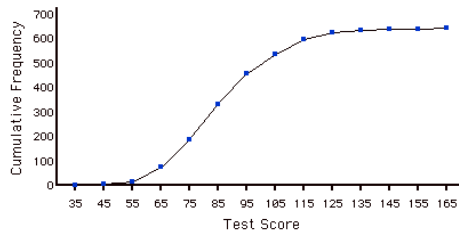
55

## Cumulative frequency polygon (累积频数折线图)

- A diagram representing grouped numerical data in which cumulative frequency is plotted against upper class boundary, and the resulting points are joined by straight line segments to form a polygon.
- The polygon starts at the point on the x-axis corresponding to the lower class boundary of the lowest class.

56

## Cumulative frequency polygon (累积频数折线图)



[https://onlinestatbook.com/2/glossary/cumulative\\_frequency\\_poly.html](https://onlinestatbook.com/2/glossary/cumulative_frequency_poly.html)

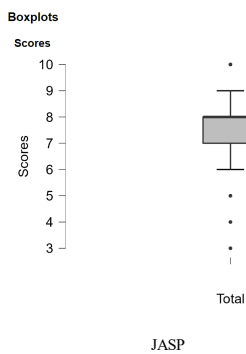
57

## Boxplot (箱型图)

- The method to summarize a set of data that is measured using an interval scale is called a box and whisker plot.
- We display the data distribution in a standardized way using 5 summary – minimum, Q1 (First Quartile), median, Q3 (third Quartile), and maximum.

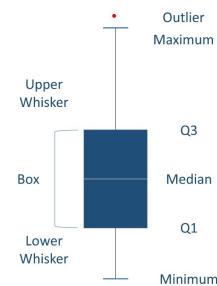
58

## Boxplot (箱型图)



59

## Boxplot (箱型图)



<https://insightoriel.com/what-is-boxplot-box-and-whisker-plot/>

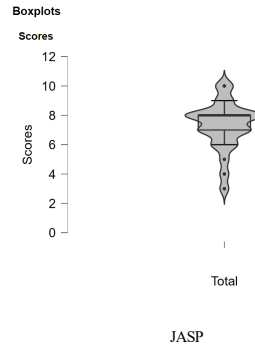
60

## Violin Plot (小提琴图)

- A violin plot depicts distributions of numeric data for one or more groups using density curves.
- The width of each curve corresponds with the approximate frequency of data points in each region.
- Densities are frequently accompanied by an overlaid chart type, such as box plot, to provide additional information.

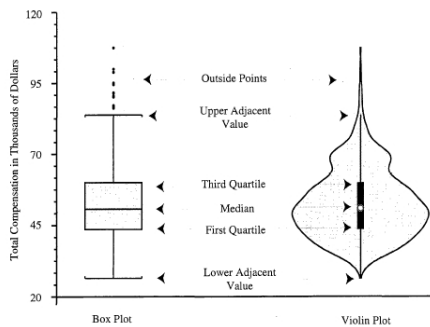
61

## Violin Plot (小提琴图)



62

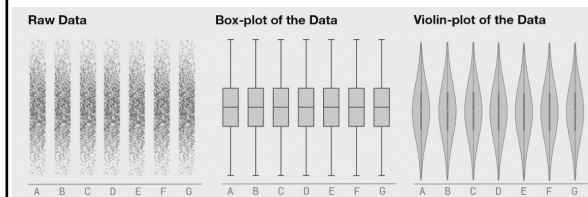
## Violin Plot (小提琴图)



<https://zhuanlan.zhihu.com/p/376055263>

63

## Violin Plot (小提琴图)



<https://zhuanlan.zhihu.com/p/376055263>

64

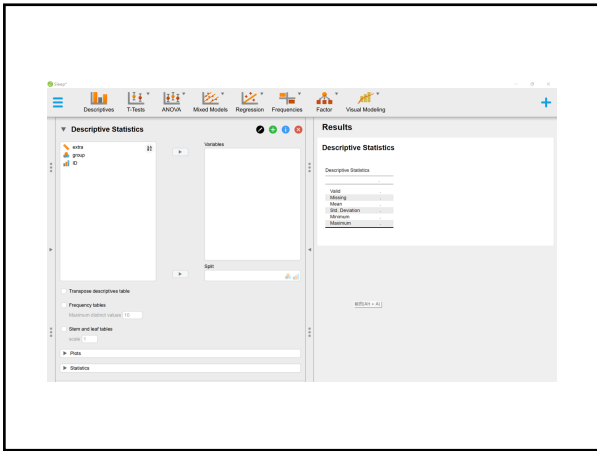


JASP is a free, open-source alternative to SPSS that allows you to perform both simple and complex analyses in a user-friendly package. JASP is a free alternative that aims to give you exactly what you want when you are analysing data.

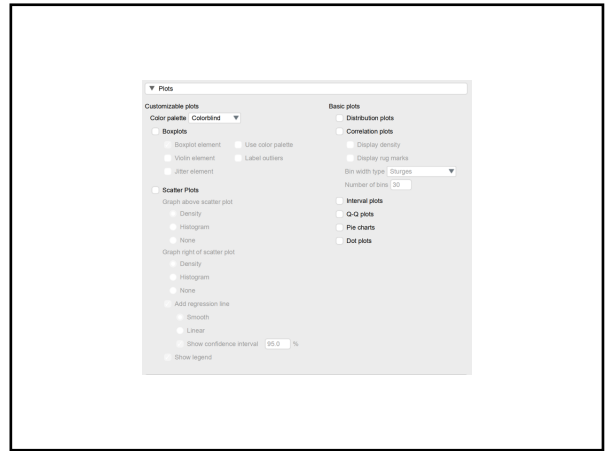
65

1	0.7	1	2
2	-0.6	1	2
3	0.2	1	3
4	-0.2	1	4
5	0.1	1	5
6	0.4	1	6
7	0.7	1	7
8	0.9	1	8
9	0	1	9
10	2	1	10
11	1.9	2	1
12	0.8	2	2
13	1.1	2	3
14	0.1	2	4
15	-0.1	2	5
16	0.4	2	6
17	0.5	2	7
18	1.8	2	8
19	1.6	2	9
20	0.4	2	10

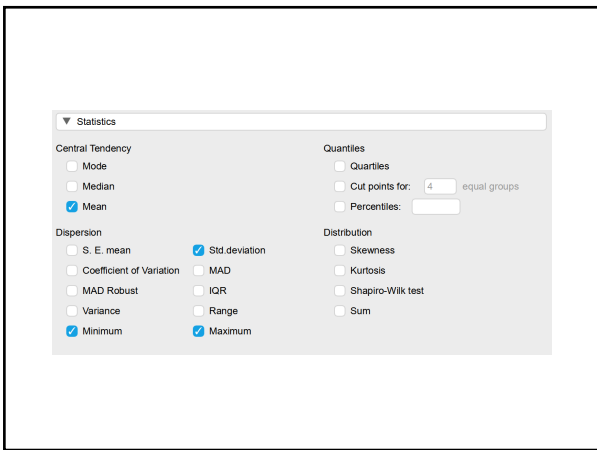
66



67



68



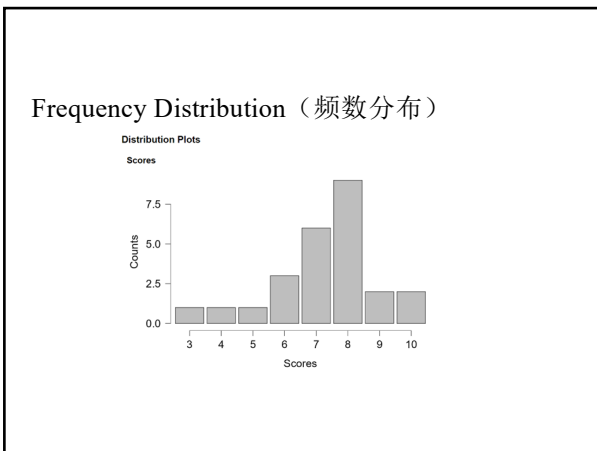
69

**Frequency Tables ▼**

Frequencies for Scores

Scores	Frequency	Percent	Valid Percent	Cumulative Percent
3	1	4.000	4.000	4.000
4	1	4.000	4.000	8.000
5	1	4.000	4.000	12.000
6	3	12.000	12.000	24.000
7	6	24.000	24.000	48.000
8	9	36.000	36.000	84.000
9	2	8.000	8.000	92.000
10	2	8.000	8.000	100.000
Missing	0	0.000		
Total	25	100.000		

70



71

**测查**

1. 一位研究者作大学生的智商研究。他实验前先收集了如下一般资料：

- 学生的家庭收入状况：低，中，高；
- 学生的考入地区：北方，中部，南方；
- 学生的专业；
- 学生的GPA

72

## 测查

学生	家庭收入	考入地区	专业	GPA
S1	低	北方	社会学	3.11
S2	中	中部	生理学	1.88
S3	高	中部	人类学	2.64
S4	低	中部	教育学	1.12
S5	中	北方	社会学	2.53
S6	高	南方	社会学	2.96
S7	中	北方	英语	3.16
S8	中	北方	工程	2.32
S9	中	中部	法语	1.96
S10	中	中部	化学	4.00
S11	低	南方	音乐	1.27
S12	低	中部	社会学	2.42
S13	中	南方	心理学	4.00
S14	中	北方	物理学	0.76
S15	中	中部	社会学	2.70
S16	高	北方	英语	2.10
S17	中	北方	社会学	3.83
S18	中	南方	社会学	0.09
S19	中	中部	心理学	2.65
S20	低	北方	人类学	2.17
S21	中	南方	生理学	2.73
S22	高	中部	心理学	0.13
S23	低	中部	社会学	1.58
S24	中	中部	经济学	3.68
S25	中	中部	管理学	3.25

73

## 测查

1. 一位研究者作大学生的智商研究。他实验前先收集了如下一般资料：

- a) 学生的家庭收入状况：低，中，高；
- b) 学生的考入地区：北方，中部，南方；
- c) 学生的专业；
- d) 学生的GPA

- 1) 该研究的样本量是多大？
- 2) 对a) -d) 的4个变量，其量度的类型是什么？它们是离散的还是连续的？
- 3) 作出家庭收入状况，考入地区和专业的次数分布表
- 4) 作GPA 的分组次数分布表

74

## 测查

2、研究者评价4种品牌的咖啡的味道，方法是让被试逐一品尝并对其进行做5点量表的评价（1——很糟糕，5——非常好）。结果如下：

品牌	平均评价值
品牌A	2.5
品牌B	4.1
品牌C	3.2
品牌D	3.6

- 1) 指出本研究中的自变量和因变量
- 2) 本测量中的自变量是什么类型的数据（命名、顺序、等距、等比）
- 3) 如果要用图来表示自变量和因变量的关系，应该用什么图（线图、直方图、棒图）？
- 4) 用图表示本实验的结果。

75

## 测查

3、研究者研究失眠者接受放松训练的次数对失眠治疗的效果。有4个实验组，分别接受2, 4, 8次的训练，控制组不接受训练（0次）。研究者测量被试入睡所需的时间，结果如下：

训练次数	入睡所需时间（分钟）
0	72
2	58
4	31
8	14

- 1) 判断本研究的自变量和因变量
- 2) 自变量是什么数据类型（命名、顺序、等距、等比）？
- 3) 如果要用图来表示自变量和因变量的关系，应该用什么图（线图、直方图、棒图）？
- 4) 用图表示本实验的结果。

76