

经济统计学

刁莉男

diaoln@jlu.edu.cn

2012.03.07

- 课件下载: <http://diaolinan.weebly.com>
- 期中考试: 2012.04.25(第八周)
- 期末考试: 2012.06.25-2012.06.29 (第十七周)
- 习题课与上机课: 第3周、6周、9周、12周、15周
 - 习题课: 10课时 (5*2课时)
 - 上机课: 20课时 (5*4课时)

参考书:

双语教材: Lind, Douglas A., Marchal, William G., and Wathen, Samuel A., "Basic Statistics for Business & Economics", 6th Edition, McGraw-Hill Company, Inc. 东北财经大学出版社。

汉语译本: 道格拉斯·A·林德（作者），威廉·G·马夏尔（作者），塞缪尔·A·沃森（作者），《商务与经济统计学（第6版）》，东北大学出版社，王维国（译者）。

其它参考: 宋廷山，葛金田，《统计学以Excel为分析工具》，北京大学出版社，21世纪全国高等院校财经管理类规划教材。

第一章：什么是统计学

提纲

- 为什么要学习统计学？
- 统计的含义，
- 统计的类型
- 变量的类型
- 测量的尺度

为什么要学习统计学？

1. 数字信息随处可见，经济增长、股票市场、房地产市场、消费、就业等。

例1：数据来源：国家统计局网站

http://www.stats.gov.cn/tjfx/jdfx/t20120218_402785249.htm

表1：2012年1月70个大中城市新建住宅价格指数

城市	新建住宅价格指数			城市	新建住宅价格指数		
	环比	同比	定基		环比	同比	定基
北 京	99.9	100.1	102.5	唐 山	100.0	100.6	101.7
天 津	99.8	100.2	103.2	秦 皇 岛	99.8	100.5	106.3
石 家 庄	99.7	101.3	107.7	包 头	99.8	100.0	103.9
太 原	100.0	100.9	101.7	丹 东	100.0	99.5	107.7
呼 和 浩 特	100.0	102.2	105.0	锦 州	100.0	101.6	105.2
沈 阳	99.9	101.9	106.1	吉 林	99.7	100.0	105.7
大 连	99.9	102.1	105.4	牡 丹 江	100.0	100.5	106.9
长 春	99.9	101.4	104.0	无 锡	99.8	99.5	101.4
哈 尔 滨	100.0	100.0	103.5	扬 州	99.8	100.8	103.6
上 海	99.9	100.7	101.6	徐 州	99.9	101.2	103.0
南 京	99.6	98.7	99.5	温 州	99.4	92.4	92.7

为什么要学习统计学？

1. 数字信息随处可见， 经济增长、 股票市场、 房地产市场、 消费、 就业等。

例2: 数据来源: 国家统计局网站

http://www.stats.gov.cn/tjfx/jdfx/t20120209_402782759.htm

2012年1月居民消费价格主要数据

	2012年1月	
	环比涨跌幅(%)	同比涨跌幅(%)
居民消费价格总指数	1.5	4.5
其中:城市	1.5	4.5
农村	1.5	4.6
其中:食品	4.2	10.5
非食品	0.2	1.8
其中:消费品	1.8	5.5
服务项目	0.8	2.2
分类别		
一、食品	4.2	10.5
粮 食	0.3	6.1
油 脂	0.2	6.1
肉禽及其制品	3.2	18.7
其中:猪肉	3.9	25.0
蛋	-0.8	-2.0

为什么要学习统计学？

2. 做决策所用到的统计技术影响着我们的日常生活。 例如：

- 保险公司利用统计数据制定赔付率。
- 环保部门根据统计数据确定空气质量，江河湖泊的污染程度。
- 医疗工作者根据统计数据确定采取哪种药物或者治疗手段。

为什么要学习统计学？

3.学习统计方法将帮助我们明白决策是怎样制定的以及更好地帮助我们理解统计是怎样影响我们的。

例如：

- 决定现有数据是否充分，是否需要额外的信息。
- 如果有必要，以不会产生误导性的结果为前提继续搜集信息。
- 使用有效地途径总结信息。
- 分析信息。
- 得出结论，并且要考虑如果结论不正确的风险是什么。

统计的含义

几乎所有的教材都写到，统计有或三个含义：统计活动（工作），统计资料和统计学。

统计（Statistic, pl. Statistics）牛津英汉双解词典

- Statistic: 一项统计数据。

An important statistic is that 94% of crime relates to property.

- Statistics: 一组统计数据。

crime/unemployment statistics.

美国汽车的行驶距离是11,099英里/年，公共汽车的行驶距离是9,353英里/年，卡车是13,942英里/年；而加拿大汽车的是10,371英里/年，公共汽车是19,823英里/年，卡车是7,001英里/年。

- Statistics: 统计学。There is a compulsory course in statistics.

统计学绝不仅仅是收集和列出数字信息这么简单。

统计学

统计学是一门收集、组织、提供、分析和解释数据以协助做出有效决定的科学。

【Lind版】

- 根据定义，调查问题的第一步是收集相关数据。
- 数据以某种方式进行组织（例如图表）。
- 只有数据在被组织之后我们才能分析和解释它们。

例如：Frito-Lay公司主要零食和薯片的销售额以及市场份额

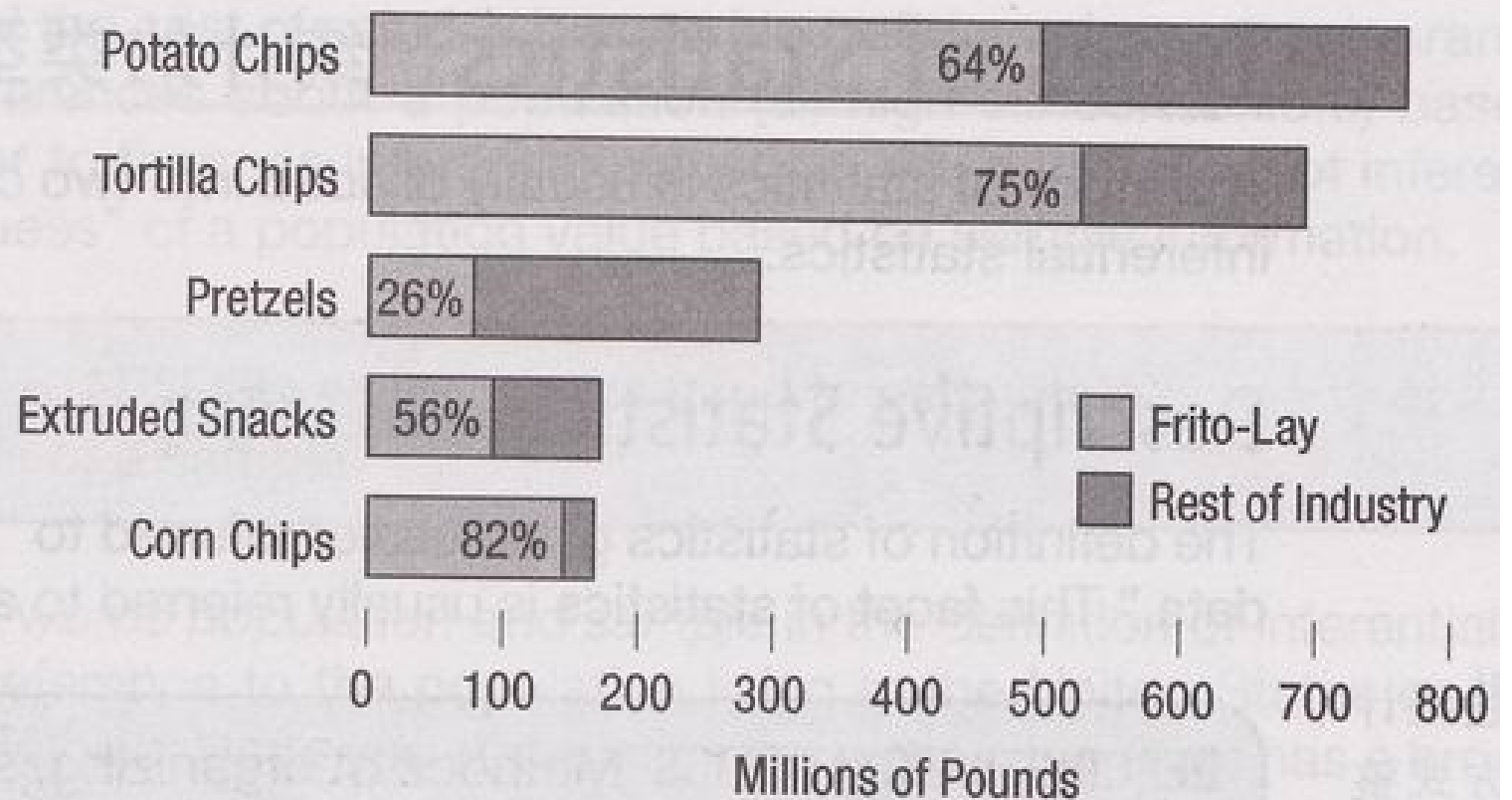


CHART 1-1 Frito-Lay Volume and Share of Major Snack Chip Categories in U.S. Supermarkets

统计的类型

从方法的功能来看:

- 描述统计
- 推断统计

从研究的重点来看:

- 理论统计
- 应用统计

- 描述统计（学）(Descriptive Statistics)

【宋廷山版】 研究如何取得反映客观现象的数据，并通过图表形式对所搜集的数据进行加工处理和显示，进而通过综合、概括与分析得出反映客观现象的规律性数量特征。

【Lind版】 以直观有效的方式组织(organize)、汇总(summarize)以及显示(present)数据的一种方法。

例如：

美国政府公布美国1960年的人口总数为179.3mil.； 1970年为203.3mil.； 1980年为226.5mil.； 1990年为248.7mil.； 2000年为265mil.。

这些信息是描述性统计，如果我们计算年代之间的增长率则是描述统计学。然而，如果我们用这些数据来估计2010年美国人口或者2000年到2010年人口的增长率，则不属于描述统计学的范畴。因为，数据没有被用来总结过去的人口而用来估计未来的人口。

- 推断统计（学）(Inferential Statistics)

【宋廷山版】研究如何根据样本数据去推断总体数量特征，它是在对样本数据进行描述的基础上，对统计总体的未知数量特征做出以概率形式表述的推断。

【Lind版】基于样本基础上对总体属性进行估计的一种方法。

例如：

美国一项最近的统计调查表明，只有46%的高中生可以解决带有分数、小数以及百分比的数学题，77%的高中生可以正确地加总菜单上沙拉、汉堡、薯条和可乐的价格。

因为，这些是基于一些样本数据而对总体数据的一个推断，因此我们将此称为推断统计学。

推断统计中的总体（Population）和样本（Sample）

- 总体：

【宋廷山版】凡是客观存在、在某一共同性质基础上结合起来的许多个别事物的整体，叫做统计总体（简称总体）。

【Lind版】个体或者全部研究对象的集合，或者是指所获得的对全部个体或研究对象的测量结果。

总体可以由人组成，例如：吉林大学所有的学生；也可以由物体组成，例如：某工厂去年生产的所有产品。

- 样本：

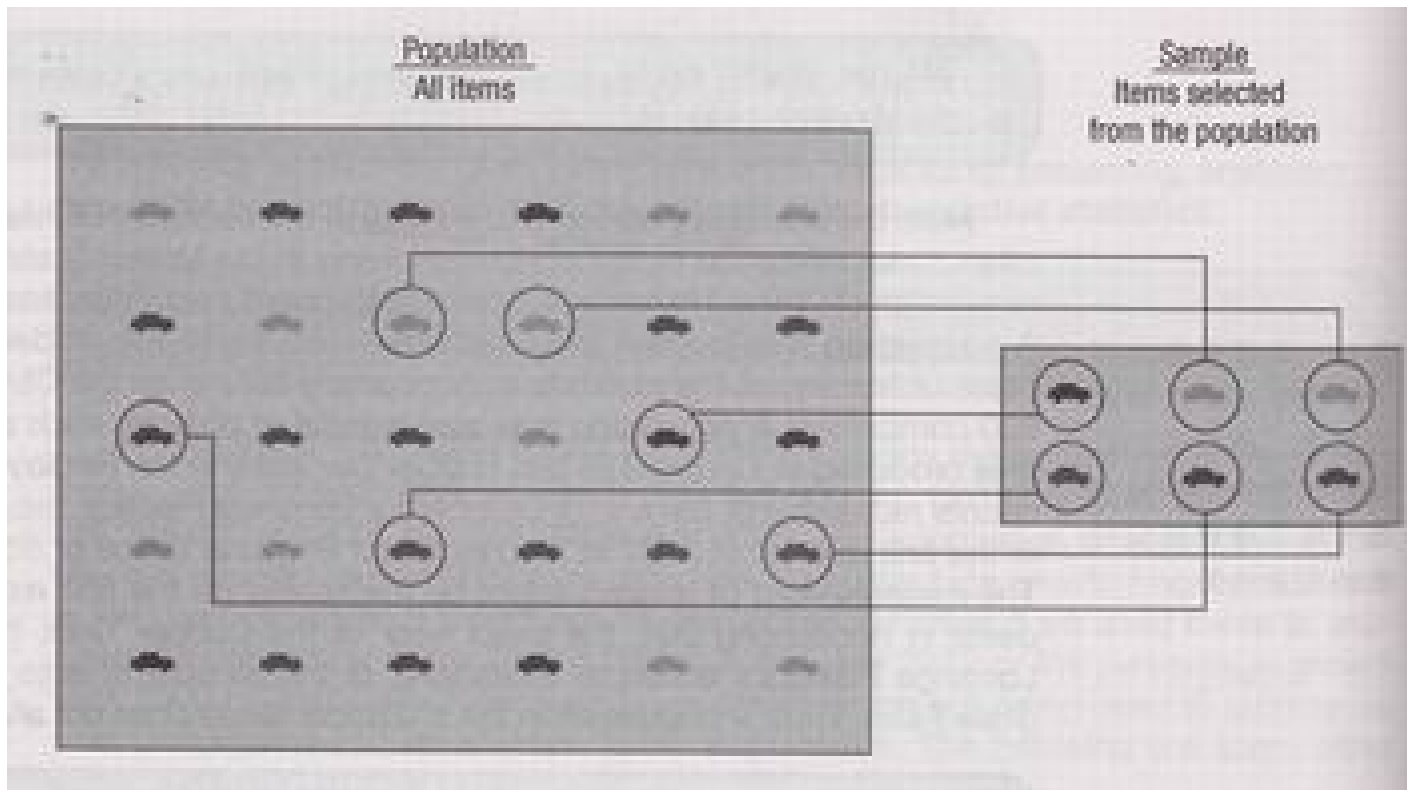
【宋廷山版】按照一定的概率从总体中抽取并作为总体代表的一部分总体单位的集合体。

【Lind版】总体的一部分。

为什么要抽样：成本；可行性等（以后还会详细介绍）。

总体与样本关系：

例子： 我们想要估计SUV汽车的平均每英里耗油量（Mean Miles per Gallon, MPG）



- 理论统计：是研究统计理论和统计方法的。

可分为：参数统计、非参数统计、数理统计、多元统计、贝叶斯统计等。

- 应用统计：是统计理论和方法在各部门各方面的应用。

可分为工业统计、农业统计、人口统计、经济统计、环境统计、医学统计、外贸统计、税务统计、科技统计、教育统计等。

习题1

一广告公司（Brandon and Associates）请1960个消费者尝试一种Boston Market公司新开发的鸡肉晚餐。1176个消费者表明他们愿意购买这种晚餐。

- 1) Brandon and Associate将如何向Boston Market报告消费总体对鸡肉晚餐的接受程度？
- 2) 此例是描述统计还是推断统计？为什么？

变量的类型

- 变量：表示现象某种特征的概念称为变量。例如：性别，产品等级，销售额等等。

- 定性变量 (qualitative variable)

当所研究的变量特征是非数值型的时候 (nonnumeric)，该变量就是定性变量。例如：性别、宗教、出生地以及眼睛的颜色。

- 定量变量 (quantitative variable)

当所研究的变量可以用数值表示时，该变量就是定量变量。例如：账户中的余额、年龄、汽车电池的使用寿命等。

- 离散型变量 (discrete variable)，例如：别墅中的房间数，某一停车场出口1小时内通过的汽车数，统计学来上课的学生数。
- 连续型变量 (continuous variable)，例如：轮胎内部的气压，一船货物的重量，两地之间的飞行时间。

测量的尺度 (Levels of Measurement)



例1: 一袋M&M's有6种颜色巧克力豆。我们设定棕色为1, 黄色为2, 蓝色为3, 橙色为4, 绿色为5, 红色为6。把一袋巧克力豆按照颜色将数值相加, 除以巧克力豆总数, 平均值为3.56。

这一袋M&M's的平均颜色是蓝色或者是橙色?

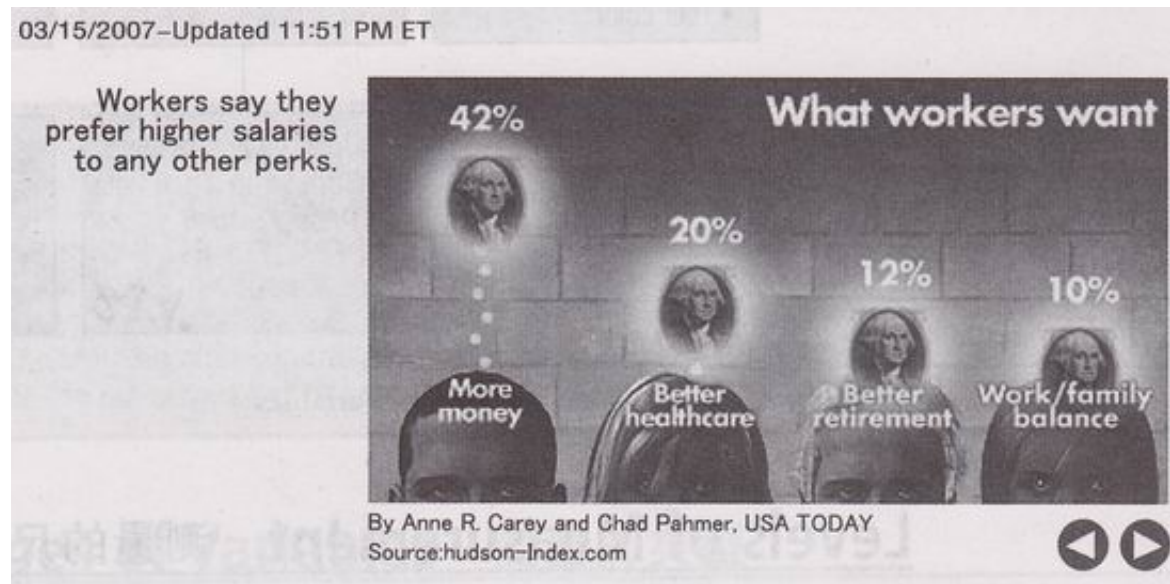
例2: 400米田径赛有八名运动员参加, 比赛结束后我们除了报告名次, 还有平均完成是 $4.5(=(1+2+3+4+5+6+7+8)/8)$ 。这里的4.5有意义吗?

测量的尺度 (Levels of Measurement)

- Nominal-Level Data: 定类尺度、名类尺度、分类尺度、列明尺度...
- Ordinal-Level Data: 定序尺度、顺序尺度、序数尺度、顺位尺度...
- Interval-Level Data: 定距尺度、区间尺度、间隔尺度、等距尺度...
- Ratio-Level Data: 定比尺度、比尺度、比率尺度...

定类尺度 (Nominal-Level Data)

只能用来分类和计数的定性变量的观测值。例如：M&M's巧克力豆颜色，性别等。



定类数据有以下性质：

- 只能将所研究变量分成不同类别或结果；
- 这些类别或结果之间不存在自然顺序。

定序尺度 (Ordinal-Level Data)

比定类尺度更高一级的测量结果，它除了能提供变量的分类信息外，还能提供变量的排序信息。

Rating of a Finance Professor

Rating	Frequency
Superior	6
Good	28
Average	25
Poor	12
Inferior	3

定类数据有以下性质：

- 因此分类数据可以被排序（例如：高、中、低）。
- 然而，只能提供排序信息，而不能提供各分类结果之间的距离信息，不能进行加、减、乘、除等运算。

定距尺度 (Interval-Level Data)

比定序尺度更高一级的测量结果，除了能提供变量的分类信息和排序信息外，还能提供变量分类和排序结果之间的距离信息。例如：温度、衣服的尺码。我们可以对不同的温度值进行分类，还可以进行排序，还可以计算不同温度值之间的温度差。

区间尺度没有绝对0点，0只是标尺上的一个点，0度并不代表温度不存在。

定距数据有以下性质：

- 数据分类被排序；
- 特性之间的差可以由测度之间的差来表示，例如：10度到15度之间相差5度。
- 不能进行乘、除等运算。

定比尺度 (Ratio-Level Data)

最高级别的测量结果。除了拥有定类尺度、定序尺度和定距尺度的所有特点外，基点0是有意义的，并且两个数字之间的比率也是有意义的。例如：工资、产量、股票价格的变换、重量、身高等。

例：4对父子工资组合

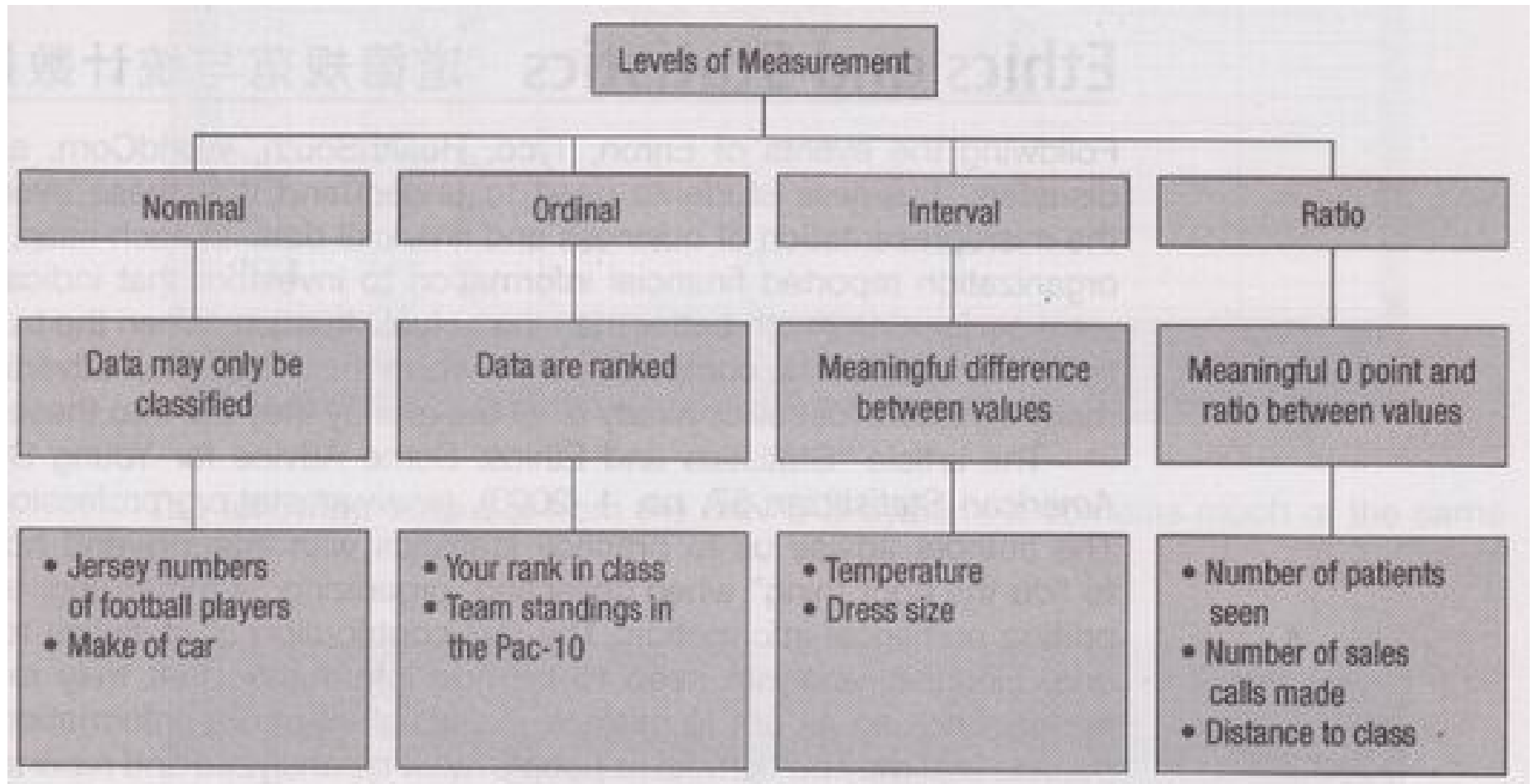
Father-Son Income Combinations

Name	Father	Son
Lahey	\$80,000	\$ 40,000
Nale	90,000	30,000
Rho	60,000	120,000
Steele	75,000	130,000

定比数据有以下性质:

- 数据分类被排序;
- 特性之间的差可以由测度之间的差来表示。
- 0作为绝对起点, 两个值之间的比例是有意义的。

下图总结了不同计量尺度的主要特性:



习题2

下列数据属于哪种尺度数据:

- 下图中列出了收听美国1, 230个谈话频道之一的50个成年人的年龄。

35	29	41	34	44	46	42	42	37	47
30	36	41	39	44	39	43	43	44	40
47	37	41	27	33	33	39	38	43	22
44	39	35	35	41	42	37	42	38	43
35	37	38	43	40	48	42	31	51	34

- 一项统计调查调查了200个名牌轿车的车主，其中100个来自于上海，50个来自于天津，30个来自于南京，20个来自于广州。

道德规范与统计

Enron(安然公司), Tyco, HealthSouth, WorldCom, 破产、会计丑闻。

Article “Statistics and Ethics: Some Advice for Young Statisticians” ,
American Statistician 57, no. 1 (2003).

正直、诚实。

计算机应用

- Excel
- R
- Stata
- Matlab
- Eviews :.....

第二章、描述数据：频数表、频数分布以及图形表示

提纲：

- 构建频数表

-

-

构建频数表

描述一个数据集合称为描述统计(descriptive statistics)。描述统计通过组织数据来展示数据的形状、数据趋向于集中在哪一个值，并且发现极值和非正常值。

组织和归总数据的第一个步骤则为建立频数表 (frequency table)。

频数表：把定性数据分成若干相互排斥的组，各组中显示该组的观测值个数，这样的分类数据表称为频数表。在各组中的观测值个数称为组频数(class frequency)。

一汽车销售商统计了关于每一辆销售汽车的三个变量：销售价格、购买者年龄和汽车类别。汽车类别以国产（记为1）和进口（记为0）。

CarType	Number of Cars
Domestic	50
Foreign	30

相对组频率 (Relative Class Frequencies)

在频数表中，我们可以把组频数转换成相对组频率，以反映各组中的组频数占全部观测值个数的比重，这样的表称为**相对频率表**。相对频率表是用各组的观测值的个数除以总观测值个数得到的，因此它反映组观测值之和与总观测值之和之间的关系。

CarType	Number sold	Relative Frequency
Domestic	50	0.625
Foreign	30	0.375
Total	80	1.000

定性数据的图形表示

- 柱形图 (Bar Chart) : 变量表示在横轴, 频数或频率表示在纵轴; 组频数或组频率与柱形高度成正比。
- 饼图 (Pie Chart) : 饼图的每一个部分代表着定性变量各组频数占总观测值个数的比重或百分比。

Ohio State Lottery Expenses in 2006

Use of Sales	Amount(\$million)	Percent of Share
Prizes	1,310.0	59
Payments to Education	646.3	29
Bonuses/Commissions	139.8	6
Operating Expenses	109.8	5

习题3

SkiLodges.com正在测试他们的网站，感兴趣的问题是网站的设计是否容易操作。他们随机选取了200个internet用户，请他们在网站上完成一个搜索任务，然后对网站操作的容易程度给出评分：不好，一般，很好，非常好。统计数据见下图：

非常好	102
很好	58
一般	30
很差	10

- 1) 网站使用的容易程度是哪种测量尺度？
- 2) 为调查数据画出柱形图。
- 3) 为调查数据画出圆饼图。

