

认识数据

慕课网《专为程序员设计的统计学》
讲师：angelayuan
版权所有 侵权必究

数据和变量

慕课网《专为程序员设计的统计学》
讲师：angelayuan
版权所有 侵权必究

数据

性别：男，女

颜色：赤，橙，黄，绿，青，蓝，紫

教育程度：小学，初中，高中，本科，硕士，博士

评价：好评，中评，差评



定性数据

数据

年份：2019, 2018, 2017, 2016, 2015

温度：10, 15, 20, 25, 30, 35, 40, 45

身高：1.60, 1.65, 1.70, 1.75, 1.80

体重：40, 45, 50, 55, 60, 65, 70



定量数据

数据和变量

- ◆ 数据(data)是由变量(variable)组成的

性别 颜色

教育程度 评价

年份 温度

身高 体重

- ◆ 一个变量(variable)应该包含至少2个不同的取值

定性/分类变量

性别：男(1)，女(2)

颜色：赤 (1)，橙(2)，黄(3)，绿(4)，青(5)，蓝(6)，紫(7)

- ➡ 类别；性质上的差异，而非数量上的差异
- ➡ 只能进行**是否相等**的比较
- ➡ **无顺序(不能比较大小)或等级(不能比较优劣)**

定性/分类变量

教育程度：小学(1)，初中(2)，高中(3)，本科(4)，硕士(5)，博士(6)

评价：好(5)，良好(4)，中(3)，中下(2)，差评(1)

- ➡ 类别；性质上的差异，而非数量上的差异
- ➡ 可以进行是否相等的比较
- ➡ 有顺序(可以比较大小)或等级(可以比较优劣)
- ➡ 好评(5) 不等于 5倍差评(1)

定量/数值变量

年份：2019, 2018, 2017, 2016, 2015

温度：10, 15, 20, 25, 30, 35, 40, 45



数值



可以进行是否相等的比较



有顺序(可以比较大小)



等间距, 可以进行加减运算

定量/数值变量

身高：1.60, 1.65, 1.70, 1.75, 1.80

体重：40, 45, 50, 55, 60, 65, 70



数值



可以进行是否相等的比较；有顺序(可以比较大小)



等间距，可以进行加减运算



可以进行乘除运算

观测，变量，与数据矩阵

性别	教育程度	出生年份	身高
女	本科	1999	165 cm
男	本科	2000	175 cm
男	高中	2003	170 cm
女	研究生	1995	168 cm

一次观测
(observation)

一个变量
(variable)

数据矩阵
(data matrix)

测量尺度

慕课网《专为程序员设计的统计学》
讲师：angelayuan
版权所有 侵权必究

抽象数字系统的属性

◆ 四个属性

每一个数字有其独特的含义 (identity) 性别 颜色

数字具有从小到大的固有顺序 (magnitude) 教育程度 评价

任意位置的单位间距是相同的 (equal intervals) 年份 温度

零(0)表示被测量的属性不存在 (absolute/true zero) 身高 体重

测量 (measurement)

- ◆ 测量是指根据一定的规则，把数字(numbers)分配给观测(observations)
- ◆ 把数字分配给观测的方式，决定了测量尺度(scale of measurement)
- ◆ 每个测量尺度代表抽象数字系统的特定属性或属性集合

名目尺度 (nominal scale)

- ◆ 名目尺度是根据事物的特征对其进行分类的一种尺度

只具备第一个属性，即每一个数字有其独特的含义

名目变量 (nominal variable) 性别 颜色

次序尺度 (ordinal scale)

- ◆ 次序尺度是根据事物的特征对其进行等级排序的一种尺度

具备前两个属性 每一个数字有其独特的含义

数字具有从小到大的固有顺序

次序变量 (ordinal variable)

教育程度 评价

等距尺度 (interval scale)

- ◆ 等距尺度是指数值之间的单位间隔是等距的一种尺度

具备前三个属性 每一个数字有其独特的含义

数字具有从小到大的固有顺序

任意位置的单位间距是相同的

等距尺度 (interval scale)

- ◆ 等距尺度是指数值之间的单位间隔是等距的一种尺度

等距变量 (interval variable) 年份 温度 时间

慕课网《专为程序员设计的统计学》
讲师：angelayuan
版权所有 侵权必究

等比尺度 (ratio scale)

- ◆ 等比尺度是具有等距尺度的所有特性并且零点有意义的尺度

具备所有四个属性

每一个数字有其独特的含义

数字具有从小到大的固有顺序

任意位置的单位间距是相同的

零(0)表示被测量的属性不存在

等比尺度 (ratio scale)

- ◆ 等比尺度是具有等距尺度的所有特性并且零点有意义的尺度

等比变量 (ratio variable) 身高 体重

有真零点, 因此可以进行乘除运算

没有真零点, 则乘除运算结果不唯一(无意义)

温度(摄氏度): 10, 25 vs 温度(华氏度): 50, 77

测量尺度的重要性

- ◆ 测量尺度决定了可以进行哪些逻辑与数学运算和使用哪种统计方法

尺度	举例	逻辑与数学运算	类别
名目	性别、颜色	$=, \neq$	定性/(无序)分类变量
次序	教育程度、评价	$=, \neq, >, <$	定性/(有序)分类变量
等距	温度、年份、时间	$=, \neq, >, <, +, -$	定量/数值变量 离散型
等比	身高、体重、年龄	$=, \neq, >, <, +, -, \times, \div$	定量/数值变量 连续型

