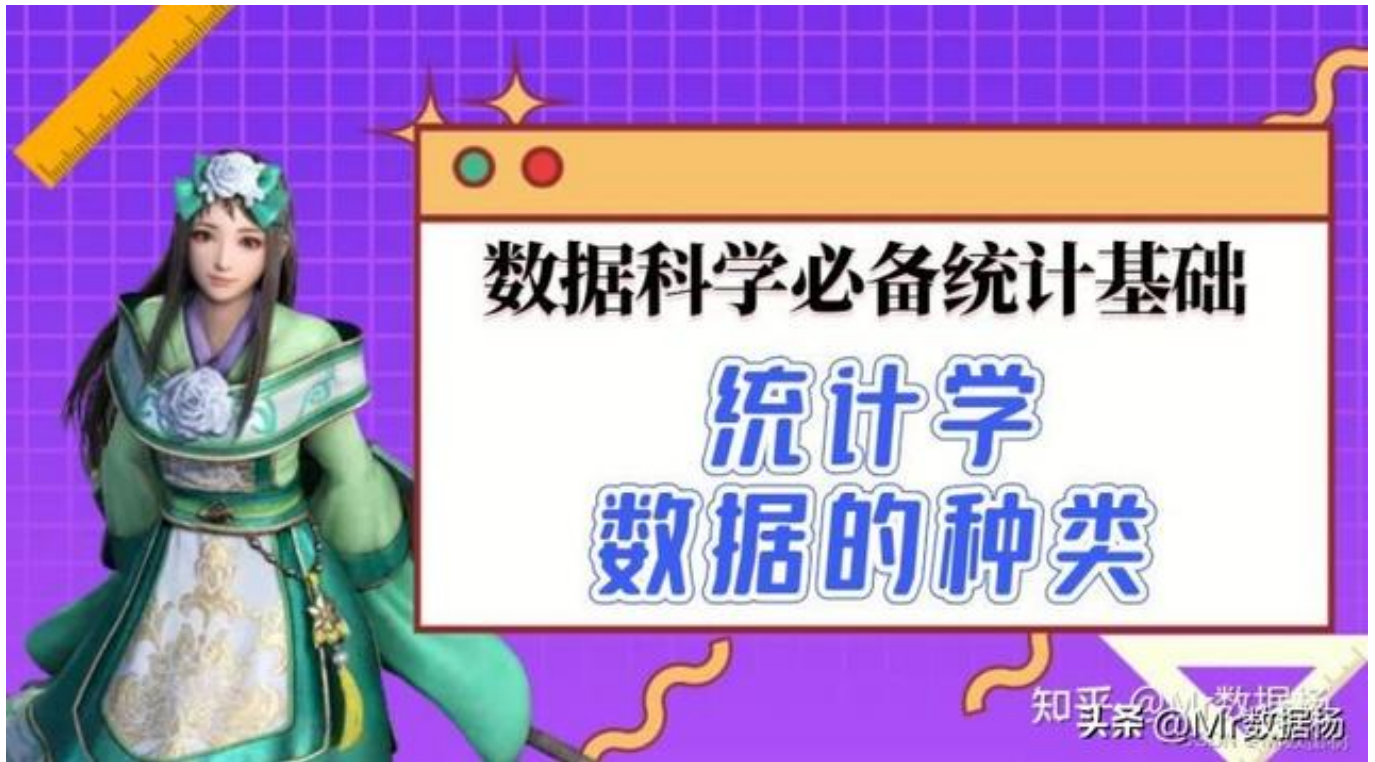


统计中有不同类型的数据，用于被收集、分析、解释和呈现。数据是记录的个别事实信息，用于分析过程的目的。数据分析的两个过程是解释和呈现。统计是数据分析的结果。数据分类和数据处理是重要的过程，因为它涉及大量标签和标签来定义数据、其完整性和机密性。

统计数据中只有两类数据，即定性数据和定量数据。

通过日常的例子可以学习到以下知识：

1. 分类数据和数值型数据的区分和判断。
2. 分类数据和数值型数据的互相转化。



## 分类数据和数值数据

俗称离散型数据（分类数据）、连续型数据（数值数据）。

例如一份问卷调查中

从读者处获得的数据

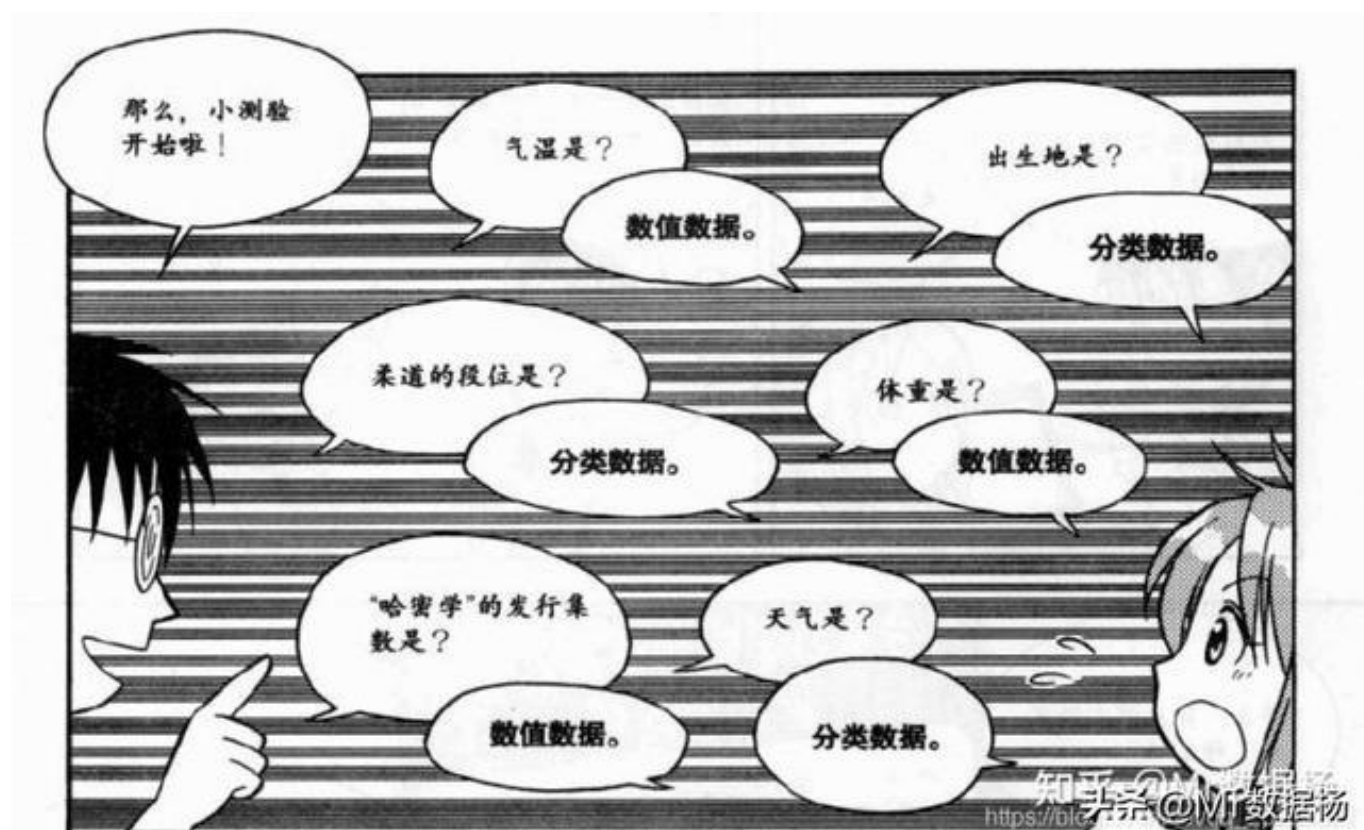
	Q1 读“哈密 学”的感觉	Q2 性别	Q3 年龄（岁）	Q4 平均一个月购 买的杂志数 （册）
琬衣	非常有趣	女	17	2
A	有点有趣	女	17	1
B	一般	男	18	5
C	有点无趣	男	22	7
D	有点有趣	女	25	4
E	非常无趣	男	20	3
F	非常有趣	女	16	1
G	有点有趣	女	17	2
H	一般	男	18	0
I	一般	女	21	

对于表格内的数据看看有什么区别？有些可以用具体数值测量有的不能测量，因此对于数据种类的理解再次进行深入的理解为：

不可测量的数据，也称离散型数据，在实际应用中称分类数据。

可以测量的数据，也称连续型数据，在实际应用中称数值数据。

## 如何进行数据种类的判断



以尺度标准来区分数据的种类，可以理解为：

- 定类尺度：即无序列的类别数据，如技能名称乾坤一掷、迅速果断这种。
- 定序尺度：即有序排列的类别数据，如品质等级SSR、SR+、SR这种。
- 定距尺度：即数值为0时存在意义，如人物经验具体数值或者等级。
- 定比尺度：即数值为0时无意义，如人的体重，0kg是不存在的。

结合之前的数据分类，我们可以看出来分类数据较多存在定类、定序尺度属性。而数值数据较多存在定距、定比尺度属性。

但是在某些情况下存在相互转化的情况。

## 数据类型的相互转化

依旧以之前的问卷调查举例来说，类别数据可以转化成数值数据，数值数据也可以转化成类别数据，但是实际上表达的意义并不会发生多大的变化。

针对下面的内容数据可以进行对应的量化处理。

从读者处获得的数据

	Q1 读“哈密 学”的感觉	Q2 性别	Q3 年龄（岁）	Q4 平均一个月购 买的杂志数 （册）
琉衣	非常有趣	女	17	2
A	有点有趣	女	17	1
B	一般	男	18	5
C	有点无趣	男	22	7
D	有点有趣	女	25	4
E	非常无趣	男	20	3
F	非常有趣	女	16	1
G	有点有趣	女	17	2
H	一般	男	18	0
I	一般	女	21	

### 1.类别数据进行量化

对于文字 非常有趣、有点兴趣、一般 可以使用 1、2、3 来进行表述。对于文字 男、女 可以使用 0、1 来进行表示。虽然表达的形式变化了，但是实际的意义并不发生改变。

### 2.数值数据进行量化

对于连续性数据转化，其实是转化对应成理解的业务类别，即区间分类。如 年龄 类别分为 少年、青少年、青年、中年 等等。对应例如青年定义的年龄是 21-25 岁

，则将表格中的对应的数值转化成对应的类别文字，然后对应的文字类别有根据类别数据进行量化处理的方式进行数值的转化。

需要注意的是，这块处理的业务逻辑是针对数据分析师对于业务的逻辑理解。不同的观点对数据的理解都是不一样的。

## 定性（分类）数据

定性数据，也称为分类数据，描述适合类别的数据。定性数据不是数字。分类信息涉及描述诸如人的性别、家乡等特征的分类变量。分类度量是根据自然语言规范定义的，而不是根据数字定义的。

有时分类数据可以保存数值（定量值），但这些值没有数学意义。分类数据的示例是生日、最喜欢的运动、学校邮政编码。在这里，出生日期和学校邮政编码保留了数量值，但没有给出数字含义。

## 标称数据

名义数据是定性信息的一种，它有助于在不提供数值的情况下标记变量。标称数据也称为标称尺度。它不能被订购和测量。但有时，数据可以是定性和定量的。名义数据的示例是字母、符号、单词、性别等。

使用分组方法检查标称数据。在这种方法中，数据被分组，然后可以计算数据的频率或百分比。这些数据使用饼图直观地表示。

## 序数数据

有序数据/变量是一种遵循自然顺序的数据。标称数据的显着特点是数据值之间的差异是不确定的。这个变量多见于调查、金融、经济学、问卷调查等。

序数数据通常使用条形图表示。这些数据通过许多可视化工具进行调查和解释。可以使用表格来表达信息，其中表格中的每一行都显示不同的类别。

## 定量（数值）数据

定量数据也称为数值数据，它表示数值（即，多少、多久、多少）。数值数据提供有关特定事物数量的信息。数字数据的一些示例是高度、长度、大小、重量等。定

量数据可以根据数据集分为两种不同的类型。数值数据的两种不同分类是离散数据和连续数据。

## 离散数据

离散数据只能采用离散值。离散信息仅包含有限数量的可能值。这些值不能有意义地细分。在这里，事情可以用整数来计算。

## 连续数据

连续数据是可以计算的数据。它具有无限数量的可能值，可以在给定的特定范围内选择。