

数据整理通常包括数据的预处理、分类或分组、汇总等几个方面的内容。

对品质数据主要是做分类整理，对数值型数据主要是做分组整理。

品质数据的整理与显示

分类数据的整理与显示

频数与频数分布

比例

是一个总体中各个部分的数量占总体数量的比重，通常用于反映总体的构成或结构。比例的值是不会大于1的。

百分比

是一个更为标准化的数值，很多相对数都用百分比表示。

比率

是各不同类别的数量的比值。它可以是一个总体中各不同部分的数量对比。

比值可能大于1，比率也可以是同一客观现象在不同时间或空间上的数量之比，例如经济增长率

分类数据的图示

条形图（也称柱状图）

圆形图（也称饼图）

是将各类别的频数逐级累加起来。方法有两种：

累积频数

一是从类别顺序的开始一方向类别顺序的最后一方累加频数，称为向上累积。

二是从类别顺序的最后一方向类别顺序的开始一方累加频数，称为向下累积。

累积频率（百分比）

是将各类别的百分比逐级累加起来，也有向上累积和向下累积两种方法。

品质数据的整理与图示方法，也都适用于数值型数据的整理与显示。但数值型数据还有一些特定的整理和图示方法，并不适用于品质数据，如组距、极差、直方图、折线图。

数据的分组

单变量分组

第一步，确定分组组数。

要求如下：

- (1) 划分的组数，不应太多也不应太少。
- (2) 组数的确定要尽量保证组间资料的差异性与组内资料的同质性
- (3) 采用的分组方法，要能够成分显示客观现象本身存在的状态

第二步，对原始资料进行排序。

第三步，求极差（最大值-最小值）

组距分组

第四步，确定各组组距（组距=极差/组数）

组距组成接近于能被5除尽的一个数。

组距与组数成反比关系，组数越多，组距越小，组数越少，组距越大。组距是每组观察值的最大差，即每组上限值与下限值直接的差。

第五步，确定组限。

第六步，确定各组观察值出现的频数。

第七步，制作频数分布表。

数值型数据的图示

直方图

与条形图的区别

1、条形图是用条形的长度，表示各类别频数的多少，其宽度则是固定的；直方图是用面积表示各组频数的多少，矩形的高度表示每一组的频数或百分比，宽度则表示各组的组距，因此其高度与宽度均有意义。

2、由于分组数据具有连续性，直方图的各矩形通常是连续排列，而条形图则是分开排列。

折线图

也称频数多边形图，把直方图顶部的中点（即组中值）用直线连接起来。折线图的两个终点要与横轴相交。

数值型数据的整理显示

第二十章 统计数据的整理与显示

统计表

统计表的构成

一般由四个主要部分组成，即表头、行标题、列标题和数字资料，必要时可在统计表的下方加上表外附加。

表外附加主要包括资料来源、指标的注释和必要的说明等内容。

应符合科学、实用、简练、美观的要求，注意以下几点：

统计表的设计

第一，要合理安排统计表的结构

第二，表头一般应包括表号、总标题和表中数据的单位等内容，标题内容应满足3W要求。

第三，表中的上下两条横线一般用粗线，中间有细线，统计表左右两边不封口，表中的数据一般是右对齐。对于没有数字的表格单元，一般用“—”表示，一张填好的统计表不应出现空白单元格。

第四，必要时可在表的下方加上注释，特别要注意注明资料来源。