

第二十一章 数据特征的测度

如何对统计数据进行测度

- 一是分布的集中趋势，反映各数据向其中心值靠拢或聚集的程度；
- 二是分布的离散程度，反映各数据远离其中心值的趋势；
- 三是分布的偏态和峰度，反映数据分布的形状。

集中趋势的测度

集中趋势的含义 是指一组数据向某一中心值靠拢的倾向，测度集中趋势也就是寻找数据一般水平的代表值或中心值。

- 位置平均数**
 - 众数** 众数是一组数据中出现频数最多的那个数值，用 M_0 表示。
用众数反映集中趋势，非常直观，不仅适用于品质数据，也适用于数值型数据。众数不受极端值的影响，抗干扰性强。
 - 中位数** 把一组数据按从小到大的顺序进行排列，位置居中的数值叫做中位数，用 Me 表示。
主要用于顺序数据，也适用于数值型数据，但不适用于分类数据。中位数也是一个位置代表值，不受极端值的影响，抗干扰性强。
- 数值平均数**
 - 算术平均数** 算术平均数是全部数据的算术平均，又称均值。算术平均数是集中趋势最主要的测度值，它主要适用于数值型数据，但不适用于品质数据。
加权算术平均数主要用于处理经分组整理的数据。
计算和运用算术平均数的注意事项
 - 第一，算术平均数同时受到两个因素的影响，一个是各组数值的大小，另一个是各组分布频数的多少。
 - 第二，算术平均数易受极端值的影响。
 - 几何平均数** 如何计算： n 个观察值连乘积的 n 次方根就是几何平均数
主要用途
 - (1) 对比率、指数等进行平均，如产品平均合格率
 - (2) 计算平均发展速度

离散程度的测度

- 含义** 是指数据之间的差异程度或频数分布的分散程度。
离散程度越大，集中趋势的测度值对该组数据代表性就越差
离散程度越小，集中趋势的测度值对该组数据代表性就越好
- 极差** 极差是最简单的变异指标。它就是总体或分布中最大的标志值与最小的标志值之差，又称全距，用 R 表示。易受极端值影响。
- 方差** 方差就是标准差的平方。
- 标准差** 标准差是总体所有单位标志值与其平均数离差之平方的平均数的平方根。
- 离散系数** 通常就是标准差来计算的，因此也称标准差系数，它是一组数据的标准差与其相应的算术平均数之比，是测度数据离散程度的相对指标。