

第二章 气象资料及其表示方法

本章教学重难点

本章将教授气象资料、正态分布统计检验、气候资料的审核和订正的相关内容。

1.1 单个要素的气象资料

补充内容：气象资料的基本概念

用统计方法作气象要素的分析和预报是依据大量的气象观测、模式资料来进行的。
从概论论或统计学的观点来看，某个气象要素及其变化可看成为一个变量（或随机变量），它的全体在概率论中称为总体，而把收集到的该要素的资料称为样本。气象统计分析是利用统计学方法对样本进行分析来估计和推测总体的规律性。气象中单个或多个要素可以看作统计学中单个或多个变量。

1.1.1 数据资料

- 描述** 气象资料绝大多数是以数据形式给出的。
- 表示方法** 某气象要素 x 有 n 次观测值，向量表达为： $x = (x_1x_2x_3 \dots x_n)^T$ 或 $x = (x_t)^T, t = 1, 2, 3, \dots, n$
- 时间序列** 气象要素 x 在一段时间内的 n 个观测数据是随时间 t 变化的序列，称为时间序列。
- 几何意义** ① n 维空间中的一个点 ② 一维空间(单坐标)中的 n 个点，散点图

1.1.2 一般统计特征

1.1.2.1 描述平均状况(中心趋势)的统计量

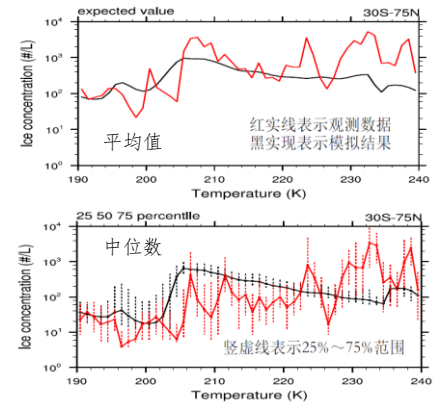
- 平均值** $\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t$ 是要素总体数学期望的一个估计，反映了该要素的**平均(气候)状况**。气候态
- 时间平均** 包括日平均、月平均、年平均、多年平均值。
- 空间平均** 纬向平均、经向平均、区域平均、半球平均、全球平均等。

大数定律
随机事件的大量重复出现中(30 年以上的平均)，往往呈现几乎必然的规律。

- 中位数**
 - 概念**：将变量值按大小顺序排列，处于中间位置的那个数就是中位数。
 - 意义**：表征变量的中心趋势，反应研究对象的一般水平。
 - 计算**： $Position = (n + 1)/2$ 若 n 为偶数，取中间两个数的平均。
 - 优点**：不容易受到异常值的干扰，在样本量较小的情况下，这一优点尤为显著。
- 众数**
 - 概念**：要素变量值中出现次数最多的数，可以有多个，或者没有
 - 意义**：表征研究要素的一般水平，即最容易发生的情况。若变量取值次数较少或取值次数多但无明显集中趋势，计算众数就没有意义。
 - 注意**：众数和中位数又称为**位置平均数**，不受极端变量的影响。

1.1.2.2 描述异常状况(变化幅度)的统计量

- 距平** $x'_t = (x_t - \bar{x})$ 反映数据偏离平均值的状况，也是通常所说的异常。
- 距平序列** 单个要素样本中每个样本资料点的距平值组成的序列称为**距平序列(距平向量)**。
- 中心化** 把资料处理为距平的方法叫做**中心化**。气象上常用距平值代替原样本中的资料值作为研究对象，便于比较且更加直观。平均气温为 36°C或平均气温偏高 2°C。



必要性：① 气象要素在不同周期下平均值不同，为了使他们在同一水平下比较，使用距平值。

② **距平值的平均值为零**，使用方便，直接作为预报值，**呈现直观(偏高/偏低)**

方差 $s_x^2 = \frac{1}{n} \sum_{t=1}^n (x_t - \bar{x})^2$ 变量 x 减去常数，方差/均方差不变。

标准差 $s_x = \sqrt{\frac{1}{n} \sum_{t=1}^n (x_t - \bar{x})^2}$ 标准差与变量值同量纲，一般用标准差表示变量取值变化的大小。

它们描述样本中**资料与平均值差异的平均状况**，反映变量围绕平均值的平均变化程度(**离散程度**)。可以用来评估**波动活跃程度**，方差大的活跃程度大。

绝对变率 $V_a = \frac{1}{n} \sum_{t=1}^n |x_t - \bar{x}|$ 距平绝对值的平均，说明变量值变化的大小

相对变率 $V_r = \frac{V_a}{\bar{x}}$ 绝对变率与平均值的比值，避免平均值不同的影响。

变差系数 $V_p = \frac{s_x}{\bar{x}} = \frac{1}{\bar{x}} \sqrt{\frac{1}{n} \sum_{t=1}^n (x_t - \bar{x})^2}$ 表示**变量的相对变化**，是标准差与平均值的比。

① 绝对变率和标准差的数量级与平均值的量级有关（例如：3 岁的儿童和成年人体重变化）。

② 有些同类型变量，彼此之间平均值差别大（例如不同气候下降水的情况），若要比较它们的变化性用绝对变率和标准差不恰当，应当利用**相对变率或变差系数**。