

总体取值规律的估计---学习指南

一、学习目标：

1. 通过对居民用水案例的分析，学会列频率分布表、画频率分布直方图；
2. 经历数据处理的过程，理解数据分析的思路，学习数据分析的概念和方法；
3. 在利用频率分布表和频率分布直方图描述样本数据的取值规律的过程中，体会用样本估计总体的思想，正确运用统计结果解释实际问题，提升数据分析、数学建模、逻辑推理和数学运算素养。

二、学法指导：

统计学是通过收集数据和分析数据来认识未知现象的一门科学。面对一个统计问题，首先要根据实际需求，通过恰当的方法获取数据。因为实际问题中数据多而且杂乱，往往无法直接从原始数据中发现规律，所以需要根据问题的背景特点，选择合适的统计图表对数据进行整理和直观描述。在此基础上用各种统计方法对数据进行分析，从样本数据中提取需要的信息，推断总体的情况，进而解决相应的实际问题。

总体取值规律描述了总体的所有可能的取值，以及取每个值的个体数占总体数的比例。在很多实际问题中，经常需要根据样本的取值规律来估计总体的取值规律。因此，在实际问题的解决中，要理解数据分析的思路，体会用样本估计总体的思想。

三、学习过程：

前面研究学习了两种抽样方法来收集数据，还知道了一些常见的获取数据的途径，数据收集后，必须从中寻找包含的信息，以使我们能通过样本的规律估计总体的规律，解决相应的实际问题。但由于数据多而杂，所以需要通过一定的方法去处理数据。

问题：我国是世界上严重缺水的国家之一，城市缺水问题较为突出，某市政府为了减少水资源的浪费，计划对居民生活用水费用实施阶梯式水价制度，即确定一户居民月均用水量标准 a ，用水量不超过 a 的部分按平价收费，超出 a 的部分按议价收费.

如果希望确定一个比较合理的标准，以使大部分居民用户的水费支出不受影响，你认为需要做哪些工作？

问题 1：如何收集用水量数据？

(1) 全面调查;

(2) 抽样调查

由于全市居民用户很多,通常采用抽样调查的方式,通过分析样本观测数据,来估计全市居民用户月均用水量的分布情况。

假设通过简单随机抽样,获得了 100 户居民用户的月均用水量数据(单位: t)

9.0	13.6	14.9	5.9	4.0	7.1	6.4	5.4	19.4	2.0
2.2	8.6	13.8	5.4	10.2	4.9	6.8	14.0	2.0	10.5
2.1	5.7	5.1	16.8	6.0	11.1	1.3	11.2	7.7	4.9
2.3	10.0	16.7	12.0	12.4	7.8	5.2	13.6	2.6	22.4
3.6	7.1	8.8	25.6	3.2	18.3	5.1	2.0	3.0	12.0
22.2	10.8	5.5	2.0	24.3	9.9	3.6	5.6	4.4	7.9
5.1	24.5	6.4	7.5	4.7	20.5	5.5	15.7	2.6	5.7
5.5	6.0	16.0	2.4	9.5	3.7	17.0	3.8	4.1	2.3
5.3	7.8	8.1	4.3	13.3	6.8	1.3	7.0	4.9	1.8
7.1	28.0	10.2	13.8	17.9	10.1	5.5	4.6	3.2	21.6

探究一: 数据分析的基本方法

问题 2: 从这些数据中,你能得到哪些信息?如何整理数据以获得我们想要的信息?

(1) 借助于表格:用表格整理数据是通过改变数据的组织方式,为数据的解释提供新方式;

(2)借助于图形:用图表示数据不仅有利于_____,还可以_____。

探究二: 绘制频率分布表与频率分布直方图

思考 1: 上述 100 个数据是极差是多少?由此说明样本数据的变化范围是什么?

思考 2: 如果将上述 100 个数据按组距为 3 进行分组,那么这些数据共分为多少组?

思考 3: 以组距为 3 进行分组,上述 100 个数据共分为 9 组,各组数据的取值范围可以如何设定?

思考 4: 如何统计上述 100 个数据在各组中的频数? 如何计算样本数据在各组中的频率?
你能将这些数据用表格反映出来吗?

思考 5: 如何更加直观的看出数据的分布?

思考 6: 作图时, 横轴代表什么? 纵轴代表什么?

思考 7: 频率分布直方图中小长方形的高表示什么?

小长方形的面积表示什么?

所有小长方形的面积和是多少?

根据频率分布表和频率分布直方图可以推测该市全体居民月均用水量分布的大致情况,
这里体现了一种什么统计思想?

思考 8: 作频率分布直方图的步骤

(1) **求极差:** 即一组数据中_____和_____的差;

(2) **决定组距与组数:** 将数据分组时, 组数应力求合适, 以使数据的分布规律能较清楚地呈现出来。这时应注意: ①一般数据的个数越多, 所分组数_____; ②当样本容量不超过 100 时, 常分成 5~12 组;

(3) **将数据分组:** 按组距将数据分组, 分组时, 各组均为左闭右开区间, 最后一组是闭区间;

(4) _____: 一般分四列: 分组、频数累计、频数、频率, 最后一行是合计。其中频数合计应是样本容量, 频率合计是 1;

(5) **画频率分布直方图:** 画图时, 应以横轴表示分组, 纵轴表示_____。其相应组距上的频率等于该组上的小长方形的面积。

$$\text{即每个小长方形的面积} = \text{组距} \times \frac{\text{频率}}{\text{组距}} = \text{频率}$$

探究三: 分析对比频率分布直方图

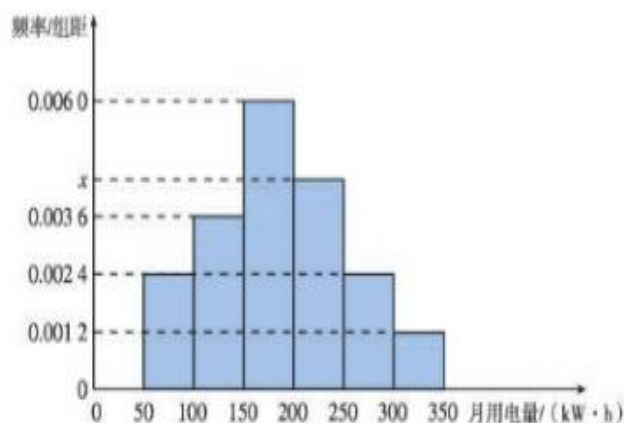
问题 3: 从频率分布直方图中, 你能发现居民用户月均用水量的哪些分布规律?

思考 1: 样本的取值规律一定与总体的取值规律完全相同吗?

思考 2: 哪些环节可能会导致结论出现偏差?

思考 3: 不同的组数对于直方图呈现数据分布规律有什么影响?

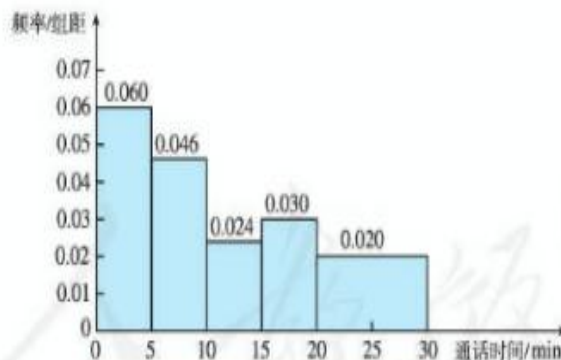
练习 1: 从某小区抽取 100 户居民进行月用电量调查，发现他们的用电量都在 $50 \sim 350 \text{ kW} \cdot \text{h}$ 之间，进行适当分组后（每组为左闭右开的区间），画出频率分布直方图如图所示。



- (1) 直方图中的 x 的值为_____；
- (2) 在被调查的用户中，用电量落在区间 $[100, 250)$ 内的户数为_____。

练习 2: 如图，胡晓统计了他和爸爸 9 月的手机通话明细单，发现他爸爸该月共通话 60 次。胡晓按每次通话时间长短进行分组（每组为左闭右开区间），画出了频率分布直方图。

在区间 $[20, 30)$ 上的小长方形高度低于 $[15, 20)$ 上的小长方形的高度，说明什么？



小结：

1. 频率分布表和频率分布直方图，是对相同数据的两种不同表达方式。用紧凑的表格改变数据的排列方式和构成形式，可展示数据的分布情况；通过作图既可以从数据中提取信息，又可以利用图形传递信息。

2. 画频率分布直方图的一般步骤为：

- (1) 求极差；
- (2) 决定组距与组数；
- (3) 将数据分组；
- (4) 列频率分布表；
- (5) 画频率分布直方图。

3. 频率分布直方图具有随机性和规律性两方面的特征。