# 总体和样本

## 概念

**总体**：研究对象的全体称为总体，构成总体的每个成员称为个体。

**样本**：从总体中抽取的部分个体组合称为样本，样本中所含有的个体称为样品，样本中样品的个数称为样本容量。

**通过总体和样本的：通过总体来推测样本；通过样本来认识总体。就是这两个方向进行处理。**

## 抽样

通过抽样，得到抽样数据，计算抽样数据的概率属性。通过样本的均值和方差来推测总体的均值和方差。

当然对于样本的质量就很重要了，如果样本数据发生了偏倚，那么对总体的预测也就会造成估计错误。

### 如何设计样本

* 确定目标总体

首先需要明确目标总体，才能从总体中抽取样本，对于目标总体是正在研究，并且打算为其采集结果的群体，说选择的目标总体很大程度上取决于研究的目的。目标总体要尽可能准确，这样更容易的得出可能代表总体的样本。

* 确定抽样单位

一旦确定了目标总体，就需要确定抽取哪一类对象。通常，要抽取的对象类型就是在确定目标总体时描述的对象类型。

* 确定抽样空间

在目标总体的范围中所有抽样单位编号，这样单位构成抽样空间，有时不可能得出覆盖整个目标总体的抽样空间。

### 对于样本数据偏倚的原因

样本有时会发生偏倚，使得最终的记过发生误差。对于产生偏倚的原因：

* 抽样空间中条目不齐全，未包含目标总体中的所有对象。
* 抽样单位不正确
* 为样本选择的一个个抽样单位，未出现在实际样本中。
* 调查问卷问题设计不当
* 样本缺乏随机性

### 如何选择样本

* 简单随机抽样

即随机选择抽样单位并形成样本，包括重复抽样和不重复抽样，简单随机抽样具体方式包括抽签或使用随机编号生成器。

* 分层抽样

即将总体划分为几组或者叫做几层，组或层中的单位相似，每一层都尽可能不一样，分好层后，对每一层进行简单随机抽样。

* 整群抽样

将整体划分为几个群,其中每个群尽可能与其他群相似，可通过简单随机抽样抽取几个群,然后利用这些群的一个抽样单位形成样本。

* 系统抽样

即选取一个数值K，没后每个到K的抽样单位就抽样一次。

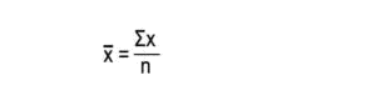
## 总体和样本的估计

### 样本对总体估计

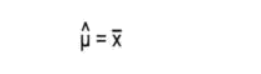
获取样本，计算各项的统计量，通过一个样本点的统计量进行预估总体。点估计*(point estimation)*是依据样本估计总体分布中所含的未知参数或未知参数的函数。通常它们是[总体](https://baike.baidu.com/item/%E6%80%BB%E4%BD%93)的某个特征值，如数学期望、[方差](https://baike.baidu.com/item/%E6%96%B9%E5%B7%AE)和[相关系数](https://baike.baidu.com/item/%E7%9B%B8%E5%85%B3%E7%B3%BB%E6%95%B0)等。

* 估计均值

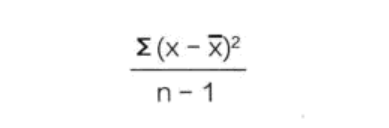
计算样本的均值用



其中X代表各个样本的数值，n为样本的个数，通过计算得到的均值可得到均值的点估计量



计算样本的方差，依据方差求解的方法中，相关的公式，但是由于使用样本方差来估计总体的方法，往往时会偏小的。所以需要进行修正，修正后的公式如下：



### 总体对样本估计