1. 样本和总体

总体数量太大，无法做到完全统计，因此对总体进行随机抽样，对样本进行统计分析代替对总体进行统计分析，例如计算平均值。

μ：总体均值

：样本均值

1. 总体方差

总体方差表征总体离中趋势,计算公式：

1. 样本方差

通过样本方差来估计总体方差，计算公式：

总体方差的无偏估计：

1. 标准差

总体标准差：

样本标准差：

引入标准差的目的之一是为了统一单位。

1. 诸方差公式

=

1. 随机变量介绍

随机变量并非传统意义上的变量，而更像是从随机过程映射到数值的函数。由于过程是随机的，因此这个数值也是随机的。

随机变量分为离散随机变量和连续随机变量。离散随机变量的概率分布图用柱状图表示，连续随机变量的概率分布图用曲线图表示。

1. 概率密度函数

连续随机变量恰好等于某一个数值的概率值为0。

连续型随机变量的概率密度函数（在不至于混淆时可以简称为密度函数）是一个描述这个随机变量的输出值，在某个确定的取值点附近的可能性的函数。而随机变量的取值落在某个区域之内的概率则为概率密度函数在这个区域上的积分。

设f(X)为连续随机变量的概率密度函数，则：

1. 二项分布

二项分布是n个独立的是/非试验中成功的次数的离散概率分布。如果事件发生的[概率](https://baike.baidu.com/item/%E6%A6%82%E7%8E%87)是P，不发生的概率q=1-p，用X表示[事件发生的次数](https://baike.baidu.com/item/%E9%9A%8F%E6%9C%BA%E8%AF%95%E9%AA%8C)，N次[独立重复试验](https://baike.baidu.com/item/%E7%8B%AC%E7%AB%8B%E9%87%8D%E5%A4%8D%E8%AF%95%E9%AA%8C)中发生K次的概率是：

二项分布的期望：

二项分布在趋于无穷的情况下概率密度曲线将趋于正态分布概率密度曲线。

1. 期望值

随机变量的期望值其实就是总体整体的均值，但是当总体数量是无穷的时候，无法通过先求和再除以样本数量的方法计算出期望，因此通过计算可能结果的概率乘以其结果的总和来计算期望。

离散随机变量的期望计算公式：

连续随机变量的期望计算公式：

1. 泊松分布

泊松分布是无限个二项分布的叠加。泊松分布的概率公式：

其中λ为随机变量X的期望。

1. 大数定律

在数学与统计学中，大数定律又称大数法则、大数律，是描述相当多次数重复实验的结果的定律。根据这个定律知道，样本数量越多，则其平均就越趋近期望值。

《魔鬼数学》对大数定律的解释：硬币没有记忆，因此，再次抛出硬币时，正面朝上的概率仍然是 50%。幸运女神就会青睐反面。 实际的情况是，随着抛硬币的次数越来越多，前 10次结果的影响力就会越来越小。简而言之，大数定律发挥作用，是靠大数对小数的稀释作用。

1. 正态分布

正态分布概率密度函数：

标准正态分布，。

正态分布图像的性质（对称性，曲线下面积）。

正态分布性质（68-95-99.7）。

现实生活中许多数据都能近似表示为正态分布。

z分数表示随机变量距离均值有多少个标准差，z分数计算公式：

正偏态分布、负偏态分布。

正峰态分布、负峰态分布。