1. 扑克牌54张，平均分成2份，求这2份都有2张A的概率。
2. 王炸概率 
3. 男生点击率增加，女生点击率增加，总体为何减少?
4. **参数估计**:

点估计: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(矩估计\极大\最小)

***矩法***：不需要知道分布

***极大似然估计***：总体分布已知，构造似然->对数->求导

区间估计:有百分之多少的把握保证某值在某个范围内

1. 参数估计和假设检验是统计推断的两个组成部分，它们都是利用样本对总体进行某种推断，但推断的角度不同。

参数估计讨论的是用**样本估计总体参数的方法，总体参数μ在估计前是未知的**。

假设检验中，则是**先对μ的值提出一个假设，然后利用样本信息去检验这个假设是否成立**。

*# 检验硬币(抛了100次，90次正)*

*假设：硬币是公平的*

*检验：认为假设是成立的，然后扔十次，看结果与假设是否相符（应该符合），由中心极限定理，正态分布是二次分布的极限，即服从可计算相应概率。*

*显著水平： 认为假设不正确*

*计算正常情况下 ，发生概率太小了，拒绝他*

1. **置信区间**是我们所计算出的变量存在的范围，**置信水平**是我们对于这个数值存在于计算出的范围的可信程度。

举例来讲，有95%的把握，真正的数值在我们所计算的范围里。在这里，95%是置信水平，而计算出的范围，就是置信区间。如果置信度为95%， 则抽取100个样本来估计总体的均值，由100个样本所构造的100个区间中，约有95个区间包含总体均值。

1. 协方差：

协方差表示的是两个变量的总体的误差，这与只表示一个变量误差的方差不同。 如果两个变量的变化趋势一致，也就是说如果其中一个大于自身的期望值，另外一个也大于自身的期望值，那么两个变量之间的协方差就是正值。 如果两个变量的变化趋势相反，即其中一个大于自身的期望值，另外一个却小于自身的期望值，那么两个变量之间的协方差就是负值。

1. 相关系数：

研究变量之间线性相关程度的量，取值范围是[-1,1]。相关系数也可以看成协方差：一种剔除了两个变量量纲影响、标准化后的特殊协方差。

9. 中心极限定理定义（多组样本）：

（1）任何一个样本的平均值将会约等于其所在总体的平均值。

（2）不管总体是什么分布，任意一个总体的样本平均值都会围绕在总体的平均值周围，并且呈正态分布。

中心极限定理作用：

（1）在没有办法得到总体全部数据的情况下，我们可以用样本来估计总体。

（2）根据总体的平均值和标准差，判断某个样本是否属于总体。

10. **大数定律讲的是样本均值收敛到总体均值**（一组样本）

**区别**：大数定律是说，n只要越来越大，我把这n个独立同分布的数加起来去除以n得到的这个样本均值（也是一个随机变量）会依概率收敛到真值u，但是样本均值的分布是怎样的我们不知道。

中心极限定理是说，n只要越来越大，这n个数的样本均值会趋近于正态分布，并且这个正态分布以u为均值，为方差。

综上所述，这两个定律都是在说样本均值性质。随着n增大，大数定律说样本均值几乎必然等于均值。中心极限定律说，均值越来越趋近于正态分布。并且这个正态分布的方差越来越小。

**检验**--利用观测数据判断总体是否服从正态分布的检验称为正态性检验。

**检验**--主要是用来检验频数问题的，即检验各类实际观察的频数是否显著不同于建设的期望频数。

**检验**--判断两类样本在某一变量上的均值差异是否显著（两块大麦田）

**检验**--检验几个分布的方差是否相同（学生的教学方法）

t检验最常见的四个用途：

1.**单样本均值检验**（One-sample t-test）

用于检验 总体方差未知、正态数据或近似正态的 单样本的均值 是否与 已知的总体均值相等

2.**两独立样本均值检验**（Independent two-sample t-test）

用于检验 两对独立的 正态数据或近似正态的 样本的均值 是否相等，这里可根据总体方差是否相等分类讨论

3.**配对样本均值检验**（Dependent t-test for paired samples）

用于检验 一对配对样本的均值的差 是否等于某一个值

4.**回归系数的显著性检验**（t-test for regression coefficient significance）

用于检验 回归模型的解释变量对被解释变量是否有显著影响

