1. 描述集中趋势的几种方式

- 均值：(x1 + x2 + x3 + x4 + ... + xn) / n
- 众数： 数据中出现次数最多的数，如 2，3，4，4，4，5，众数是4
- 中位数：如果数据集个数是偶数，则取中间两个数的平均值，1，3，4，5，6，7，中位数是(4 + 5) / 2 = 4.5；如果是奇数，则取中间数，如1，3，4，6，7，众数是4。要注意的是，取中位数时，数据集是要有序的。
- 极差：最大值 - 最小值
- 中程数：(最大数 + 最小值) / 2
  思考：均值会受到异常点的影响；极差表征数据的范围；以上几种描述感觉都没有办法体现数据本身的离散程度。

2.几种图形

- 象形图：用图形表示单位数量，如统计捐血，一滴血代表8个人，6滴血代表48个人；
- 棒图：归类、同比和环比
- 曲线：反映数据的趋势
- 饼图：反映各部分的占比
- 茎叶图：反映数据在不同层级的分布
- 箱线图：通过四分卫对数据划分，可以表征数据的最大、最小、中位数。左边是最小值，右边是最大值，长方形的那根线是中位数。感觉可以一定程度上反馈数据的离散程度。

3.样本和总体

- 总体：所有潜在的情况或者取值，可能是无穷多的，一般总体是很难统计或者是不可统计
- 样本：从总体中取出来的某一个子集
- 总体方差：((x1 - xavg)^2 + (x2-xavg)^2 + ... + (xN - xavg)^2) / N
- 样本方差：((x1 - xavg)^2 + (x2-xavg)^2 + ... + (xn - xavg)^2) / n
  由于样本可能不能完成反映总体的情况，样本方差可能比实际的总计方差要小(证明还没搞懂)，所以样本的无偏方差 = ((x1 - xavg)^2 + (x2-xavg)^2 + ... + (xn - xavg)^2) / (n - 1)
  个人觉得方差可以反映数据的离散程度，方差越大说明数据在均值两侧左右波动。
- 标准差：方差的开根方，单位和均值等一致

4.随机变量

- 随机变量：表示随机试验各种结果的函数。随机变量分了离散随机变量和连续随机变量，对应离散随机变量函数和概率密度函数。
  对于概率密度函数，某一点的取值为0，某一范围的概率取值等于该范围内的函数积分(面积)
- 二项分布：n次抛硬币得到k次结果的概率 = n! / (k! (n-k)!) * p ^ k * (1-p)^(n-k)
  二项分布的期望= np，证明的话还需要多练一下
  二项分布的方差= np(1-p)

- 泊松分布：结合二项分布和极限定理推导而来，用处是可以使用期望(均值)来计算概率。
  推导过程还记不住，需要复习。
- 大数定理：如果样本的数量足够大，那么期望和方差和总体就差别不大了，一定程度可以代表总体。感觉和最大似然估计是反过来的，最大似然估计是使概率最大时的参数解。
- 正态分布：勉强记得住公式，均值控制钟形曲线的位置，方差控制形状。方差越大，曲线形状越扁平。这块看的有点晕。需要多复习下。