1. 统计学基本知识

1．1 集中趋势

a. 均值

总体均值$\mu=\frac{1}{N}\sum_{i=1}^{N}X_i$

样本均值$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$

b. 众数

c. 中位数

1.2 离中趋势

a. 方差

总体方差$\sigma^2=\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$

样本方差${S_{n-1}}^2=\frac{1}{n-1}\sum_{i=1}^{N}(x_i-\bar{X})^2$

b. 标准差

1.3 随机变量

随机变量不是传统意义上的变量，更像是从随机过程映射到数值的函数。它包括离散型变量（个数有限）和连续型变量（个数无限）

2. 二项及泊松分布

2.1 二项分布

含义：重复 n 次独立的伯努利试验，其概率公式为$P=C_n^k p^k(1-p)^{n-k}$, p 为成功概率

其期望$E(X)=np$, 方差$D(X)=np(1-p)$

2.2 泊松分布（来自二项分布）

其概率分布函数$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$,λ是单位时间（或单位面积）内随机事件的平均发生率

其期望和方差均为λ

泊松分布与二项分布联系

当二项分布 n 很大 p 很小时，泊松分布可作为二项分布近似，其中$\lambda=np$

3. 大数定律

它是随机变量的平均值向数学期望收敛的定律，是一种描述当实验次数很大时所呈现的概率性质的定律。

4. 正态分布

其概率密度函数$P(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$,标准 Z 分数为$\frac{x-\mu}{\sigma}$，累计分布函数$CDF(x)=\int_{-\infty}^{x}P(x)\,dx$

正态分布图像特征：以 x=μ为对称轴，标准差越小，图像越窄

正态分布数字特征：$\mu\pm\sigma$~68%　　　　$\mu\pm2\sigma$~95%

$\mu\pm3\sigma$~99.7%