



统计知识主线

统计二级知识线索

三级知识线索

● 一级知识点
○ 二级知识点
● 三级知识点

统计

随机抽样

简单随机抽样

特点

- ①被抽取样本的总体中个体数是有限的;
- ②从总体中逐个抽取;
- ③属于不放回抽样;
- ④总体中每个个体被抽到的可能性相同.

方法

抽签法(抓阄法);
随机数法.

系统抽样

步骤

- ①将总体的 N 个个体随机编号;
- ②确定分段间隔 k , 对编号进行分段. 当 $\frac{N}{n}$ (n 是样本容量) 是整数时, 取 $k = \frac{N}{n}$; 当 $\frac{N}{n}$ 不是整数时, 可利用简单随机抽样先从总体中剔除几个个体, 使得总体中剩余的个体数 N' 能被样本容量 n 整除, 此时 $k = \frac{N'}{n}$;
- ③在第一段用简单随机抽样确定起始个体编号 l ($1 \leq l \leq k$);
- ④按照一定的规则抽取样本.

分层抽样

步骤

- ①分层;
- ②按比例确定每层抽取个体的个数;
- ③各层抽样(方法可以不同);
- ④汇合成样本.

三种抽样方法的比较

类别	简单随机抽样	系统抽样	分层抽样
各自特点	从总体中逐个抽取	将总体均匀分成几部分, 按事先确定的规则分别在各部分中抽取	将总体分成几层, 按比例分层进行抽取
相互联系	最基本的抽样方法	在起始部分抽样时采用简单随机抽样	各层抽样时采用简单随机抽样或系统抽样
适用范围	总体中的个体数较少	总体中的个体数较多	总体由差异明显的几部分组成
共同点	均为不放回抽样, 且抽样过程中每个个体被抽取的机会相等		

用样本估计总体

用样本的频率分布估计总体分布

茎叶图

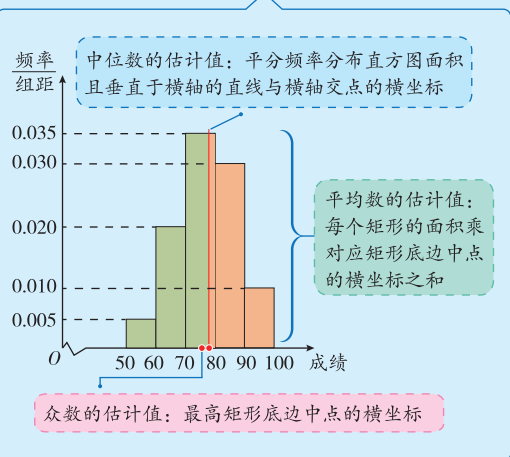
频率分布直方图的相关公式

- ① 小矩形的面积 = $\frac{\text{频率}}{\text{组距}} \times \text{组距} = \text{频率}$
- ② $\frac{\text{频数}}{\text{样本容量}} = \text{频率}$, 此关系式的变形为: 样本容量 $\times$ 频率 = 频数
- ③ 频率分布直方图中各个小矩形的面积之和为 1, 即各组的频率之和为 1
- ④ 频率分布直方图中各小矩形高之比也就是各组的频率之比

用样本的数字特征估计总体的数字特征

变量间的相关关系及回归分析

频率分布直方图中众数、中位数、平均数的估计值的几何意义

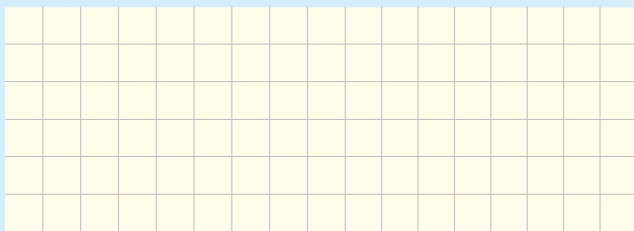


众数、中位数和平均数

平均数、中位数和众数刻画了一组数据的集中趋势, 其中平均数受极端值影响大, 中位数对极端值不敏感.

标准差与方差

一组数据的标准差(方差)越大, 说明这组数据的离散程度越大;
一组数据的标准差(方差)越小, 说明这组数据的离散程度越小.



随机事件的概率

古典概型

特征

试验中所有可能出现的基本事件只有有限个；
每个基本事件出现的可能性相等。

概率公式

$$P(A) = \frac{\text{事件 } A \text{ 包含的基本事件的个数 } (m)}{\text{基本事件的总数 } (n)}$$

利用公式法求解古典概型概率的步骤

- ① 算出所有基本事件的个数 n ;
- ② 求出事件 A 包含的基本事件个数 m ;
- ③ 代入公式 $P(A) = \frac{m}{n}$ 求解。

事件的关系及符号表示

- ① 包含关系: $B \subseteq A$ (或 $A \supseteq B$);
- ② 相等关系: $A = B$;
- ③ 并事件 (和事件): $A \cup B$ (或 $A + B$);
- ④ 交事件 (积事件): $A \cap B$ (或 AB);
- ⑤ 互斥事件: $A \cap B = \emptyset$;
- ⑥ 对立事件: $A \cap B = \emptyset$ 且 $P(A \cup B) = P(A) + P(B) = 1$.

概率的几个基本性质

- ① $0 \leq P(A) \leq 1$;
- ② 必然事件发生的概率为 1, 不可能事件发生的概率为 0;
- ③ 事件 A, B 互斥, 则 $P(A \cup B) = P(A) + P(B)$;
事件 A, B 对立, 则 $A \cup B$ 为必然事件, $P(A \cup B) = 1, P(A) = 1 - P(B)$.

复杂事件的概率求法

直接法: 将所求的事件表示成一些彼此互斥的事件的和, 运用互斥事件的概率加法公式计算。

间接法: 先求复杂事件的对立事件的概率, 再用对立事件的概率公式 $P(A) = 1 - P(\bar{A})$ 计算, 即运用逆向思维 (正难则反), 特别是“至多”“至少”型题目, 用间接法求解一般会比较简便。

几何概型

定义

如果某个事件发生的概率只与构成该事件区域的长度 (面积或体积) 成比例, 则称这样的概率模型为几何概型。

特征

- ① 无限性: 试验结果的个数是无限的;
- ② 等可能性: 每一个结果发生的可能性相等。

概率公式

$$P(A) = \frac{\text{构成事件 } A \text{ 的区域长度 (面积或体积)}}{\text{试验的全部结果所构成的区域长度 (面积或体积)}}$$

利用公式法求解几何概型概率的步骤

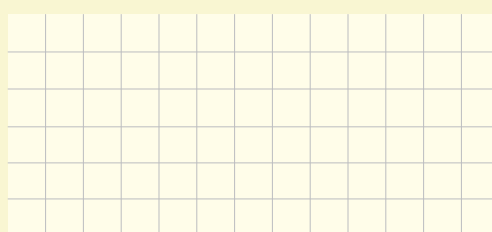
第一步 明确取点的区域 Ω ;

第二步 确定所求事件对应的点的区域 A ;

第三步 计算区域 Ω 和区域 A 的几何度量 μ_{Ω} 和 μ_A ;

第四步 代入公式 $P(A) = \frac{\mu_A}{\mu_{\Omega}}$ 求解。

随机变量及其分布列



二项分布

条件概率

定义

一般地, 设 A, B 为两个事件, 称 $P(B|A) = \frac{P(AB)}{P(A)}$ 为在事件 A 发生的条件下, 事件 B 发生的概率.

性质

- $0 \leq P(B|A) \leq 1$;
- 如果 B 和 C 是两个互斥事件, 则 $P(B \cup C|A) = P(B|A) + P(C|A)$.

离散型随机变量及其分布列

离散型随机变量的分布列的性质

- ① $p_i \geq 0, i = 1, 2, 3, \dots, n$;
- ② $p_1 + p_2 + \dots + p_i + \dots + p_n = 1$;
- ③ $P(x_i \leq X \leq x_j) = p_i + p_{i+1} + \dots + p_j$ ($i < j \leq n$ 且 $i, j \in \mathbf{N}^*, x_i < x_j$).

求离散型随机变量 X 的分布列和数学期望的步骤

确定 X 的所有可能取值 $x_i (i = 1, 2, 3, \dots, n)$, 并明确每个取值代表的意义;

求出相应的概率 $P(X = x_i) = p_i (i = 1, 2, 3, \dots, n)$;

列出分布列, 如下表所示, 根据公式 $E(X) = x_1 p_1 + x_2 p_2 + \dots + x_i p_i + \dots + x_n p_n$ 求数学期望.

X	x_1	x_2	$\dots$	x_i	$\dots$	x_n
P	p_1	p_2	$\dots$	p_i	$\dots$	p_n

注意

列出分布列后, 可根据分布列的性质 2 验证所求是否正确, 或者利用该性质计算随机变量的某一取值的概率.

离散型随机变量的方差

$$D(X) = \sum_{i=1}^n (x_i - E(X))^2 p_i, i = 1, 2, \dots, n.$$

正态分布

相互独立事件

定义

设 A, B 为两个事件, 若 $P(AB) = P(A)P(B)$, 则称事件 A 与事件 B 相互独立.

性质

如果事件 A 与 B 相互独立, 那么 A 与 $\bar{B}$, $\bar{A}$ 与 B , $\bar{A}$ 与 $\bar{B}$ 也都相互独立.

求相互独立事件同时发生的概率的方法

- ① 直接法: 利用相互独立事件的概率乘法公式直接求解;
- ② 间接法: 从对立事件入手计算.

独立重复试验

特征

- 每次试验的条件完全相同, 有关事件的概率保持不变;
- 每次试验只有两种结果: 事件发生或者不发生;
- 每次试验的结果互不影响, 即每次试验相互独立.

二项分布

一般地, 在 n 次独立重复试验中, 用 X 表示事件 A 发生的次数, 设每次试验中事件 A 发生的概率为 p , 则 $P(X = k) = C_n^k p^k (1-p)^{n-k} (k = 0, 1, 2, 3, \dots, n)$, 此时称随机变量 X 服从二项分布, 记作 $X \sim B(n, p)$, 并称 p 为成功概率.

超几何分布

一般地, 在含有 M 件次品的 N 件产品中, 任取 n 件, 其中恰有 X 件次品,

则 $P(X = k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}, k = 0, 1, 2, \dots, m$, 即

X	0	1	$\dots$	m
P	$\frac{C_M^0 C_{N-M}^{n-0}}{C_N^n}$	$\frac{C_M^1 C_{N-M}^{n-1}}{C_N^n}$	$\dots$	$\frac{C_M^m C_{N-M}^{n-m}}{C_N^n}$

其中 $m = \min\{M, n\}$, 且 $n \leq N, M \leq N, n, M, N \in \mathbf{N}^*$. 如果随机变量 X 的分布列具有上表的形式, 则称随机变量 X 服从超几何分布.

注意: 超几何分布的应用有明显的分类, 如“成品”和“次品”, “及格”与“不及格”等.

求概率问题的步骤

第一步

确定事件的性质: 古典概型、互斥事件、独立事件、条件概率等;

第二步

综合运用相应公式求解.