



# 第四部分

## 统计



## 考点：统计与数据科学

### 统计概念

概念	概括来讲，统计学是关于收集、整理、分析数据和从数据中得出结论的科学
描述统计	是研究数据收集、整理和描述的统计学方法
	①如何取得所需要的数据
	②如何用图表或数学方法对数据进行整理和展示 ③如何描述数据的一般性特征
推断统计	是研究如何利用样本数据来推断总体特征的统计学方法
	①参数估计—利用样本信息推断总体特征 ②假设检验—利用样本信息判断对总体的假设是否成立



## 考点：统计与数据科学

普查特点	抽样调查特点
<p>(1) 普查通常是一次性的或者周期性的</p> <p>(2) 普查一般需要规定统一的标准调查时间，以避免调查数据的重复或遗漏，保证普查结果的准确性</p>	<p>(1) 经济性：调查的工作量小，可节省大量的人力、物力、财力和时间。是抽样调查的一个最显著优点</p> <p>(2) 时效性强：可以迅速、及时地获得所需要的信息，提高数据的时效性</p>



## 考点：统计与数据科学

普查特点	抽样调查特点
<p>(3) 普查的数据一般比较准确，规范化程度也较高，可以为抽样调查或其他调查提供基本依据</p> <p>(4) 普查的使用范围比较窄，只能调查一些最基本及特定的现象</p>	<p>(3) 适应面广：适用于对各个领域、各种问题的调查。如产品质量检验、农产品试验、医药的临床实验等</p> <p>(4) 准确性高：由于工作量小，可使各环节的工作做得更仔细，误差往往很小</p>



## 考点：数据科学与大数据

### 一、大数据4V 特性

(1) 数据量大

(2) 数据多样性：结构化数据（二维表）、非结构化数据、半结构化数据（员工简历）

(3) 价值密度低：大数据价值密度的高低与数据总量的大小成反比。

(4) 数据的产生和处理速度快：“1 秒定律”

### 二、数据挖掘

监督学习（对目标需求的概念）、无监督学习：（没有明确的标识变量）；半监督学习



## 考点：描述统计

	指标 (测度值)	是否受极 端值影响	数值型数据 (定量数据)	分类 数据	顺序 数据
测度数据的 集中趋势	均值	受影响	适用	不适用	不适用
	中位数	不受影响	适用	不适用	适用
	众数	不受影响	不适用	适用	适用
测度数据的 离散程度	方差	受影响	适用	不适用	不适用
	标准差				
	离散系数				



## 考点：抽样调查

概率抽样（随机抽样）				
简单随机抽样	分层抽样	系统抽样	整群抽样	多阶段抽样
非概率抽样				
判断抽样	方便抽样	自愿样本	配额抽样	—
人为确定样本单元	“拦截式”调查	网上调查	无严格限制 方便抽样	—



## 考点：决定系数

决定系数	含义	决定系数 $R^2$ ，也称为拟合优度或判定系数，可以测度回归模型对样本数据的拟合程度
	取值	决定系数的取值范围在0到1之间 决定系数越接近于1，回归模型的拟合效果越好 $R^2=1$ ，说明回归直线可以解释因变量的所有变化 $R^2=0$ ，说明回归直线无法解释因变量的变化，因变量的变化与自变量无关