



变量类型和所适用的集中趋势测度值				
变量类型	定类变量	定序变量	定距变量	定比变量
适用的测度值	众数	中位数	均值	均值
		众数	众数	众数
			中位数	中位数

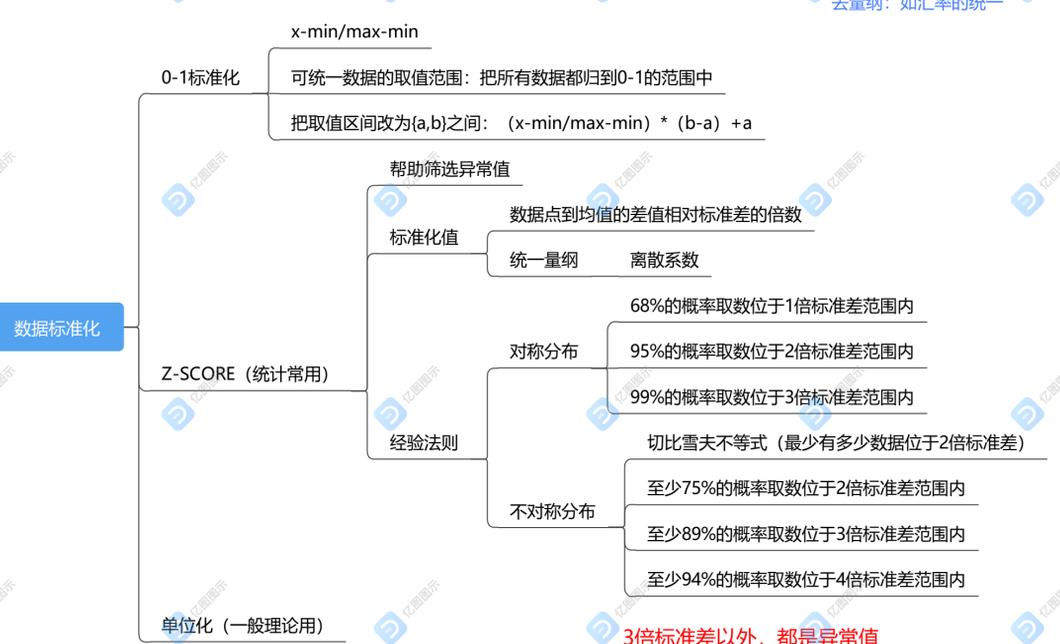
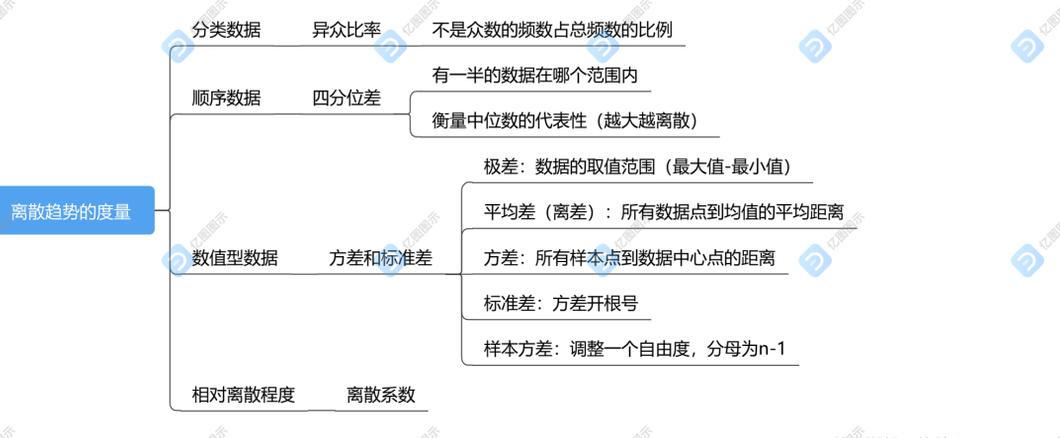
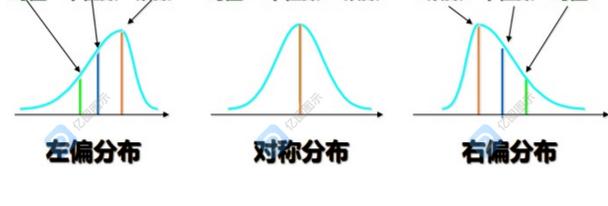
样本中必须提前处理的数据:

- 重复值
- 缺失值 (先处理异常值, 用均值填补)
- 异常值 (盖帽法, 处理成80%以上数据的均值)
- 编码

均值是到所有样本点平方和最小的点  
x的偏导数指平面上的点平行于x平面的斜率  
y的偏导数指平面上的点平行于y平面的斜率

决策树: 生成二分类树进行判断 分类型变量  
欠拟合: 决策树生长不充分, 分类效果差 (回归易出现)  
过拟合: 因错误值导致决策树分类过多 (神经网络易出现)

左偏分布: 尾巴在左边 偏度系数 < 0  
右偏分布: 尾巴在右边 偏度系数 > 0



EXCEL中函数后缀为".P"算的总体, ".S"算的样本  
代数运算, 矩阵运算

- 矩阵乘法
  - 1.要先选中作为结果的区域
  - 2.MMULT函数: 直接选取要计算的两个区域
  - 3.矩阵运算的执行: **ctrl+shift+enter**
- 矩阵的转置: A'  
 1.选定结果区域;  
 2.TRANSPOSE函数: 直接选取要计算的两个区域
- 矩阵的逆, 逆矩阵: A^-1  
 逆矩阵可让空间变回原始状态, 但并非一定存在 (空间降维则无法复原)  
 单位矩阵: 斜对角线值为1, 其他为0  
 1.选定结果区域;  
 2.MINVERSE函数: 直接选取要计算的两个区域
- 矩阵的行列式: |A| |A| < 0 空间发生翻转 |A| = 0 空间被降维  
 行列式的值用于度量空间的面积、体积等经过线性变换后增加了多少倍  
 1.选取单元格  
 2.MDETERM函数, 直接回车

线性相关: 两个点无论怎么算都只能在一条线上, 生成一维空间  
 线性无关: 由两个点的计算可生成一个二维空间  
 线性变换: 原点不动, 使基向量变化从而使空间发生变换  
 (矩阵描述数据空间的线性变换, 且唯一)

非方阵:  
 3\*2矩阵 可将二维升到三维  
 2\*3矩阵 可将三维降到二维

内积:  
 同侧投影的积为正, 异侧投影的积为负  
 用于判断两个向量是否垂直 内积=0