



专注于商业智能BI和大数据的垂直社区平台

一元线性回归模型的概念

Allen

www.hellobi.com

课程目录

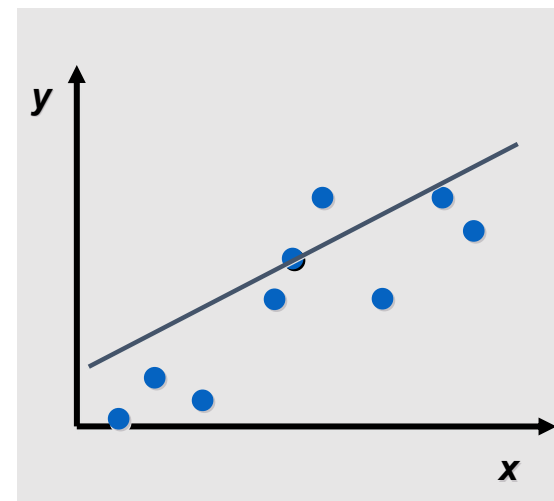
- 相关分析与回归分析
- 回归分析的模型及概念
- 一元线性回归模型及概念
- 回归方程、回归值的概念
- 小结

相关分析与回归分析

- 1. **相关分析**：主要研究两个变量之间相关的方向和相关的密切程度，但是相关分析不能指出两变量相互关系的具体形式，也无法从一个变量的变化来推测另一个变量的变化关系

例：产品销售收入与广告费用之间的关系

- 变量间关系不能用函数关系精确表达
- 一个变量的取值不能由另一个变量唯一确定
- 各观测点分布在直线周围

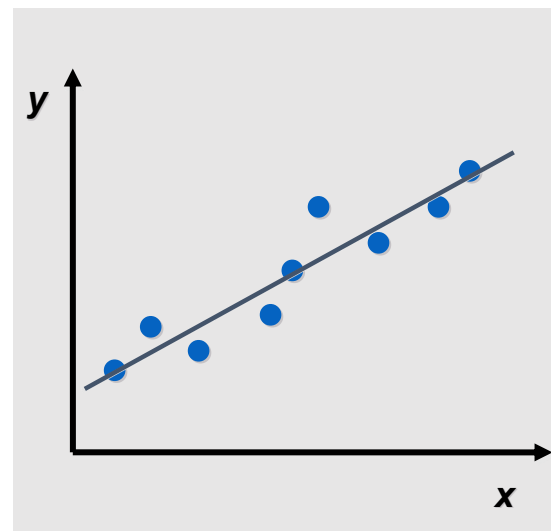


相关分析与回归分析

- 2. **回归分析**：是通过一定的数学方程来反映变量之间相互关系的具体形式，以便从一个已知量来推测另一个未知量

例：圆面积与圆半径之间的关系

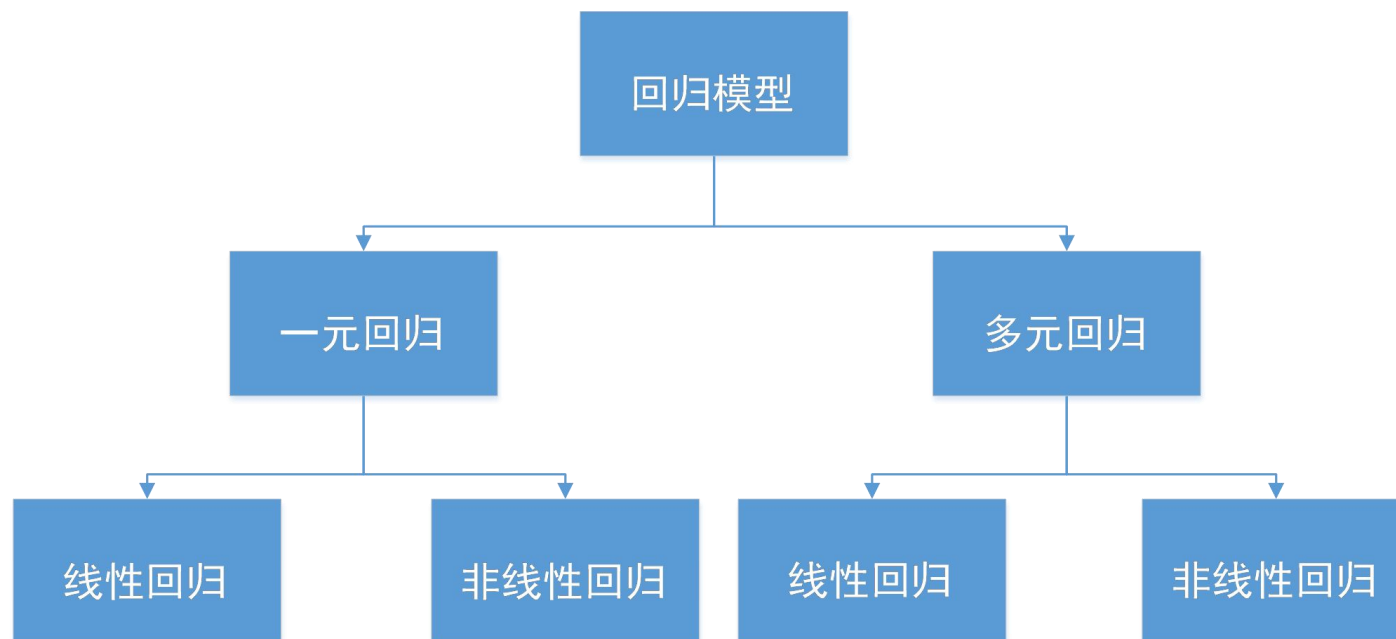
- 变量间关系可用函数关系表达
- 一个变量的取值能由另一个变量唯一确定
- 各观测点分布在直线附近



回归分析的概念

- 1.从一组样本数据出发，确定变量之间的数学关系式
- 2.对这些关系式的可信程度进行各种统计检验，并从影响某一特定变量的诸多变量中找出哪些变量的影响显著，哪些不显著
- 3.利用所求的关系式，根据一个或几个变量的取值来预测或控制另一个特定变量的取值，并给出这种预测或控制的精确程度

回归分析的模型



一元线性回归模型及概念

- 涉及一个自变量的回归
- 因变量 y 与自变量 x 之间为线性关系
 - 被预测或被解释的变量称为因变量，用 y 表示
 - 用来预测或用来解释因变量的一个或多个变量称为自变量，用 x 表示
- 因变量与自变量之间的关系用一个线性方程来表示

一元线性回归模型及概念

- 描述因变量 y 如何依赖于自变量 x 和误差项 ε 的方程称为回归模型
- 一元线性回归模型可表示为: $y = \beta_0 + \beta_1 x + \varepsilon$
 - y 是 x 的线性函数(部分)加上误差项
 - 线性部分反映了由于 x 的变化而引起的 y 的变化
 - 误差项 ε 是随机变量
 - 反映了除 x 和 y 之间的线性关系之外的随机因素对 y 的影响
 - 是不能由 x 和 y 之间的线性关系所解释的变异性
 - β_0 和 β_1 称为模型的参数

一元线性回归模型——随机误差的几个假设

- 误差项 ε 是一个期望值为0的随机变量，即 $E(\varepsilon)=0$ 。对于一个给定的 x 值， y 的期望值为： $E(y) = \beta_0 + \beta_1 x$
- 对于所有的 x 值， ε 的方差 σ^2 都相同
- 误差项 ε 是一个服从正态分布的随机变量，且相互独立。即 $\varepsilon \sim N(0, \sigma^2)$
 - 独立性意味着对于一个特定的 x 值，它所对应的 ε 与其他 x 值所对应的 ε 不相关
 - 对于一个特定的 x 值，它所对应的 y 值与其他 x 所对应的 y 值也不相关

回归方程、回归值的概念

- 描述 y 的平均值或期望值如何依赖于 x 的方程称为回归方程
- 一元线性回归方程可表示为: $E(y) = \beta_0 + \beta_1 x$
 - 方程的图示是一条直线，也称为直线回归方程
 - β_0 是回归直线在 y 轴上的截距，是当 $x=0$ 时 y 的期望值
 - β_1 是直线的斜率，称为回归系数，表示当 x 每变动一个单位时， y 的平均变动值

回归方程、回归值的概念

- 一元线性回归中估计的回归方程为: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$
- 其中: $\hat{\beta}_0$ 是估计的回归直线在 y 轴上的截距, $\hat{\beta}_1$ 是直线的斜率, \hat{y} 是 y 的估计值
- 将观察值 x_i 带入回归方程得到的值 y_i 称为回归值 (也称预测值、拟合值等)

小结

- 相关分析与回归分析
- 回归分析的模型及概念
- 一元线性回归模型及概念
- 回归方程、回归值的概念
- 小结