

线性回归分析双变量模型



目录

Contents

- 引言
- 双变量模型基本概念
- 数据收集与预处理
- 线性回归分析建模过程
- 模型诊断与优化
- 实例分析与应用展示
- 总结与展望

01

引言



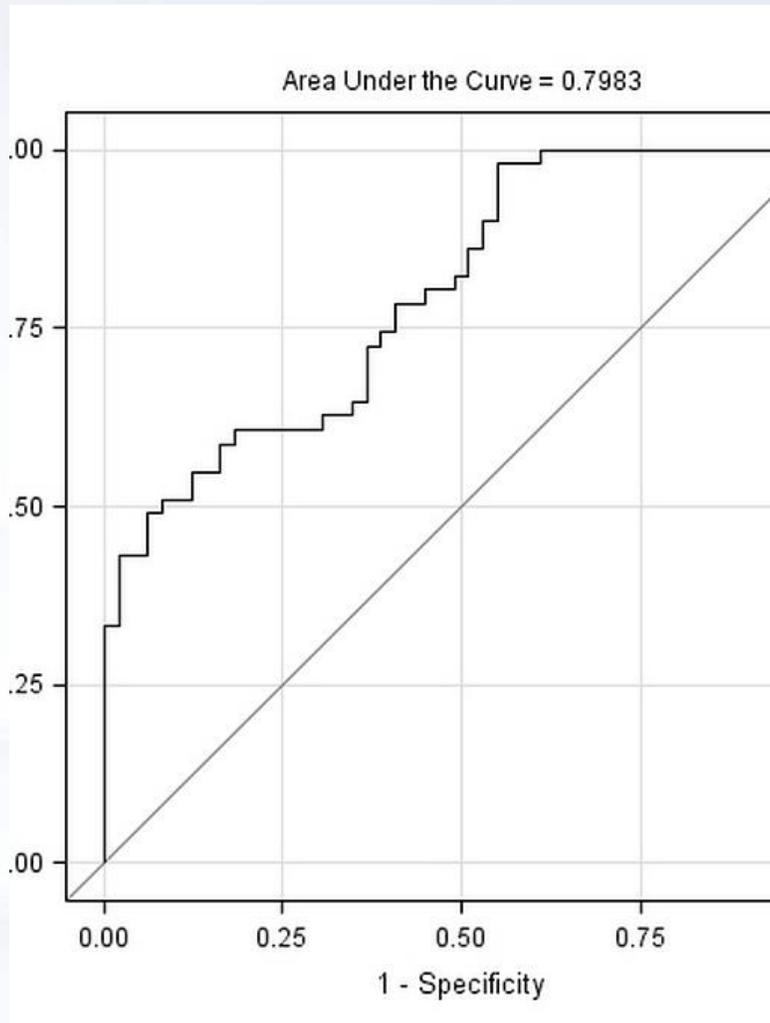
目的和背景

探究两个变量之间的关系

线性回归分析双变量模型的主要目的是探究两个变量之间的线性关系，通过建立一个数学模型来描述这种关系。

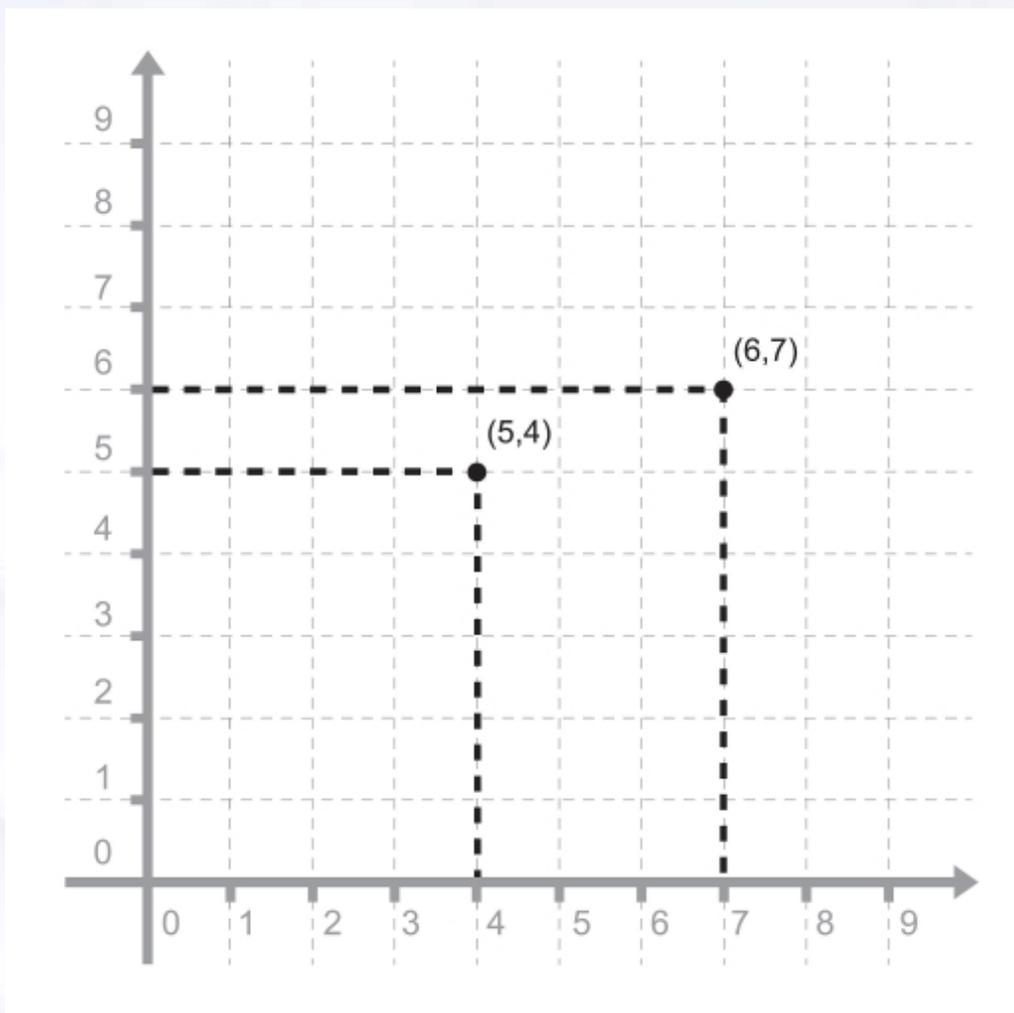
预测和决策支持

该模型可以用于预测一个变量基于另一个变量的值，为决策提供支持，例如在商业、经济、医学等领域中。





线性回归分析简介



线性回归定义

线性回归是一种统计分析方法，用于研究因变量和一个或多个自变量之间的线性关系。在线性回归分析中，数据被拟合到一条直线上，使得这条直线能够最好地描述变量之间的关系。

双变量线性回归

双变量线性回归特指只涉及两个变量的线性回归模型，其中一个变量作为因变量，另一个变量作为自变量。通过建立一个线性方程来描述这两个变量之间的关系，可以进一步进行预测和分析。

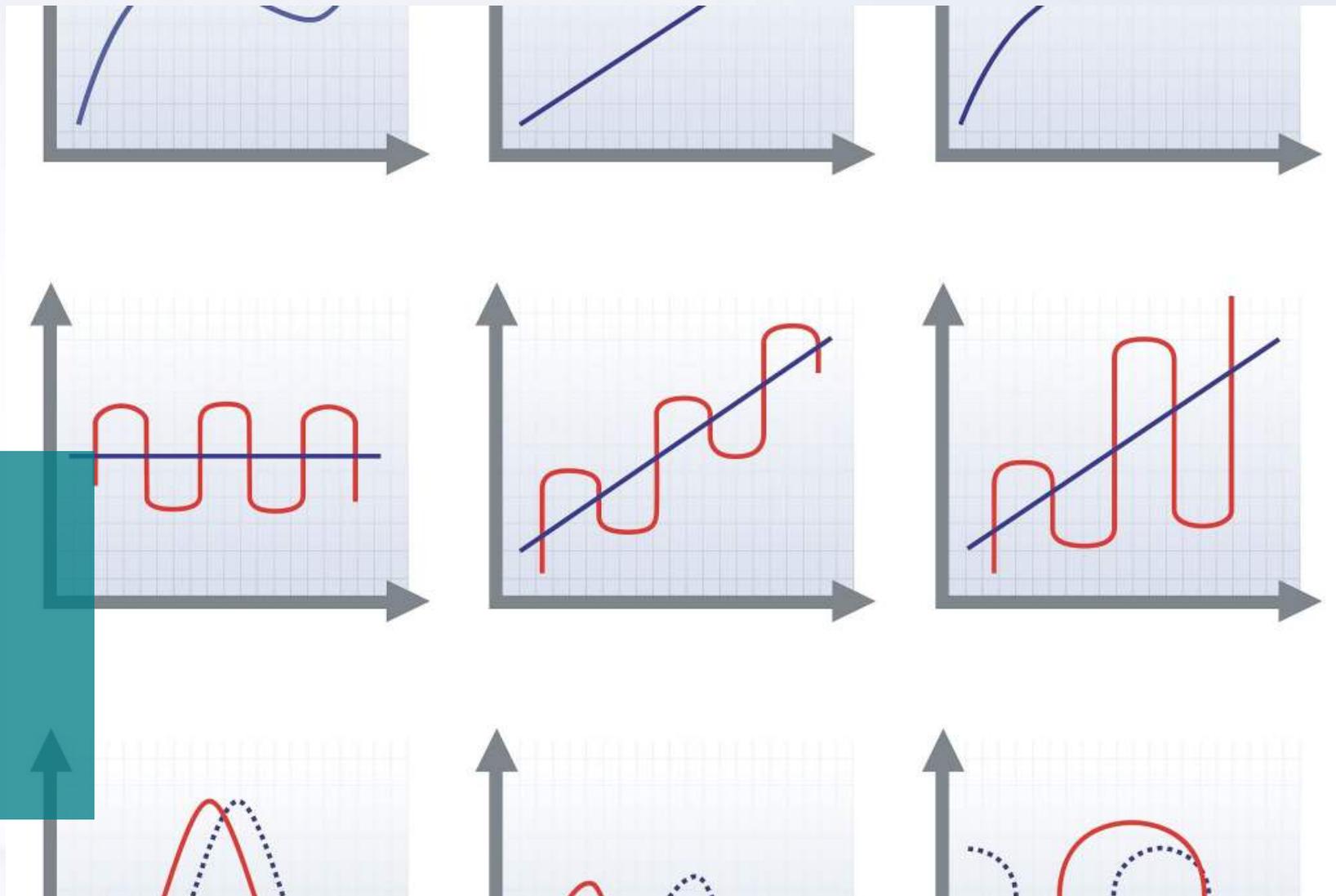
02

双变量模型基本概念

双变量模型定义

双变量模型是指探讨两个变量之间关系的统计模型，其中一个变量作为自变量（解释变量），另一个变量作为因变量（被解释变量）。

在线性回归分析中，双变量模型主要用于研究两个变量之间的线性关系，并通过建立回归方程来描述这种关系。





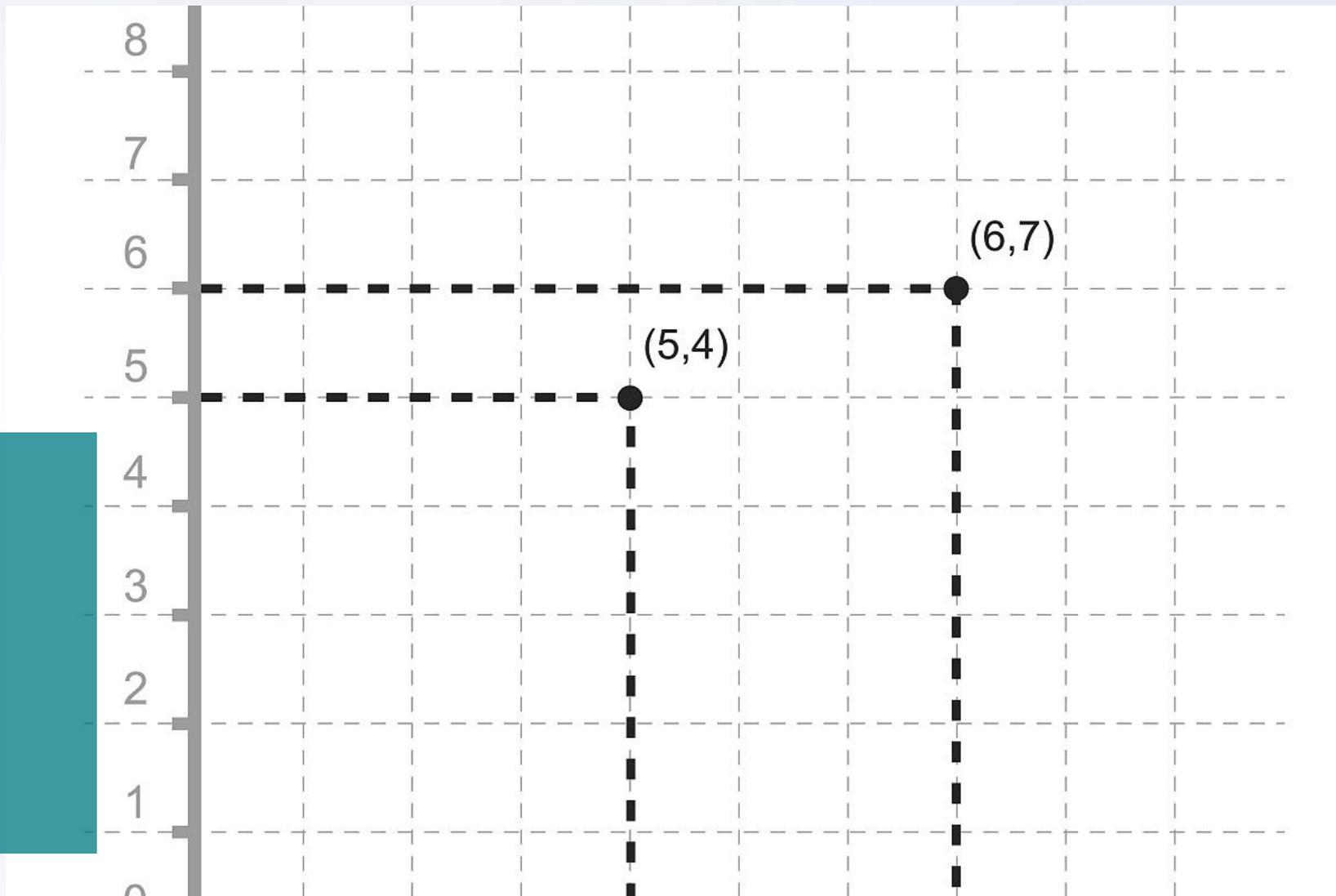
自变量与因变量

自变量 (X)

在双变量模型中，自变量是引起因变量变化的因素或条件，它是可以独立观测或控制的变量。

因变量 (Y)

因变量是受到自变量影响而发生变化的变量，也称为响应变量或结果变量。





线性关系与非线性关系

线性关系

当两个变量之间的关系可以通过一条直线近似表示时，称这两个变量之间存在线性关系。在线性回归分析中，我们假设自变量和因变量之间存在线性关系。

VS

非线性关系

如果两个变量之间的关系不能通过一条直线近似表示，而是呈现出曲线或其他复杂形态，那么这两个变量之间存在非线性关系。在非线性情况下，可能需要采用其他类型的回归模型进行分析。

03

数据收集与预处理



数据来源及收集方法

● 实验数据

通过设计和实施实验，直接获取与双变量相关的观测数据。

● 调查数据

通过问卷调查、访谈等方式收集实际场景中的双变量数据。

● 公开数据集

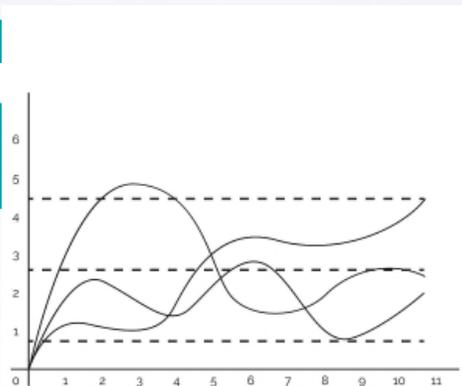
利用已有的公开数据库或数据集，筛选与双变量模型相关的数据进行研究。





数据质量评估与清洗

01

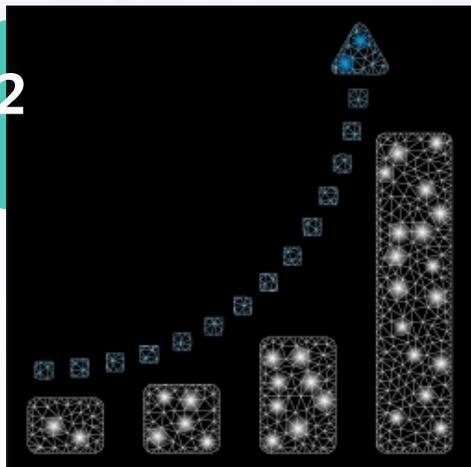


缺失值处理



检查数据中的缺失值，采用插值、删除或基于模型的方法进行处理。

02

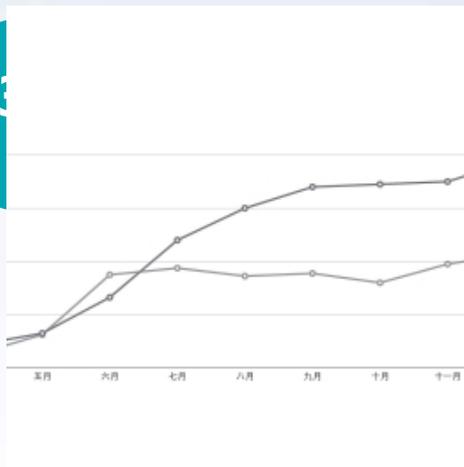


异常值检测与处理



利用统计方法识别异常值，并根据实际情况进行删除、替换或保留。

03



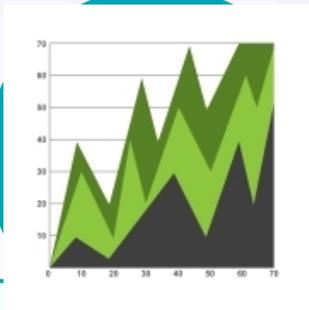
数据一致性检查



确保数据的逻辑性和一致性，处理可能存在的矛盾或错误数据。

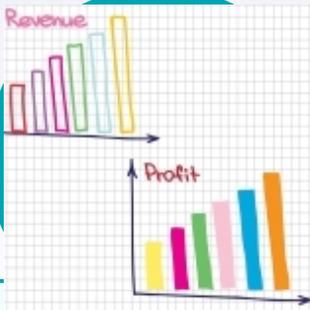


数据变换与标准化



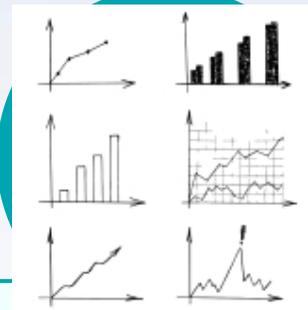
数据变换

根据分析需求，对数据进行对数变换、Box-Cox变换等，以改善数据的分布特性。



数据标准化

将数据按比例缩放，使之落入一个小的特定区间，如 $[0,1]$ 或 $[-1,1]$ ，以消除量纲对分析的影响。常用的方法有最小-最大标准化和Z-score标准化。



数据离散化

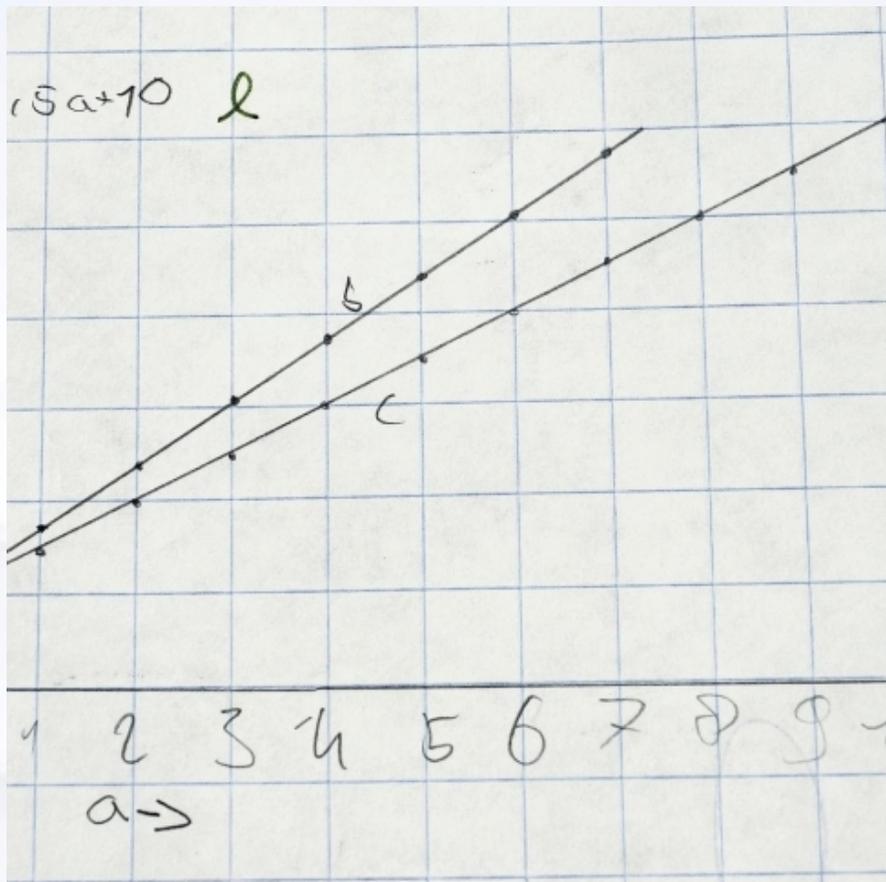
对于连续型变量，有时需要将其转换为离散型变量，以便于分析和解释。可采用等宽分箱、等频分箱等方法进行离散化。

04

线性回归分析建模过程



模型假设条件检验



线性关系假设

通过散点图初步判断两个变量之间是否存在线性关系。

误差项独立性假设

检验残差是否独立，可以通过Durbin-Watson检验等方法。

同方差性假设

检验误差项是否具有相同的方差，可以通过残差图、White检验等方法。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：
<https://d.book118.com/606015024052010110>