



回归分析：广义线性模型技术教程

回归分析基础

1. 线性回归模型简介

线性回归是统计学中一种常用的数据分析方法，用于研究一个或多个自变量与一个因变量之间的线性关系。在简单线性回归中，我们关注一个自变量与一个因变量之间的关系，模型可以表示为：

$$[y = \beta_0 + \beta_1 x + \epsilon]$$

其中， y 是因变量， x 是自变量， β_0 和 β_1 是模型参数， ϵ 是误差项，假设服从正态分布。

1.1 示例代码

假设我们有一组数据，表示学生的学习时间与考试成绩之间的关系，我们可以使用Python的statsmodels库来拟合一个简单线性回归模型。

```
import numpy as np
import pandas as pd
import statsmodels.api as sm

# 创建示例数据
data = {
    'study_hours': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10],
    'exam_scores': [65, 67, 70, 72, 75, 78, 80, 83, 85, 88]
}
df = pd.DataFrame(data)

# 添加常数项
df['intercept'] = 1

# 定义模型
model = sm.OLS(df['exam_scores'], df[['intercept', 'study_hours']])

# 拟合模型
results = model.fit()

# 输出模型参数
print(results.params)
```

2. 最小二乘法与最大似然估计

2.1 最小二乘法

最小二乘法是一种用于估计线性回归模型参数的方法，其目标是 minimized 残差平方和（RSS），即模型预测值与实际观测值之间的差异的平方和。最小二乘估计（LSE）可以解析求解，适用于较小的数据集。

2.2 最大似然估计

最大似然估计（MLE）是另一种参数估计方法，它基于假设模型的误差项服从特定分布（如正态分布），然后找到使观测数据出现概率最大的参数值。MLE适用于更广泛的情况，包括非线性模型和非正态分布的误差项。

3. 假设检验与模型评估

3.1 假设检验

在回归分析中，我们通常进行假设检验来确定模型参数是否显著，即参数是否与零有显著差异。常用的检验包括t检验和F检验。

3.2 模型评估

模型评估涉及检查模型的拟合优度和预测能力。常用的评估指标包括R方（R-squared）、调整R方（Adjusted R-squared）、均方误差（MSE）和均方根误差（RMSE）。

3.3 示例代码

使用statsmodels库，我们可以进行假设检验和模型评估。

```
# 输出模型的详细统计信息
print(results.summary())

# 计算R方
r_squared = results.rsquared
print('R-squared:', r_squared)

# 计算均方误差
mse = results.mse_resid
print('Mean Squared Error:', mse)
```

4. 多元线性回归

多元线性回归扩展了简单线性回归，允许同时考虑多个自变量对因变量的影响。模型可以表示

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/308002004024006111>