**

**HUNAN UNIVERSITY**

《DM&BID》

Report 4

|  |  |
| --- | --- |
| **报告名称：** | 线性回归模型 |
| **学生姓名：** | 杨超然 |
| **学生学号：** | 202106060220 |
| **专业班级：** | 电商2102班 |
| **学 院：** | 工商管理学院 |
| **指导老师：** | 江资斌 |
| **日 期：** | 2023.2.20 |

目录

[一、一元线性回归 3](#_Toc1962557932)

[（一）一元线性回归的代码实现 3](#_Toc1509828851)

[1.1绘制散点图 3](#_Toc1026725805)

[1.2引入sklearn库搭建模型 3](#_Toc1149500990)

[1.3模型预测与数据可视化 3](#_Toc900250518)

[1.4构造回归方程 3](#_Toc1488162911)

[（二）案例：不同行业工龄与薪水的线性回归模型 4](#_Toc1959492215)

[2.1案例背景 4](#_Toc1523930760)

[2.2读取数据与模型搭建 4](#_Toc1814309198)

[2.3模型可视化 5](#_Toc974387033)

[2.4构造线性回归方程 5](#_Toc1960055256)

[二、 线性回归模型评估 6](#_Toc249542612)

[（一） 模型评估的编程实现 6](#_Toc27117293)

[三、 多元线性回归 7](#_Toc493810287)

[（一）案例：客户价值预测模型 7](#_Toc1592681601)

[1.1案例背景 7](#_Toc1963491799)

[1.2读取数据与模型搭建 7](#_Toc25462344)

[1.3构造回归方程 7](#_Toc596369855)

[1.4模型评估 8](#_Toc881972436)

[四、 知识小记 9](#_Toc1378600258)

[（一） 模型评估 9](#_Toc933468723)

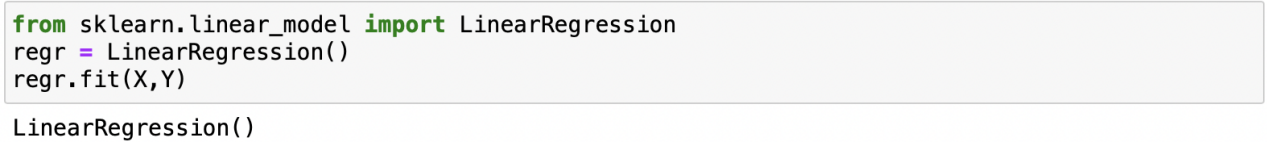
# 一、一元线性回归

## （一）一元线性回归的代码实现

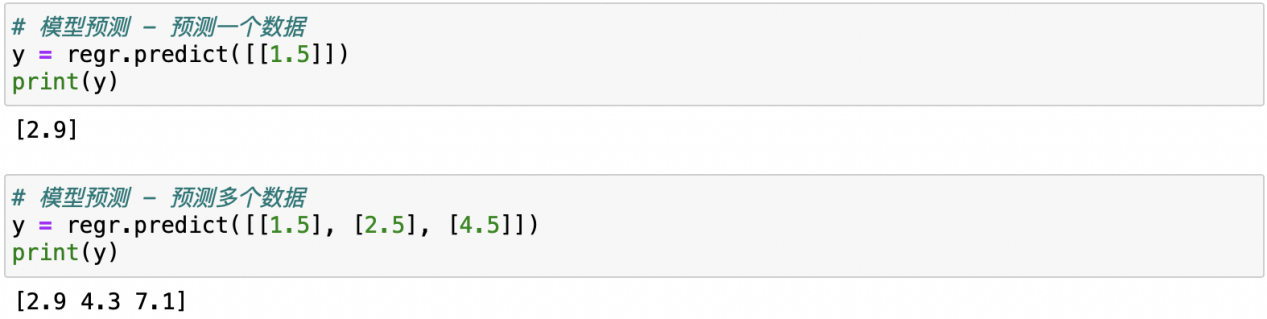
1.1绘制散点图

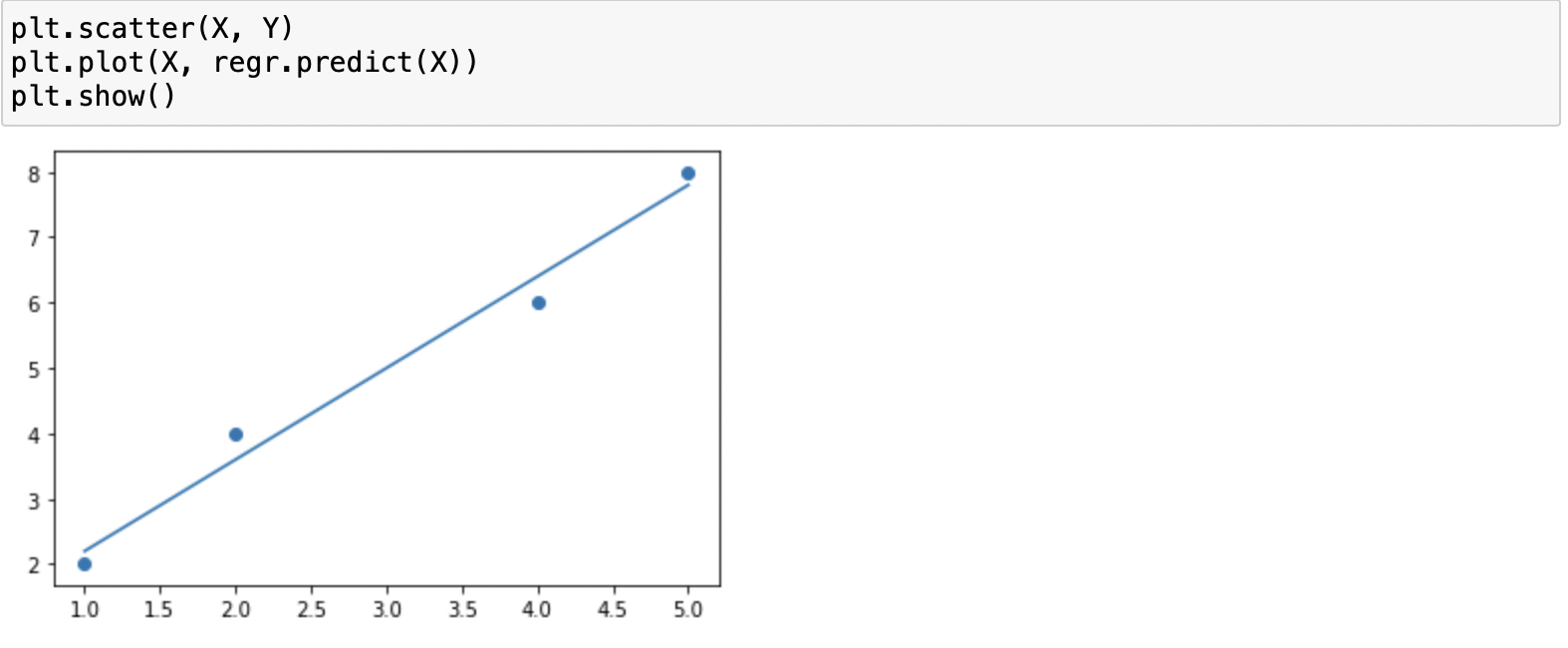


1.2引入sklearn库搭建模型

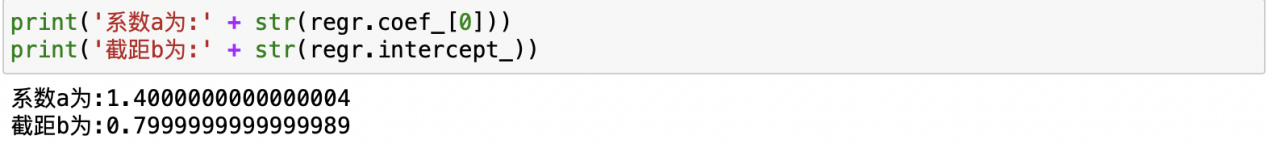


1.3模型预测与数据可视化





1.4构造回归方程



那么此时的一元线性回归得到的线性回归方程就可以表示为如下形式：y = 1.4\*x + 0.8

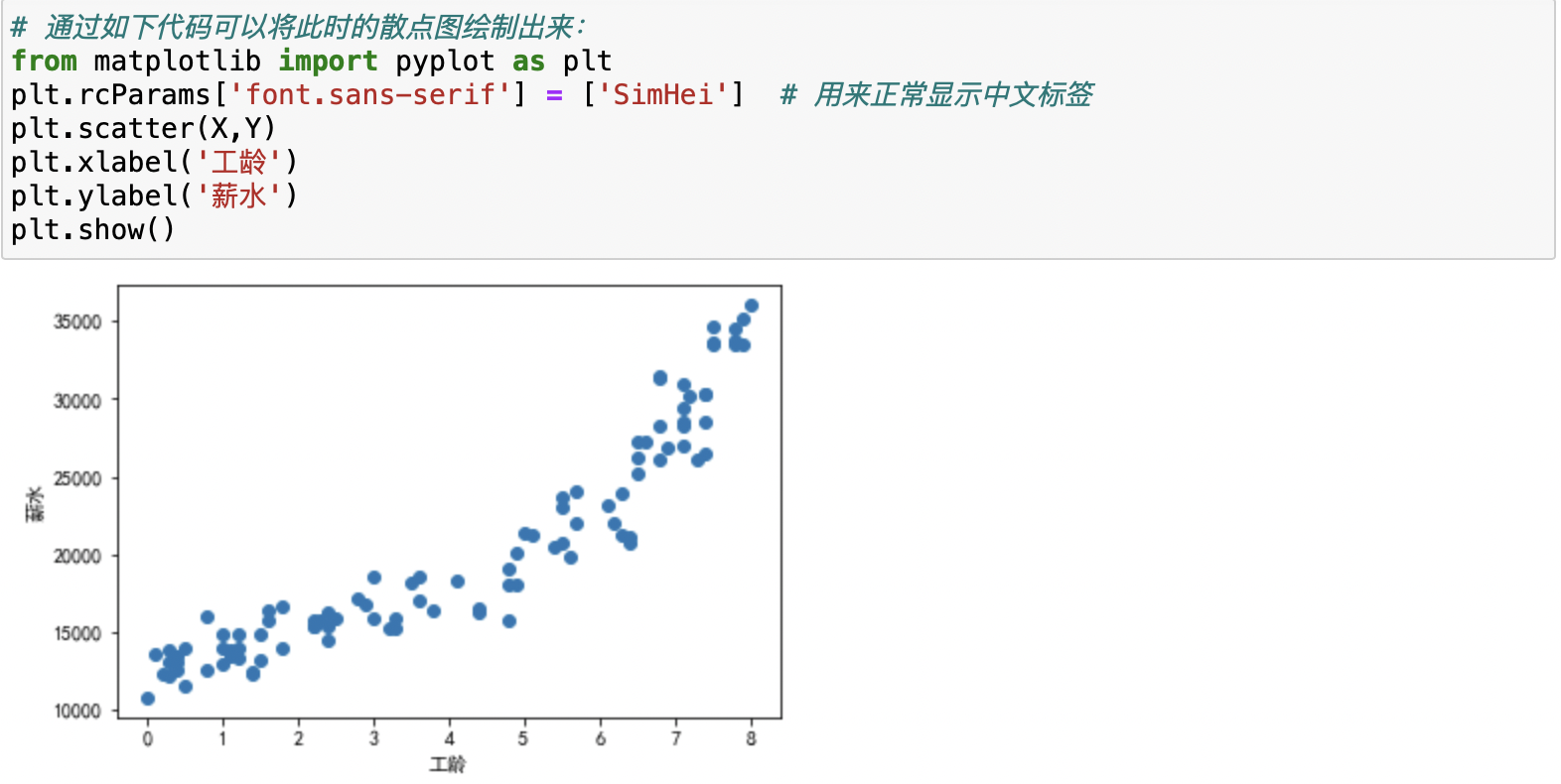
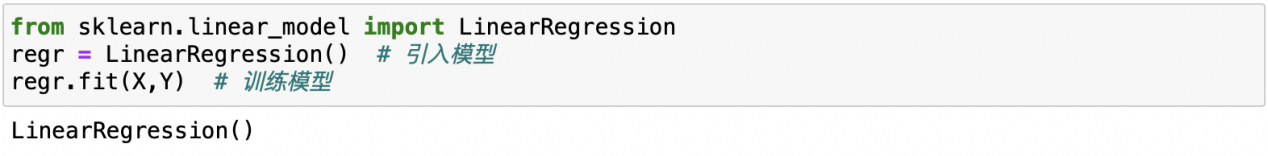
## （二）案例：不同行业工龄与薪水的线性回归模型

2.1案例背景

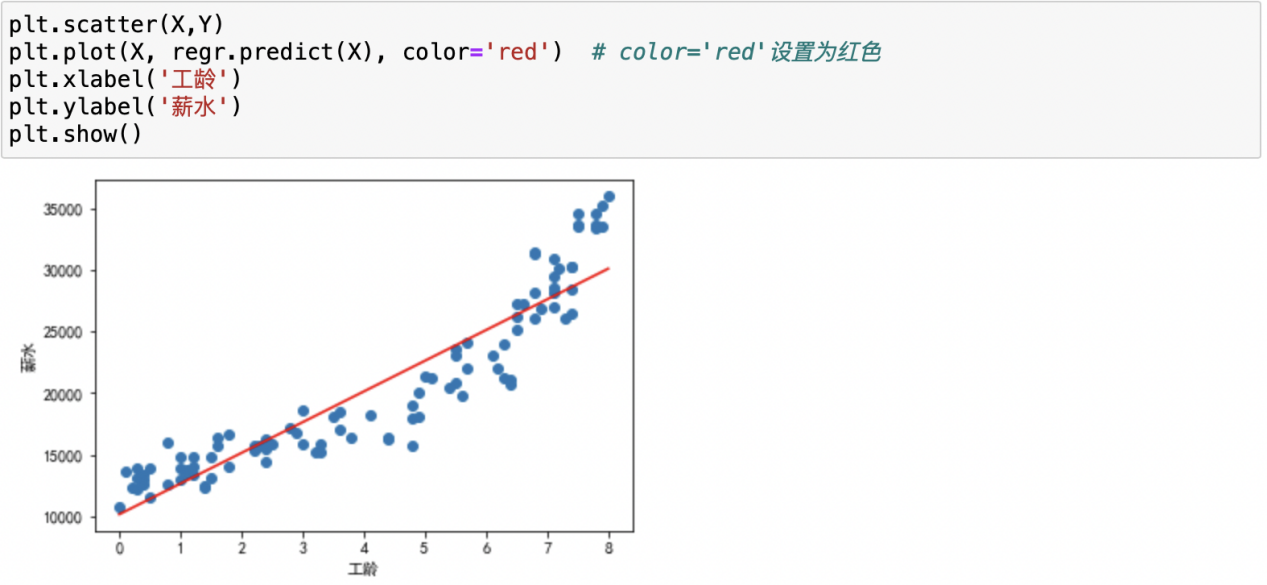
通常来说，薪水会随着工龄的增长而增长，而在不同的行业中薪水的增长速度都会有所不同，本案例通过一元线性回归模型来探寻工龄对薪水的影响，也即搭建薪水预测模型，同时比较多个行业的薪水预测模型来分析各个行业的特点。

2.2读取数据与模型搭建





2.3模型可视化



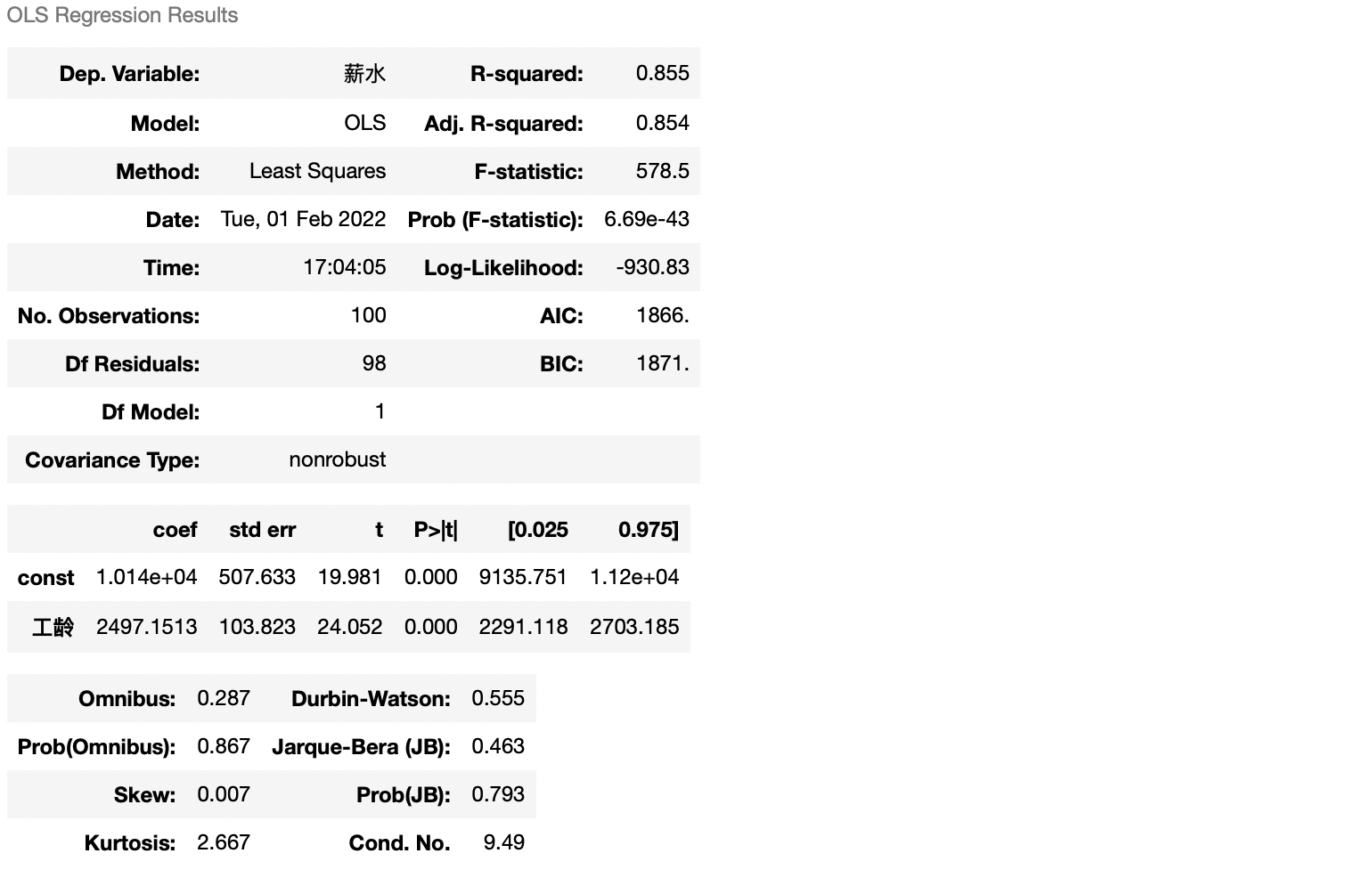
2.4构造线性回归方程



# 线性回归模型评估

# 模型评估的编程实现





此外，还可以用另外一种方法来获取R-Squared值，代码如下：



# 多元线性回归

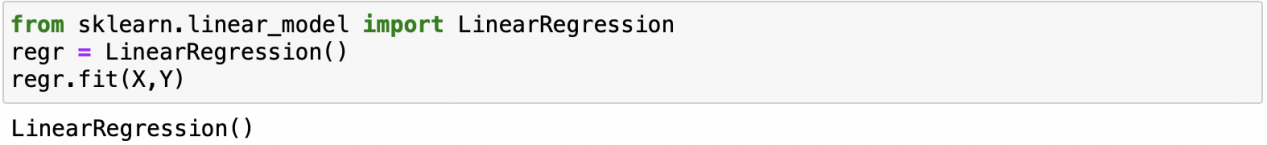
# （一）案例：客户价值预测模型

1.1案例背景

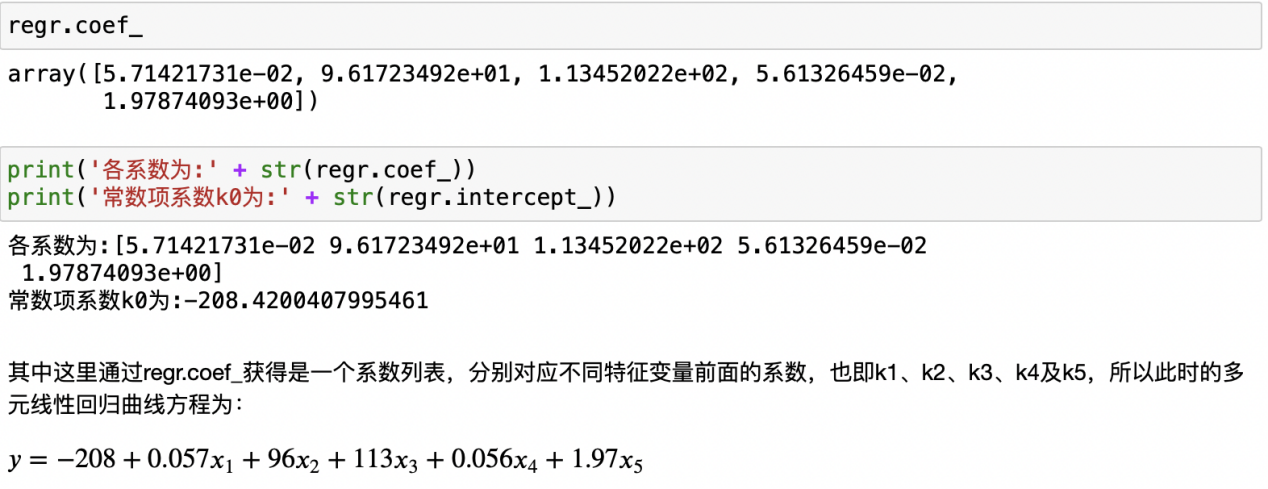
这里以信用卡客户的客户价值来解释下客户价值预测的具体含义：客户价值预测就是指客户未来一段时间能带来多少利润，其利润的来源可能来自于信用卡的年费、取现手续费、分期手续费、境外交易手续费用等。而分析出客户的价值后，在进行营销、电话接听、催收、产品咨询等各项服务时，就可以针对高价值的客户进行区别于普通客户的服务，有助于进一步挖掘这些高价值客户的价值，并提高这些高价值客户的忠诚度。

1.2读取数据与模型搭建

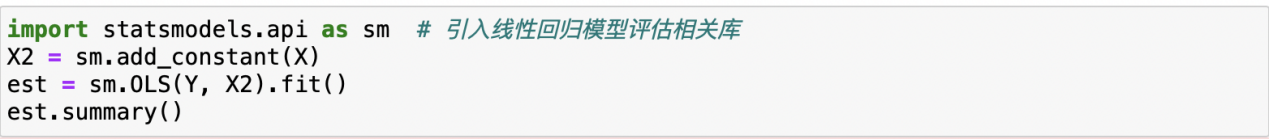




1.3构造回归方程



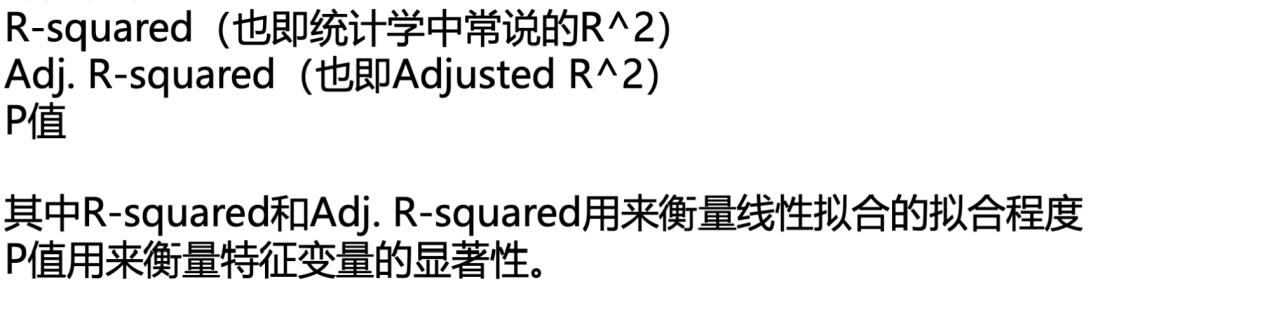
1.4模型评估



# 知识小记

## 模型评估

我们主要以三个值作为评判标准：



R squared或者Adj. R-squared越高，那么模型的拟合程度越高；如果P值越低，那么该特征变量的显著性越高，也即真的和预测变量有相关性。R-squared和Adj.R-squared的取值范围为0-1 。P值本质是个概率值，其取值范围也为0-1