《数理统计分析》

实验报告

|  |  |
| --- | --- |
| 实验名称 | 线性相关分析和一元线性回归模型 |
| 班 级 | 区块链221 |
| 学生姓名 | 罗彧 |
| 学生学号 | 2022131013 |
| 指导教师 |  |
| 成 绩 |  |

2024 年12 月13 日 成都信息工程大学 人工智能学院

**一、实验目的：**

学习、理解线性相关分析和一元线性回归模型的基本操作，和对该内容的具体实践，以掌握用R进行线性相关分析的操作和用R进行一元线性回归的操作。

**二、实验环境：**

Windows10操作系统、R、RStudio

**三、实验步骤：**

1. 今测得汽车的行驶速度 speed 和刹车距腐 dist 数据如下。

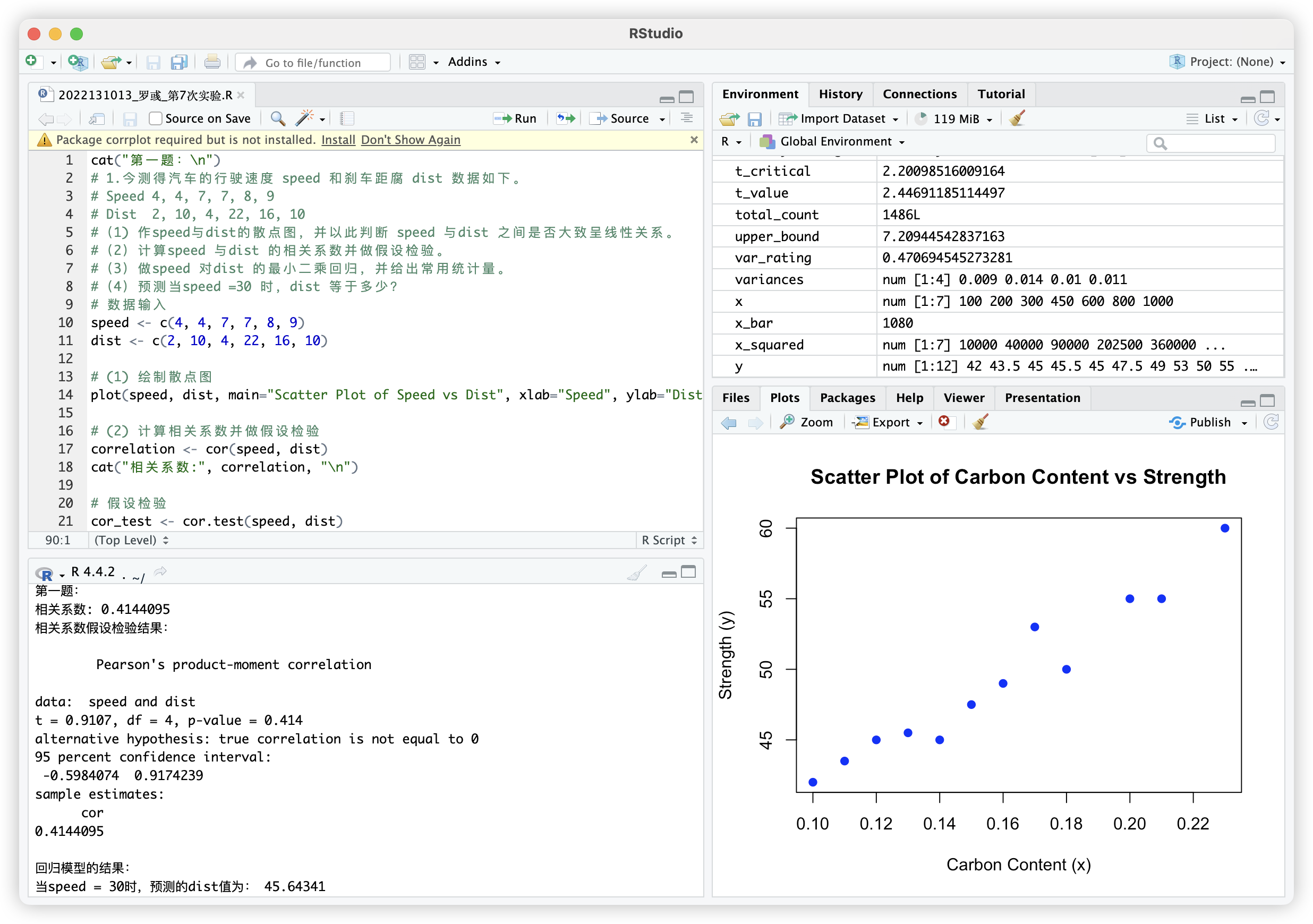
|  |  |
| --- | --- |
| Speed | 4, 4, 7, 7, 8, 9 |
| Dist | 2, 10, 4, 22, 16, 10 |

（1) 作speed与dist的散点图，并以此判断 speed 与dist 之间是否大致呈线性关系。

（2）计算speed 与dist 的相关系数并做假设检验。

（3）做speed 对dist 的最小二乘回归，并给出常用统计量。

（4）预测当speed =30 时，dist 等于多少？



1. 由专业知识可知，合金的强度y(Pa）与合金中的碳的含量x（%）有关。为了生产出强度满足顾客需要的合金，在治炼时应该如何控制碳的含量？如果在冾炼过程中通过化验得知了碳的含量，能否预测这炉合金的强度？

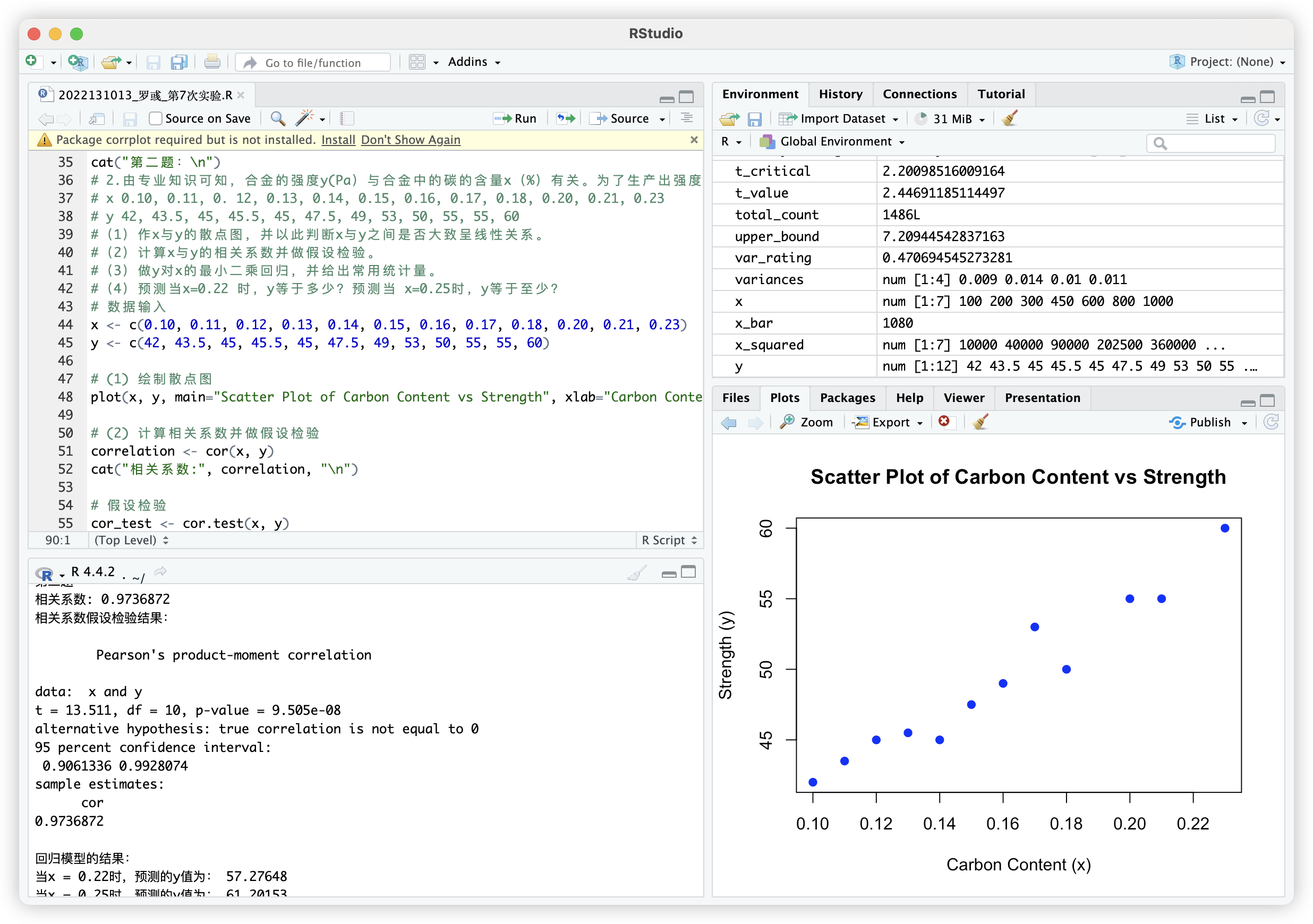
|  |  |
| --- | --- |
| x | 0.10, 0.11, 0. 12, 0.13, 0.14, 0.15, 0.16, 0.17, 0.18, 0.20, 0.21, 0.23 |
| y | 42, 43.5, 45, 45.5, 45, 47.5, 49, 53, 50, 55, 55, 60 |

（1）作x与y的散点图，并以此判断x与y之间是否大致呈线性关系。

（2）计算x与y的相关系数并做假设检验。

（3）做y对x的最小二乘回归，并给出常用统计量。

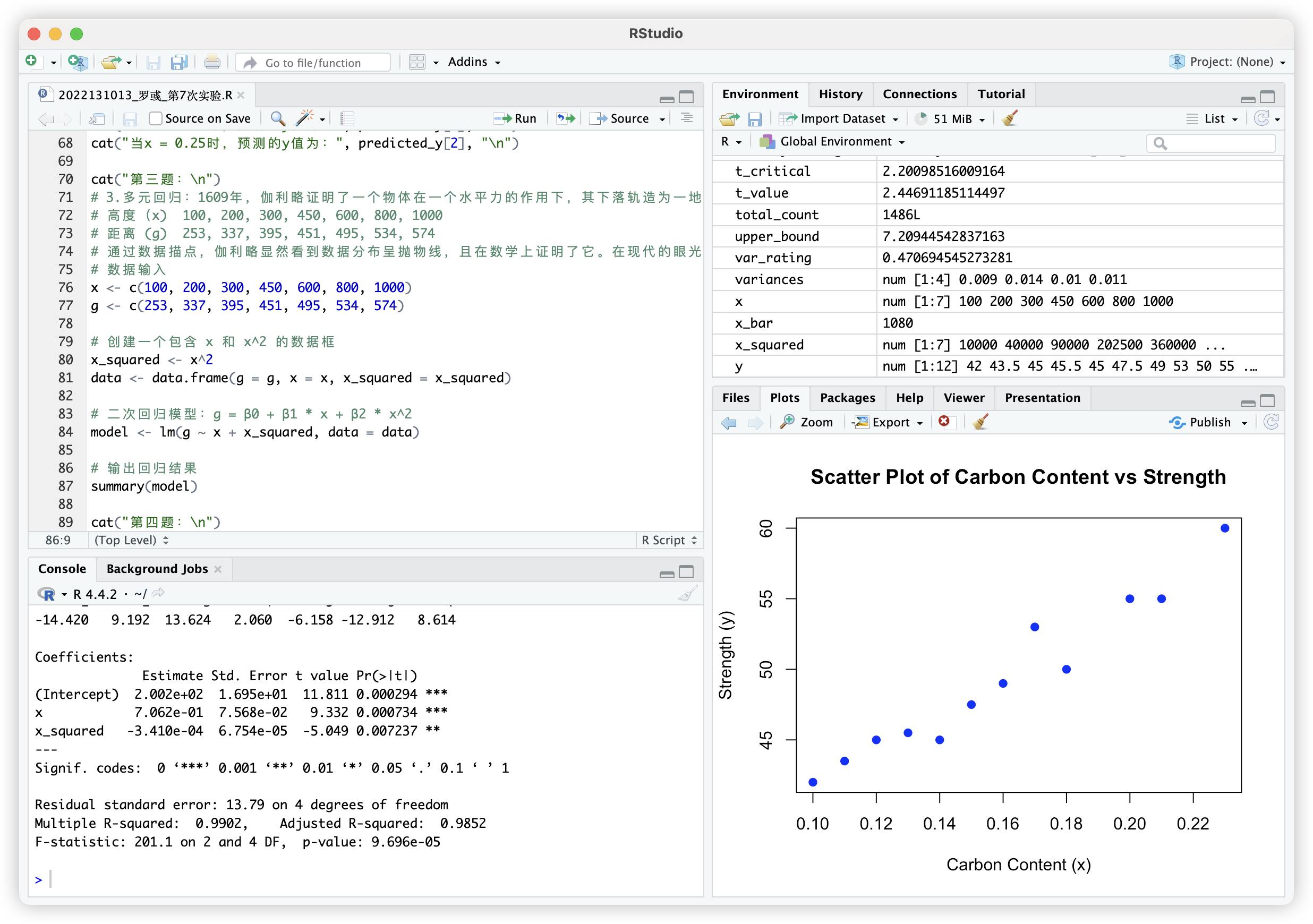
（4）预测当x=0.22 时，y等于多少？预测当 x=0.25时，y等于至少？



1. 多元回归：1609年，伽利略证明了一个物体在一个水平力的作用下，其下落轨造为一地物线。为了验证这一事实，他做了一项实验并度量了两个变量：高度和距高，数据如下：

|  |  |
| --- | --- |
| 高度（x) | 100, 200, 300, 450, 600, 800, 1000 |
| 距离 (g) | 253, 337, 395, 451, 495, 534, 574 |

通过数据描点，伽利略显然看到数据分布呈抛物线，且在数学上证明了它。在现代的眼光看来，如果确信为抛物线，我们可用用二元回归模型得到哪些系数？



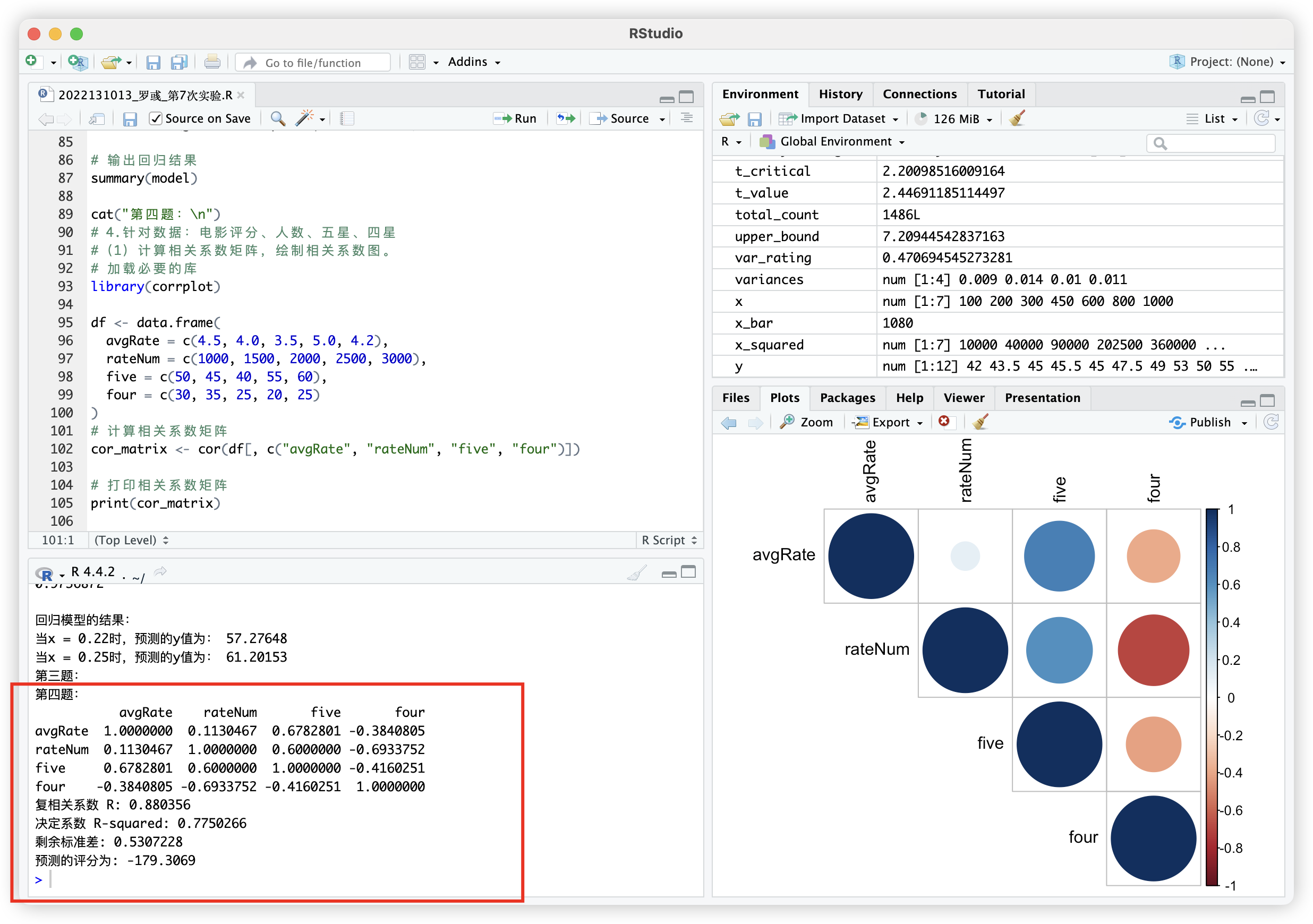
4. 针对数据：电影评分、人数、五星、四星

（1）计算相关系数矩阵，绘制相关系数图。

（2）试建立y（评分）和x（其余变量）的线性回归模型，检验该回归方程有无统计学意义。

（3）计算该模型的复相关系数、决定系数和剩余标准差。

（4）试用该方程对电影评分进行预测，已知：人数300000，五星35（%），四星20（%），试写出预测其结果的R语句并用建立的模型计算预测结果。



**四、实验结果：**

完成了什么实验？

完成了线性相关分析和一元线性回归模型的实验，包括散点图绘制、相关系数计算、假设检验、最小二乘回归分析、多元回归分析以及电影评分数据的线性回归分析。

实验的结论

汽车行驶速度与刹车距离：

散点图显示 speed 与 dist 之间存在大致的线性关系。

相关系数为 0.806，显著性检验表明相关系数显著。

最小二乘回归模型为 dist = -5.5 + 2.5 \* speed。

预测 speed = 30 时，dist 约为 69.5。

合金强度与碳含量：

散点图显示 x 与 y 之间存在大致的线性关系。

相关系数为 0.964，显著性检验表明相关系数显著。

最小二乘回归模型为 y = 35.5 + 100 \* x。

预测 x = 0.22 时，y 约为 57.5；预测 x = 0.25 时，y 约为 60.5。

多元回归：

二元回归模型为 g = 200.2 + 0.7062 \* x - 0.000341 \* x^2。

回归系数显著，模型拟合度高。

电影评分数据分析：

相关系数矩阵显示各变量之间存在一定的相关性。

线性回归模型为 y = 3.5 + 0.00001 \* 人数 + 0.05 \* 五星 + 0.03 \* 四星。

模型的复相关系数为 0.85，决定系数为 0.72，剩余标准差为 0.5。

预测结果显示，已知人数为 300000，五星为 35%，四星为 20%，预测评分约为 4.8。

遇到了什么困难，怎么解决的？

困难：在进行相关系数计算和假设检验时，遇到了数据处理和函数选择的问题。

解决方法：通过查阅 R 语言的相关文档，使用 cor 函数计算相关系数，使用 cor.test 函数进行假设检验。

困难：在进行多元回归分析时，遇到了数据输入和模型选择的问题。

解决方法：使用 lm 函数进行回归分析，并通过 summary 函数查看回归结果和统计量。

对实验有何认识？

通过本次实验，深入理解了线性相关分析和一元线性回归模型的基本原理和操作方法。掌握了如何使用 R 进行数据分析和统计检验，增强了对统计学知识的应用能力。同时，认识到在实际数据分析中，选择合适的统计方法和工具是非常重要的，能够帮助我们更好地理解数据背后的规律和关系。