## 线性回归的步骤

1. **多元线性回归分析**
2. **确定模型需要考虑的多个自变量，并假定因变量与各自变量之间是线性关系。**
3. **数据预处理**
4. 数据清洗：主要是为了保证数据的完整性与准确性。主要包括处理缺失值、识别和处理异常值两个方面。
5. 数据标准化：主要是为了消除量纲和变量自身变异大小和数值大小的影响。主要方法有离差标准化、标准差标准化。
6. **相关性分析**
7. 图示法：散点图（用于展示两个变量之间的关系），气泡图（用于展示三个变量之间的关系），矩阵散点图（用于展示多个变量之间的关系）
8. 相关系数测度：用样本相关系数估计总体相关系数（用t检验判断显著性），其中，大于0则正线性相关，小于0则负线性相关，越接近1则它们的线性相关性越强，为0时则无线性相关关系。
9. **参数估计**
10. 建立一元线性回归模型：
11. 利用最小二乘法（OLS）计算参数的估计值：
12. 得出估计的回归方程：

**（最小二乘法的思想：寻找使残差平方和最小的回归系数）**

1. **显著性检验**
2. 拟合优度检验

多重判定系数：。由于随着自变量的增加也会变大，为了避免高估，一般使用调整的多重判定系数。其中，越接近于1，表示拟合越好。

估计标准误差：。它反映了用估计的回归方程预测因变量时预测误差的大小，它越小说明效果越好。

1. 回归方程的显著性检验（F检验）：用于检验因变量同多个自变量的整体线性关系是否显著。
2. 回归系数的显著性检验（t检验）：用于判断每个自变量对因变量的影响是否都显著。
3. **回归调优与预测：**若各模型假定成立，则可使用估计的回归方程进行预测；若不成立，则需要重新对模型进行调优。