



绿色表示第一个变量的情况，黄色表示第二个变量的情况。任务是准备数据，以模拟公路汽油消耗量（MPG_highway）（或城市汽油消耗量（MPG_city））与数字变量 "车长 重量 轴距 马力 发票 发动机尺寸 汽缸数" 以及分类变量 "产地" 和 "类型" 之间的关系。任务要求

1. 对所有变量构建 MNC 回归，估计 R^2 、AIC 信息准则（BIC）和调整后的 R^2 （ $MSE=SSE/df_{error}$ ）。绘制残差与预测值的关系图和残差分布直方图。使用学生标准选择最显著和最不显著的变量（如果 p 值匹配，可任选其一）。根据这些数据，您认为您的问题陈述是否满足线性回归的假设？
2. 使用 CookD 统计量（DFBETA）搜索并排除异常值。有必要绘制这些统计数据（样本中所有观测值的值），并标出 3 个最异常的值。将它们从样本中剔除，重建 MNC 回归。第 1 项中的所有统计数据、图表和变量的重要性有何变化。对这些变化发表评论。
3. 对于第 2 项之后的样本，使用配对 t 检验对分类变量的水平进行分组。重组 ANC 回归。第 1 项中所有变量的统计量、绘图和重要性有何变化。评述分类变量的 t 检验在转换前后的变化。
4. 对于第 3 项之后的样本，使用讲座中讨论过的任何方法（正向、反向、组合、LARS）进行逐步选择，绘制系数轨迹图，并绘制每一步模型质量对 AIC（BIC）标准的依赖关系图。请注意，LARS 是按级别选择分类变量的，而您需要的是 "整体"。第 1 项中的所有统计数据、图表和变量的重要性发生了哪些变化。对这些变化发表评论。
5. 对于第 3 项之后的样本，通过尝试所有可能的分量数来构建一系列 PCR（PLS）回归，绘制 AIC（BIC）与分量数的关系图，并根据这一标准选择最佳模型。第 1 项、第 4 项和第 5 项中的所有统计数据、图表和变量的重要性有何变化。对这些变化发表评论。

将整个功能作为一个支持方法的类来实现：

- `model<-fit(train_data, step=1)` - 基于 train_data 建立模型，包括所有必要的模型建立步骤。
- **摘要（模型）** - 输出模型的所有必要统计信息，同时考虑到建立模型的步骤
- `plot(model)` - 根据建立模型的步骤，绘制所有必要的模型图