

# 多元统计分析—R 与 Python 的实现

## 回归

吴喜之

June 28, 2019

皇帝的新衣——“回归模型的可解释性”：

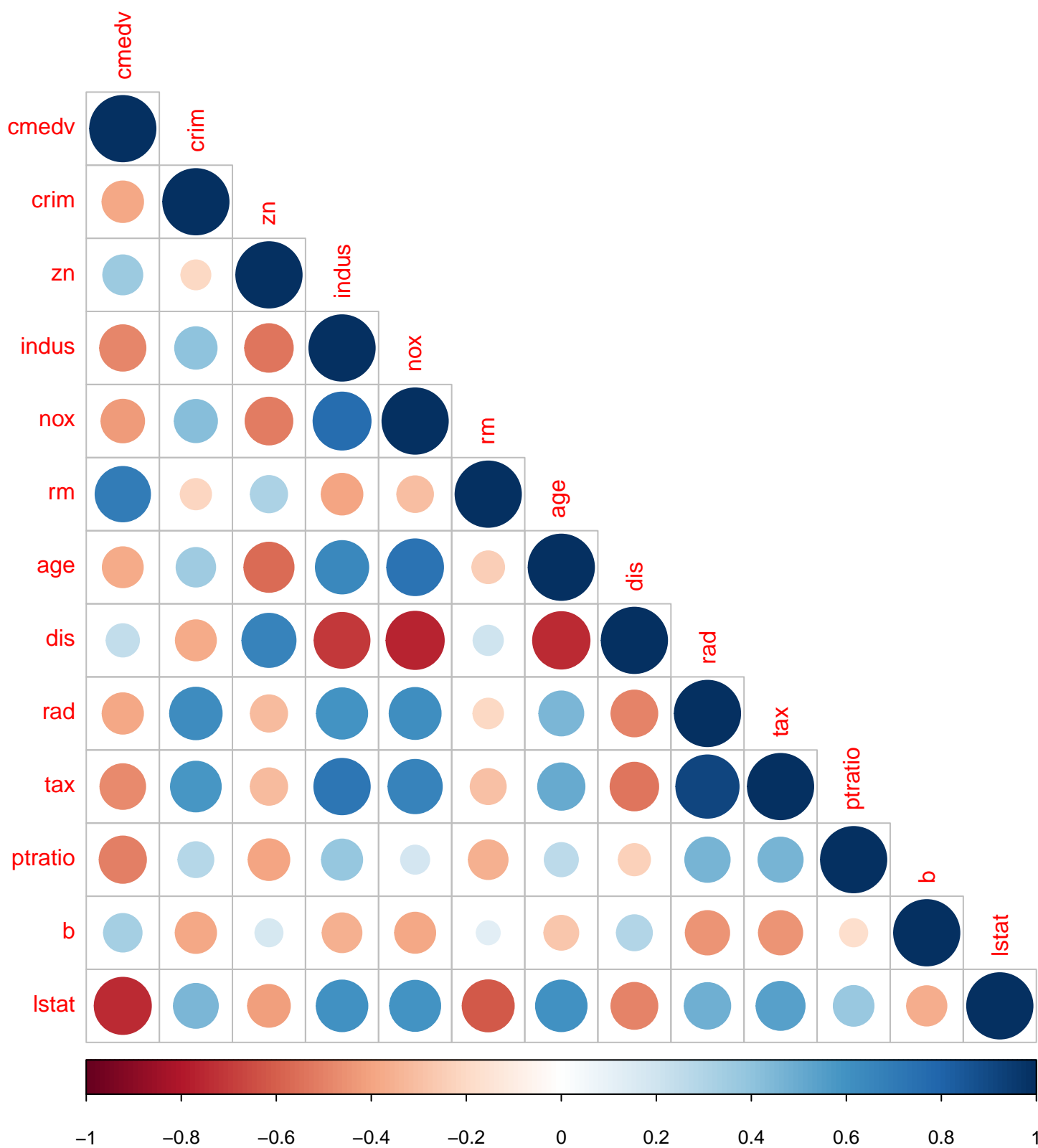
“多自变量线性回归某变量系数的值为其他变量不变时该变量增加一个单位时它对因变量的贡献”

如果其他变量可以不变, 而某一个变量可以自由变动, 说明这些变量独立.

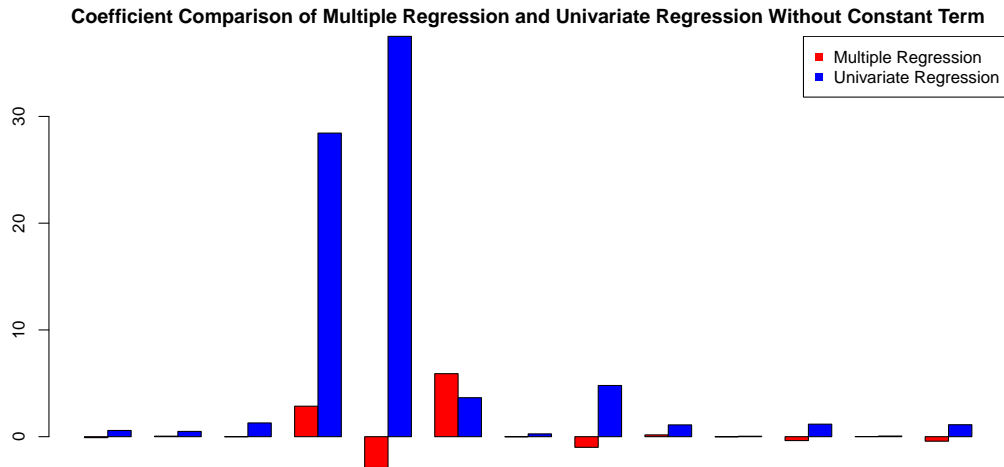
如果这些变量独立, 单独回归不就行了, 还为什么要做多自变量回归?

做多自变量回归的理由就是因为它们不独立!

下面看波士顿住房例子



我们来比较一下每一个自变量单独回归时的系数估计与多重回归时该变量系数估计的不同. 图29中的条形图展示了每个系数的两种回归的估计值, 这里的两种回归模型均未包含常数项. 结果表明对全部自变量和单个自变量回归时系数估计之间的巨大差异 (很多正负号都相反).



经典多重回归模型的所谓“可解释性”，是那些不情愿接受机器学习方法的人的一个最常见的借口。实际上，这种“可解释性”仅仅是一件皇帝的新衣而已。“在其他变量不变时，一个变量系数为该变量增加一个单位对因变量的贡献”的说法是多重回归“可解释性”的最主要的依据。由于历史原因，目前在国内尚有一些市场，但在整个统计界，很少有人这么认识了。

对于多重回归，当自变量具有多重共线性时，单独一个变量的作用掩盖在其他变量的共同作用之中，其系数大小没有任何解释意义，根本不值得探讨。

许多机器学习回归方法的可解释性相当好，而且不对数据做任何假定。