

引言

马学俊 (主讲) 沈琳 (助教)

苏州大学
数学科学学院

<https://xuejunma.github.io/>



1 人大真题

人大真题

- 在建立多元线性回归模型进行数据分析时，对数据进行“标准化”是一种常用的做法，请问为什么需要进行数据标准化？标准化的具体步骤是什么？请证明标准化回归系数和原始回归系数之间的关系，并阐述如何解读标准化回归系数？(2022,20 分)
- 现欲调查某大学学生对于手机依赖症对成绩绩点 (GPA) 的影响，设自变量手机依赖症为哑变量 (X)，成绩绩点 (GPA) 为因变量，高考成绩 (GK)，回归模型如下：

$$GPA_i = \beta_0 + \beta_1 GK_i + \beta_2 X_i + \varepsilon_i$$

请问随机误差项 ε_i 和手机依赖症 (X) 有相关性吗？为什么？倘若从该学校抽取 500 分简单随机样本，参数的最小二乘估计量有没有偏差？你认为该模型有哪些需要改进的研究建议？(2021,25)

- 线性回归模型的随机误差项存在序列相关时，会对统计推断带来什么问题？如何处理？(2020)
- 在回归分析中，说明如何判别是否存在异方差？(2019)
- 选用全部的自变量纳入回归方程的估计模型是否合理？用什么方法选择自变量？(2017)
- 自变量：出口总额、社会消费品总额、还有一个变量不记得了；因变量：地区生产总值给出了以上变量的线性回归模型分析结果，都是表格，有可决系数，调整的可决系数，方差分析表，回归系数检验结果，VIF（其中两个变量小于 10，一个变量为 18）（30 分）问题：1. 这个模型有哪些假定 2. 这个模型中哪个自变量对因变量的影响更重要？3. 请对这个模型做出综合评价。(2016)

- 设地区生产总值 (亿元) 为因变量, 固定资产投资 (亿元)、社会消费品零售总额 (亿元)、出口总额 (亿美元)、地方财政收入 (亿元)、电力消费量 (亿千瓦时)、居民消费水平 (元) 为自变量, 根据 31 个样本数据得到回归结果如下:

	Estimate	Std. Error	t value	Pr(t)
	-2377	1.166	-2.038	0.05270
固定资产投资	0.4504	0.08.166	5.515	1.14 e-05
社会消费品零售总额	1.110	0.1572	7.060	2.68 e-0.7
出口总额	18.87	6.379	2.958	0.00686
地方财政收入	0.9596	0.6959	1.379	0.18061
电力消费量	0.6683	0.5671	1.178	0.25016
居民消费水平	0.1194	0.06949	1.718	0.09868

Residual standard error:	1526	自由度	24
Multiple R-Squared:	0.9944	Adjusted R-squared	0.993
F -statistic:	708.8	P-Value	< 2.2 e-16

对该回归模型进行综合分析，评价是否需要改进，并给出思路(2018)。

- 简述多元线性回归建模的步骤. (2015)
- 多重共线性的影响以及处理方法 (2014)