

计量经济学 第二次作业

2020 年秋

注意事项:

1. 请于 **2020 年 11 月 6 日下午 5 点**前提交电子版至网络学堂, 不接受以其他形式提交的作业。对作业提交有疑问请发邮件联系助教齐伟 (qi.w.16@sem.tsinghua.edu.cn)。
2. 上机作业的数据集和代码请从网络学堂 → 课程文件 → 习题集下载。
3. 请将所有作业的答案生成一份 PDF 文档 (手写扫描或者打印皆可) 提交。在提交前, 请确认文件是否可以正常打开。
4. 欢迎在提交作业前参加助教答疑, 也可以和同学讨论作业思路, 但请提交你自己写的答案。

[不包含截距回归] 考虑回归方程

$$y = \beta x + u. \quad (1)$$

(1) 用最小二乘法估计 β , 将所得的估计量 $\hat{\beta}$ 化简到最简单的形式. (10 分)

(2) 在讨论多元线性回归的参数估计时, 我们介绍了一种得到 OLS 参数 $\hat{\beta}_1$ 的方法: 首先将 x_1 对其他所有自变量 (包括截距) 回归, 将得到的残差记为 \hat{r}_1 ; 然后将 y 对 \hat{r}_1 回归, 得到的斜率参数就是 $\hat{\beta}_1$ 。请利用上述推导证明, 在第二步中是否包括截距得到的估计值是一样的 (10 分)

(3) 记 $\hat{u}_i = y_i - \hat{\beta}x_i$, 其中 $\hat{\beta}$ 是上一步中得到的 OLS 估计量。请问如下性质是否对任意的回归方程 $y = \beta x + u$ 都成立? 如果是, 请证明。如果不是, 请举出一个反例. (10 分)

$$\sum_{i=1}^n \hat{u}_i = 0.$$

(4) 考虑另外一个回归方程:

$$y = \beta_0 + \beta_1 x + u. \quad (2)$$

将这个式子中对 β_1 的 OLS 估计量记为 $\tilde{\beta}_1$, 对 β_0 的 OLS 估计量记为 $\tilde{\beta}_0$ 。证明如下等式 (10 分) 并讨论在什么情况下 $\tilde{\beta}_1 = \hat{\beta}_1$ 。(10 分)

$$\hat{\beta}_1 = \tilde{\beta}_1 + \tilde{\beta}_0 \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2}.$$

[上机题：吸烟率与禁烟政策] 甲国为了降低吸烟率，在某些省份推行了禁烟政策。具体来说，推行了该政策的省份不允许居民在饭店里吸烟，没有推行的省份则无此限制。为了研究这一政策的影响，我们收集了一个数据集，每一个观测代表了一个人。包括以下变量：

- `cigs` 每天抽烟数（根）
- `age` 年龄
- `agesq` 年龄的平方
- `educ` 受教育年数
- `restaurn` 该个体所在地是否实施了上述禁烟政策

上机作业的数据集和对应的 Stata do file 可以从网络学堂 → 课程文件 → 习题集下载，Stata do file 文件名为 “`cigs.do`”，数据集文件名为 “`cigs.csv`”。注意：

- **此题不要求上传 Stata do file**
- 请直接下载代码后研究清楚每个步骤的作用（遇到看不懂的代码可以利用 `help` 指令），并在自己的计算机上运行结果，然后回答下列问题。
- 课堂展示的数据集和作业不同，因此**切勿照搬课堂展示结果**，否则不得分。

(1) 新建一个名为 `agegrp` 的变量代表不同的年龄组，其中 30 岁及以下的人群取值为 0，大于 30 小于等于 50 的取值为 1，大于 50 小于等于 70 的取值为 2，大于 70 岁的取值为 3。计算每一个年龄组的平均日吸烟量和每组的样本数，将 Stata 结果复制到下面（截屏、复制或者手抄数字皆可）。(2 分) 在你看来，年龄和抽烟数目是线性的关系吗？(3 分)

(2) 估计下面这个回归方程并给出回归结果（附上回归结果的截图）。(2 分)

$$\text{cigs} = \beta_0 + \beta_1 \text{age} + \beta_2 \text{age}^2 + \beta_3 \text{educ} + \beta_4 \text{restaurn} + u.$$

根据回归结果，在控制了教育水平和禁烟政策之后，平均来说年龄多大的人群吸烟数目最多？(3 分)

(3) 我们想检验在上述回归中， β_1 , β_2 和 β_3 是否都为零. 写出 H_0 和 H_1 (5 分). 解释检验的思路 (5 分)。在 1% 的显著性水平下是否可以拒绝原假设？在 5% 的水平上呢？为什么？(5 分)

(4) 解释 β_4 的含义.(5 分)

我们想检验在上述回归中 β_4 是否为零. 写出 H_0 和 H_1 (5 分). 在 1% 的显著性水平下是否可以拒绝原假设？在 5% 的水平上呢？为什么？(5 分)

(5) 我们想要研究禁言政策对不同年龄的人群影响是否不同。请估计如下模型并报告回归结果 (5 分)：

$$\text{cigs} = \beta_0 + \beta_1 \text{age} + \beta_2 \text{age}^2 + \beta_3 \text{educ} + \beta_4 \text{restaurn} + \beta_5 \text{restaurn} \cdot \text{educ} + u.$$

请写出 `restaurn` 对 `cigs` 偏效应的表达式。如何理解 β_5 的含义？(5 分)