

# 计量经济学-作业7-多重共线性



## P148 习题1

- 2 针对加利福尼亚州北部一个教学区的小学教师薪水进行的一项最新研究得出如下的估计方程（括号中的值为  $t$  统计量）：

$$\widehat{\ln SAL_i} = 10.5 - 0.006 EMP_i + 0.002 UNITS_i + 0.079 LANG_i + 0.0020 EXP_i$$

(−0.98)            (2.39)            (2.08)            (4.97)

$$\bar{R}^2 = 0.866 \quad N = 25 \quad (8-21)$$

式中， $SAL_i$  代表第  $i$  位教师的工资（单位：美元）； $EMP_i$  代表第  $i$  位教师在该教学区的教龄； $UNITS_i$  代表第  $i$  位教师完成的毕业生工作量； $LANG_i$  代表第  $i$  位教师是否会讲两种语言，代表虚拟变量，如果会两种为 1，否则为 0； $EXP_i$  代表第  $i$  位教师的教龄。

- 在 5% 的显著性水平下，对方程中的参数建立合适的假设并进行检验。
- 该方程的函数形式是什么？它是否恰当？请说明理由。
- 该方程存在什么样的计量经济学问题（在变量不相关、变量遗漏以及多重共线性中进行选择）？请说明理由。
- 假如你现在被告知，变量  $EMP$  与  $EXP$  之间的简单相关系数为 0.89，且它们的方差膨胀因子都恰好刚刚大于 5。这会改变你对问题 c 的回答吗？如何改变？
- 对于在问题 d 中所发现的问题，你建议采取什么样的补救措施？请说明理由。
- 如果在方程 (8-21) 剔除变量  $EMP$ ，那么，估计方程将变为方程 (8-22)。根据第 6 章的四个模型设定准则，你是更倾向于选择方程 (8-21)，还是方程 (8-22)？为什么？

$$\widehat{\ln SAL_i} = 10.5 + 0.002 UNITS_i + 0.081 LANG_i + 0.015 EXP_i \quad (8-22)$$

(2.47)            (2.09)            (8.65)

$$\bar{R}^2 = 0.871 \quad N = 25$$

## 回答

- a. 在显著性水平 5% 的条件下分别对 4 个参数进行假设检验。其中，样本总数为 25，可以求得自由度为  $n - k - 1 = 25 - 4 - 1 = 20$ ，查表得单侧  $t_c = 1.725$
- EMP**: 做出假设  $H_0: \beta \leq 0$   $H_A: \beta > 0$ ，且预测参数为正。  $t_1 = -0.98$ ，可知  $|-0.98| < 1.725$  即  $|t_1| < t_c$ ，因此**不能拒绝**  $H_0$
  - UNITS**: 做出假设  $H_0: \beta \leq 0$   $H_A: \beta > 0$ ，且预测参数为正。  $t_2 = 2.36$ ，可知  $|2.36| > 1.725$  即  $|t_2| > t_c$ ，且其预期符号和对立假设相同（均为正数），因此**可以拒绝**  $H_0$

- **LANG**:做出假设  $H_0: \beta \leq 0$   $H_A: \beta > 0$ ,且预测参数为正。  $t_3 = 2.08$  , 可知  $|2.08| > 1.725$  即  $|t_3| > t_c$  , 且其预期符号和对立假设相同 (均为正数) , 因此**可以拒绝**  $H_0$
- **EXP**:做出假设  $H_0: \beta \leq 0$   $H_A: \beta > 0$ ,且预测参数为正,  $t_4 = 4.97$  , 可知  $|4.97| > 1.725$  即  $|t_4| > t_c$  , 且其预期符号和对立假设相同 (均为正数) , 因此**可以拒绝**  $H_0$

- 方程的函数形式是**左侧半对数**。这是合理的, 因为当类似工作经验的解释变量每增加一个单位时, 被解释变量工资也通常会增加一定的比率。这种呈比率增长的符合对数的形式。
- 该方程可能存在**多重共线性**的问题, 因为方程的EMP变量的预期符号不正确, 且EMP的显著性不高。很有可能在EMP和EXP中存在多重共线性。当然也可能是遗漏变量的问题
- EMP和EXP的相关系数较高, 具有明显相关性, 且其方差膨胀因子大于5, **可以认为存在多重共线性**, 符合上一题的推论。即方程存在**多重共线性**的问题
- 可以**舍弃EMP这一变量, 因为EXP包含了EMP**, 且从生活常识来讲, EXP比EMP更能度量教师的教学经验。
- 我们更倾向选择方程8-22, 原因如下:
  - 理论: EMP (教龄) 对工资有很强的影响, 但是EMP和EXP是重复的, 所以我们只应保留EXP
  - t检验: 变量参数的估计值在预期的方向上显著。
  - $\overline{R^2}$  拟合优度: 当变量EMP被剔除出方程后, 方程总体的拟合优度变得更高了。(增长了0.005)
  - 偏误: 剔除参数后, 参数估计值改变并不明显

综上, 以上四点都指出EMP和EXP是重复的, 并且EMP应该被剔除, 因此8-22更优

## P148 习题6

- 考察下述来自斯坦福商学院招生办公室一项研究的回归结果 (括号中的数值为标准差):

$$\hat{G}_i = 1.00 + 0.005M_i + 0.20B_i - 0.10A_i + 0.25S_i$$

$$(0.001) \quad (0.20) \quad (0.10) \quad (0.10)$$

$$\overline{R^2} = 0.20 \quad N = 1000$$

式中,  $G_i$  代表第  $i$  位学生在斯坦福商学院的 GPA 成绩 (4 分制);  $M_i$  代表第  $i$  位学生的 GMAT 成绩 (800 分制);  $B_i$  代表第  $i$  位学生的商业从业年限;  $A_i$  代表第  $i$  位学生的年龄;  $S_i$  代表第  $i$  位学生主修的是不是经济学, 代表虚拟变量, 如果是则为 1, 否则为 0。

- 从理论上推导所有参数的预期符号 (尽量不要看上面的结果), 并设定适当的假设对这些预期符号进行检验 (包括选择显著性水平)。
- 这一方程存在问题吗? 请说明理由。
- 如果有人建议用一个多项式函数形式来表示变量  $A$ , 你对此有何反应? 为什么?
- 如果对该方程再进行一次回归, 你有什么建议 (如果有的话)?

## 回答

- 我们选择显著性水平 5%, 分别对 4 个参数进行假设检验。其中, 样本总数为 1000, 可以求得自由度为  $1000 - k - 1 = 1000 - 4 - 1 = 995$ , 查表得单侧  $t_{\infty} = 1.645$ ,  $t_{120} = 1.658$ 
  - M**: 做出假设  $H_0: \beta \leq 0$   $H_A: \beta > 0$ , 且预测参数为正。  $t_1 = \frac{0.005}{0.001} = 5$ , 可知  $|5| > t_{120} > t_{995}$  即  $|t_1| > t_c$ , 且其预期符号和对立假设相同 (均为正数), 因此**可以拒绝**  $H_0$
  - B**: 做出假设  $H_0: \beta \leq 0$   $H_A: \beta > 0$ , 且预测参数为正。  $t_2 = \frac{0.20}{0.20} = 1$ , 可知  $|1| < t_{\infty} < t_{995}$  即  $|t_2| < t_c$ , 因此**不能拒绝**  $H_0$
  - A**: 做出假设  $H_0: \beta \leq 0$   $H_A: \beta > 0$ , 且预测参数为正。  $t_3 = \frac{-0.10}{0.10} = -1$ , 知  $|-1| < t_{\infty} < t_{995}$  即  $|t_3| < t_c$ , 因此**不能拒绝**  $H_0$
  - EXP**: 做出假设  $H_0: \beta \leq 0$   $H_A: \beta > 0$ , 且预测参数为正,  $t_4 = \frac{0.25}{0.10} = 2.5$ , 知  $|2.5| > t_{120} > t_{995}$  即  $|t_4| > t_c$ , 且其预期符号和对立假设相同 (均为正数), 因此**可以拒绝**  $H_0$
- A 和 B 参数的值存在不显著性, **可能是具有多重共线性**。因为大多数学生都是 20 多岁, 商业从业年限近乎相同, 和年龄成正比, A 和 B 之间可能存在严重的多重共线性。当然, 也可能是遗漏变量的原因。
- 我认为是**可以的**, 变量 A (年龄) 对 GPA 的影响可能来自多方面, 成熟的人 (年龄较大) 一般因为思考力的下降, 获得高 GPA 的能力较弱; 年龄低的人因为涉及知识能力不足也无法获得较高的 GPA, 处在中间的人往往能取得较高 GPA, 其影响也许符合一个多项式函数的形式
- 在保持不变的情况下, 舍去 A、B 参数中的其中一个 (因为其可能存在潜在的多重共线性), 我会选择去除 A, 因为 B 对 GPA 的影响可能更显著, 年龄对 GPA 的影响似乎稍弱。