

Homework 4

姓名: 方嘉聪 学号: 2200017849 得分: 93

Problem 1. (1) 变量分布见表1:

表 1: 相关变量的均值及标准差分布 (N = 15862)

变量	均值	方差
居民收入对数	7.44	0.43
性别(女性)(男 =0)	0.48	0.50
教育年限	10.67	3.12
工作年限	19.72	10.25
党员身份(党员)(非党员 =0)	0.24	0.43

(2) 回归分析结果见表2: 类似上述表格修改一下

表 2: 居民收入影响因素的多元线性回归模型回归表 (N = 15862)

变量	模型 1	模型 2	模型 3	模型 4
性别 (男 = 0)	−0.163** (.006)	−0.125** (.006)	−0.114** (.006)	−0.113** (.006)
工作年限	0.044** (.001)	0.046** (.001)	0.044** (.001)	0.046** (.001)
工作年限的平方项	−0.001** (.000)	−0.001** (.000)	−0.001** (.000)	−0.001** (.000)
教育年限	/	0.035** (.001)	0.031** (.001)	0.036** (.001)
党员身份 (非党员 = 0)	/	/	0.071** (.008)	0.230** (.026)
教育年限 × 党员身份	/	/	/	−0.014** (.002)
截距项	7.001** (.011)	6.558** (.017)	6.591** (.017)	6.538** (.019)
R ²	0.2045	0.2575	0.2614	0.2633

注：* : $p < 0.05$, ** : $p < 0.01$. 括号中的数字为标准误. / 表示模型不包含该变量. (教育年限 × 党员身份) 表示教育年限与党员身份的交互项.

(3) 为检验教育年限和党员身份对因变量的影响的联合效应, 考虑使用以下两个模型:

限制性模型: 模型 1 $\text{logearn} = \beta_0 + \beta_1\text{sex} + \beta_2\text{exp} + \beta_3\text{exp}^2 + \varepsilon,$

非限制性模型: 模型 3 $\text{logearn} = \beta_0 + \beta_1\text{sex} + \beta_2\text{exp} + \beta_3\text{exp}^2 + \beta_4\text{edu} + \beta_5\text{cpc} + \varepsilon.$

考虑显著性水平 $\alpha = 0.01$ 下, 原假设与备择假设分别为:

$H_0 : \beta_4 = \beta_5 = 0, \quad H_1 : \beta_4 \neq 0 \vee \beta_5 \neq 0.$

F 值为

$$F = \frac{(SSE_R - SSE_U)/q}{SSE_U/(n - K)} = \frac{(2347.24 - 2179.22)/2}{2179.22/(15862 - 6)} = 611.26 > F_{0.01}(2, 15856) = 4.61.$$

故拒绝原假设. 即在显著性水平 $\alpha = 0.01$ 下, 教育年限和党员身份对因变量的影响具有显著联合效应.

- (4) 斜率系数 β_1 表示取值等于 1 的类别 (虚拟变量) 与参照组之间在因变量上的均值差, 若取值对调, 则差值的符号会变为相反数.

在模型 1 中将男女取值对调后, 性别对应的变量估计的斜率系数会变为原有的相反数, 同时截距项变为新参照组的截距. 具体地, 模型 1 可改写为

$$\text{logearn} = \beta'_0 + \beta'_1 \text{sex} + \beta_2 \text{exp} + \beta_3 \text{exp}^2 + \varepsilon, \text{ 其中 } \beta'_1 = -\beta_1, \beta'_0 = \beta_0 + \beta_1.$$

若带入回归结果, 则有

$$\text{logearn} = 6.838 + 0.163 \text{sex} + 0.044 \text{exp} - 0.001 \text{exp}^2 + \varepsilon.$$

- (5) 根据回归模型 (即模型 4) 的估计结果, 解释如下:

- 对于主效应要使用还未添加交互项的模型 (模型 3) 进行分析!!!, 注意强调显著性水平.
- 教育年限变量的主效应 ($\beta_4 = 0.036$): 在控制其他变量不变的情况下, 教育年限每增加一个单位, 居民收入的对数期望值将增加 0.036 个单位. 说明教育年限对居民收入有 正向 影响.
- 党员身份虚拟变量 (参照组 非党员 = 0, $\beta_5 = 0.230$) 的主效应: 在控制其他变量不变的情况下, 党员收入的对数期望值相较于非党员将增加 0.230 个单位. 说明党员身份对居民收入有 正向 影响.
- 教育年限与党员身份的交互效应: 偏回归系数 $\beta_6 = -0.014$ 刻画了这两个变量对因变量的非线性作用, 表明教育年限对收入的作用和党员身份对收入的作用间存在着 相互削弱 的关系.

