抽样调查：第九周作业

蒋翌坤 20307100013

Exercise 4.8: 18

a

令

$$y_i = \begin{cases} 1, & \text{if 女作家} \\ 0, & \text{if 男作家} \end{cases} \qquad x_i = \begin{cases} 1, & \text{if 男作家} \\ 0, & \text{if 女作家} \end{cases}$$

由 $\hat{B} = \bar{y}_{\text{str}}/\bar{x}_{\text{str}}$ 可得 $\hat{B} = 2.21$

由 $\hat{\mathrm{V}}(\hat{B}) = \frac{1-f}{n\bar{x}^2} s_e^2$ 可得 $\hat{\mathrm{V}}(\hat{B}) = 0.0014$，于是 95% 置信区间为 $\hat{B} \mp 1.96\sqrt{\hat{\mathrm{V}}(\hat{B})} = [2.13, 2.28]$

b

与 a 类似，可得 $\hat{B} = 0.26$，95% 置信区间为 $[-0.68, 1.19]$

2、补充

a

根据总体数据，可以得到 $\bar{x}_{\mathcal{U}}$ 总计均值 $\bar{x}_{\mathcal{U}} = 3.13 \times 10^5$，各层总体均值 $\bar{x}_{h\mathcal{U}}$ 如下表所示：

region	$\bar{x}_{h\mathcal{U}}$
NC	$3.32 \cdot 10^5$
NE	$1 \cdot 10^5$
S	$2.03 \cdot 10^5$
W	$7.36 \cdot 10^5$

b

1. 分别比估计

$$\hat{\bar{y}}_{\mathcal{U}} = \bar{y}_{\text{rs}} = \frac{1}{N}\sum_{h=1}^{H} \hat{B}_h t_{xh} = 310050$$

$$\mathrm{SE}(\bar{y}_{\text{rs}}) = \sqrt{\sum_{h=1}^{H} \frac{W_h^2(1-f_h)}{n_h}\left(s_{yh}^2 - 2\hat{B}_h s_{yxh} + \hat{B}_h^2 s_{xh}^2\right)} = 1810$$

2. 联合比估计

$$\hat{\bar{y}}_{\mathcal{U}} = \bar{y}_{\mathrm{rc}} = \hat{B}\hat{\bar{x}}_{\mathcal{U}} = 309306$$

$$\mathrm{SE}(\bar{y}_{\mathrm{rc}}) = \sqrt{\sum_{h=1}^{H} \frac{W_h^2(1-f_h)}{n_h}(s_{yh}^2 - 2\hat{B}s_{yxh} + \hat{B}^2 s_{xh}^2)} = 1850$$

c

1. 分别回归估计

$$\hat{\bar{y}}_{\mathcal{U}} = \bar{y}_{\mathrm{reg,s}} = \sum_{h=1}^{H} W_h[\bar{y}_h + \hat{B}_{1h}(\bar{x}_{h\mathcal{U}} - \bar{x}_h)] = 310512$$

$$\mathrm{SE}(\bar{y}_{\mathrm{reg,s}}) = \sqrt{\sum_{h=1}^{H} \frac{W_h^2(1-f_h)}{n_h} \frac{n_h-1}{n_h-2} s_{yh}^2 (1-r_h^2)} = 1258$$

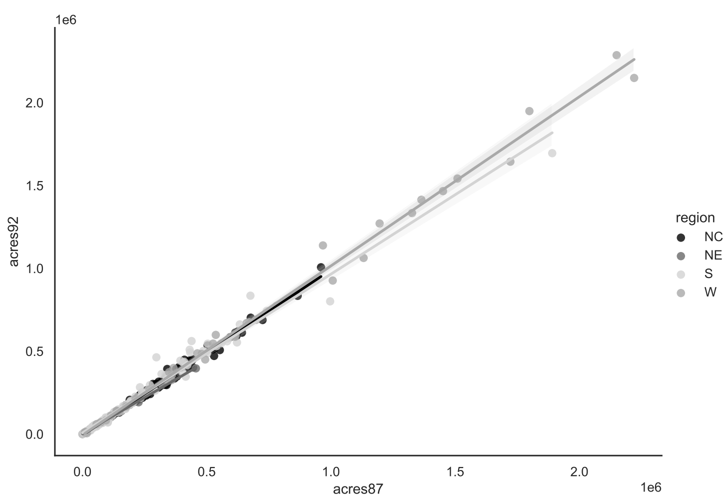
2. 联合回归估计

$$\hat{\bar{y}}_{\mathcal{U}} = \bar{y}_{\mathrm{reg,c}} = \bar{y}_{\mathrm{str}} + \hat{B}_1(\bar{x}_{\mathcal{U}} - \bar{x}_{\mathrm{str}}) = 310015$$

$$\mathrm{SE}(\bar{y}_{\mathrm{reg,c}}) = \sqrt{\sum_{h=1}^{H} \frac{W_h^2(1-f_h)}{n_h}(s_{yh}^2 - 2\hat{B}_1 s_{yxh} + \hat{B}_1^2 s_{xh}^2)} = 1842$$

d

对 acres87 和 acres92 画散点图如下图所示，



acres87 和 acres92 两变量各层间线性关系明显，各层的总体比率都比较接近，因此联合比估计或者联合回归估计更加适用。

3、补充

a

课件 p9 点三条性质可改写为：

1. $\bar{p}$ 是 p 的无偏估计
2. $\mathrm{V}(\bar{p}) = \frac{1-f}{nM^2}S_t^2 = \frac{1-f}{n}\frac{1}{N-1}\sum_{i=1}^{N}(p_i - p)^2$
3. $\mathrm{SE}(\bar{p}) = \frac{1}{M}\sqrt{\frac{1-f}{n}s_t^2} = \sqrt{\frac{1-f}{n}\frac{1}{n-1}\sum_{i\in\mathcal{S}}(p_i - \bar{p})^2}$

b

由 $\mathrm{SSW} = \sum_{i=1}^{N}\sum_{j=1}^{M}(y_{ij} - p_i)^2 = \sum_{i=1}^{N} Mp_i(1-p_i)$ 以及 $\mathrm{SSTO} = \mathrm{SSB} + \mathrm{SSW}$，可得，

$$\mathrm{SSTO} = \sum_{i=1}^{N} M(p_i - p)^2 + \sum_{i=1}^{N} Mp_i(1-p_i) = MNp^2 + (1-p)\sum_{i=1}^{N} p_i$$

$$\mathrm{Deff} = \frac{\mathrm{MSB}}{\mathrm{SSTO}/(NM-1)} = \frac{M(NM-1)\sum_{i=1}^{N}(p_i - p)^2}{MNp^2 + (1-p)\sum_{i=1}^{N} p_i}$$

$$\mathrm{ICC} = \frac{1}{M-1}\Big(\frac{M\cdot\mathrm{MSB}}{\mathrm{SSTO}} - 1\Big) = \frac{1}{M-1}\Big(\frac{M^2\sum_{i=1}^{N}(p_i - p)^2}{MNp^2 + (1-p)\sum_{i=1}^{N} p_i} - 1\Big)$$

Exercise 5.8: 11

a

可以使用刚刚补充题 3 中的结论，令 p 为总体的错误率，p_i 为第 i 个 PSU 的错误率，用 $\bar{p}$ 来估计 p

$$\bar{p} = \frac{1}{n}\sum_{i\in\mathcal{S}} p_i = 2.46\times 10^{-3}$$

$$\mathrm{SE}(\bar{p}) = \frac{1}{M}\sqrt{\frac{1-f}{n}s_t^2} = \sqrt{\frac{1-f}{n}\frac{1}{n-1}\sum_{i\in\mathcal{S}}(p_i - \bar{p})^2} = 3.57\times 10^{-4}$$

b

$$\hat{t} = \frac{N}{n}\sum_{i\in\mathcal{S}} t_i = 360$$

$$\mathrm{SE}(\hat{t}) = NM\mathrm{SE}(\bar{p}) = 63.6$$

c

$$\hat{\mathrm{V}}(\hat{p}_{\mathrm{SRS}})=\frac{1-f}{nM-1}\hat{p}(1-\hat{p})=9.92\times 10^{-8}$$

而（a）中的方差估计 $\hat{\mathrm{V}}(\bar{p})=1.27\times 10^{-7}$，比 $\hat{\mathrm{V}}(\hat{p}_{\mathrm{SRS}})$ 大。

5、补充

a

$s_t^2=\frac{1}{n-1}\sum_{i\in\mathcal{S}}(\bar{y}_{iU}-\bar{y})^2$ 是 S_t^2/M^2 的无偏估计。因此，

$$\mathrm{E}(\mathrm{msb})=\frac{\mathrm{E}(\mathrm{ssb})}{n-1}=\frac{M\mathrm{E}(\sum_{i\in\mathcal{S}}(\bar{y}_{iU}-\bar{y})^2)}{n-1}=\frac{M(n-1)S_t^2}{M^2(n-1)}=\frac{S_t^2}{M}=\mathrm{MSB}$$

由于 $S_i^2=\frac{1}{M-1}\sum_{j=1}^{M}(y_{ij}-\bar{y}_{iU})^2$，所以 $\mathrm{ssw}=\sum_{i\in\mathcal{S}}(M-1)S_i^2$，$\mathrm{SSW}=\sum_{i=1}^{N}(M-1)S_i^2$。因此，$\mathrm{E}(\mathrm{ssw})=\frac{n}{N}\mathrm{SSW}$

$$\mathrm{E}(\mathrm{msw})=\frac{\mathrm{E}(\mathrm{ssw})}{n(M-1)}=\frac{\mathrm{SSW}}{N(M-1)}==\mathrm{MSW}$$

因此，msw, msb 分别是 MSW, MSB 的无偏估计。

b

由于 $\mathrm{SSTO}=\mathrm{SSB}+\mathrm{SSW}$，$S^2=\mathrm{SSTO}/(NM-1)$，而 msw, msb 分别是 MSW, MSB 的无偏估计，所以可以构造总体方差 S^2 的无偏估计 s^2

$$s^2=\frac{(N-1)\mathrm{msb}+N(M-1)\mathrm{msw}}{NM-1}$$

c

$$\begin{aligned}\mathrm{E}(\mathrm{msto})&=\frac{\mathrm{E}(\mathrm{ssb}+\mathrm{ssw})}{nM-1}=\frac{(n-1)\mathrm{MSB}+n(M-1)\mathrm{MSW}}{nM-1}\\&=\frac{\frac{n-1}{N-1}\mathrm{SSB}+\frac{n}{N}\mathrm{SSW}}{nM-1}\approx\frac{n}{N}\frac{\mathrm{SSTO}}{nM-1}\approx\frac{\mathrm{SSTO}}{NM-1}=S^2\end{aligned}$$

附录

解答题目所使用的代码及输出请见：`https://thisiskunmeng.github.io/sampling/hw9.html`