抽样调查 第六周作业

20307100013 蒋翌坤

§3.8 Exercises：6

（a）利用 Neyman 最优配置，可得$n_1 = 504, n_2 = 324, n_3 = 72$

（b）$\widehat{p_{srs}} = \widehat{p_{str}} = \sum_{h=1}^{3} W_h \widehat{p_h} = 0.303$
$\frac{V_{prop}(\widehat{p_{str}})}{V_{SRS}(\widehat{p_{srs}})} = 0.914$

§3.8 Exercises：8

（a）利用 Neyman 最优配置，可得$n_{tel} = 4396, n_{no_tel} = 423$

（b）利用 Neyman 最优配置，可得$n_{tel} = 8249, n_{no_tel} = 458$

§3.8 Exercises：25

由$r = 10\%$，可得$V_{max}(\bar{y}) = \left(\frac{r\overline{y_u}}{1.96}\right)^2 = 3.98 \times 10^7$,再通过遍历$n$的取值，可得满足$V(\bar{y}) \le V_{max}(\bar{y})$的最小$n$为 666，各层的样本量为$[4, 76, 180, 406]$

§3.8 Exercises：27

（a）女、男的样本量为$[405, 495]$

（b）利用 Neyman 最优配置，女、男的样本量为$[618, 482]$

§3.8 Exercises：37

令$\boldsymbol{a} = (a_1, \dots, a_K)'$，$C'(n_1, \dots, n_H) = C(n_1, \dots, n_H) - c_0 = \sum_{h=1}^{H} c_h n_h$，
$V_k(n_1, \dots, n_H) = V\left(\widehat{t_{k,str}}\right) + \frac{1}{N}\sum_{h=1}^{H} N_h^2 S_{kh}^2 = \sum_{h=1}^{H} \frac{N_h^2 S_{kh}^2}{n_h}$
$\boldsymbol{V}(n_1, \dots, n_H) = \left(V_1(n_1, \dots, n_H), \dots, V_K(n_1, \dots, n_H)\right)$
问题等价于：

$$\min w = \boldsymbol{aV}(n_1, \dots, n_H)$$
$$s.t. C'(n_1, \dots, n_H) = C - c_0$$

由 Cauchy-Schwarz 不等式：

$$\boldsymbol{aV}(n_1, \dots, n_H)C'(n_1, \dots, n_H) = \left[\sum_{h=1}^{H}\left(\sqrt{\frac{N_h^2}{n_h}\left(\sum_{k=1}^{K} a_k S_{kh}^2\right)}\right)^2\right] \cdot \left[\sum_{h=1}^{H}\left(\sqrt{c_h n_h}\right)^2\right]$$

$$\ge \sum_{h=1}^{H} \sqrt{\frac{N_h^2}{n_h}\left(\sum_{k=1}^{K} a_k S_{kh}^2\right)} \cdot \sqrt{c_h n_h}$$

等号成立当且仅当

$$\frac{\sqrt{c_h n_h}}{N_h\sqrt{\frac{1}{n_h}(\sum_{k=1}^{K} a_k\, S_{kh}^2)}} = \text{constant } K, h = 1, \dots, H$$

此时，$n_h = K \cdot N_h\sqrt{\frac{(\sum_{k=1}^{K} a_k S_{kh}^2)}{c_h}}$

又，$n = \sum_{h=1}^{H} n_h = \sum_{h=1}^{H} K \cdot N_h\sqrt{\frac{(\sum_{k=1}^{K} a_k S_{kh}^2)}{c_h}}$

得证，

$$n_h = n \cdot \frac{N_h\sqrt{\frac{(\sum_{k=1}^{K} a_k\, S_{kh}^2)}{c_h}}}{\sum_{j=1}^{H} N_j\sqrt{\frac{(\sum_{k=1}^{K} a_k\, S_{kj}^2)}{c_j}}}$$

§3.8 Exercises：50

（a）首先计算出各层的层权$W_h = N_h/N$，然后计算出各层应抽取的样本量$n_h = W_h n$，最后在各层中以简单随机抽样的方式抽取n_h个样本，组合成 150 个分层抽样的样本。

（b）$\widehat{p_{str}} = \sum_{h=1}^{H} W_h \widehat{p_h} = 0.396$，95%置信区间为$[\widehat{p_{str}} \pm 1.96SE(\widehat{p_{str}})] = [0.330, 0.462]$

（c）$\overline{y_{str}} = \sum_{h=1}^{H} W_h \overline{y_h} = 13.85$，95%置信区间为$[\overline{y_{str}} \pm 1.96SE(\overline{y_{str}})] = [13.68,14.02]$

（d）第二章练习 37 中得到的$\hat{p} = 0.44$，$\bar{y} = 13.77$，都比分层抽样所估计的$\widehat{p_{str}}$与$\overline{y_{str}}$小。

（e）可以发现各层 *logsal* 方差最小值为 0.026，最大值为 3.685，差距很大，使用最优配置是有价值的。

（f）利用 Neyman 最优配置以及总样本量为 150，可以得到各层的样本量。为了表示最优配置和比例配置的差距，可以利用$diff = \sum_{h=1}^{H}\left(n_{h,prop} - n_{h,opt}\right)^2$来表示，最终计算得到$diff = 279$

补充：

可以构造以下的例子：$W_1 = 0.4, W_2 = 0.6; S_1 = 2, S_2 = 1; \widetilde{S_1} = 1.5, \widetilde{S_2} = 1.5$

$$V(\overline{y_{str}}) - V_{prop}(\overline{y_{str}}) = \frac{1}{n}\left[\left(\frac{0.4n}{0.5n} - 1\right) \times 0.4 \times 2^2\right] + \frac{1}{n}\left[\left(\frac{0.6n}{0.5n} - 1\right) \times 0.6 \times 1^2\right] = -\frac{1}{n} < 0$$

附录：

解答题目所使用的代码及输出请见：
https://thisiskunmeng.github.io/jupyterlab_lite/retro/notebooks/?path=sampling/hw6.ipynb