抽样调查: 第十周作业

蒋翌坤 20307100013

Exercise 5.8: 1

估计值 $\hat{p}$ 和 $\hat{\mathrm{V}}(\hat{p})$ 不合理。这是因为这个调查是一个群规模不同的单阶整群抽样调查，而题目中估计值的计算是假设了抽样方式为 SRS，而 SRS 和整群抽样方式不同，不能用同一个估计量来估计 p 和 $\mathrm{V}(\hat{p})$。

Exercise 5.8: 4

a

这是一个整群抽样因为

1. 总体（关于社会和行为科学的学术文章）首先被分割成了 1285 个单元（期刊）。
2. 然后对 1285 个期刊进行抽样，抽中了 26 个期刊，接着对这 26 个期刊进行了全面调查。

b

利用比估计，$\hat{\bar{y}}_r = 0.926$, $SE(\hat{\bar{y}}_r) = 3.399 \times 10^{-2}$，即使用 nonprobability sampling 的文章比例有 0.926，标准误为 3.399×10^{-2}

c

估计值为 0.926，且估计的标准误较小，说明估计的偏差不会很大，因此，作者说使用 nonprobability sampling 的文章比例非常大是没有问题的。

Exercise 5.8: 5

a

利用无偏估计，$\hat{t}_{\text{unb}} = 454$，95% 置信区间为 $[234, 673]$

b

利用比估计，$\hat{\bar{y}}_r = 66.80$, 95% 置信区间为 $[63.50, 70.09]$

Exercise 5.8: 6

利用无偏估计，$\hat{t}_{\text{unb}} = 50653$，95% 置信区间为 $[24618, 76689]$

通过 (5.27) 计算可得 $\hat{\mathrm{V}}(\hat{t}_{\text{unb}}) = 1.764 \times 10^8$，而通过 (5.29) 计算可得 $\hat{\mathrm{V}}_{\text{WR}}(\hat{t}_{\text{unb}}) = 1.797 \times 10^8$，两者差距不是很大。

Exercise 5.8: 7

对于 6 个城市的超市，sampling weight 分别是 $39, 35.625, 39.643, 36.563, 30, 35$

利用无偏估计，$\hat{t}_{\text{unb}} = 152972$, $SE(\hat{t}_{\text{unb}}) = 56781.08$

利用比估计，$\hat{y}_r = 120.69$, $SE(\hat{y}_r) = 20.05$

Exercise 5.8: 27

a

$$
\begin{aligned}
\mathrm{V}(\hat{\bar{y}}_{\text{unb}}) &= \frac{1-n/N}{n}\frac{S_t^2}{M^2} + \frac{1-m/M}{nm}\frac{1}{N}\sum_{i=1}^{N} S_i^2 \\
&= \left(1-\frac{n}{N}\right)\frac{\text{MSB}}{nM} + \left(1-\frac{m}{M}\right)\frac{\text{MSW}}{nm}
\end{aligned}
$$

b

利用 R_a^2 的定义：$R_a^2 = 1 - \frac{\text{MSW}}{S^2}$，可得 $\text{MSW} = S^2(1-R_a^2)$

$$
\text{MSB} = \frac{\text{SSB}}{N-1} = \frac{\text{SSTO}-\text{SSW}}{N-1} = \frac{(NM-1)S^2 - N(M-1)S^2(1-R_a^2)}{N-1} = S^2\left[\frac{N(M-1)R_a^2}{N-1} + 1\right]
$$

c

$$
\begin{aligned}
\mathrm{V}(\hat{\bar{y}}) &= \left(1-\frac{n}{N}\right)\frac{\text{MSB}}{nM} + \left(1-\frac{m}{M}\right)\frac{\text{MSW}}{nm} \\
&= \left[\left(1-\frac{n}{N}\right)\frac{1}{nM}\left(\frac{N(M-1)R_a^2}{N-1}+1\right) + \left(1-\frac{m}{M}\right)\frac{1-R_a^2}{nm}\right]S^2
\end{aligned}
$$

d

$$
\frac{\partial}{\partial R_a^2}\mathrm{V}(\hat{\bar{y}}) = S^2\left[\left(1-\frac{n}{N}\right)\frac{N(M-1)}{nM(N-1)} - \left(1-\frac{m}{M}\right)\frac{1}{nm}\right] = S^2\left(\frac{(N-n)(M-1)m-(M-m)(N-1)}{nMm(N-1)}\right)
$$

当 $(m-1)/m > n/N$ 时，$\frac{\partial}{\partial R_a^2}\mathrm{V}(\hat{\bar{y}}) > 0$，因此，$\mathrm{V}(\hat{\bar{y}})$ 随着 R_a^2 的增大而增大。

e

二阶整群抽样的设计效应公式为

$$
\text{Deff} = \frac{\text{MSB}}{S^2} = \left[\frac{N(M-1)R_a^2}{N-1}+1\right]
$$

Exercise 5.8: 28

a

$$\hat{t}_r = M_0 \frac{\hat{t}_{\text{unb}}}{\hat{M}_0} = \frac{NM}{NM}\hat{t}_{\text{unb}} = \hat{t}_{\text{unb}}$$

所以，$\hat{t}_{\text{unb}} = \hat{t}_r$，$\hat{\bar{y}}_{\text{unb}} = \hat{\bar{y}}_r$

b

令 $\hat{\bar{y}} = \hat{\bar{y}}_{\text{unb}} = \hat{\bar{y}}_r$，可以得到

Source	df	Sum of Squares	Mean Squares
Between psus	$n-1$	$\sum_{i\in\mathcal{S}}\sum_{j\in\mathcal{S}_i}(\bar{y}_i - \hat{\bar{y}})^2$	msb
Within psus	$n(m-1)$	$\sum_{i\in\mathcal{S}}\sum_{j\in\mathcal{S}_i}(y_{ij} - \bar{y}_i)^2$	msw
Total	$nm-1$	$\sum_{i\in\mathcal{S}}\sum_{j\in\mathcal{S}_i}(y_{ij} - \hat{\bar{y}})^2$	msto

c

由于 $\frac{1}{m-1}\sum_{j\in\mathcal{S}_i}(y_{ij} - \bar{y}_i)^2$ 是 S_i^2 的无偏估计，于是

$$\text{E}[\text{msw}] = \frac{\text{E}[\text{ssw}]}{n(m-1)} = \frac{1}{n}\frac{\text{E}[\sum_{i\in\mathcal{S}}\sum_{j\in\mathcal{S}_i}(y_{ij} - \bar{y}_i)^2]}{m-1} = \frac{1}{n}\text{E}\Big[\sum_{i\in\mathcal{S}} S_i^2\Big] = \frac{1}{N}\sum_{i=1}^{N} S_i^2 = \text{MSW}$$

由于 $\text{ssb} = m\sum_{i\in\mathcal{S}}(\bar{y}_i - \hat{\bar{y}})^2 = m\sum_{i\in\mathcal{S}}(\bar{y}_i^2 - 2\bar{y}_i\hat{\bar{y}} + \hat{\bar{y}}^2) = m\sum_{i\in\mathcal{S}}(\bar{y}_i^2) - mn\hat{\bar{y}}^2$，于是

$$\begin{aligned}
\text{E}[\text{msb}] = \frac{\text{E}[\text{ssb}]}{n-1} &= \frac{\text{E}[m\sum_{i\in\mathcal{S}}(\bar{y}_i^2) - mn\hat{\bar{y}}^2]}{n-1} = \frac{mn}{n-1}\Big[\frac{1}{N}\sum_{i=1}^{N}(\text{V}(\bar{y}_i) + \text{E}(\bar{y}_i)^2) - (\text{V}(\hat{\bar{y}}) + \text{E}(\hat{\bar{y}})^2)\Big] \\
&= \frac{mn}{n-1}\Big[\frac{1}{N}\sum_{i=1}^{N}\Big(1-\frac{m}{M}\Big)\frac{S_i^2}{m} + \bar{y}_{iU}^2 - \frac{1}{M^2}\Big(1-\frac{n}{N}\Big)\frac{S_t^2}{n} - \frac{1}{nN}\sum_{i=1}^{N}\Big(1-\frac{m}{M}\Big)\frac{S_i^2}{m} - \bar{y}_U^2\Big] \\
&= \frac{1}{N}\Big(1-\frac{m}{M}\Big)\sum_{i=1}^{N} S_i^2 + \frac{mn}{N(n-1)}\sum_{i=1}^{N}(\bar{y}_{iU}^2 - \bar{y}_U^2) - \frac{m}{M^2(n-1)}\Big(1-\frac{n}{N}\Big)S_t^2 \\
&= \Big(1-\frac{m}{M}\Big)\text{MSW} + \frac{mn(N-1)}{MN(n-1)}\text{MSB} + \frac{m}{M(n-1)}\Big(1-\frac{n}{N}\Big)\text{MSB} = \Big(1-\frac{m}{M}\Big)\text{MSW} + \frac{m}{M}\text{MSB}
\end{aligned}$$

d

$$\text{E}[\widehat{\text{MSB}}] = \frac{M}{m}\Big[\Big(1-\frac{m}{M}\Big)\text{MSW} + \frac{m}{M}\text{MSB}\Big] + \Big(\frac{M}{m} - 1\Big)\text{MSW} = \text{MSB}$$

e

$s_t^2 = \frac{M^2}{m}\text{msb}$, $\sum_{i\in\mathcal{S}} s_i^2 = n\cdot\text{msw}$，所以

$$\hat{\mathrm{V}}(\hat{\bar{y}}_{\text{unb}}) = \frac{1}{N^2M^2}\hat{\mathrm{V}}(\hat{t}_{\text{unb}}) = \frac{1}{M^2}\left(1-\frac{n}{N}\right)\frac{s_t^2}{n} + \frac{1}{Nn}\left(1-\frac{m}{M}\right)\sum_{i\in\mathcal{S}}\frac{s_i^2}{m} = \left(1-\frac{n}{N}\right)\frac{\text{msb}}{nm} + \frac{1}{N}\left(1-\frac{m}{M}\right)\frac{\text{msw}}{m}$$

Exercise 5.8: 29

a

$$\begin{aligned}
\mathrm{E}[\text{msto}] &= \frac{1}{nm-1}\mathrm{E}[\text{ssto}] = \frac{1}{nm-1}\mathrm{E}[\text{ssw}+\text{ssb}] = \frac{1}{nm-1}\Big[n(m-1)\mathrm{E}(\text{msw}) + (n-1)\mathrm{E}(\text{msb})\Big] \\
&= \frac{1}{nm-1}\left[n(m-1)\text{MSW} + (n-1)\left(\left(1-\frac{m}{M}\right)\text{MSW} + \frac{m}{M}\text{MSB}\right)\right] \\
&= \frac{nmM+m-M-nm}{M(nm-1)}\text{MSW} + \frac{m(n-1)}{M(nm-1)}\text{MSB}
\end{aligned}$$

b

$$\mathrm{E}[\text{msto}] = \frac{nmM+m-M-nm}{MN(M-1)(nm-1)}\text{SSW} + \frac{m(n-1)}{M(N-1)(nm-1)}\text{SSB} \approx \frac{1}{NM-1}\text{SSTO} = S^2$$

所以当 n 和 N 充分大时，$\mathrm{E}[\text{msto}] \approx S^2$

c

$$\begin{aligned}
\mathrm{E}[\hat{S}^2] &= \frac{M(N-1)}{m(NM-1)}\left(\left(1-\frac{m}{M}\right)\text{MSW} + \frac{m}{M}\text{MSB}\right) + \frac{(m-1)NM+M-m}{m(NM-1)}\text{MSW} \\
&= \frac{1}{NM-1}(\text{MSW}+\text{MSB}) = \frac{1}{NM-1}\text{SSTO} = S^2
\end{aligned}$$

补充

a

通过计算可得如下 ANOVA 表

Source	df	Sum of Squares	Mean Squares
Between psus	11	149.64	13.60
Within psus	24	108.67	4.53
Total	11	258.31	7.38

b

根据 28.(c) 可知 $\widehat{\text{MSW}} = \text{msw} = 4.53$；根据 28.(d) 可知 $\widehat{\text{MSB}} = \frac{M}{m}\text{msb} - \left(\frac{M}{m} - 1\right)\text{msw} = 77.13$；根据 29.(c) 知 $\hat{S}^2$ 的表达式，计算得 $\hat{S}^2 = 7.548$

$$\widehat{\text{ICC}} = \frac{1}{M-1}\left[\frac{(N-1)M \cdot \widehat{\text{MSB}}}{(NM-1)\cdot \hat{S}^2} - 1\right] = 0.400108$$

$$\widehat{R_a^2} = 1 - \frac{\widehat{\text{MSW}}}{\hat{S}^2} = 0.400121$$

c

$$\widehat{\text{Deff}} = \frac{\widehat{\text{MSB}}}{\hat{S}^2} = 10.22$$

附录

解答题目所使用的代码及输出请见：`https://thisiskunmeng.github.io/sampling/hw10.html`