抽样调查：第十二周作业

蒋翌坤 20307100013

Exercise 6.9: 6

a

当概率和总体人口成比例时，$\psi_i, \hat{t}_\psi$ 如下表所示

i	1	2	3	4	5	6	7	8	9	10	11	12	13
ψ_i	0.04	0.08	0.08	0.03	0.02	0.01	0.01	0.12	0.06	0.25	0.03	0.13	0.12
$\hat{t}_\psi(\times 10^5)$	7.67	8.07	7.81	10.47	5.93	7.81	13.15	9.25	8.89	6.83	6.56	8.58	7.38

$\mathrm{V}(\hat{t}_\psi) = 1.36 \times 10^{10}$

b

当概率相等时，$\psi_i, \hat{t}_\psi$ 如下表所示

i	1	2	3	4	5	6	7	8	9	10	11	12	13
ψ_i	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
$\hat{t}_\psi(\times 10^5)$	4.27	7.91	8.51	4.34	1.74	0.58	2.12	14.8	7.51	22.28	2.37	14.98	11.92

$\mathrm{V}(\hat{t}_\psi) = 3.87 \times 10^{11}$。相等概率抽样的方差比成比例抽样的方差大了近 30 倍。成比例抽样更好是因为总体人口和住户数量有很强的正相关性，这意味着 $\hat{t}_\psi$ 比较接近。

c

利用 Lahiri 方法抽取到样本 (2,7,1)，可得 $\hat{t}_\psi = 9.631 \times 10^5, \hat{\mathrm{V}}(\hat{t}_\psi) = 3.113 \times 10^{10}$

Exercise 6.9: 8

样本总和估计 $\hat{t}_\psi = 185$，标准误 SE $= 79$

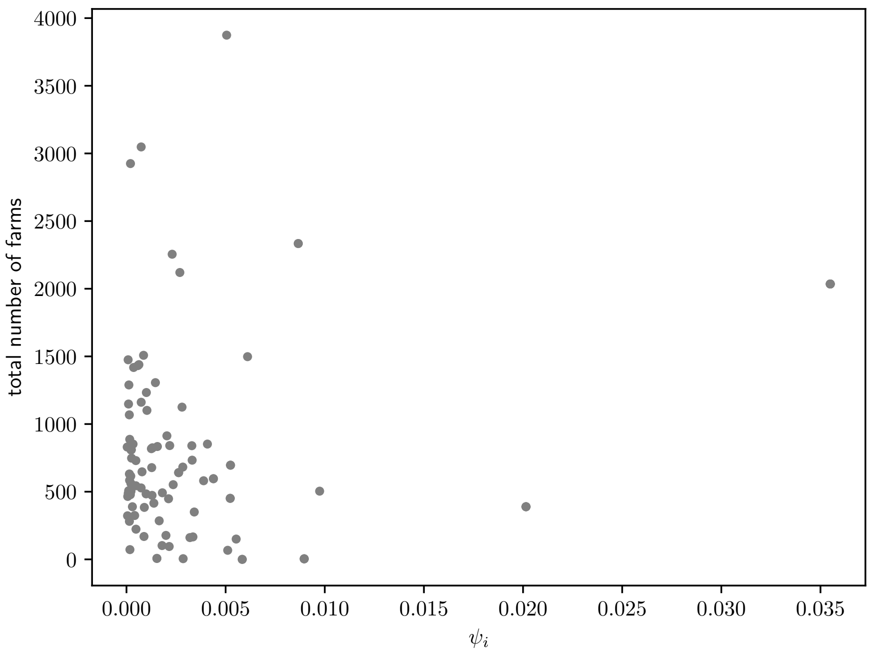
补充

样本均值估计 $\hat{\bar{y}}_\psi = 0.230$，标准误 SE $= 0.098$

Exercise 6.9: 13

a

农场总数和 ψ_i 的图像如下所示。从图中可以看出，农场总数和 ψ_i 并没有明显的相关性，用不等概率抽样的效果不会很好。



b

美国农场总数估计 $\hat{t}_\psi = 1.896 \times 10^6$，95%CI 为 $[1.176 \times 10^6, 2.616 \times 10^6]$

补充

用 SRSWR 方法，设 $N = 3078$，美国农场总数估计 $\hat{t} = 2.486 \times 10^6$，标准误 SE $= 2.211 \times 10^6$。SRSWR 与 HH 相比，得到的估计值大很多，标准误也大了很多。SRSWR 用在不等概率抽样场合中，他的结果由于没有把不同的抽样概率包含进去，因此估计不是无偏的，也是不合理的。

Exercise 6.9: 21

如果 psus 是以 SRS 方式抽取的，则 $\pi_i \equiv \dfrac{n}{N}, \pi_{ij} \equiv \dfrac{n(n-1)}{N(N-1)}$，由于 (6.28) 和 (6.29) 式中有相同项 $\displaystyle\sum_{i\in\mathcal{S}} \frac{\hat{\mathrm{V}}(\hat{t}_i)}{\pi_i}$，所以只看前面的，可得

$$
\begin{aligned}
\frac{1}{2}\sum_{i\in\mathcal{S}}\sum_{k\in\mathcal{S},k\neq i}\frac{\pi_i\pi_k-\pi_{ik}}{\pi_{ik}}\left(\frac{\hat{t}_i}{\pi_i}-\frac{\hat{t}_k}{\pi_k}\right)^2 &= \frac{1}{2}\frac{\pi_1^2-\pi_{12}}{\pi_{12}}\frac{1}{\pi_1^2}\sum_{i\in\mathcal{S}}\sum_{k\in\mathcal{S},k\neq i}\left(\hat{t}_i-\hat{t}_k\right)^2 \\
&= \frac{\pi_1^2-\pi_{12}}{\pi_1^2\pi_{12}}\sum_{i\in\mathcal{S}}(n-1)\hat{t}_i^2 - \frac{\pi_1^2-\pi_{12}}{\pi_1^2\pi_{12}}\sum_{i\in\mathcal{S}}\sum_{k\in\mathcal{S},k\neq i}\hat{t}_i\hat{t}_k \\
&= \frac{1-\pi_1}{\pi_1^2}\sum_{i\in\mathcal{S}}\hat{t}_i^2 + \frac{\pi_{12}-\pi_1^2}{\pi_1^2\pi_{12}}\sum_{i\in\mathcal{S}}\sum_{k\in\mathcal{S},k\neq i}\hat{t}_i\hat{t}_k
\end{aligned}
$$

因此，在 SRS 的情况下，(6.28) 和 (6.29) 式是一样的。而不在 SRS 的情况下，π_i 并不是一个恒定值，所以 (6.28) 和 (6.29) 式会不一样。

Exercise 6.9: 27

a

$$\mathrm{V}\Big(\sum_{i=1}^{N} Z_i\Big) = \sum_{i=1}^{N} \mathrm{V}(Z_i) = \sum_{i=1}^{N} \pi_i(1-\pi_i)$$

b

$$P(\sum_{i=1}^{N} Z_i = 0) = \prod_{i=1}^{N} P(Z_i = 0) = \prod_{i=1}^{N}(1-\pi_i)$$

c

$$P(\sum_{i=1}^{N} Z_i = 0) = \prod_{i=1}^{N}(1-n/N) = (1-n/N)^N \approx e^{-n}$$

d

$$\mathrm{E}\Big[\hat{t}_{\mathrm{HT}}\Big] = \mathrm{E}\Big[\sum_{i=1}^{N} \frac{Z_i y_i}{\pi_i}\Big] = \sum_{i=1}^{N} \frac{\mathrm{E}(Z_i)}{\pi_i} y_i = \sum_{i=1}^{N} \frac{\pi_i}{\pi_i} y_i = \sum_{i=1}^{N} y_i = t$$

e

$$\mathrm{V}\Big(\hat{t}_{\mathrm{HT}}\Big) = \mathrm{V}\Big(\sum_{i=1}^{N} \frac{Z_i y_i}{\pi_i}\Big) = \sum_{i=1}^{N} \frac{\mathrm{V}(Z_i)}{\pi_i^2} y_i^2 = \sum_{i=1}^{N} \frac{\pi_i(1-\pi_i) y_i^2}{\pi_i^2} = \sum_{i=1}^{N} \frac{y_i^2}{\pi_i}(1-\pi_i)$$

f

$$\hat{\mathrm{V}}\Big(\hat{t}_{\mathrm{HT}}\Big) = \sum_{i \in \mathcal{S}} \frac{y_i^2}{\pi_i^2}(1-\pi_i) = \sum_{i=1}^{N} \frac{y_i^2}{\pi_i^2}(1-\pi_i) Z_i$$

$$\mathrm{E}\Big[\hat{\mathrm{V}}\Big(\hat{t}_{\mathrm{HT}}\Big)\Big] = \sum_{i=1}^{N} \frac{y_i^2}{\pi_i^2}(1-\pi_i)\mathrm{E}(Z_i) = \sum_{i=1}^{N} \frac{y_i^2}{\pi_i^2}(1-\pi_i)\pi_i = \sum_{i=1}^{N} \frac{y_i^2}{\pi_i}(1-\pi_i) = \mathrm{V}\Big(\hat{t}_{\mathrm{HT}}\Big)$$

Exercise 6.9: 32

(1)

下表为所有 PSU 的 π_i, π_{ij}

i, j	1	2	3	4	5
1	-	0.068	0.193	0.090	0.049
2	0.068	-	0.148	0.068	0.036
3	0.193	0.148	-	0.193	0.107
4	0.090	0.068	0.193	-	0.049
5	0.049	0.036	0.107	0.049	-
π_i	0.400	0.320	0.640	0.400	0.240

(2)

$$\mathrm{V}(\hat{\bar{y}}_{\mathrm{HT}}) = \frac{1}{M_0^2}\mathrm{V}(\hat{t}_{\mathrm{HT}}) = \frac{1}{2M_0^2}\sum_{i=1}^{N}\sum_{k=1,k\neq i}^{N}(\pi_i\pi_k - \pi_{ik})\Big(\frac{t_i}{\pi_i} - \frac{t_k}{\pi_k}\Big)^2 \approx 0.389$$

(3)

a

$$\mathrm{V}(\hat{\bar{y}}_{\mathrm{HT}}) = \frac{1}{M_0^2}\mathrm{V}(\hat{t}_{\mathrm{HT}}) = \frac{1}{M_0^2}\Big[\frac{1}{2}\sum_{i=1}^{N}\sum_{k=1,k\neq i}^{N}(\pi_i\pi_k - \pi_{ik})\Big(\frac{t_i}{\pi_i} - \frac{t_k}{\pi_k}\Big)^2 + \sum_{i=1}^{N}\frac{\mathrm{V}(\hat{t}_i)}{\pi_i}\Big] \approx 1.093$$

b

$$\hat{t}_3 = 48, \hat{t}_4 = 20，\bar{y}_U = \hat{\bar{y}}_{\mathrm{HT}} = \frac{1}{M_0}\sum_{i\in\mathcal{S}}\frac{\hat{t}_i}{\pi_i} = \frac{1}{25}\times\Big(\frac{48}{0.64} + \frac{20}{0.4}\Big) = 5$$

$\hat{\mathrm{V}}_{\mathrm{HT}}(\hat{\bar{y}}_{\mathrm{HT}}) = 3.76, \hat{\mathrm{V}}_{\mathrm{SYG}}(\hat{\bar{y}}_{\mathrm{HT}}) = 0.40, \hat{\mathrm{V}}_{\mathrm{WR}}(\hat{\bar{y}}_{\mathrm{HT}}) = 1.00$，可得这三种方差估计量计算所得的标准误为 $\mathrm{SE}_{\mathrm{HT}}(\hat{\bar{y}}_{\mathrm{HT}}) = 1.94, \mathrm{SE}_{\mathrm{SYG}}(\hat{\bar{y}}_{\mathrm{HT}}) = 0.64, \mathrm{SE}_{\mathrm{WR}}(\hat{\bar{y}}_{\mathrm{HT}}) = 1.00$

补充题

(1)

令 Z_{ij} 表示第 i 个 PSU 中第 j 个 SSU 被抽中的次数，则有

$$\hat{t}_i = \frac{M_i}{m_i}\sum_{j=1}^{M_i} Z_{ij}y_{ij}, \mathrm{E}(Z_{ij}) = \mathrm{E}\Big(\mathrm{E}(Z_{ij}|Q_i)\Big) = \mathrm{E}\Big(\frac{Q_i v}{M_i}\Big) = \frac{nv\psi_i}{M_i}$$

$$\mathrm{E}(\hat{t}) = \frac{1}{n}\sum_{i=1}^{N}\frac{1}{\psi_i}\mathrm{E}(\frac{M_i}{v}\sum_{j=1}^{M_i}Z_{ij}y_{ij}) = \frac{1}{n}\sum_{i=1}^{N}\frac{1}{\psi_i}\frac{M_i}{v}\sum_{j=1}^{M_i}y_{ij}\frac{nv\psi_i}{M_i} = \sum_{i=1}^{N}t_i = t$$

(2)

$$\begin{aligned}\mathrm{V}(Z_{ij}) &= \mathrm{E}\Big(\mathrm{V}(Z_{ij}|Q_i)\Big) + \mathrm{V}\Big(\mathrm{E}(Z_{ij}|Q_i)\Big) = \mathrm{E}\Big[\frac{Q_i v}{M_i}\Big(1-\frac{Q_i v}{M_i}\Big)\Big] + \mathrm{V}\Big(\frac{Q_i v}{M_i}\Big) \\ &= \frac{v}{M_i}\mathrm{E}(Q_i) - \frac{v^2}{M_i^2}\mathrm{E}(Q_i^2) + \frac{v^2}{M_i^2}\mathrm{V}(Q_i) = \frac{nv\psi_i}{M_i} - \frac{n^2v^2\psi_i^2}{M_i^2}\end{aligned}$$

$$\begin{aligned}\mathrm{V}(\hat{t}) &= \frac{1}{n^2}\sum_{i=1}^{N}\frac{1}{\psi_i^2}\mathrm{V}(Q_i\hat{t}_i) = \frac{1}{n^2}\sum_{i=1}^{N}\frac{1}{\psi_i^2}\frac{M_i^2}{v^2}\sum_{j=1}^{M_i}y_{ij}^2\mathrm{V}(Z_{ij}) \\ &= \frac{1}{n}\Big[\sum_{i=1}^{N}\frac{M_i}{\psi_i v}\sum_{j=1}^{M_i}y_{ij}^2 - n\sum_{i=1}^{N}\sum_{j=1}^{M_i}y_{ij}^2\Big]\end{aligned}$$

而

$$\begin{aligned}\frac{1}{n}\Big[\sum_{i=1}^{N}\psi_i\Big(\frac{t_i}{\psi_i}-t\Big)^2 + \sum_{i=1}^{N}\frac{M_i^2}{\psi_i v}\sigma_{2i}^2\Big] &= \frac{1}{n}\Big[\sum_{i=1}^{N}\frac{t_i^2}{\psi_i} - t^2 + \sum_{i=1}^{N}\frac{M_i}{\psi_i v}\sum_{j=1}^{M_i}y_{ij}^2 - \sum_{i=1}^{N}\frac{t_i^2}{\psi_i v}\Big] \\ &= \frac{1}{n}\Big[\sum_{i=1}^{N}\frac{M_i}{\psi_i v}\sum_{j=1}^{M_i}y_{ij}^2 - n\sum_{i=1}^{N}\sum_{j=1}^{M_i}y_{ij}^2\Big]\end{aligned}$$

因此

$$\mathrm{V}(\hat{t}) = \frac{1}{n}\Big[\sum_{i=1}^{N}\psi_i\Big(\frac{t_i}{\psi_i}-t\Big)^2 + \sum_{i=1}^{N}\frac{M_i^2}{\psi_i v}\sigma_{2i}^2\Big]$$

附录

解答题目所使用的代码及输出请见: `https://thisiskunmeng.github.io/sampling/hw12.html`