抽样调查：第十周作业

蒋翌坤 20307100013

Exercise 5.8: 13

a

这里可以使用整群抽样因为

1. 县里的认证低收入住户可以作为初级单元进行抽样。
2. 由于对每个县里的住户进行抽样和对住户内一个人进行抽样的成本是一样的，所以可以对次级单元进行全面调查而不会增加调查费用。
3. 在这里，使用的是单阶整群抽样。

b

如果利用 SRS 对县里的低收入人群进行抽样，要达到不超过 0.03 的绝对误差，$n_0 = 1.96^2 \times 0.1 \times (1-0.1)/0.03^2 = 385$，需要的样本量为 $n_{\text{SRS}} = n_0/(1+n_0/N) = 310$，而使用整群抽样，$n = n_{SRS}\frac{\text{V}(\hat{t}_{\text{cluster}})}{\hat{t}_{\text{SRS}}}\frac{1}{M} \approx 310 \times (\frac{1}{M} + \text{ICC})$

Exercise 5.8: 14

a

利用比估计，区域内抽烟的女高中生比例估计为 0.43, 95% 置信区间为 $[0.24, 0.62]$。

b

利用无偏估计，区域内抽烟的女高中生总人数估计为 7971, 95% 置信区间为 $[7688, 8254]$。

c

通过计算，$msw = 0.22, s^2 = 0.25$，$R_a^2 = 1 - \frac{msw}{s^2}$ 的估计值为 0.10，于是，

$$m_{\text{opt}} \approx \sqrt{(\frac{c_1}{c_2}\frac{1-R_a^2}{R_a^2})} \approx 26 \qquad n_{\text{opt}} = \frac{C}{c_1 + c_2 m_{\text{opt}}} \approx 4$$

Exercise 5.8: 19

利用无偏估计，学生总人数估计为 90680, 95% 置信区间为 $[31950, 149411]$

Exercise 5.8: 22

a

通过计算可得如下 ANOVA 表

Source	df	Sum of Squares	Mean Squares
Between psus	41	2.00×10^{13}	4.88×10^{11}
Within psus	258	1.55×10^{13}	5.99×10^{10}
Total	299	258.31	1.19×10^{11}

$R_a^2 \approx 1 - \mathrm{msw}/s^2 \approx 0.50$，由于 $R_a^2 > 0$，所以有整群效应。

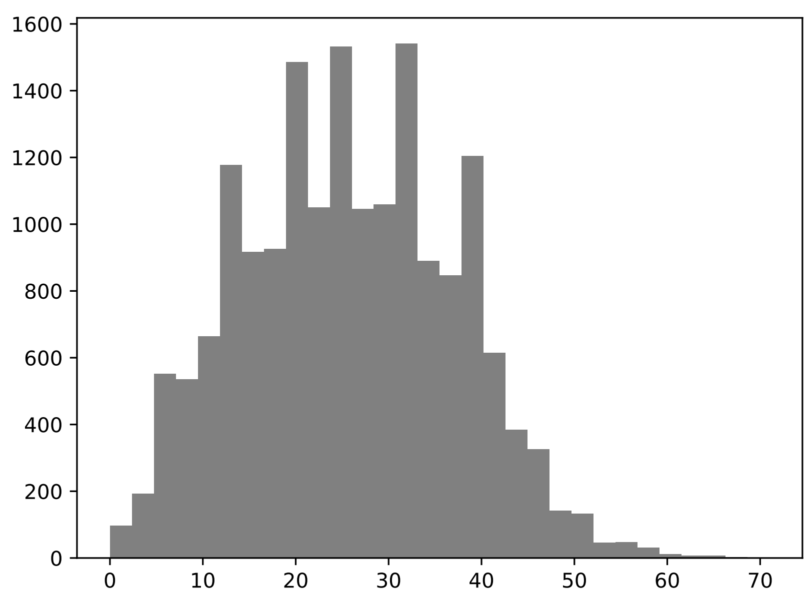
b

$$m_{\mathrm{opt}} \approx \sqrt{\left(\frac{c_1}{c_2}\frac{1-R_a^2}{R_a^2}\right)} \approx 4 \qquad n_{\mathrm{opt}} = \frac{300}{4} = 75$$

Exercise 5.8: 24

a

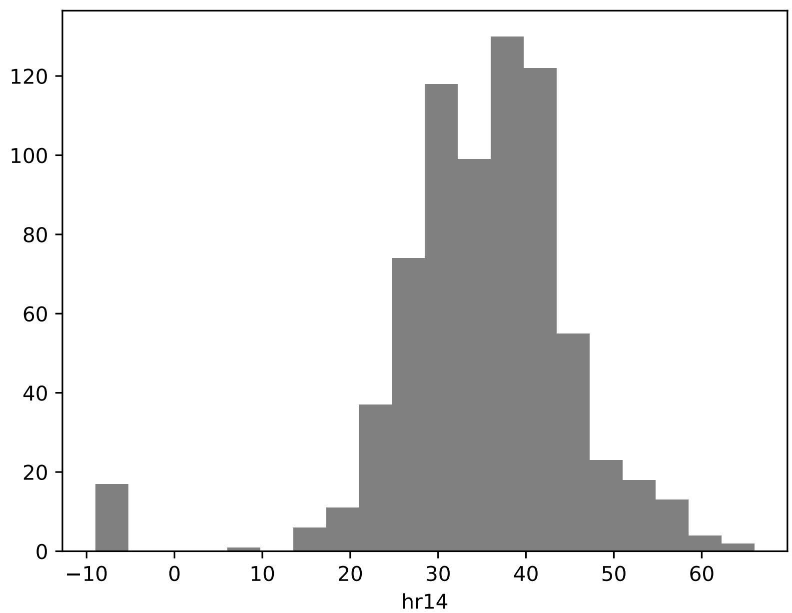
总体直方图如下所示



总体均值为 25.78，标准差 11.37，中位数 26

b

随机选择到整数 14，即选取第 14 个小时的臭氧读数，其直方图如下所示



c

样本均值 $\bar{y}=34.96$，标准差 10.73，中位数 36，样本均值的 95% 置信区间为 $[34.2,35.7]$，此置信区间不包含总体均值。

d

随机选择 0 至 95 中的 4 个数，得 [92611586]，将总体看成由 96 个初级单元构成，每个初级单元有 183 个次级单元（臭氧读数），根据给出的数据，可以计算 $\hat{y}=24.49$, 95% 置信区间为 $[17.11,31.88]$，此置信区间包含总体均值。

补充

a

$$\mathrm{E}(s_t^2)=\mathrm{E}_1\Big(\mathrm{E}_2(s_t^2)\Big)=S_t^2+\frac{1}{N}\sum_{i=1}^{N}\Big(1-\frac{m_i}{M_i}\Big)M_i^2\frac{S_i^2}{m_i}=S_t^2+\frac{1}{N}\sum_{i=1}^{N}\mathrm{V}(\hat{t}_i)$$

b

$$\mathrm{E}\Big(\hat{\mathrm{V}}_{\mathrm{WR}}(\hat{t}_{\mathrm{unb}})\Big) = \frac{N^2}{n}\mathrm{E}(s_t^2) = \frac{N^2}{n}\Big[S_t^2 + \frac{1}{N}\sum_{i=1}^{N}\Big(1-\frac{m_i}{M_i}\Big)M_i^2\frac{S_i^2}{m_i}\Big] = \frac{N^2}{n}S_t^2 + \frac{N}{n}\sum_{i=1}^{N}\Big(1-\frac{m_i}{M_i}\Big)M_i^2\frac{S_i^2}{m_i}$$

$$\mathrm{Bias} = \mathrm{E}\Big(\hat{\mathrm{V}}_{\mathrm{WR}}(\hat{t}_{\mathrm{unb}})\Big) - \mathrm{V}(\hat{t}_{\mathrm{unb}}) = \frac{N^2}{n}S_t^2 - N^2\Big(1-\frac{n}{N}\Big)\frac{S_t^2}{n} = NS_t^2$$

c

$$\mathrm{Bias} = \mathrm{E}\Big(\hat{\mathrm{V}}_{\mathrm{WR}}(\hat{\bar{y}}_{\mathrm{unb}})\Big) - \mathrm{V}(\hat{\bar{y}}_{\mathrm{unb}}) = \frac{N}{M_0^2}S_t^2$$

附录

解答题目所使用的代码及输出请见：`https://thisiskunmeng.github.io/sampling/hw11.html`