Stat 257: Solutions to Assignment #4

(4.1) From the data given in the problem, we have $N = 235$ public libraries which are grouped into three strata: strata 1 consists of $N_1 = 74$ small libraries, strata 2 consists of $N_2 = 133$ medium libraries, and strata 3 consists of $N_3 = 28$ large libraries. Let $\overline{Y}_i$ denote the (sub-)population proportion of the number of books borrowed from libraries in the ith strata so that $\overline{Y}_1 = 1/6$, $\overline{Y}_2 = 2/6$, and $\overline{Y}_3 = 3/6$. Furthemore, if the (sub-)population variance for the number of books borrowed from libraries in the ith strata is proportional to the square root of the corresponding mean, then $S_i^2 = \alpha\sqrt{\overline{Y}_i}$ where α is the constant of proportionality. Hence,

$$S_1^2 = \alpha\sqrt{\overline{Y}_1} = \alpha\sqrt{\frac{1}{6}}, \quad S_2^2 = \alpha\sqrt{\overline{Y}_2} = \alpha\sqrt{\frac{2}{6}}, \quad S_3^2 = \alpha\sqrt{\overline{Y}_3} = \alpha\sqrt{\frac{3}{6}}.$$

Since sampling costs are the same per strata, we have $c_1 = c_2 = c_3$. Since we require a sample size of about $n = 50$, we find that the optimum allocation has

$$n_i = \frac{W_i S_i}{\sum_{i=1}^{k} W_i S_i} \cdot n.$$

It now follows that

$$\sum_{i=1}^{3} W_i S_i = \frac{74}{235} \cdot \sqrt{\alpha\sqrt{\frac{1}{6}}} + \frac{133}{235} \cdot \sqrt{\alpha\sqrt{\frac{2}{6}}} + \frac{28}{235} \cdot \sqrt{\alpha\sqrt{\frac{3}{6}}} = \frac{\sqrt{\alpha}}{235\sqrt{\sqrt{6}}}\left(74 + 133\sqrt{\sqrt{2}} + 28\sqrt{\sqrt{3}}\right)$$

so that

$$n_1 = \frac{\frac{74}{235} \cdot \sqrt{\alpha\sqrt{\frac{1}{6}}}}{\frac{\sqrt{\alpha}}{235\sqrt{\sqrt{6}}}\left(74 + 133\sqrt{\sqrt{2}} + 28\sqrt{\sqrt{3}}\right)} \cdot 50 = \frac{74 \cdot 50}{74 + 133\sqrt{\sqrt{2}} + 28\sqrt{\sqrt{3}}} \approx 13.75,$$

$$n_2 = \frac{\frac{133}{235} \cdot \sqrt{\alpha\sqrt{\frac{2}{6}}}}{\frac{\sqrt{\alpha}}{235\sqrt{\sqrt{6}}}\left(74 + 133\sqrt{\sqrt{2}} + 28\sqrt{\sqrt{3}}\right)} \cdot 50 = \frac{133\sqrt{\sqrt{2}} \cdot 50}{74 + 133\sqrt{\sqrt{2}} + 28\sqrt{\sqrt{3}}} \approx 29.39,$$

$$n_3 = \frac{\frac{28}{235} \cdot \sqrt{\alpha\sqrt{\frac{3}{6}}}}{\frac{\sqrt{\alpha}}{235\sqrt{\sqrt{6}}}\left(74 + 133\sqrt{\sqrt{2}} + 28\sqrt{\sqrt{3}}\right)} \cdot 50 = \frac{28\sqrt{\sqrt{3}} \cdot 50}{74 + 133\sqrt{\sqrt{2}} + 28\sqrt{\sqrt{3}}} \approx 6.84.$$

Thus, the optimum allocation requires that we sample $n_1 = 14$ of the small libraries, $n_2 = 29$ of the medium libraries, and $n_3 = 7$ of the large libraries. (Note that $n_1 + n_2 + n_3 = 14 + 29 + 7 = 50$).

(4.2) We begin by finding the overall mean $\overline{Y}$ which is given by

$$\overline{Y} = \sum_{i=1}^{k} \frac{N_i}{N} \cdot \overline{Y}_i = \frac{117}{375} \cdot 7.3 + \frac{98}{375} \cdot 6.9 + \frac{74}{375} \cdot 11.2 + \frac{41}{375} \cdot 9.1 + \frac{45}{375} \cdot 9.6 \approx 8.438,$$

and the overall sample variance which is given by

$$S^2 = \frac{1}{N-1}\left(\sum_{i=1}^{k}(N_i-1)S_i^2 + \sum_{i=1}^{k} N_i\left(\overline{Y}_i - \overline{Y}\right)^2\right)$$

$$= \frac{1}{374}\Big(116\cdot 1.31 + 97\cdot 2.03 + 73\cdot 1.13 + 40\cdot 1.96 + 44\cdot 1.74$$

$$+\,117(7.3-8.438)^2 + 98(6.9-8.438)^2 + 74(11.2-8.438)^2 + 41(9.1-8.438)^2 + 45(9.6-8.438)^2\Big)$$

$$\approx 4.313.$$

For proportional allocation we simply require that $N_i/N = n_i/n$. Thus,

$$n_1 = \frac{N_1}{N}\cdot n = \frac{117}{375}\cdot 80 \approx 25,\quad n_2 = \frac{N_2}{N}\cdot n = \frac{98}{375}\cdot 80 \approx 21,\quad n_3 = \frac{N_3}{N}\cdot n = \frac{74}{375}\cdot 80 \approx 16,$$

$$n_4 = \frac{N_4}{N}\cdot n = \frac{41}{375}\cdot 80 \approx 9,\quad n_5 = \frac{N_5}{N}\cdot n = \frac{45}{375}\cdot 80 \approx 10.$$

Note that $n_1+n_2+n_3+n_4+n_5 = 81$. This is acceptable since we are allowed a stratified random sample size of "about 80."

For the Neyman allocation we require that

$$n_i = \frac{W_i S_i}{\sum_{i=1}^{k} W_i S_i}\cdot n.$$

It now follows that

$$\sum_{i=1}^{5} W_i S_i = \frac{117}{375}\cdot\sqrt{1.31} + \frac{98}{375}\cdot\sqrt{2.03} + \frac{74}{375}\cdot\sqrt{1.13} + \frac{41}{375}\cdot\sqrt{1.96} + \frac{45}{375}\cdot\sqrt{1.74} \approx 1.251$$

so that

$$n_1 \approx \frac{\frac{117}{375}\cdot\sqrt{1.31}}{1.251}\cdot 80 \approx 23,\quad n_2 \approx \frac{\frac{98}{375}\cdot\sqrt{2.03}}{1.251}\cdot 80 \approx 24,\quad n_3 \approx \frac{\frac{74}{375}\cdot\sqrt{1.13}}{1.251}\cdot 80 \approx 13,$$

$$n_4 \approx \frac{\frac{41}{375}\cdot\sqrt{1.96}}{1.251}\cdot 80 \approx 10,\quad n_5 \approx \frac{\frac{45}{375}\cdot\sqrt{1.74}}{1.251}\cdot 80 \approx 10.$$

Note that for the Neyman allocation, in contrast to the proportional allocation, we find $n_1+n_2+n_3+n_4+n_5 = 80$.

Since the proportional allocation, and the Neyman allocation each give different total sample sizes, we need to make two separate calculations for the variance of the simple random sampling mean. If $n = 80$, then

$$\text{var}(\overline{y}) = \frac{(1-f)}{n}\cdot S^2 \approx \frac{(1-80/375)}{80}\cdot 4.313 \approx 0.0424,$$

while if $n = 81$ then

$$\text{var}(\overline{y}) = \frac{(1-f)}{n}\cdot S^2 \approx \frac{(1-81/375)}{81}\cdot 4.313 \approx 0.0417.$$

Note that since we actually know the population variance S^2 we do not need to estimate $\text{var}(\overline{y})$ by $s^2(\overline{y})$.

For stratified random sampling, we find that

$$\text{var}(\overline{y}_{ST}) = \sum_{i=1}^{k} W_i^2 \frac{(1-f_i)}{n_i} S_i^2.$$

Note that this variance depends on the sample size of each stratum. This means the formula will be different for proportional allocation and Neyman allocation.

For proportional allocation we have

$$\begin{aligned}
\text{var}(\overline{y}_{ST}) &= \sum_{i=1}^{k} W_i^2 \cdot \frac{(1-f_i)}{n_i} \cdot S_i^2 \\
&= \left(\frac{117}{375}\right)^2 \cdot \frac{(1-25/117)}{25} \cdot 1.31 + \left(\frac{98}{375}\right)^2 \cdot \frac{(1-21/98)}{21} \cdot 2.03 + \left(\frac{74}{375}\right)^2 \cdot \frac{(1-16/74)}{16} \cdot 1.13 \\
&\quad + \left(\frac{41}{375}\right)^2 \cdot \frac{(1-9/41)}{9} \cdot 1.96 + \left(\frac{45}{375}\right)^2 \cdot \frac{(1-10/45)}{10} \cdot 1.74 \\
&\approx 0.0153
\end{aligned}$$

and for Neyman allocation we have

$$\begin{aligned}
\text{var}(\overline{y}_{ST}) &= \sum_{i=1}^{k} W_i^2 \cdot \frac{(1-f_i)}{n_i} \cdot S_i^2 \\
&= \left(\frac{117}{375}\right)^2 \cdot \frac{(1-23/117)}{23} \cdot 1.31 + \left(\frac{98}{375}\right)^2 \cdot \frac{(1-24/98)}{24} \cdot 2.03 + \left(\frac{74}{375}\right)^2 \cdot \frac{(1-13/74)}{13} \cdot 1.13 \\
&\quad + \left(\frac{41}{375}\right)^2 \cdot \frac{(1-10/41)}{10} \cdot 1.96 + \left(\frac{45}{375}\right)^2 \cdot \frac{(1-10/45)}{10} \cdot 1.74 \\
&\approx 0.0153
\end{aligned}$$

Thus, the relative efficiency of the simple random sampling mean relative to the Neyman allocation stratified random sampling mean is

$$\frac{0.0153}{0.0424} \approx 36.1\%$$

and the relative efficiency of the simple random sampling mean relative to the proportional allocation stratified random sampling mean is

$$\frac{0.0153}{0.0417} \approx 36.7\%.$$

(4.5) For proportional allocation we simply require that $N_i/N = n_i/n$. Thus,

$$n_1 = \frac{N_1}{N} \cdot n = \frac{368}{1908} \cdot 100 \approx 19.3, \quad n_2 = \frac{N_2}{N} \cdot n = \frac{425}{1908} \cdot 100 \approx 22.3,$$

$$n_3 = \frac{N_3}{N} \cdot n = \frac{389}{1908} \cdot 100 \approx 20.4, \quad n_4 = \frac{N_4}{N} \cdot n = \frac{316}{1908} \cdot 100 \approx 16.6,$$

$$n_5 = \frac{N_5}{N} \cdot n = \frac{174}{1908} \cdot 100 \approx 9.1, \; n_6 = \frac{N_6}{N} \cdot n = \frac{98}{1908} \cdot 100 \approx 5.1, \; n_7 = \frac{N_7}{N} \cdot n = \frac{138}{1908} \cdot 100 \approx 7.2$$

Note that if we round each n_i, then $n = n_1 + \cdots + n_7 = 99$. However, we are required to have $n = 100$. Therefore, we round n_3 up to $n_3 = 21$ since of all the sample sizes to be rounded down, n_3 had the largest fractional part. In summary, for proportional allocation take

$$n_1 = 19, \quad n_2 = 22, \quad n_3 = 21, \quad n_4 = 17, \quad n_5 = 9, \quad n_6 = 5, \quad n_7 = 7.$$

For the Neyman allocation we require that

$$n_i = \frac{W_i S_i}{\sum_{i=1}^{k} W_i S_i} \cdot n.$$

It now follows that

$$\sum_{i=1}^{7} W_i S_i = \frac{368}{1908} \cdot 2.1 + \frac{425}{1908} \cdot 3.6 + \frac{389}{1908} \cdot 3.9 + \frac{316}{1908} \cdot 5.1 + \frac{174}{1908} \cdot 6.1 + \frac{98}{1908} \cdot 6.5 + \frac{138}{1908} \cdot 9.1 \approx 4.395.$$

so that

$$n_1 \approx \frac{\frac{368}{1908} \cdot 2.1}{4.395} \cdot 100 \approx 9.2, \quad n_2 \approx \frac{\frac{425}{1908} \cdot 3.6}{4.395} \cdot 100 \approx 18.2, \quad n_3 \approx \frac{\frac{389}{1908} \cdot 3.9}{4.395} \cdot 100 \approx 18.1,$$

$$n_4 \approx \frac{\frac{316}{1908} \cdot 5.1}{4.395} \cdot 100 \approx 19.2, \quad n_5 \approx \frac{\frac{174}{1908} \cdot 6.1}{4.395} \cdot 100 \approx 12.7,$$

$$n_6 \approx \frac{\frac{98}{1908} \cdot 6.5}{4.395} \cdot 100 \approx 7.6, \quad n_7 \approx \frac{\frac{138}{1908} \cdot 9.1}{4.395} \cdot 100 \approx 15.0.$$

If this case, if we round each n_i, then $n_1 + \cdots + n_7 = 100$. In summary, for Neyman allocation take

$$n_1 = 9, \quad n_2 = 18, \quad n_3 = 18, \quad n_4 = 19, \quad n_5 = 13, \quad n_6 = 8, \quad n_7 = 15.$$

For stratified random sampling, we find that

$$\text{var}(\overline{y}_{ST}) = \sum_{i=1}^{k} W_i^2 \frac{(1 - f_i)}{n_i} S_i^2.$$

Note that this variance depends on the sample size of each stratum. This means the formula will be different for proportional allocation and Neyman allocation.

For proportional allocation we have

$$\begin{aligned}
\text{var}(\overline{y}_{ST}) &= \sum_{i=1}^{k} W_i^2 \cdot \frac{(1 - f_i)}{n_i} \cdot S_i^2 \\
&= \left(\frac{368}{1908}\right)^2 \cdot \frac{(1 - 19/368)}{19} \cdot 2.1^2 + \left(\frac{425}{1908}\right)^2 \cdot \frac{(1 - 22/425)}{22} \cdot 3.6^2 + \left(\frac{389}{1908}\right)^2 \cdot \frac{(1 - 21/389)}{21} \cdot 3.9^2 \\
&\quad + \left(\frac{316}{1908}\right)^2 \cdot \frac{(1 - 17/316)}{17} \cdot 5.1^2 + \left(\frac{174}{1908}\right)^2 \cdot \frac{(1 - 9/174)}{9} \cdot 6.1^2 \\
&\quad + \left(\frac{98}{1908}\right)^2 \cdot \frac{(1 - 5/98)}{5} \cdot 6.5^2 + \left(\frac{138}{1908}\right)^2 \cdot \frac{(1 - 7/138)}{7} \cdot 9.1^2 \\
&\approx 0.2166
\end{aligned}$$

and for Neyman allocation we have

$$
\begin{aligned}
\text{var}(\overline{y}_{ST}) &= \sum_{i=1}^{k} W_i^2 \cdot \frac{(1-f_i)}{n_i} \cdot S_i^2 \\
&= \left(\frac{368}{1908}\right)^2 \cdot \frac{(1-9/368)}{9} \cdot 2.1^2 + \left(\frac{425}{1908}\right)^2 \cdot \frac{(1-18/425)}{18} \cdot 3.6^2 + \left(\frac{389}{1908}\right)^2 \cdot \frac{(1-18/389)}{18} \cdot 3.9^2 \\
&\quad + \left(\frac{316}{1908}\right)^2 \cdot \frac{(1-19/316)}{19} \cdot 5.1^2 + \left(\frac{174}{1908}\right)^2 \cdot \frac{(1-13/174)}{13} \cdot 6.1^2 \\
&\quad + \left(\frac{98}{1908}\right)^2 \cdot \frac{(1-8/98)}{8} \cdot 6.5^2 + \left(\frac{138}{1908}\right)^2 \cdot \frac{(1-15/138)}{15} \cdot 9.1^2 \\
&\approx 0.1813.
\end{aligned}
$$

Thus, the relative efficiency of the proportional allocation stratified random sampling mean relative to the Neyman allocation stratified random sampling mean is

$$\frac{0.1813}{0.2166} \approx 83.7\%.$$