Stat 257.01 Fall 2004
Assignment #5 Solutions

(5.1) We begin by computing $\hat{p}_i$ from the given data:

$$\hat{p}_1 = \frac{4}{14} \approx 0.286 \qquad \hat{p}_2 = \frac{2}{9} \approx 0.222 \qquad \hat{p}_3 = \frac{8}{21} \approx 0.381 \qquad \hat{p}_4 = \frac{1}{6} \approx 0.167.$$

We can now use equation (5.13) to determine $\hat{p}_{\text{ST}}$ as our estimator of p, the proportion of deliquent accounts for the chain. Thus,

$$\begin{aligned}
\hat{p}_{\text{ST}} &= \frac{1}{N}\left(N_1\hat{p}_1 + N_2\hat{p}_2 + N_3\hat{p}_3 + N_4\hat{p}_4\right)\\
&\approx \frac{1}{(65+42+93+25)}\left(65\cdot 0.286 + 42\cdot 0.222 + 93\cdot 0.381 + 25\cdot 0.167\right)\\
&\approx 0.30.
\end{aligned}$$

Using equation (5.14), we find the estimated variance of $\hat{p}_{\text{ST}}$ to be

$$\begin{aligned}
\hat{V}(\hat{p}_{\text{ST}}) &= \frac{1}{N^2}\sum_{i=1}^{4} N_i^2 \left(\frac{N_i - n_i}{N_i}\right)\left(\frac{\hat{p}_i\hat{q}_i}{n_i - 1}\right)\\
&= \frac{1}{225^2}\left[65^2\left(\frac{65-14}{65}\right)\left(\frac{0.286\cdot 0.714}{14-1}\right) + 42^2\left(\frac{42-9}{42}\right)\left(\frac{0.222\cdot 0.778}{9-1}\right)\right.\\
&\qquad \left. + 93^2\left(\frac{93-21}{93}\right)\left(\frac{0.381\cdot 0.619}{21-1}\right) + 25^2\left(\frac{25-6}{25}\right)\left(\frac{0.167\cdot 0.833}{6-1}\right)\right]\\
&\approx 0.0034397.
\end{aligned}$$

Thus, the error of estimation is

$$2\sqrt{\hat{V}(\hat{p}_{\text{ST}})} \approx 0.117.$$

(5.2) Recall that the Neyman allocation is used if there is no difference in cost per observation between strata. Thus, from equation (5.9), we find that the allocation proportions w_i are

$$w_i = \frac{n_i}{n} = \left(\frac{N_i\sigma_i}{N_1\sigma_1 + N_2\sigma_2 + N_3\sigma_3}\right).$$

Hence, $N_1\sigma_1 + N_2\sigma_2 + N_3\sigma_3 = 132\cdot 6 + 92\cdot 5 + 27\cdot 3 = 1333$ so that

$$\begin{aligned}
n_1 &= n\cdot w_1 = 30\cdot(132\cdot 6/1333) \approx 17.82 \approx 18,\\
n_2 &= n\cdot w_2 = 30\cdot(92\cdot 5/1333) \approx 10.35 \approx 10,\\
n_3 &= n\cdot w_3 = 30\cdot(27\cdot 3/1333) \approx 1.82 \approx 2.
\end{aligned}$$

Note that after rounding the n_i to the nearest integer, we have $18 + 10 + 2 = 30$, as required.

(5.5) We are told that the variance must be fixed at $V(\overline{y}_{\text{ST}}) = 0.1 = D$. Thus, the optimal sample size n is given by equation (5.8) so that

$$
\begin{aligned}
n &= \frac{\left(\sum_{i=1}^{3} N_i\sigma_i/\sqrt{c_i}\right)\left(\sum_{i=1}^{3} N_i\sigma_i\sqrt{c_i}\right)}{N^2 D + \sum_{i=1}^{3} N_i\sigma_i^2} \\
&= \frac{\left(112\cdot\sqrt{2.25}/3 + 68\cdot\sqrt{3.24}/5 + 39\cdot\sqrt{3.24}/6\right)\left(112\cdot\sqrt{2.25}\cdot 3 + 68\cdot\sqrt{3.24}\cdot 5 + 39\cdot\sqrt{3.24}\cdot 6\right)}{(112+68+39)^2\cdot(0.1) + (112\cdot 2.25 + 68\cdot 3.24 + 39\cdot 3.24)} \\
&\approx \frac{(92.18)(1537.2)}{(219^2)(0.1) + 598.68} \\
&\approx 26.3 \approx 27
\end{aligned}
$$

Note that if we round down to 26, then we DO NOT achieve the desired bound of $V(\overline{y}_{\text{ST}}) = 0.1 = D$; it is acceptable to have smaller variance than 0.1, but not larger! Now that we have $n = 27$, we use equation (5.7) to determine the n_i; hence,

$$n_i = n\left(\frac{N_i\sigma_i/\sqrt{c_i}}{N_1\sigma_1/\sqrt{c_1} + N_2\sigma_2/\sqrt{c_2} + N_3\sigma_3/\sqrt{c_3}}\right)$$

so that

$$
\begin{aligned}
n_1 &\approx 27\cdot\left(\frac{(112)(\sqrt{2.25})/3}{92.18}\right) \approx 16.40, \\
n_2 &\approx 27\cdot\left(\frac{(68)(\sqrt{3.24})/5}{92.18}\right) \approx 7.17, \\
n_3 &\approx 27\cdot\left(\frac{(39)(\sqrt{3.24})/6}{92.18}\right) \approx 3.43.
\end{aligned}
$$

Rounding off yields $n_1 = 16$, $n_2 = 7$, $n_3 = 3$, which do not add to 27. We can add 1 to stratum 3 to make $n_3 = 4$ since 3.43 is closer to the next higher integer than any of the other approximate sample size (16.40 or 7.17).

(5.13) From the problem description, we find that there are four natural strata identified. We also find that $c_1 = 4$, $c_2 = 4$, $c_3 = 8$, $c_4 = 8$ and $\hat{p}_1 = 0.9$, $\hat{p}_2 = 0.9$, $\hat{p}_3 = 0.5$, $\hat{p}_4 = 0.5$. Furthermore, if we want the error of estimation to satisfy $B = 0.05$, then this is equivalent to specifying that the estimated variance of $\hat{p}_{\text{ST}}$ satisfy $\hat{V}(\hat{p}_{\text{ST}}) = 0.05^2/4 = 0.000625 = D$. The next step is to find the allocation proportions w_i which can be accomplished with equation (5.16); hence,

$$w_i = \frac{n_i}{n} = \frac{N_i\sqrt{p_i q_i/c_i}}{\sum_{i=1}^{4} N_i\sqrt{p_i q_i/c_i}}.$$

(continued)

Since we no not know the exact values of p_i, we can use the *a priori* values given above. Hence,

$$\begin{aligned}\sum_{i=1}^{4} N_i\sqrt{p_iq_i/c_i} &= 97\sqrt{0.9\cdot 0.1/4} + 43\sqrt{0.9\cdot 0.1/4} + 145\sqrt{0.5\cdot 0.5/8} + 68\sqrt{0.5\cdot 0.5/8} \\ &\approx 14.55 + 6.45 + 25.63 + 12.02 \approx 58.65\end{aligned}$$

so that

$$\begin{aligned}w_1 &\approx 14.55/58.65 \approx 0.248, \quad w_2 \approx 6.45/58.65 \approx 0.110, \\ w_3 &\approx 25.63/58.65 \approx 0.437, \quad w_4 \approx 12.02/58.65 \approx 0.205.\end{aligned}$$

We are now in a position to find the sample size n, which can be done with equation (5.15):

$$\begin{aligned}n &= \frac{\sum_{i=1}^{4} N_i^2 p_i q_i / w_i}{N^2 D + \sum_{i=1}^{4} N_i p_i q_i} \\ &\approx \frac{97^2\cdot 0.9\cdot 0.1/0.248 + 43^2\cdot 0.9\cdot 0.1/0.110 + 145^2\cdot 0.5\cdot 0.5/0.437 + 68^2\cdot 0.5\cdot 0.5/0.205}{(353^2)(0.000625) + (97\cdot 0.9\cdot 0.1 + 43\cdot 0.9\cdot 0.1 + 145\cdot 0.5\cdot 0.5 + 68\cdot 0.5\cdot 0.5)} \\ &\approx 157.2 \approx 158.\end{aligned}$$

Finally, we find the values of $n_i = nw_i$ must therefore be

$$n_1 \approx 39,\ \ n_2 \approx 17,\ \ n_3 \approx 69,\ \ n_4 \approx 33.$$

(5.14) In order to estimate the population proportion p, we use $\hat{p}_{\mathrm{ST}}$ as our estimator of p, so that equation (5.13) yields

$$\begin{aligned}\hat{p}_{\mathrm{ST}} &= \frac{1}{N}\left(N_1\hat{p}_1 + N_2\hat{p}_2 + N_3\hat{p}_3 + N_4\hat{p}_4\right) \\ &= \frac{1}{(97+43+145+68)}\left(97\cdot 0.87 + 43\cdot 0.93 + 145\cdot 0.60 + 68\cdot 0.53\right) \\ &\approx 0.701\end{aligned}$$

Using equation (5.14), and our solution to **(5.13)**, we find the estimated variance of $\hat{p}_{\mathrm{ST}}$ to be

$$\begin{aligned}\hat{V}(\hat{p}_{\mathrm{ST}}) &= \frac{1}{N^2}\sum_{i=1}^{4} N_i^2\left(\frac{N_i - n_i}{N_i}\right)\left(\frac{\hat{p}_i\hat{q}_i}{n_i - 1}\right) \\ &= \frac{1}{353^2}\left[97^2\cdot\left(\frac{97-39}{97}\right)\cdot\frac{(0.87)(0.13)}{39-1} + 43^2\cdot\left(\frac{43-17}{43}\right)\cdot\frac{(0.93)(0.07)}{17-1}\right. \\ &\qquad\left. + 145^2\cdot\left(\frac{145-69}{145}\right)\cdot\frac{(0.60)(0.40)}{69-1} + 68^2\cdot\left(\frac{68-33}{68}\right)\cdot\frac{(0.53)(0.47)}{33-1}\right] \\ &\approx 0.0006325.\end{aligned}$$

Thus, the error of estimation is

$$2\sqrt{\hat{V}(\hat{p}_{\mathrm{ST}})} \approx 0.0503.$$

(5.15) If the total cost of sampling is fixed at \$400, then

$$c_1n_1 + c_2n_2 + c_3n_3 + c_4n_4 = 400.$$

Substituting $n_i = nw_i$, $c_1 = c_2 = 4$, $c_3 = c_4 = 8$, and factoring yields

$$n(w_1 + w_2 + 2w_3 + 2w_4) = 100.$$

However, we computed the allocation proportions w_i in **(5.13)** so that

$$n \approx \frac{100}{0.248 + 0.110 + 2(0.437) + 2(0.205)} \approx 60.9 \approx 61.$$

Finally, we can use the w_i to find the n_i:

$$n_1 = 15, \quad n_2 = 7 \quad n_3 = 27 \quad n_4 = 12.$$

(It is important to also check that these values of n_i do, in fact, satisfy the total cost requirement:

$$15 \cdot 4 + 4 \cdot 4 + 27 \cdot 8 + 12 \cdot 8 = 388.$$

This is okay since our rounding gives us a total cost that is less than \$400.)