Semester 2	Applied Statistics	2015

Solution to Tutorial 12

1. Note that a single sample of $n = 15$ customers is drawn from a population of $N = 300$. We have $n = n_1 + n_2 = 6 + 9 = 15$, $X = X_1 + X_2 = 24500 + 21200 = 45700$ and $\bar{X}' = X_1/N = 24500/300 = 81.6667$. The data are

C_l	i	x_i	y_i	x'_i	y'_i	y_i/x_i	$y'_i - r_1 x'_i$
C_1	1	204	210	204	210	1.0294	-4.0415
	2	143	160	143	160	1.1189	9.9611
	3	82	75	82	75	0.9146	-11.0363
	4	256	280	256	280	1.0938	11.3990
	5	275	300	275	300	1.0909	11.4637
	6	198	190	198	190	0.9596	-17.7461
C_2	7	137	150	0	0	1.0949	0.0000
	8	189	200	0	0	1.0582	0.0000
	9	119	125	0	0	1.0504	0.0000
	10	63	60	0	0	0.9524	0.0000
	11	103	110	0	0	1.0680	0.0000
	12	107	100	0	0	0.9346	0.0000
	13	159	180	0	0	1.1321	0.0000
	14	63	75	0	0	1.1905	0.0000
	15	87	90	0	0	1.0345	0.0000
Mean		145.6667	153.6667	77.2	81	1.0482	0.0000
Var.		4454.952	5398.095	11411.17	12957.86	0.0062	58.1168

(a) The total sales estimate using the Hartley Ross estimator is

$$\begin{aligned}\widehat{Y}_{hr} &= X\bar{r}^* + (N-1)\frac{n(\bar{y} - \bar{r}^*\bar{x})}{n-1} \\ &= 45700(1.0482) + (300-1)\frac{15(153.6667 - 1.0482(145.6667)}{15-1} \\ &= 48216.35\end{aligned}$$

(b) The estimate for the rate of increase of sales of Brand I product and its s.e. are

$$\begin{aligned}r_1 &= \frac{\sum\limits_{i\in C_1} y_i}{\sum\limits_{i\in C_1} x_i} = \frac{81}{77.2} = 1.049223 \\ \text{var}(r_1) &= \frac{1}{(\bar{X}')^2}\left(1 - \frac{n}{N}\right)\frac{{s'_{r1}}^2}{n} = \frac{1}{81.6667^2}\left(1 - \frac{15}{300}\right)\frac{58.1168}{15} = 0.000552 \\ \text{se}(r_1) &= \sqrt{0.000552} = 0.023492\end{aligned}$$

The rate of increase is 4.9%.

(c) The estimate for the total sales of Brand I product and its s.e. are

$$\begin{aligned}\widehat{Y}_{r_1} &= X_1 r_1 = 24500(1.049223) = 25705.96 \\ \text{se}(Y_{r_1}) &= X_1 \text{se}(r_1) = 24500(0.023492) = 575.5568\end{aligned}$$

2. (a) **Separate ratio estimate:** $n_A = n_B = 10$; $N_A = 1,000$ and $N_B = 1,500$; $X_A = 16,300$ and $X_B = 12,800$; $W_A = \frac{1000}{2500}$ and $W_B = \frac{1500}{2500}$;

$$\begin{aligned}\text{For A:}\quad s_{y,A} &= 10.36; s_{x,A} = 9.99; s_{xy,A} = 101.822; \sum_i y_i = 187; \sum_i x_i = 17.8; \\ r_A &= \frac{\overline{y}_A}{\overline{x}_A} = \frac{18.7}{17.8} = 1.05. \\ s^2_{sr,A} &= s^2_{y,A} - 2\, r_A s_{xy,A} + r^2_A s^2_{x,A} \\ &= 10.36^2 - 2 \times \frac{18.7}{17.8} \times 101.822 + \left(\frac{18.7}{17.8}\right)^2 \times 9.99^2 \\ &= 3.477 \\ s_{sr,A} &= 1.86.\end{aligned}$$

$$\begin{aligned}\text{For B:}\quad s_{y,B} &= 3.41; s_{x,B} = 5.45; s_{xy,B} = 10.356; \sum_i y_i = 78; \sum_i x_i = 46; \\ r_B &= \frac{\overline{y}_B}{\overline{x}_B} = \frac{4.6}{7.8} = 0.59. \\ s^2_{sr,B} &= s^2_{y,B} - 2\, r_B s_{xy,B} + r^2_B s^2_{x,B} \\ &= 3.42^2 - 2 \times \frac{4.6}{7.8} \times 10.356 + \left(\frac{4.6}{7.8}\right)^2 \times 5.45^2 \\ &= 9.727 \\ s_{sr,B} &= 3.12.\end{aligned}$$

$$\begin{aligned}\widehat{\overline{Y}}_{st,sr} &= W_A \widehat{R}_A \overline{X}_A + W_B \widehat{R}_B \overline{X}_B \\ &= \left(\frac{1000}{2500}\right)\left(\frac{18.7}{17.8}\right) \times 16.3 + \left(\frac{1500}{2500}\right)\left(\frac{4.6}{7.8}\right) \times 8.53 = 9.87. \\ \text{var}(\widehat{\overline{Y}}_{st,sr}) &= W^2_A \left(1 - \frac{n_A}{N_A}\right)\frac{s^2_{sr,A}}{n_A} + W^2_B \left(1 - \frac{n_B}{N_B}\right)\frac{s^2_{sr,B}}{n_B} \\ &= \left(\frac{1000}{2500}\right)^2 \left(1 - \frac{10}{1000}\right)\frac{1.86^2}{10} + \left(\frac{1500}{2500}\right)^2 \left(1 - \frac{10}{1500}\right)\frac{3.12^2}{10} \\ &= 0.40.\end{aligned}$$

(b) **Combine ratio estimate:**

$$\begin{aligned}
\overline{X} &= \frac{X_A + X_B}{N_A + N_B} = \frac{16,300 + 12,800}{1,000 + 1,500} = 11.64 \\
\widehat{\overline{X}}_{st} &= W_A \overline{x}_A + W_B \overline{x}_B = 0.4 \times 17.8 + 0.6 \times 7.8 = 11.80. \\
\widehat{\overline{Y}}_{st} &= W_A \overline{y}_A + W_B \overline{y}_B = 0.4 \times 18.7 + 0.6 \times 4.6 = 10.24. \\
r_c &= \frac{10.24}{11.80} = 0.86779661.
\end{aligned}$$

$$\begin{aligned}
\text{For A: } s_{y,A} &= 10.36; s_{x,A} = 9.99; s_{xy,A} = 101.822; \sum_i y_i = 187; \sum_i x_i = 17.8; \\
s^2_{cr,A} &= s^2_{y,A} - 2r_C\, s_{xy,A} + r_C^2 s^2_{x,A} \\
&= 10.36^2 - 2 \times \frac{10.24}{11.80} \times 101.822 + \left(\frac{10.24}{11.80}\right)^2 \times 9.99^2 \\
&= 5.729 \\
s_{cr,A} &= 2.39.
\end{aligned}$$

$$\begin{aligned}
\text{For B: } s_{y,B} &= 3.41; s_{x,B} = 5.45; s_{xy,B} = 10.356; \sum_i y_i = 78; \sum_i x_i = 46; \\
s^2_{cr,B} &= s^2_{y,B} - 2r_C\, s_{xy,B} + r_C^2 s^2_{x,B} \\
&= 3.42^2 - 2 \times \frac{10.24}{11.80} \times 10.356 + \left(\frac{10.24}{11.80}\right)^2 \times 5.45^2 \\
&= 16.018 \\
s_{cr,B} &= 4.00.
\end{aligned}$$

$$\begin{aligned}
\widehat{\overline{Y}}_{st,cr} &= \widehat{R}_{st,cr}\overline{X} = 0.86779661 \times 11.64 = 10.10. \\
\text{var}(\widehat{\overline{Y}}_{st,cr}) &= W_A^2 \left(1 - \frac{n_A}{N_A}\right) \frac{s^2_{cr,A}}{n_A} + W_B^2 \left(1 - \frac{n_B}{N_B}\right) \frac{s^2_{cr,B}}{n_B} \\
&= \left(\frac{1000}{2500}\right)^2 \left(1 - \frac{10}{1000}\right) \frac{2.39^2}{10} + \left(\frac{1500}{2500}\right)^2 \left(1 - \frac{10}{1500}\right) \frac{4.00^2}{10} \\
&= 0.66 > 0.40 = \text{var}(\widehat{\overline{Y}}_{st,sr}).
\end{aligned}$$

Note that $\text{var}(\widehat{\overline{Y}}_{st,sr}) < \text{var}(\widehat{\overline{Y}}_{st,cr})$. Hence separate ratio estimator is preferred since 1. $n_A = n_B = 10$ is not too small and 2. $\widehat{R}_A = 1.05$, $\widehat{R}_B = 0.59$ is not similar.

(c) **Discussion**

If the stratum sample sizes n_h are large enough (say, 20) so that the separate ratio estimator $\widehat{\overline{Y}}_{st,sr}$ does not have large biases and that the variance approximation works adequately, use the separate ratio estimator.

If the stratum sample sizes n_h are very small and the stratum ratio $R_h = \frac{\overline{Y}_h}{\overline{X}_h}$ is constant over strata, the combined ratio estimator $\widehat{\overline{Y}}_{st,cr}$ may perform better.

3. We have $n_s = 3$, $n = 4$, $k = 9$, $N = 36$ and total sample size $n' = 3(4) = 12$.

(a) The sample means and variances are

Samples	1	2	3	4	5	6	7	8	9	Overall
1	0	0	0	1	1	0	2	1	2	
2	2	5	6	6	5	6	5	5	5	
3	4	3	4	2	2	2	2	2	1	
4	2	1	1	1	1	1	1	0	0	
Mean	2.00	2.25	2.75	2.50	2.25	2.25	2.50	2.00	2.00	2.28
Var.	2.667	4.917	7.583	5.667	3.583	6.917	3.000	4.667	4.667	3.806

The true variance of the mean estimator is the variance of these 9 sample means which is

$$\mathrm{Var}(\bar{Y}) = \frac{1}{k}\left(\sum_{i=1}^{k} \bar{y}_i^2 - k\,\bar{\bar{y}}^2\right) = \frac{1}{9}[47.475 - 9(2.28^2)] = 0.06173.$$

(b) The sum of squares and mean sum of squares are

$$\begin{aligned}
SST_o &= \sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}^2 - N\bar{\bar{y}}^2 = 320 - 36(2.28^2) = 133.222 \\
SSW &= \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_i)^2 = \sum_{i=1}^{k}(n-1)s_i^2 = 3(2.667 + \cdots + 4.667) = 131 \\
SSB &= n\sum_{i=1}^{k} \bar{y}_i^2 - N\bar{\bar{y}}^2 = 4(2^2 + 2.25^2 + \cdots + 2^2) - 36(2.28^2) = 2.222 \\
&\overset{or}{=} SST_o - SSW = 133.222 - 131 = 2.222 \\
S_w^2 &= \frac{SSW}{k(n-1)} = \frac{131}{27} = 4.852 \\
S^2 &= \frac{SST_o}{N-1} = \frac{133.222}{36-1} = 3.806 < 4.852
\end{aligned}$$

Since $S_w^2 > S^2$, systematic sampling is more efficient.

(c) With random starts of 2, 4 and 8, the selected sample means are $\bar{y}_i = 2.25, 2.5, 2$. We have $\sum_{i=1}^{n_s} \bar{y}_i^2 = 15.3125$. The estimate for the average level of dieldrin in this stretch of the river and its variance are

$$
\begin{aligned}
\hat{\bar{\bar{Y}}} &= \frac{1}{n_s}\sum_{i=1}^{n_s}\bar{y}_i = \frac{1}{3}(2.25+2.5+2) = 2.25 \\
s_{\bar{y}}^2 &= \frac{1}{n_s-1}\left(\sum_{i=1}^{n_s}\bar{y}_i^2 - n_s\bar{\bar{y}}^2\right) = \frac{1}{2}[15.3125 - 3(2.25^2)] = 0.0625 \\
\text{var}(\hat{\bar{\bar{Y}}}) &= \left(1-\frac{n'}{N}\right)\frac{s_{\bar{y}}^2}{n_s} = \left(1-\frac{12}{36}\right)\frac{0.0625}{3} = 0.0139
\end{aligned}
$$

Extra exercise

1. (a) Estimate and its s.e. for the total commission received using post-stratification when the data is stratified according to branches:

$$
\begin{aligned}
\widehat{Y}_{pst,1} &= \sum_l N_l\bar{y}_l \\
&= 12\times 61.34 + 10\times 56.30 + 15\times 55.90 = 2137.58 \text{ thousands} \\
\text{var}(\widehat{Y}_{pst,1}) &= N^2\left[\left(1-\frac{n}{N}\right)\sum_{l=1}^{L}W_l\frac{S_l^2}{n} + \frac{1}{n^2}\sum_{l=1}^{L}(1-W_l)S_l^2\right] \\
&= 37^2\left[\left(1-\frac{15}{37}\right)\left(\frac{12}{37}\frac{819.328}{15} + \frac{10}{37}\frac{833.2}{15} + \frac{15}{37}\frac{799.045}{15}\right)\right. \\
&\quad \left. + \frac{1}{15^2}\left(\frac{25}{37}\times 819.328 + \frac{27}{37}\times 833.2 + \frac{22}{37}\times 799.045\right)\right] \\
&= 37^2\left[\left(1-\frac{15}{37}\right)(17.7152 + 15.0126 + 21.5958) + \right. \\
&\quad \left. \frac{1}{15^2}(553.6 + 608.0108 + 475.1078)\right] \\
&= 37^2(32.30051892 + 7.274305105) = 54,177.93409 \\
\text{se}(\widehat{Y}_{pst,1}) &= \sqrt{54,177.93409} = 232.7615391
\end{aligned}
$$

(b) Estimate and its s.e. for the total commission received using post-stratification when the data is stratified according to 'the length of stay in the company:

$$
\begin{aligned}
\widehat{Y}_{pst,2} &= \sum_l N_l \bar{y}_l \\
&= 17 \times 31.28 + 12 \times 64.74 + 8 \times 89.08 = 2021.28 \text{ thousands} \\
\mathrm{var}(\widehat{Y}_{pst,2}) &= N^2 \left[\left(1 - \frac{n}{N}\right) \sum_{l=1}^{L} W_l \frac{S_l^2}{n} + \frac{1}{n^2} \sum_{l=1}^{L} (1 - W_l) S_l^2 \right] \\
&= 37^2 \left[\left(1 - \frac{15}{37}\right) \left(\frac{17}{37}\frac{128.942}{15} + \frac{12}{37}\frac{174.613}{15} + \frac{8}{37}\frac{60.989}{15} \right) \right. \\
&\quad \left. + \frac{1}{15^2} \left(\frac{20}{37} \times 128.942 + \frac{25}{37} \times 174.613 + \frac{29}{37} \times 60.989 \right) \right] \\
&= 37^2 \left[\left(1 - \frac{15}{37}\right) (3.9496 + 3.7754 + 0.8791) + \right. \\
&\quad \left. \frac{1}{15^2} (69.6984 + 117.9818 + 47.8022) \right] \\
&= 37^2 (5.115958315 + 1.046588108) = 8,436.526053 \\
\mathrm{se}(\widehat{Y}_{pst,2}) &= \sqrt{8,436.526053} = 91.85056371
\end{aligned}
$$

(c) Estimator in (b) is better. The auxiliary variable of 'the length of stay in the company' leads to more efficient estimator because the resulting strata is more internally homogeneous w.r.t. the commission received.

(d) For SRS, the sample mean $= 57.84\dot{6}$ and the sample variance=707.0155238. Hence

$$
\begin{aligned}
\widehat{Y} &= N\bar{y} = 37 \times 57.84\dot{6} = 2,140.32\dot{6} \\
\mathrm{var}(\widehat{Y}) &= N^2 \left(1 - \frac{n}{N}\right) \frac{s_y^2}{n} = 37^2 \left(1 - \frac{15}{37}\right) \frac{707.0155238}{15} = 38,367.37576 \\
\mathrm{se}(\widehat{Y}) &= \sqrt{38,367.37576} = 195.8759193.
\end{aligned}
$$

Note that $\mathrm{se}(\widehat{Y}_{pst,2}) < \mathrm{se}(\widehat{Y}) < \mathrm{se}(\widehat{Y}_{pst,1})$. Only the stratification using 'the length of service in the company' improves the efficiency of the estimator.

(e) Variance reduction in poststratification

$$
\begin{aligned}
\mathrm{var}(\widehat{Y}_{srs}) - \mathrm{var}(\widehat{Y}_{pst}) &= N^2 \left[\left(1 - \frac{n}{N}\right) \frac{s_y^2}{n} - \left(1 - \frac{n}{N}\right) \sum_{l=1}^{L} W_l \frac{S_l^2}{n} - \frac{1}{n^2} \sum_{l=1}^{L} (1 - W_l) S_l^2 \right] \\
&= N^2 \left[\left(1 - \frac{n}{N}\right) \frac{1}{n} (s^2 - \sum_{l=1}^{L} W_l S_l^2) - \frac{1}{n^2} \sum_{l=1}^{L} (1 - W_l) S_l^2 \right] \\
&\approx N^2 \left[\frac{1}{n} (s^2 - \sum_{l=1}^{L} W_l S_l^2) \right]
\end{aligned}
$$

Assuming $\frac{n}{N}$ is sufficiently small so that $(1 - \frac{n}{N}) \approx 1$ and n is sufficiently large so that $\frac{1}{n^2}$ is negligible as compared with $\frac{1}{n}$.

$$\begin{aligned}
\text{For (a)} \quad s^2 - \textstyle\sum_{l=1}^{L} W_l S_l^2 &= 707.0155 - \left(\frac{12}{37} \times 819.328 + \frac{10}{37} \times 833.2 + \frac{15}{37} \times 799.045\right) \\
&= 707.0155 - 814.8543514 = -107.8389 \\
\text{For (b)} \quad s^2 - \textstyle\sum_{l=1}^{L} W_l S_l^2 &= 707.0155 - \left(\frac{17}{37} \times 128.942 + \frac{12}{37} \times 174.613 + \frac{8}{37} \times 60.989\right) \\
&= 707.0155 - 129.0617 = 577.9538
\end{aligned}$$

Hence Δs^2 in (b) is much larger and leads to larger reduction of variance for the estimator. For (a), $s^2 - \sum_{l=1}^{L} W_l S_l^2$ is negative which implies an increase of variance for the estimator using post-stratification. This is due to the fact that post-stratification does not help in reducing the sample variance in each stratum but the sample size in each stratum is much smaller.

2. (a) **Method A:** If the stratification leads to more internally homogeneous strata, the stratified SRS will be preferred. Since S_l will be small if the resulting strata are internally homogeneous and hence $\text{var}(\widehat{\overline{Y}}_{st}) = \sum_{l=1}^{L} W_l^2 \left(1 - \frac{n_l}{N_l}\right) \frac{s_l^2}{n_l}$ will be small as well. The mean estimator using the stratified SRS and proportional allocation method is the same as the mean estimator using SRS because $\widehat{\overline{Y}}_{st} = \sum_l \frac{N_l}{N} \bar{y}_l = \sum_i \frac{n_i}{n} \bar{y}_i = \bar{y}$.

   Method B: If the auxiliary variable X_i is positively and highly correlated to Y_i, the variable of interest and the population mean $\overline{X}$ or total X is known for the auxiliary variable, this method is preferred. Note that $s_r^2 = \frac{1}{n-1}(\sum_{i=1}^{100} y_i^2 - 2 \times \widehat{R} \times \sum_{i=1}^{100} x_i y_i + \widehat{R}^2 \sum_{i=1}^{100} x_i^2$ will be small if X and Y are highly correlated.

   (b) (i) For Neyman allocation,

$$\begin{aligned}
\sum_{l=1}^{L} N_l s_l &= 2,940 \times 3.2 + 3,530 \times 6.3 + 2,110 \times 10.1 + 1070 \times 16.5 = 70,613 \\
n_1 &= n w_1 = n \left(\frac{N_1 s_1}{\sum_{i=1}^{L} N_i s_i}\right) = 100 \left[\frac{2940 \times 3.2}{70,613}\right] = 13.32 = 13 \\
n_2 &= n w_2 = 100 \left[\frac{3,530 \times 6.3}{70,613}\right] = 32.49 = 32 \\
n_3 &= n w_3 = 100 \left[\frac{2,110 \times 10.1}{70,613}\right] = 30.18 = 30 \\
n_4 &= n w_4 = 100 \left[\frac{1070 \times 16.5}{70,613}\right] = 25.00 = 25
\end{aligned}$$

   (ii) Estimate of the total annual profit last year for all the trading firms in that industry and its CI estimate:

$$
\begin{aligned}
\widehat{Y}_{st} &= \sum_l N_l \bar{y}_l \\
&= (2,940)(8.7) + (3,530)(15.2) + (2,110)(20.4) + (1,070)(44.8) \\
&= 170,214 \\
\text{var}(\widehat{Y}_{st}) &= \sum_{l=1}^{L} N_l^2 \left(1 - \frac{n_l}{N_l}\right) \frac{s_l^2}{n_l} \\
&= 2,940^2 \left(1 - \frac{13}{2,940}\right) \frac{16.1604}{13} + 3,530^2 \left(1 - \frac{32}{3,530}\right) \frac{67.6996}{32} + \\
&\quad\; 2,110^2 \left(1 - \frac{30}{2,110}\right) \frac{151.5361}{30} + 1,070^2 \left(1 - \frac{25}{1,070}\right) \frac{232.2576}{25} \\
&= 69,377,544.89 \\
\text{se}(\widehat{Y}_{st}) &= \sqrt{69,377,544.89} = 8,329.318393 \\
95\% \text{ CI for } Y_{st} &= (\widehat{Y}_{st} - z_{0.025} \times \text{se}(\widehat{Y}_{st}),\ \widehat{Y} + z_{0.025} \times \text{se}(\widehat{Y}_{st})) \\
&= (170,214 - 1.96 \times 8,329.318393,\ 170,214 + 1.96 \times 8,329.318393) \\
&= (153,888.5359,\ 186,539.4641)
\end{aligned}
$$

(c) Estimate of the total annual profit last year for all the trading firms in that industry using ratio estimation and its s.e. estimate:

$$
\begin{aligned}
\widehat{R} &= \frac{\bar{y}}{\bar{x}} = \frac{1,926.5}{825} = 2.335\dot{1}\dot{5} \\
\widehat{Y}_r &= X\widehat{R} = 72,730 \times 2.335\dot{1}\dot{5} = 169,835.5697 \\
s_r^2 &= \frac{1}{n-1}\left(\sum_i y_i^2 - 2\widehat{R}\sum_i x_i y_i + \widehat{R}^2 \sum_i x_i^2\right) \\
&= \frac{1}{99}(82,671 - 2 \times 2.335\dot{1}\dot{5} \times 25,707 + 2.335\dot{1}\dot{5}^2 \times 8,674) = 100.1036097 \\
\text{var}(\widehat{Y}_r) &= N^2 \left(1 - \frac{n}{N}\right) \frac{s_r^2}{n} = 9,650^2 \left(1 - \frac{100}{9,650}\right) \frac{100.1036097}{100} = 92,252,984.12 \\
\text{se}(\widehat{Y}_r) &= \sqrt{92,252,984.12} = 9,604.841702 > se(\widehat{Y}_{st}) = 8,329.318393.
\end{aligned}
$$

(d) Would prefer the total estimator using stratified SRS rather than the total estimator using SRS and ratio estimate. It is possible that the correlation between X_i and Y_i may not be strong enough as large firm size does not necessary lead to higher annual profit. However the relationship between the firm size and annual profit should generally be positive so that stratification using the firm size should lead to more internally homogeneous strata w.r.t. the annual profit.

3. We ignore fpc since N is not given.

(a) For sample mean using SRS

$$
\begin{aligned}
\widehat{\overline{Y}}_{srs,4} &= \bar{y} = \bar{y}_{st} = 9,333.\dot{3} \\
\sum_i y_{i1}^2 &= s_1^2(n_1 - 1) + n_1 \times \bar{y}_1^2 = 400^2 \times 99 + 100 \times 12,000^2 = 1.441584 \times 10^{10} \\
\sum_i y_{i2}^2 &= s_2^2(n_2 - 1) + n_2 \times \bar{y}_2^2 = 100^2 \times 199 + 200 \times 8,000^2 = 1.280199 \times 10^{10} \\
\sum_{i,j} y_{ij}^2 &= \sum_i y_{i1}^2 + \sum_i y_{i2}^2 = 1.441584 \times 10^{10} + 1.280199 \times 10^{10} = 2.721783 \times 10^{10} \\
s_y^2 &= \frac{1}{n-1}\left(\sum_{i,j} y_{ij}^2 - n\bar{y}^2\right) = \frac{1}{299}(2.721783 \times 10^{10} - 300 \times 9,333.\dot{3}^2) \\
&= 3,627,079.159 \\
\mathrm{var}(\widehat{\overline{Y}}_{srs,4}) &= \frac{s_y^2}{n} = \frac{3,627,079.159}{300} = 12,090.26386 \\
\mathrm{se}(\widehat{\overline{Y}}_{srs,4}) &= \sqrt{12,090.26386} = 109.96.
\end{aligned}
$$

Since there is large variation between the 2 strata, the s.e. of SRS estimator is large. Post-stratification estimator is not feasible as the population sizes are unknown.

(b) For double sampling for stratification

$$
\begin{aligned}
\widehat{\overline{Y}}_{st,ds} &= \sum_{i=1}^{L} w_i \bar{y}_i = \frac{100}{300} \times 12,000 + \frac{200}{300} \times 8,000 = 9,333.\dot{3} \\
\mathrm{var}(\widehat{\overline{Y}}_{st}) &= \frac{1}{n_1}(w_1 s_1)^2 + \frac{1}{n_2}(w_2 s_2)^2 + \frac{1}{n'}[w_1(\bar{y}_1 - \widehat{\overline{Y}}_{st,ds})^2 + w_2(\bar{y}_2 - \widehat{\overline{Y}}_{st,ds})^2] \\
&= \frac{1}{10}(0.\dot{3} \times 400)^2 + \frac{1}{20}(0.\dot{6} \times 100)^2 + \\
&\quad \frac{1}{300}[0.\dot{3}(12,000 - 9,333.\dot{3})^2 + 0.\dot{6}(8,000 - 9,333.\dot{3})^2] \\
&= 2,000 + 11,851.85185 = 13,851.8519 \\
\mathrm{se}(\widehat{\overline{Y}}_{st,ds}) &= \sqrt{13,851.8519} = 117.69389
\end{aligned}
$$

4. (a) (i) The total inventory X in thousand dollars last year for all dealers of the juice

drink company

$$\begin{aligned}
\widehat{X}_{pst} &= N(W_1\overline{y}_1 + W_2\overline{y}_2 + W_3\overline{y}_3) \\
&= 19,400(0.3 \times 8.5 + 0.55 \times 17.2 + 0.15 \times 27) \\
&= 19,400 \times 16.06 = 311,564 \\
\text{var}(\overline{y}_{pst}) &= N^2 \left[\left(1 - \frac{n}{N}\right) \sum_{l=1}^{L} W_l \frac{s_l^2}{n} + \frac{1}{n^2} \sum_{l=1}^{L} (1 - W_l) s_l^2 \right] \\
&= 19,400^2 \left[\frac{1}{970} \left(1 - \frac{970}{19,400}\right) \right. \\
&\quad (0.3 \times 16.1 + 0.55 \times 25.2 + 0.15 \times 55.5) + \\
&\quad \left. \frac{1}{970^2} (0.7 \times 16.1 + 0.45 \times 25.2 + 0.85 \times 55.5) \right] \\
&= 19,400^2 (0.026458 + 0.000074) = 9,985,643
\end{aligned}$$

(ii) The population size N_i for each stratum

$$\begin{aligned}
N_1 &= N \times \frac{n_1'}{n'} = 19,400 \times \frac{272}{970} = 5,440 \\
N_2 &= N \times \frac{n_2'}{n'} = 19,400 \times \frac{510}{970} = 10,200 \\
N_3 &= N \times \frac{n_3'}{n'} = 19,400 \times \frac{188}{970} = 3,760
\end{aligned}$$

(iii) Neyman allocation for a sample of size $n = 97$ into the three strata

$$\begin{aligned}
\sum_{l=1}^{L} N_l s_l &= N_1 s_1 + N_2 s_2 + N_3 s_3 \\
&= 5,440\sqrt{16.1} + 10,200\sqrt{25.2} + 3,760\sqrt{55.5} = 101,042.857
\end{aligned}$$

$$\begin{aligned}
n_1 &= n w_1 = n \left(\frac{N_1 s_1}{\sum_{i=1}^{L} N_i s_i} \right) \\
&= 97 \left[\frac{5,440\sqrt{16.1}}{101,042.857} \right] = 97(.216) = 20.96 = 21 \\
n_2 &= n w_2 = 97 \left[\frac{10,200\sqrt{25.2}}{101,042.857} \right] = 97(.507) = 49.16 = 49 \\
n_3 &= n w_3 = 97 \left[\frac{3,760\sqrt{55.5}}{101,042.857} \right] = 97(.277) = 26.89 = 27
\end{aligned}$$

(b) The total inventory Y in thousand dollars this year for all dealers of the juice drink company

$$
\begin{aligned}
\widehat{Y}_{st,ds} &= N\left(\sum_{i=1}^{L} w_i \bar{y}_i\right) \\
&= 19,400\left(\frac{272}{970}\times 9.7+\frac{510}{970}\times 18.3+\frac{188}{970}\times 31.5\right) \\
&= 19,400\times 18.45 = 357,868
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{var}(\widehat{Y}_{st,ds}) &= N^2\left[\sum_i \frac{1}{n_i} w_i^2 s_i^2 + \frac{1}{n'}\sum_i w_i(\bar{y}_i - \widehat{\overline{Y}}_{st,ds})^2\right] \\
&= 19,400^2\left\{\frac{1}{21}\left(\frac{272}{970}\right)^2\times 15.3+\frac{1}{49}\left(\frac{510}{970}\right)^2\times 28.5+\right. \\
&\qquad \frac{1}{27}\left(\frac{188}{970}\right)^2\times 60.2+\frac{1}{97}\left[\frac{272}{970}(9.7-18.45)^2\right. \\
&\qquad \left.\left. +\frac{510}{970}(18.3-18.45)^2+\ \frac{188}{970}(31.5-18.45)^2\right]\right\} \\
&= 19,400^2(0.3018+0.0562) = 134,737,058.5
\end{aligned}
$$