Stat 257.01 Fall 2004
Assignment #8 Solutions

(6.25) As noted, each brand serves as a strata. In order to use consistent notation with the text, we will write strata A for brand I and strata B for brand II. In order to estimate μ_y we perform both a separate ratio estimate and a combined ratio estimate.

For the separate ratio estimate, we have

$$\hat{\mu}_y = \frac{N_A}{N}\hat{\mu}_{yA} + \frac{N_B}{N}\hat{\mu}_{yB}$$

where $\hat{\mu}_{yA} = r_A \mu_{xA}$ and $\hat{\mu}_{yB} = r_B \mu_{xB}$ with $r_A = \overline{y}_A/\overline{x}_A = \sum y_{iA}/\sum x_{iA}$ and $r_B = \overline{y}_B/\overline{x}_B = \sum y_{iB}/\sum x_{iB}$.

From the data supplied in the problem, we find that $N_A = 120$, $\tau_{xA} = 24500$, and $n_A = 6$. Furthermore, $N_B = 180$, $\tau_{xB} = 21200$, and $n_B = 9$. Hence, we calculate $N = N_A + N_B = 300$, $\mu_{xA} = \tau_{xA}/N_A = 24500/120 \approx 204.17$ and similarly $\mu_{xB} = 21200/180 \approx 117.78$.

We can also calculate $\sum y_{iA} = 1215$ and $\sum x_{iA} = 1158$, so that $\overline{y}_A = 1215/6 = 202.5$ and $\overline{x}_A = 1158/6 = 193$. Similarly, $\sum y_{iB} = 1090$ and $\sum x_{iB} = 1027$, so that $\overline{y}_B = 1090/9 \approx 121.1$ and $\overline{x}_B = 1027/9 \approx 114.1$. Hence,

$$r_A = \frac{1215}{1158} \approx 1.049 \quad \text{and} \quad r_B = \frac{1090}{1027} = 1.061.$$

This now enables us to determine

$$\hat{\mu}_{yA} = r_A \mu_{xA} \approx 1.049 \cdot 204.17 \approx 214.17 \quad \text{and} \quad \hat{\mu}_{yB} = r_B \mu_{xB} \approx 1.061 \cdot 117.78 \approx 124.96$$

so that our separate ratio estimate for μ_y is

$$\hat{\mu}_y \approx \frac{120}{300} \cdot 214.17 + \frac{180}{300} \cdot 124.96 \approx 160.65.$$

Conversely, for the combined ratio estimate we have

$$\hat{\mu}_y = r_c \mu_x$$

where r_c denotes the combined ratio given by

$$r_c = \frac{\overline{y}_{\text{st}}}{\overline{x}_{\text{st}}}.$$

As in chapter 5, we can calculate $\overline{y}_{\text{st}}$ and $\overline{x}_{\text{st}}$:

$$\overline{y}_{\text{st}} = \frac{N_A}{N}\overline{y}_A + \frac{N_B}{N}\overline{y}_B \approx \frac{120}{300} \cdot 202.5 + \frac{180}{300} \cdot 121.1 = 153.66$$

and

$$\overline{x}_{\text{st}} = \frac{N_A}{N}\overline{x}_A + \frac{N_B}{N}\overline{x}_B \approx \frac{120}{300} \cdot 193 + \frac{180}{300} \cdot 114.1 = 145.66.$$

This tells us that

$$r_c = \frac{\overline{y}_{\text{st}}}{\overline{x}_{\text{st}}} \approx \frac{153.66}{145.66} \approx 1.055.$$

If we wish to calculuate μ_x then we ignore our stratification and view both brands as being pooled into one. As such, we have $\tau_x = \tau_{xA} + \tau_{xB} = 24500 + 21200 = 45700$ and $N = N_A + N_B = 120 + 180 = 300$ so that

$$\mu_x = \frac{\tau_x}{N} = \frac{45700}{300} \approx 152.33.$$

Hence, our estimate of μ_y in this case is given by

$$\hat{\mu}_y \approx 1.055 \cdot 152.33 \approx 160.71.$$

Now, the question is asking for the ratio estimate of the *total* potential sales for next quarter. Since $\tau = N\mu$ in general, we can give both our ratio estimates. Notice that both procedures give virtually the same estimate:

- **separate ratio estimate:** $\hat{\tau}_y = N\hat{\mu}_y \approx 300 \cdot 160.65 = 48195$,
- **combined ratio estimate:** $\hat{\tau}_y = N\hat{\mu}_y \approx 300 \cdot 160.71 = 48213$.

Next, we need to calculate the variances for both these estimators. As a first step, note that

$$\hat{V}(\hat{\tau}_y) = \hat{V}(N\hat{\mu}_y) = N^2\hat{V}(\hat{\mu}_y).$$

The text gives formulæ for the estimated variance of $\hat{\mu}_y$ in both the separate case and the combined case in section 6.5.

Hence, for the separate ratio estimate, we have

$$\hat{V}(\hat{\mu}_y) = \frac{N_A^2}{N^2} \cdot \left(\frac{N_A - n_A}{n_A N_A}\right) \cdot s_{rA}^2 + \frac{N_B^2}{N^2} \cdot \left(\frac{N_B - n_B}{n_B N_B}\right) \cdot s_{rB}^2$$

where

$$s_{rA}^2 = \frac{1}{n_A - 1}\sum_{i=1}^{n_A}(y_{iA} - r_A x_{iA})^2 \quad \text{and} \quad s_{rB}^2 = \frac{1}{n_B - 1}\sum_{i=1}^{n_B}(y_{iB} - r_B x_{iB})^2.$$

A straightforward calculation gives $s_{rA}^2 \approx 162.73$ and $s_{rB}^2 \approx 56.61$ so that

$$\hat{V}(\hat{\mu}_y) \approx \frac{120^2}{300^2} \cdot \left(\frac{120 - 6}{6 \cdot 120}\right) \cdot 162.73 + \frac{180^2}{300^2} \cdot \left(\frac{180 - 9}{9 \cdot 180}\right) \cdot 56.61 \approx 6.27.$$

Hence, we conclude that for the separate ratio estimate,

$$\hat{V}(\hat{\tau}_y) = N^2\hat{V}(\hat{\mu}_y) \approx 300^2 \cdot 6.27 = 564630.6.$$

On the other hand, for the combined ratio estimate, we have

$$\hat{V}(\hat{\mu}_y) = \frac{N_A^2}{N^2} \cdot \left(\frac{N_A - n_A}{n_A N_A}\right) \cdot s_{rAc}^2 + \frac{N_B^2}{N^2} \cdot \left(\frac{N_B - n_B}{n_B N_B}\right) \cdot s_{rBc}^2$$

where

$$s_{rAc}^2 = \frac{1}{n_A - 1}\sum_{i=1}^{n_A}(y_{iA} - r_c x_{iA})^2 \quad \text{and} \quad s_{rBc}^2 = \frac{1}{n_B - 1}\sum_{i=1}^{n_B}(y_{iB} - r_c x_{iB})^2.$$

Since $r_c \approx 1.055$ a straightforward compuation yields $s^2_{rAc} \approx 159.21$ and $s^2_{rBc} \approx 58.32$. Hence, for the combined ratio estimate,

$$\hat{V}(\hat{\mu}_y) \approx \frac{120^2}{300^2} \cdot \left(\frac{120-6}{6 \cdot 120}\right) \cdot 159.21 + \frac{180^2}{300^2} \cdot \left(\frac{180-9}{9 \cdot 180}\right) \cdot 58.32 \approx 6.25$$

so that we may conclude

$$\hat{V}(\hat{\tau}_y) = N^2 \hat{V}(\hat{\mu}_y) \approx 300^2 \cdot 6.25 = 562453.2.$$

Again, notice that both procedures give virtually the same estimated variance:

- **separate ratio estimate:** $\hat{V}(\hat{\tau}_y) \approx 564630.6$,
- **combined ratio estimate:** $\hat{V}(\hat{\tau}_y) \approx 562453.2$.

By definition, the relative efficiency of two estimators is the ratio of those estimators. In this case, the relative efficiency of $\hat{\tau}_y$ in the separate ratio estimate case (now called E_1) to $\hat{\tau}_y$ in the combined ratio estimate case (now called E_2) is

$$\mathrm{RE}(E_1, E_2) = \frac{\hat{V}(E_2)}{\hat{V}(E_1)} \approx \frac{562453.2}{564630.6} \approx 0.996.$$

Since our computations are all approximate anyway (we've rounded decimals at each step) and since the relative efficiency is *virtually* 1, there is no sufficient evidence to conclude that we should prefer either the separate ratio estimate or the combined ratio estimate in this example; either one is fine to use as an estimator of the parameter τ_y.

(7.1) In this instance, I would choose a systematic sample instead of a simple random sample because the population is ordered. It also is reasonable to infer that N is large enough. (If the company has been granting mortgages for 20 years, then it is reasonable to believe that they have had a steady stream of customers.)

(7.4) In this case, we have $N = 2000$, $n = 200$, $k = 10$, $\sum y_i = 132$. Hence, our estimator of p, the proportion in favour of the new policy, is

$$\hat{p}_{\text{sy}} = \frac{1}{n} \sum_{i=1}^{200} y_i = \frac{132}{200} = 0.66.$$

A bound, therefore, on the error of estimation is given by

$$B = 2\sqrt{\hat{V}(\hat{p}_{\text{sy}})} = 2\sqrt{\frac{\hat{p}_{\text{sy}}\hat{q}_{\text{sy}}}{n-1} \cdot \left(\frac{N-n}{n}\right)} = 2\sqrt{\frac{0.66 \cdot 0.34}{200-1} \cdot \left(\frac{2000-200}{200}\right)} \approx 0.0637.$$

In other words, an approximate 95% confidence interval for p is 0.66 ± 0.0637.

(7.11) From the problem, we find that $N = 4500$, $n = 30$, $k = 150$. If we denote by y the amount spent, then

$$\sum y_i = 850, \quad \sum y_i^2 = 33904, \quad \text{and} \quad s^2 \approx 338.64.$$

Hence, we conclude that

$$\hat{\tau}_{\text{sy}} = N \cdot \overline{y}_{\text{sy}} = 4500 \cdot \frac{850}{30} = 127500.$$

A bound on the error of estimation is given by

$$2\sqrt{\hat{V}(\hat{\tau}_{\text{sy}})} = 2\sqrt{N^2 \cdot \frac{s^2}{n} \cdot \left(\frac{N-n}{N}\right)} \approx 2\sqrt{4500^2 \cdot \frac{338.64}{30} \cdot \left(\frac{4500-30}{4500}\right)} \approx 30137.06.$$

That is, an approximate 95% confidence interval for τ is 127500 ± 30137.

(7.12) In order to have a bound on the error of estimation of 10000, we need a sample of size

$$n = \frac{N\sigma^2}{(N-1)D + \sigma^2}$$

where $D = B/(4N^2) = 10000^2/(4 \cdot 4500^2) \approx 1.2345679$. Using s^2 to approximate σ^2, we conclude

$$n \approx \frac{4500 \cdot 338.64}{4499 \cdot 1.2345679 + 338.64} \approx 258.6 \approx 259.$$

I would probably choose to conduct repeated systematic samples and use successive differences. The advantage of this scheme is that reasonable estimates can be obtained when very little is known about the population as seems to be the case in this exercise.