Stat 257.01 Fall 2004
Assignment #9 (Selected) Solutions

(8.2) We let saws in industry i denote the ith cluster, so that the cluster sizes m_i, $i = 1, \ldots 20$, are given by the number of saws column, and y_i are given by the total repair cost column. Hence, we find that

$$\sum m_i = 130, \quad \sum m_i^2 = 1118, \quad \sum y_i = 2565, \quad \sum y_i^2 = 460225, \quad \sum m_i y_i = 22285.$$

Furthermore, from the problem itself, we are told that $N = 96$ and $n = 20$, but that M is unknown. We can now calculate

$$\overline{y} = \frac{\sum y_i}{\sum m_i} = \frac{2565}{130} \approx 19.73.$$

The estimated variance of $\overline{y}$ is given by

$$\hat{V}(\overline{y}) = \left(\frac{N-n}{Nn\overline{M}^2} \right) s_r^2$$

where $\overline{M}$ is unknown and is therefore approximated by

$$\overline{m} = \frac{1}{n} \sum m_i = \frac{130}{20} = 6.5$$

and

$$\begin{aligned} s_r^2 &= \frac{1}{n-1} \sum (y_i - \overline{y} m_i)^2 = \frac{1}{n-1} \left(\sum y_i^2 - 2\overline{y} \sum m_i y_i + \overline{y}^2 \sum m_i^2 \right) \\ &= \frac{1}{20-1} \left(460225 - 2 \cdot (2565/130) \cdot 22285 + (2565/130)^2 \cdot 1118 \right) \approx 845.5607. \end{aligned}$$

Hence, $\hat{V}(\overline{y}) = 0.7922$ so that a bound on the error of estimation is given by

$$B = 2\sqrt{\hat{V}(\overline{y})} = 1.78.$$

In other words, an approximate 95% confidence interval for the average repair cost per saw for the past month is given by 19.73 ± 1.78.

(8.3) Since M is NOT known, we must estimate τ using the estimator $\hat{\tau} = N\overline{y}_t$. In this case we find that

$$\hat{\tau} = N\overline{y}_t = \frac{N}{n} \sum_{i=1}^{n} y_i = \frac{96}{20} \cdot 2565 = 12312.$$

The estimated variance is

$$\hat{V}(\hat{\tau}) = \hat{V}(N\overline{y}_t) = N^2 \left(\frac{N-n}{Nn} \right) s_t^2 \approx 96^2 \left(\frac{96-20}{96 \cdot 20} \right) 6908.62 \approx 2520257.982$$

since

$$\begin{aligned} s_t^2 &= \frac{1}{n-1} \sum (y_i - \overline{y}_t)^2 = \frac{1}{n-1} \left(\sum y_i^2 - n\overline{y}_t^2 \right) \\ &= \frac{1}{20-1} \left(460225 - 20 \cdot (2565/20)^2 \right) \approx 6908.62. \end{aligned}$$

Thus, a bound on the error of estimation is given by

$$B = 2\sqrt{\hat{V}(\hat{\tau})} \approx 3175.06.$$

(8.4) When M is known, we use equation (8.4) to estimate the population total τ. Hence, for $M = 710$, we have

$$\hat{\tau} = M\overline{y} = 710 \cdot \frac{2565}{130} \approx 14008.85.$$

A bound on the error of estimation is therefore given by

$$B = 2\sqrt{\hat{V}(\hat{\tau})} = 2\sqrt{N^2\left(\frac{N-n}{Nn}\right)s_r^2} \approx 1110.80.$$

(8.5) Suppose that the manufacturer wants the bound on the error of estimation to be less than \$2. We find that

$$n = \frac{N\sigma_r^2}{NB^2\overline{M}^2/4 + \sigma_r^2} \approx \frac{Ns_r^2}{NB^2\overline{M}^2/4 + s_r^2} \approx \frac{96 \cdot 845.5607}{96 \cdot 2^2 \cdot (710/96)^2/4 + 845.5607} \approx 13.3$$

so that at least 14 clusters should be selected for his sample next month if he wants to bound the error of estimation to be less than \$2. *(Notice that this is not very surprising. Since $n = 20$ in problem **(8.2)** produced an error of 1.78, we know that we will be able to sample less than 20 in order to have an error of estimation of 2.)*

(8.21) We find from the problem description that $n = 10$, $m_i = 12$, for $i = 1, \ldots, 10$. However, N and M are unknown. Furthermore, since m_i is constant for the sample, and since we are told that each circuit board has 12 microchips on it, we can deduce that $\overline{m} = \overline{M} = 12$. Now, we are also given the listing of the y_i, the number of faulty microchips per circuit board. Hence, we find that

$$\sum y_i = 16,\ \sum y_i^2 = 44,\ \sum m_i = 12 \cdot 10 = 120,\ \sum m_i^2 = 10 \cdot 144 = 1440,\ \sum m_i y_i = 12 \cdot 16 = 192.$$

Thus,

$$\overline{y} = \frac{\sum y_i}{\sum m_i} = \frac{16}{120} \approx 0.1333$$

is an estimate for the *proportion* of defective microchips in the population. Since we do not know N, we can take $N \to \infty$, which eliminates the fpc of $(N-n)/N$, so that a bound on the error of estimation is given by

$$B = 2\sqrt{\hat{V}(\overline{y})} \approx 2\sqrt{\frac{1}{n\overline{M}^2} \cdot s_r^2}.$$

Now,

$$\begin{aligned} s_r^2 &= \frac{1}{n-1}\sum(y_i - \overline{y}m_i)^2 = \frac{1}{n-1}\left(\sum y_i^2 - 2\overline{y}\sum m_i y_i + \overline{y}^2 \sum m_i^2\right) \\ &= \frac{1}{10-1}\left(44 - 2 \cdot (16/120) \cdot 192 + (16/120)^2 \cdot 1440\right) \approx 2.04444 \end{aligned}$$

so that

$$B \approx 2\sqrt{\frac{1}{10 \cdot 12^2} \cdot 2.04444} \approx 0.0754.$$

In other words, an approximate 95% confidence interval for the proportion of defective microchips per circuit board is given by 0.1333 ± 0.0754.