Stat 257.01 Fall 2004
Assignment #7 Solutions

(6.8) There was a misprint in this problem in earlier editions of the text. The earlier editions had $\overline{y} = 169$ while the corrected editions have $\overline{y} = 2.6$. If you use $\overline{y} = 169$, though, then the problem does have a "solution." However, it should be noted that the only way that $\overline{y}$ can equal $\sum y_i^2$ is if $y_i = 0$ for all i, or there is a single data point $y_1 = 1$.

$\boxed{\overline{y} = 169:}$ We begin by finding r, the estimator of the population ratio R. Recall that $r = \overline{y}/\overline{x}$ so that

$$r = \frac{\overline{y}}{\overline{x}} = \frac{169.0}{9.2} \approx 18.4.$$

From the computational formula for $\hat{V}(r)$, and a little algebra, we find that

$$\begin{aligned}\hat{V}(r) &= \left(\frac{N-n}{nN}\right)\left(\frac{1}{\mu_x^2}\right)(s_y^2 + r^2 s_x^2 - 2r\hat{\rho} s_x s_y)\\ &= \left(\frac{N-n}{nN}\right)\left(\frac{1}{\mu_x^2}\right)\left(\frac{1}{n-1}(\sum y_i^2 - 2r\sum x_i y_i + r^2 \sum x_i^2)\right).\end{aligned}$$

Using $\overline{x}$ to approximate μ_x, and plugging in the given data, we conclude that

$$\hat{V}(r) \approx \left(\frac{275-25}{25\cdot 275}\right)\left(\frac{1}{9.2^2}\right)\left(\frac{1}{25-1}(169.0 - 2\cdot 18.4 \cdot 522 + (18.4)^2 \cdot 2240)\right) \approx 13.2.$$

Thus, a reasonable bound on the error of estimation is $2\sqrt{\hat{V}(r)} \approx 7.3$. In other words, an approximate 95% confidence interval for R is 18.4 ± 13.2.

$\boxed{\text{If } \overline{y} = 2.6:}$ We begin by finding r, the estimator of the population ratio R. Recall that $r = \overline{y}/\overline{x}$ so that

$$r = \frac{\overline{y}}{\overline{x}} = \frac{2.6}{9.2} \approx 0.283$$

From the computational formula for $\hat{V}(r)$, and a little algebra, we find that

$$\begin{aligned}\hat{V}(r) &= \left(\frac{N-n}{nN}\right)\left(\frac{1}{\mu_x^2}\right)(s_y^2 + r^2 s_x^2 - 2r\hat{\rho} s_x s_y)\\ &= \left(\frac{N-n}{nN}\right)\left(\frac{1}{\mu_x^2}\right)\left(\frac{1}{n-1}(\sum y_i^2 - 2r\sum x_i y_i + r^2 \sum x_i^2)\right).\end{aligned}$$

Using $\overline{x}$ to approximate μ_x, and plugging in the given data, we conclude that

$$\hat{V}(r) \approx \left(\frac{275-25}{25\cdot 275}\right)\left(\frac{1}{9.2^2}\right)\left(\frac{1}{25-1}(169.0 - 2\cdot 0.283 \cdot 522 + (0.283)^2 \cdot 2240)\right) \approx 0.000948$$

Thus, a reasonable bound on the error of estimation is $2\sqrt{\hat{V}(r)} \approx 0.0616$. In other words, an approximate 95% confidence interval for R is 0.283 ± 0.0616.

(6.11) Again, we begin by finding r, the estimator of the population ratio R. Recall that $r = \overline{y}/\overline{x}$ so that

$$r = \frac{\overline{y}}{\overline{x}} = \frac{\sum y_i}{\sum x_i} = \frac{11458}{10103} \approx 1.134.$$

Thus, we find that

$$\hat{\mu}_y = r\mu_x \approx (1.134)(880) \approx 997.92.$$

As for a bound on the error of estimation, we find that $\sum(y_i - rx_i)^2 \approx 46939.53$ so that $s_r^2 \approx 94.067$. Hence,

$$B = 2\sqrt{\hat{V}(\hat{\mu}_y)} = 2\sqrt{\left(\frac{N-n}{nN}\right) s_r^2} \approx 2\sqrt{\left(\frac{500-12}{12 \cdot 500}\right)(94.067)} \approx 53.65.$$

(6.23) Suppose that x denotes income (in 1982 constant billions) in 1980 and that y denotes income (in 1982 constant billions) in 1989. From the data given in the problem, we find that $n = 6$, $N = 19$, and $\tau_x = 674$. Furthermore,

$$\sum x_i = 245, \quad \sum x_i^2 = 11991, \quad \sum y_i = 327, \quad \sum y_i^2 = 22131, \quad \sum x_i y_i = 16196$$

so that

$$r = \frac{\sum y_i}{\sum x_i} = \frac{327}{245} \approx 1.3347.$$

Finally, let τ_y denote the 1989 total income.

(a) We find $\hat{\tau}_y$, a ratio estimator of τ_y. Hence,

$$\hat{\tau}_y = r\tau_x \approx 1.3347 \cdot 674 \approx 899.58.$$

As for the estimated variance of $\hat{\tau}_y$, we have

$$\begin{aligned}
\hat{V}(\hat{\tau}_y) &= \left(\frac{\tau_x^2}{\mu_x^2}\right)\left(\frac{N-n}{nN}\right)\left(\frac{1}{n-1}\right)\sum(y_i - rx_i)^2 \\
&= N^2\left(\frac{N-n}{nN}\right)\left(\frac{1}{n-1}(\sum y_i^2 - 2r\sum x_i y_i + r^2\sum x_i^2)\right) \\
&\approx 19^2\left(\frac{19-6}{6 \cdot 19}\right)\left(\frac{1}{6-1}(22131 - 2 \cdot 1.3347 \cdot 16196 + 1.3347^2 \cdot 11991)\right) \\
&\approx 2127.94
\end{aligned}$$

so that a bound on the error of estimation is given by $2\sqrt{\hat{V}(\hat{\tau}_y)} \approx 2\sqrt{2127.94} \approx 92.26$.

(b) We find $\hat{\tau}_{yL}$, a regression estimator of τ_y. Since

$$\hat{\tau}_{yL} = N\hat{\mu}_{yL} = N(\overline{y} + b(\mu_x - \overline{x}))$$

where

$$b = \frac{\sum(y_i - \overline{y})(x_i - \overline{x})}{\sum(x_i - \overline{x})^2} = \frac{\sum x_i y_i - n\overline{x}\,\overline{y}}{\sum x_i^2 - n\overline{x}^2},$$

we conclude that

$$b \approx \frac{16196 - 5 \cdot (245/6) \cdot (327/6)}{11991 - 5 \cdot (245/6)^2} \approx 1.387$$

and

$$\hat{\tau}_{yL} \approx 19 \cdot ((327/6) + 1.387((674/19) - (245/6))) \approx 894.24.$$

As for the estimated variance of $\hat{\tau}_{yL}$, we have

$$\begin{aligned}
\hat{V}(\hat{\tau}_{yL}) &= N^2 \left(\frac{N-n}{nN}\right)\left(\frac{1}{n-2}\right)\left[\sum(y_i-\overline{y})^2 - b^2\sum(x_i-\overline{x})^2\right] \\
&= N^2 \left(\frac{N-n}{nN}\right)\left(\frac{1}{n-2}\right)\left[\sum y_i^2 - b^2\sum x_i^2 - n(\overline{y}^2 - b^2\overline{x}^2)\right] \\
&\approx 19^2 \left(\frac{19-6}{6\cdot 19}\right)\left(\frac{1}{6-2}\right)\left[22131 - 1.387^2\cdot 11991 - 6((327/6)^2 - 1.387^2(245/6)^2)\right] \\
&\approx 5015.04
\end{aligned}$$

so that a bound on the error of estimation is given by $2\sqrt{\hat{V}(\hat{\tau}_{yL})} \approx 2\sqrt{5015.04} \approx 141.63$.

(c) We find $\hat{\tau}_{yD}$, a difference estimator of τ_y. Let d_i be the differences between the 1989 incomes and the 1980 incomes, so that $d_i = \{5, 28, 12, 10, 1, 26\}$. We also find

$$\sum d_i = 82, \quad \sum d_i^2 = 1730.$$

Hence,

$$\hat{\tau}_{yD} = N\hat{\mu}_{yD} = N(\mu_x + \overline{d}) = 19(674/19 + 82/6) \approx 933.67.$$

As for the estimated variance of $\hat{\tau}_{yD}$, we have

$$\begin{aligned}
\hat{V}(\hat{\tau}_{yD}) = N^2\left(\frac{N-n}{nN}\right)\left(\frac{1}{n-1}\right)\sum(d_i-\overline{d})^2 &= N^2\left(\frac{N-n}{nN}\right)\left(\frac{1}{n-1}\right)\left(\sum d_i^2 - n\overline{d}^2\right) \\
&\approx 19^2\left(\frac{19-6}{6\cdot 19}\right)\left(\frac{1}{6-1}\right)\left(1730 - 6\cdot(82/6)^2\right) \\
&\approx 5016.84
\end{aligned}$$

so that a bound on the error of estimation is given by $2\sqrt{\hat{V}(\hat{\tau}_{yD})} \approx 2\sqrt{5016.84} \approx 141.66$.

(d) The bound on the error of estimation for the ratio estimator is somewhat smaller than the bounds for the other two estimators, making it somewhat preferable.