From last time,

$$E[E[X|Y]] = E[X] \text{ and}$$

$$V[X] = E_1 V_2[X] + V_1 E_2[X].$$

Apply this to $\hat{t}$:

$$E[\hat{t}] = E_1 E_2[\hat{t}] = E_1 E_2\left[\frac{N}{n}\sum_{i=1}^{n} \hat{t}_i\right]$$

$$= N E_1\left(\frac{1}{n}\sum_{i=1}^{n} E_2[\hat{t}_i]\right)$$

$$= N E_1\left(\frac{1}{n}\sum_{i=1}^{n} E_2[M_i \bar{y}_i]\right)$$

$$= N E_1\left(\frac{1}{n}\sum_{i=1}^{n} M_i E_2[\bar{y}_i]\right)$$

$$= N E_1\left(\frac{1}{n}\sum_{i=1}^{n} M_i \bar{Y}_i\right)$$

$$= N E_1\left(\frac{1}{n}\sum_{i=1}^{n} t_i\right) = N E_1(\bar{t})$$

$$= N \bar{T} = N \frac{t}{N} = t$$

(3)

$$V[\hat{t}] = \underbrace{E_1 V_2[\hat{t}]}_{(2)} + \underbrace{V_1 E_2[\hat{t}]}_{(1)}$$

(1): $V_1 E_2[\hat{t}] = V_1 E_2\left[\frac{N}{n}\sum_{i=1}^{n}\hat{t}_i\right]$

$$= N^2 V_1\left(\frac{1}{n}\sum_{i=1}^{n} E_2[\hat{t}_i]\right)$$

$$= N^2 V_1\left(\frac{1}{n}\sum_{i=1}^{n} t_i\right) = N^2 V_1(\bar{t})$$

(4)

$$= N^2 \frac{S_t^2}{n}\left(1 - \frac{n}{N}\right)$$

(2) $E_1 V_2[\hat{t}] = E_1 V_2\left(\frac{N}{n}\sum_{i=1}^{n}\hat{t}_i\right)$

$$= \frac{N^2}{n^2} E_1\left(\sum_{i=1}^{n} V_2[\hat{t}_i]\right)$$

Note: there are no covariance terms since independent samples are taken in each cluster

$$= \frac{N^2}{n^2} E_1\left(\sum_{i=1}^{n} V_2[M_i \bar{y}_i]\right)$$

$$= \frac{N^2}{n^2} E_1\left(\sum_{i=1}^{n} M_i^2 V_2[\bar{y}_i]\right)$$

(5)

$$= \frac{N^2}{n^2} E_1\left(\sum_{i=1}^{n} M_i^2 \frac{S_i^2}{m_i}\left(1 - \frac{m_i}{M_i}\right)\right)$$

$$= \frac{N^2}{n} E_1\left(\frac{1}{n} \sum_{i=1}^{n} M_i^2 \frac{S_i^2}{m_i}\left(1 - \frac{m_i}{M_i}\right)\right)$$

$\left[E(\bar{a}) = \bar{A} \right.$

$$= \frac{N^2}{n}\left(\frac{1}{N} \sum_{i=1}^{N} M_i^2 \frac{S_i^2}{m_i}\left(1 - \frac{m_i}{M_i}\right)\right)$$

$$= \frac{N}{n} \sum_{i=1}^{N} M_i^2 \frac{S_i^2}{m_i}\left(1 - \frac{m_i}{M_i}\right)$$

(6)

We have shown:

$$V[\hat{t}] = N^2 \frac{S_t^2}{n}\left(1 - \frac{n}{N}\right) + \frac{N}{n} \sum_{i=1}^{N} M_i^2 \frac{S_i^2}{m_i}\left(1 - \frac{m_i}{M_i}\right)$$

How to estimate this?

Try this:

$$\hat{V}[\hat{t}] = \underbrace{N^2 \frac{s_t^2}{n}\left(1 - \frac{n}{N}\right)}_{(1)} + \underbrace{\frac{N}{n} \sum_{i=1}^{n} M_i^2 \frac{s_i^2}{m_i}\left(1 - \frac{m_i}{M_i}\right)}_{(2)}$$

⑦

①: $$E[s_{\hat{t}}^2] = E_1 E_2[s_{\hat{t}}^2]$$

$$= E_1 E_2\left[\frac{1}{n-1}\left(\sum_{i=1}^{n}\hat{t}_i^2 - n\bar{\hat{t}}^2\right)\right]$$

$$= \frac{1}{n-1}E_1\left[\sum_{i=1}^{n}E_2[\hat{t}_i^2] - nE_2(\bar{\hat{t}}^2)\right]$$

$$= \frac{1}{n-1}E_1\left[\sum_{i=1}^{n}[V_2[\hat{t}_i] + (E_2[\hat{t}_i])^2] - n\left(V_2(\bar{\hat{t}}) + (E_2(\bar{\hat{t}}))^2\right)\right]$$

⑧

$$= \frac{1}{n-1}E_1\left[\sum_{i=1}^{n}\left(M_i^2\frac{s_i^2}{m_i}\left(1-\frac{m_i}{M_i}\right) + t_i^2\right) - n\left(\frac{1}{n^2}\sum_{i=1}^{n}M_i^2\frac{s_i^2}{m_i}\left(1-\frac{m_i}{M_i}\right) + \left(\frac{1}{n}\sum_{i=1}^{n}t_i\right)^2\right)\right]$$

$$= \frac{1}{n-1}E_1\left[\underbrace{\left(1-\frac{1}{n}\right)}_{\frac{n-1}{n}}\sum_{i=1}^{n}M_i^2\frac{s_i^2}{m_i}\left(1-\frac{m_i}{M_i}\right) + \underbrace{\sum_{i=1}^{n}t_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n}t_i\right)^2}_{(n-1)s_t^2}\right]$$

$$= E_1\left[\frac{1}{n}\sum_{i=1}^{n}M_i^2\frac{s_i^2}{m_i}\left(1-\frac{m_i}{M_i}\right) + s_t^2\right]$$

$$= \frac{1}{N}\sum_{i=1}^{N}M_i^2\frac{S_i^2}{m_i}\left(1-\frac{m_i}{M_i}\right) + S_t^2$$ to be continued...

Stat 576 HW#5

8 A homeowner with a large library needs to estimate the purchase cost and replacement value of the book collection for insurance purposes. She has 44 shelves containing books, and selects 12 shelves at random. To prepare for the second stage of sampling, she counts the number of books M_i, on the selected shelves. She generates five random numbers between 1 and M_i for each selected shelf, to determine which specific books, numbered from left to right, to examine more closely. She then looks up the replacement value for the sampled books in *Books in Print*. The data are given in the file books.xlsx.

a Draw side-by-side boxplots for the replacement costs of books on each shelf. Does it appear that the means are about the same? The variances?

b Estimate the total replacement cost for the library, and find the standard error of your estimate. What is the estimated coefficient of variation?

c Estimate the average replacement cost per book, along with the standard error. What is the estimated coefficient of variation?