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

$= E_1 E_2 [\,\cdot\,]$

$= N E_1 E_2 \left[\frac{1}{n} \sum_{i=1}^{n} M_i^2 \frac{S_i^2}{m_i}\left(1 - \frac{m_i}{M_i}\right) \right]$

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

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

(2)

$E\left(\hat{V}[\hat{t}]\right) =$

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

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

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

So $\hat{V}[\hat{t}]$ is actually $\uparrow \frac{N}{n}\left(1 - \frac{n}{N}\right) + 1$
BUT unbiased estimator of $V[\hat{t}]$

Single-stage cluster sampling with unequal probabilities

Assume we sample n clusters out of N, WR.

Let z_i be the probability that cluster i is selected on a particular draw.

Let $\hat{t} = \frac{1}{n} \sum_{i=1}^{n} \frac{t_i}{z_i}$

Then $E[\hat{t}] = \frac{1}{n} \sum_{i=1}^{n} E\left[\frac{t_i}{z_i}\right]$

$$= \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{j=1}^{N} \frac{t_j}{z_j} \cdot z_j \right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} t = t$$

Also, $V[\hat{t}] = V\left[\frac{1}{n} \sum_{i=1}^{n} t_i/z_i\right]$

$$= \frac{1}{n^2} \sum_{i=1}^{n} V[t_i/z_i]$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} \underbrace{\left(\sum_{j=1}^{N} \left(\frac{t_j}{z_j} - t\right)^2 \cdot z_j \right)}_{\sigma_z^2} = \frac{\sigma_z^2}{n}$$

Note: if you could set $z_j = \frac{t_j}{t}\psi_j$,

then $\sigma_z^2 = 0$ and your estimator would be perfect.

Since the t_j's are unknown (as is t).

One possibility is to use $\psi_j = \frac{M_j}{K}$

This is called PPS sampling

↑ probabilities proportional to size

$$\hat{t}_{pps} = \frac{1}{n}\sum_{i=1}^{n}\frac{t_i}{\psi_i} = \frac{1}{n}\sum_{i=1}^{n}\frac{t_i}{M_i/K}$$

$$= \frac{K}{n}\sum_{i=1}^{n}\bar{y}_i = K\bar{y}$$

$$V[\hat{t}_{pps}] = \frac{1}{n}\sigma_z^2 = \frac{1}{n}\underbrace{\sum_{j=1}^{N}\left(\frac{t_j}{\psi_j} - t\right)^2\psi_j}_{\sigma_{pps}^2}$$

Estimate σ_{pps}^2 with s_{pps}^2.

s_{pps}^2 is the sample variance of the $\frac{t_i}{\psi_i}$ terms

But $\frac{t_i}{\psi_i} = K\bar{y}_i$

So s^2_{pps} is the sample variance of the $K\bar{y}_i$ terms,

which is $K^2 \cdot s^2_{\bar{y}}$

↑ sample variance of the cluster means

Summary:

$\hat{t}_{pps} = K\bar{\bar{y}}$	$\bar{y}_{pps} = \bar{\bar{y}}$
$V[\hat{t}_{pps}] = \frac{\sigma^2_{pps}}{n}$	$V[\bar{y}_{pps}] = \frac{\sigma^2_{pps}}{nK^2}$
$\hat{V}[\hat{t}_{pps}] = \frac{K^2 s^2_{\bar{y}}}{n}$	$\hat{V}[\bar{y}_{pps}] = \frac{s^2_{\bar{y}}}{n}$

A generalization of the idea behind PPS sampling

Let π_j be the probability that Y_j is in the sample.

Let π_{ij} be the joint probability that Y_i and Y_j are both in the sample

Let $\hat{t}_{HT} = \sum_{i=1}^{n} \frac{y_i}{\pi_i}$ Horvitz-Thompson estimator

Horvitz-Thompson theorem:

$$E(\hat{t}_{HT}) = t$$

Proof: $E[\hat{t}_{HT}] = E\left[\sum_{i=1}^{\hat{}} \frac{y_i}{\pi_i}\right]$

$$= E\left[\sum_{j=1}^{N} \frac{y_j}{\pi_j} u_j\right]$$

let $u_j = \begin{cases} 1 & \text{if } y_j \text{ is in sample} \\ 0 & \end{cases}$

$u_j \sim \text{Bernoulli}(\pi_j)$

$$= \sum_{j=1}^{N} \frac{y_j}{\pi_j} E(u_j)$$

$$= \sum_{j=1}^{N} \frac{y_j}{\pi_j} \pi_j = t$$