

# STAT3003 Problem Sheet 2

ZHENG Weijia (William, 1155124322)

March 8, 2021

## 1 Q1

Note that  $L = 4$ ,  $N = N_1 + N_2 + N_3 + N_4 = 225$ , with  $N_1 = 64$ ,  $N_2 = 43$ ,  $N_3 = 92$ ,  $N_4 = 26$ .  
And  $\hat{p}_1 = \frac{2}{7}$ ,  $\hat{p}_2 = \frac{1}{3}$ ,  $\hat{p}_3 = \frac{8}{21}$ ,  $\hat{p}_4 = \frac{1}{3}$ .

Hence we can have the

$$\hat{p}_{st} = \frac{1}{N} \sum_{i=1}^L N_i \hat{p}_i = 0.3393.$$

Also, by

$$\hat{Var}(\hat{p}_{st}) = \frac{1}{N^2} \sum_{i=1}^L N_i^2 \left(1 - \frac{n_i}{N_i}\right) \frac{1}{n_i - 1} \hat{p}_i (1 - \hat{p}_i) = \frac{197.979}{225^2} = 3.7724 * 10^{-3}.$$

Therefore,  $\sqrt{\hat{Var}(\hat{p}_{st})} = 0.0614$ .

Our goal is to find

$$(\hat{p}_{st} \pm t_{df, 1-\frac{\alpha}{2}} * \sqrt{\hat{Var}(\hat{p}_{st})}).$$

By Satterthwaite's Approximation, we have

$$df \approx \frac{(\sum_{i=1}^L k_i s_i^2)^2}{\sum_{i=1}^L \frac{(k_i s_i^2)^2}{n_i - 1}}, k_i = \frac{N_i(N_i - n_i)}{N^2 n_i},$$

and hence  $df = 45$ .

Then  $t_{df, 1-\frac{\alpha}{2}} = t_{45, 0.975} = 2.0141$ . Therefore, then result should be

$$(0.3393 \pm 2.0141 * 0.0614)$$

Done.

## 2 Q2

Note that  $L = 3$ ,  $c = 500$ ,  $c_0 = 0$ .

Using

$$n_i = \frac{(c - c_0) \frac{N_i S_i}{\sqrt{c_i}}}{\sum_{i=1}^L N_i S_i \sqrt{c_i}}.$$

Note that we have the  $\sigma_i$  are given, hence by  $\hat{\sigma}_i^2 = \frac{N_i}{N_i - 1} \sigma_i^2$ . Then we have  $s_i = \hat{\sigma}_i$ . With  $s_1 = 1.5067$ ,  $s_2 = 1.8134$ ,  $s_3 = 1.8235$ .

Which gives that  $n_1 = 18.1510, n_2 = 7.9581, n_3 = 3.8247$ . By doing rounding, we have  $n_1 = 18, n_2 = 8, n_3 = 4$ . Under this case, we have the cost to be 506.

For saving the cost, we need to reduce the sample size in North America by 1.

Then we have  $n_1 = 17, n_2 = 8, n_3 = 4$ . Under this case, the cost is 497, which is okay.

And by

$$Var(\bar{Y}_{st}) = \frac{1}{N^2} \sum_{i=1}^L N_i^2 \frac{\sigma_i^2}{n_i} \left( \frac{N_i - n_i}{N_i - 1} \right) = 0.08825 < 0.1.$$

Which can satisfy the corporation's requirement. Hence they can be happy.  
Done.

### 3 Q3

Recall that the function we want to optimize is

$$f(n_1, n_2, \dots, n_L) = \frac{1}{N^2} \sum_{i=1}^L N_i^2 \frac{\sigma_i^2}{n_i} \left( \frac{N_i - n_i}{N_i - 1} \right).$$

And the constraint is

$$g(n_1, n_2, \dots, n_L) = c - (c_0 + c_1 n_1 + \dots + c_L n_L) = 0.$$

The lagrangian is

$$L = f - \lambda g.$$

Note that

$$\frac{\partial f}{\partial n_j} = \frac{1}{N^2} \frac{\partial (N_j^3 \sigma_j^2)}{n_j (N_j - 1)} = - \frac{N_j^3 \sigma_j^2}{N^2 (N_j - 1) n_j^2}.$$

And

$$\frac{\partial g}{\partial n_j} = -c_j.$$

Therefore we can have

$$\lambda c_j = \frac{1}{n_j^2} \frac{N_j^3 \sigma_j^2}{N^2 (N_j - 1)}.$$

Which implies

$$n_j = \frac{1}{\sqrt{\lambda}} \sqrt{\frac{N_j}{N_j - 1} \frac{N_j \sigma_j}{\sqrt{c_j} N}}.$$

Recall the constraint  $g=0$ . We can have

$$\sum_{j=1}^L c_j n_j = (c - c_0) = \sum_{j=1}^L \frac{1}{\sqrt{\lambda}} \sqrt{\frac{N_j}{N_j - 1} \frac{N_j \sigma_j}{\sqrt{c_j} N}}.$$

Hence

$$\frac{1}{\sqrt{\lambda}} = \frac{c - c_0}{\sum_{j=1}^L \sqrt{\frac{N_j}{N_j - 1} \frac{N_j \sigma_j}{\sqrt{c_j} N}}}.$$

And therefore,

$$n_j = \frac{c - c_0}{\sum_{j=1}^L \sqrt{\frac{N_j}{N_j-1} \frac{N_j \sigma_j}{\sqrt{c_j} N}}} \sqrt{\frac{N_j}{N_j-1} \frac{N_j \sigma_j}{\sqrt{c_j} N}} = \frac{(c - c_0) \left( \sqrt{\frac{N_i}{N_i-1} \frac{N_i \sigma_i}{\sqrt{c_i}}} \right)}{\sum_{i=1}^L \sqrt{\frac{N_i}{N_i-1} \frac{N_i \sigma_i}{\sqrt{c_i}}}}.$$

Which is what we want.

Done.

#### 4 Q4

By the formulae  $\hat{\mu} = \frac{\hat{\tau}}{M}$ , we can investigate the  $\hat{\tau}$  and then go from  $\hat{\tau}$  to get  $\hat{\mu}$ .  
Note that  $M = 3500, N = 108, n = 25$ .

Note that

$$\hat{\tau} = N \frac{1}{n} \sum_{i=1}^n Y_i = 151014.24.$$

And

$$t_{n-1, 1-\frac{\alpha}{2}} = t_{24, 0.975} = 2.06.$$

Also, we have

$$\hat{Var}(\hat{\tau}) = \hat{Var}(N\bar{Y}) = N^2 \left(1 - \frac{n}{N}\right) \frac{\hat{\sigma}_c^2}{n}.$$

Where the  $\hat{\sigma}_c^2$  is the sample variance of cluster totals. Hence  $\hat{\sigma}_c^2 = 149422.2016$ . And

$$\hat{Var}(\hat{\tau}) = 53576824.61.$$

Therefore  $\sqrt{\hat{Var}(\hat{\tau})} = 7319.6192$ . Hence the 95% CI for  $\tau$  is

$$(151014.24 \pm 2.06 * 7319.6192)$$

. Therefore 95% CI for  $\mu$  is  $(43, 147 \pm 4.3081)$ .

#### 5 Q5

Note that  $\hat{\mu} = \frac{N}{M} \bar{Y}$ . Hence

$$Var(\hat{\mu}) = Var(\bar{Y}) \frac{N^2}{M^2} = \frac{N^2}{M^2} \frac{N-n}{N-1} \frac{1}{n} \sigma_c^2.$$

Hence the width  $2d = 2t_{n-1, 1-\frac{\alpha}{2}} \sqrt{Var(\hat{\mu})} \leq 4$ . Solving this inequality, we then have

$$n \geq \frac{1}{\frac{1}{N} + \frac{4(N-1)M^2}{N^3 \sigma_c^2 t_{n-1, 1-\frac{\alpha}{2}}}}.$$

Now we plug in  $M = 3500$ , and  $\sigma_c^2$  can be used by 149422.2016 and  $N$  be plugged in by 108. Then we can see the minimal  $n$  such that the above inequality hold is  $n = 44$ .

Done.

#### 6 Q6