STA 304H1F-1003H Fall 2019
Assignment 2-Question 1-Solution

Question 1. (15 marks)

Consider a hypothetical population given in the table below. Consider stratified sampling with SRSs of size $n_1 = n_2 = 3$ taken from the two strata, respectively.

stratum 1				
Unit	1	2	3	4
y	1	4	4	5

stratum 2				
Unit	5	6	7	8
y	5	6	9	11

Solution

Part a) (2 marks) What are the values of μ_1 and μ_2?

$$\text{Using} \quad \mu_{\mathbf{i}} = \frac{\sum_{\mathbf{j=1}}^{\mathbf{N_i}} \mathbf{y_j}}{\mathbf{N_i}} \quad \text{for each strata i, we find that} \quad \boldsymbol{\mu_1 = 3.5}, \text{ and } \boldsymbol{\mu_2 = 7.75}$$

(1mk each)

Part b) (2 marks) What are the values of σ_1^2 and σ_2^2?

$$\text{Using} \quad \sigma_{\mathbf{i}}^{\mathbf{2}} = \frac{\sum_{\mathbf{j=1}}^{\mathbf{N_i}} (\mathbf{y_j} - \mu_{\mathbf{j}})^{\mathbf{2}}}{\mathbf{N_i}} \quad \text{for each strata i, we find that} \quad \boldsymbol{\sigma_1^2 = 2.25}, \text{ and } \boldsymbol{\sigma_2^2 = 5.6875}$$

(1mk each)

Part c) (2 marks) What are the values of $\mathbf{V}(\hat{\tau}_1)$ and $\mathbf{V}(\hat{\tau}_2)$?

There are two approaches to find $\mathbf{V}(\hat{\tau}_1)$ and $\mathbf{V}(\hat{\tau}_2)$.

Method 1: We can simply apply the formula of the variance of $\hat{\tau}$ for SRS in each strata:

$$\mathbf{V}(\hat{\tau}_1) = N_1^2 \times \frac{N_1 - n_1}{N_1 - 1} \times \frac{\sigma_1^2}{n_1} = 4^2 \times \frac{4-3}{4-1} \times \frac{2.25}{3} = \boxed{4}$$

(1mk)

$$\mathbf{V}(\hat{\tau}_2) = N_2^2 \times \frac{N_2 - n_2}{N_2 - 1} \times \frac{\sigma_2^2}{n_2} = 4^2 \times \frac{4-3}{4-1} \times \frac{5.69}{3} = \boxed{10.111}$$

(1mk)

Method 2: The second method is based on the sampling distributions of $\hat{\tau}_1$ $\hat{\tau}_1$ and $\hat{\tau}_2$

The sampling distributions of $\hat{\tau}_1$ is given Table 1 below

Table 1: Sampling Dist. of total estimator for stratum 1

SRS	y1	y2	y3	Prob	$\widehat{\mu}_1$	$\widehat{\tau}_1 = N_1 \times \widehat{\mu}_1$	Prob*$\times\widehat{\tau}_1$	Prob$\times\widehat{\tau}_1^2$
1	1	4	4	0.25	3.00000	12.00000	3.00000	36.00000
2	1	4	5	0.25	3.33333	13.33333	3.33333	44.44444
3	1	4	5	0.25	3.33333	13.33333	3.33333	44.44444
4	4	4	5	0.25	4.33333	17.33333	4.33333	75.11111

The expected value of $\widehat{\tau}_1$, obtained by summing the column 8 of Table 1

$$\mathbf{E}(\widehat{\tau}_1) = \sum_{i=1}^{4} \text{Prob} * \widehat{\tau}_i = 14$$

and the expected value of $\widehat{\tau}_1^2$, obtained by summing the column 9 of Table 1

$$\mathbf{E}(\widehat{\tau}_1^2) = \sum_{i=1}^{4} \text{Prob} * \widehat{\tau}_i^2 = 200$$

Hence, we have that

$$\mathbf{V}(\widehat{\tau}_1) = \mathbf{E}(\widehat{\tau}_1^2) - [\mathbf{E}(\widehat{\tau}_1)]^2 = 200 - (14)^2 = \boxed{4}$$

which is the same value of $\mathbf{V}(\widehat{\tau}_1)$ as the one found in method 1 above.

Using the sampling distribution of $\widehat{\tau}_2$, given by Table 2 below,

Table 2: Sampling Dist. of total estimator for stratum 2

SRS	y1	y2	y3	Prob	$\widehat{\mu}_2$	$\widehat{\tau}_2 = N_2 \times \widehat{\mu}_2$	Prob*$\times\widehat{\tau}_2$	Prob$\times\widehat{\tau}_2^2$
1	5	6	9	0.25	6.66667	26.66667	6.66667	177.7778
2	5	6	11	0.25	7.33333	29.33333	7.33333	215.1111
3	5	9	11	0.25	8.33333	33.33333	8.33333	277.7778
4	6	9	11	0.25	8.66667	34.66667	8.66667	300.4444

the expected value of $\widehat{\tau}_2$, obtained by summing the column 8 is

$$\mathbf{E}(\widehat{\tau}_2) = \sum_{i=1}^{4} \text{Prob} * \widehat{\tau}_i = 31$$

and the expected value of $\widehat{\tau}_2^2$ obtained by summing the column 8 is

$$\mathbf{E}(\widehat{\tau}_2^2) = \sum_{i=1}^{4} \text{Prob} * \widehat{\tau}_i^2 = 971.111$$

hence the variance of $\widehat{\tau}_2$ is:

$$\mathbf{V}(\widehat{\tau}_2) = \mathbf{E}(\widehat{\tau}_2^2) - [\mathbf{E}(\widehat{\tau}_2)]^2 = 971.111 - (31)^2 = \boxed{10.111}$$

which is the same value of $\mathbf{V}(\widehat{\tau}_2)$ as the one found in method 1 above.

Part d) (2 marks) What are the values of $\mathbf{V}(\widehat{\tau}_{str})$ and $\mathbf{V}(\widehat{\mu}_{str})$?

The variance $\mathbf{V}(\widehat{\tau}_{str})$ is

$$\mathbf{V}(\widehat{\tau}_{str}) = \mathbf{V}(\widehat{\tau}_1) + \mathbf{V}(\widehat{\tau}_2) = 4 + 10.111 = \boxed{14.111}$$

(1mk)

The variance $\mathbf{V}(\widehat{\mu}_{str})$ is

$$\mathbf{V}(\widehat{\mu}_{str}) = \frac{\mathbf{V}(\widehat{\tau}_{str})}{N^2} = \frac{14.111}{8^2} = \boxed{0.22}$$

(1mk)

Part e) (3 marks) Write out all possible stratified SRSs.

For each strata, the total number of SRS of size n=3 elements is 4, so there are 4*4=16 possible stratified SRSs. Below the list of these 16 stratified SRSs.

Table 3: List of all possible stratified SRSs

Stratified SRS	y1	y2	y3	y4	y5	y6
1	1	4	4	5	6	9
2	1	4	4	5	6	11
3	1	4	4	5	9	11
4	1	4	4	6	9	11
5	1	4	5	5	6	9
6	1	4	5	5	6	11
7	1	4	5	5	9	11
8	1	4	5	6	9	11
9	1	4	5	5	6	9
10	1	4	5	5	6	11
11	1	4	5	5	9	11
12	1	4	5	6	9	11
13	4	4	5	5	6	9
14	4	4	5	5	6	11
15	4	4	5	5	9	11
16	4	4	5	6	9	11

(3 mks)

Prat f) (2 marks) For each stratified sample, calculate $\widehat{\mu}$ and $\widehat{\tau}$.

For each stratified SRS:

the estimate of μ is obtained by using the following formula

$$\widehat{\mu} = \mathbf{W_1}\widehat{\mu}_\mathbf{1} + \mathbf{W_2}\widehat{\mu}_\mathbf{2} = \frac{\mathbf{4}}{\mathbf{8}}\mathbf{\overline{y}_1} + \frac{\mathbf{4}}{\mathbf{8}}\mathbf{\overline{y}_2} = \mathbf{0.5\overline{y}_1 + 0.5\overline{y}_2}$$

and the estimate of τ is obtained by using the following formula

$$\widehat{\tau} = \widehat{\tau}_\mathbf{1} + \widehat{\tau}_\mathbf{2} = \mathbf{N_1\overline{y}_1 + N_2\overline{y}_2 = 4\overline{y}_1 + 4\overline{y}_2}$$

Below the values of $\widehat{\mu}$ and $\widehat{\tau}$ for all 16 stratified samples

Table 4: Estimates of population mean and total for each stratified sampls

Stratified SRS	y1	y2	y3	y4	y5	y6	$\widehat{\mu}_1$	$\widehat{\mu}_2$	$\widehat{\mu}$	$\widehat{\tau}$
1	1	4	4	5	6	9	3.00000	6.66667	4.83333	38.66667
2	1	4	4	5	6	11	3.00000	7.33333	5.16667	41.33333
3	1	4	4	5	9	11	3.00000	8.33333	5.66667	45.33333
4	1	4	4	6	9	11	3.00000	8.66667	5.83333	46.66667
5	1	4	5	5	6	9	3.33333	6.66667	5.00000	40.00000
6	1	4	5	5	6	11	3.33333	7.33333	5.33333	42.66667
7	1	4	5	5	9	11	3.33333	8.33333	5.83333	46.66667
8	1	4	5	6	9	11	3.33333	8.66667	6.00000	48.00000
9	1	4	5	5	6	9	3.33333	6.66667	5.00000	40.00000
10	1	4	5	5	6	11	3.33333	7.33333	5.33333	42.66667
11	1	4	5	5	9	11	3.33333	8.33333	5.83333	46.66667
12	1	4	5	6	9	11	3.33333	8.66667	6.00000	48.00000
13	4	4	5	5	6	9	4.33333	6.66667	5.50000	44.00000
14	4	4	5	5	6	11	4.33333	7.33333	5.83333	46.66667
15	4	4	5	5	9	11	4.33333	8.33333	6.33333	50.66667
16	4	4	5	6	9	11	4.33333	8.66667	6.50000	52.00000

(2 mks): 1mk for the colum of Mean and 1 mk for the column of Total

Part g) (2 marks) Use the sampling distribution of $\widehat{\mu}$ and $\widehat{\tau}$ to verify that $\widehat{\mu}$ and $\widehat{\tau}$ are an unbiased estimator of μ and τ.

The following table gives sampling distribution of $\widehat{\mu}$ and $\widehat{\tau}$, and their associated probabilities

Table 5: Sampling Distribution of muhat and tauhat with associated probabilities

Stratified SRS	$\widehat{\mu}$	$\widehat{\tau}$	Prob
1	4.83333	38.66667	0.0625
2	5.16667	41.33333	0.0625
3	5.66667	45.33333	0.0625
4	5.83333	46.66667	0.0625
5	5.00000	40.00000	0.0625
6	5.33333	42.66667	0.0625
7	5.83333	46.66667	0.0625
8	6.00000	48.00000	0.0625
9	5.00000	40.00000	0.0625
10	5.33333	42.66667	0.0625
11	5.83333	46.66667	0.0625
12	6.00000	48.00000	0.0625
13	5.50000	44.00000	0.0625
14	5.83333	46.66667	0.0625
15	6.33333	50.66667	0.0625
16	6.50000	52.00000	0.0625

Based on Table 5 above, the expected value of $\widehat{\mu}$ is given by

$$\mathbf{E}(\widehat{\mu}) = \text{Mean}_1 \times \text{Prob}_1 + \cdots + \text{Mean}_{16} \times \text{Prob}_{16} = 4.83333 \times 0.0625 + \cdots + 6.5 \times 0.0625 = \boxed{5.625}$$

In the other hand, the true population mean is given by

$$\mu = W_1 * \mu_1 + W_2 \times \mu_2 = \frac{4}{8} \times 3.5 + \frac{4}{8} \times 7.75 = \boxed{5.625}.$$

Since $\boxed{\mathbf{E}(\widehat{\mu}) = \mu}$, this means $\widehat{\mu}$ is unbiased estimator for μ.

(1mk)

Similary, based on Table 5 above, the expected value of $\widehat{\tau}$ is given by

$$\mathbf{E}(\widehat{\tau}) = \text{Total}_1 \times \text{Prob}_1 + \cdots + \text{Total}_{16} \times \text{Prob}_{16} = 38.66667 \times 0.0625 + \cdots + 52 \times 0.0625 = \boxed{45}$$

In the other hand, the true population total is given by

$$\tau = N_1 * \mu_1 + N_2 \times \mu_2 = 8 \times 3.5 + 8 \times 7.75 = \boxed{45}.$$

Since $\boxed{\mathbf{E}(\widehat{\tau}) = \tau}$, this means $\widehat{\tau}$ is unbiased estimator for μ.

(1mk)

End of Question 1