STA 304H1F-1003H Fall 2019

Assignment 2-Question 2-Solution

Question 2. **(12 marks)**

Suppose that a city has 90000 dwelling units, of which 35000 are houses, 45000 are apartments, and 10000 are condominiums. We want to estimate the overall proportion (p) of households in which energy conservation is practiced, with a bound on the error of estimation equal 0.1. The cost for obtaining an observation is \$ 9 for houses, 10\$ for appartment, and 16 \$ for condominiums. Suppose that from an earlier study, we know that 47% of house dwellers, 23% of appartment dwellers, and 3% of condominum residents practice energy conservation.

(a) (4 marks) Using a proportional allocation, find the strata sample sizes, n_1, n_2, and n_3, and the sample size n.

$N_1 = 3.5 \times 10^4$, $N_2 = 4.5 \times 10^4$, $N_3 = 10^4$, and $N = N_1 + N_2 + N_3 = 9 \times 10^4$

The bound on the error of estimation is $B = 0.1 \implies D = \frac{0.1^2}{4} = 0.0025$

Applying, $a_i = \frac{N_i}{\sum_{i=1}^{3} N_i} = \frac{N_i}{N}$, the proportional allocation fractions are:

$$a_1 = 0.389, \qquad a_2 = 0.5, \qquad \text{and} \qquad a_3 = 0.111.$$

The strata population variances are:

$$\sigma_1^2 = p_1 q_1 = 0.2491, \quad \sigma_2^2 = p_2 q_2 = 0.1771, \quad \text{and} \quad \sigma_3^2 = p_3 q_3 = 0.0291$$

To calculate the sample size n, we need the following quantities:

$$\sum_{i=1}^{3} N_i \sigma_i^2 = (3.5 \times 10^4) \times 0.2491 + (4.5 \times 10^4) \times 0.1771 + (10^4) \times 0.0291 = 1.6979 \times 10^4$$

$$\sum_{i=1}^{3} \frac{N_i^2 \sigma_i^2}{a_i} = \frac{(3.5 \times 10^4)^2 \times 0.2491}{0.389} + \frac{(4.5 \times 10^4)^2 \times 0.1771}{0.5} + \frac{(10^4)^2 \times 0.0291}{0.389} = 1.52811 \times 10^9$$

and

$$N^2 \times D = (9 \times 10^4)^2 \times 0.0025 = 2.025 \times 10^7$$

Then, **the sample size required** is:

$$n = \frac{\sum_{i=1}^{3} \frac{N_i^2 \sigma_i^2}{a_i}}{N^2 \times D + \sum_{i=1}^{3} N_i \sigma_i^2} = 75.4.$$

This means the sample size required is n=76

(1mk)

The strata 1 sample size is $n_1 = a_1 \times n = 0.38889 \times 76 = 29.55556 \sim 30$

(1mk)

The strata 2 sample size is $n_2 = a_2 \times n = 0.5 \times 76 = 38$

(1mk)

The strata 3 sample size is $n_3 = a_3 \times n = 0.11111 \times 76 = 8.44444 \sim 8$

(1mk)

Note: For the entire sample size, we round up. However, in order to sum up to n, the strata sample size should be rounded to the nearest integer.

(b) (4 marks) Using a optimal allocation, find the strata sample sizes, n_1, n_2, and n_3, and the sample size n.

The bound on the error of estimation is $B = 0.1 \implies D = \frac{0.1^2}{4} = 0.0025$.

The strata population variance for proportion are:

$\sigma_1^2 = p_1 q_1 = 0.2491, \quad \sigma_2^2 = p_2 q_2 = 0.1771$, and $\quad \sigma_3^2 = p_3 q_3 = 0.0291$

The strata population standard deviation for proportion are:

$\sigma_1 = \sqrt{p_1 q_1} = 0.4990992, \quad \sigma_2 = \sqrt{p_2 q_2} = 0.4208325$, and $\quad \sigma_3 = \sqrt{p_3 q_3} = 0.1705872$

We have

$\frac{N_1 \sigma_1}{\sqrt{c_1}} = 5822.824,$

$\frac{N_2 \sigma_2}{\sqrt{c_2}} = 5988.552$

$\frac{N_3 \sigma_3}{\sqrt{c_3}} = 426.468$

By summing, we obtain $\quad \sum_{k=1}^{3} \frac{N_k \sigma_k}{\sqrt{c_k}} = 5822.824 + 5988.552 + 426.468 = 1.2237844 \times 10^4$

We also have

$N_1 \sigma_1 \sqrt{c_1} = 5.2405415 \times 10^4,$

$N_2 \sigma_2 \sqrt{c_2} = 5.9885516 \times 10^4$

$N_3 \sigma_3 \sqrt{c_3} = 6823.489$

By summing, we obtain $\quad \sum_{i=1}^{3} N_i \sigma_i \sqrt{c_i} = 5.2405415 \times 10^4 + 5.9885516 \times 10^4 + 6823.489 = 1.1911442 \times 10^5$

To calculate the sample size n, we also need the following quantities:

$$\sum_{i=1}^{3} N_i \sigma_i^2 = (3.5 \times 10^4) \times 0.2491 + (4.5 \times 10^4) \times 0.1771 + (10^4) \times 0.0291 = 1.6979 \times 10^4$$

and

$$N^2 \times D = (9 \times 10^4)^2 \times 0.0025 = 2.025 \times 10^7$$

Then, **the sample size required** is:

$$n = \frac{\left(\sum_{k=1}^{3} \frac{N_k \sigma_k}{\sqrt{c_k}} \right) \times \left(\sum_{k=1}^{3} N_k \sigma_k \sqrt{c_k} \right)}{N^2 \times D + \sum_{i=1}^{3} N_i \sigma_i^2} = 71.93 \sim 72.$$

This means the sample size required is n=72

(1mk)

Applying, $a_i = \frac{\frac{N_i \sigma_i}{\sqrt{c_i}}}{\sum_{k=1}^{3} \frac{N_k \sigma_k}{\sqrt{c_3}}}$, the optimal allocation fractions are:

$$a_1 = 0.476, \qquad a_2 = 0.489, \qquad \text{and} \quad a_3 = 0.035$$

.

The optimal allocations are:

The strata 1 sample size is $n_1 = a_1 \times n = 0.4758 \times 72 = 34.25794 \sim 34$ **(1mk)**

The strata 2 sample size is $n_2 = a_2 \times n = 0.48935 \times 72 = 35.23298 \sim 35$ **(1mk)**

The strata 3 sample size is $n_3 = a_3 \times n = 0.03485 \times 72 = 2.50908 \sim 3$ **(1mk)**

(c) (4 marks) Using a Neyman allocation, find the strata sample sizes, n_1, n_2, and n_3, and the sample size n.

The bound on the error of estimation is $B = 0.1 \implies D = \frac{0.1^2}{4} = 0.0025$.

The strata population variance for proportion are:

$\sigma_1^2 = p_1 q_1 = 0.2491, \quad \sigma_2^2 = p_2 q_2 = 0.1771$, and $\quad \sigma_3^2 = p_3 q_3 = 0.0291$

The strata population standard deviation for proportion are:

$\sigma_1 = \sqrt{p_1 q_1} = 0.4990992, \quad \sigma_2 = \sqrt{p_2 q_2} = 0.4208325$, and $\quad \sigma_3 = \sqrt{p_3 q_3} = 0.1705872$

We have

$$N_1\sigma_1 = 1.7468472 \times 10^4, \qquad N_2\sigma_2 = 1.8937463 \times 10^4, \qquad N_3\sigma_3 = 1705.872$$

By summing, we obtain

$$\sum_{k=1}^{3} \frac{N_k \sigma_k}{\sqrt{c_k}} = 1.7468472 \times 10^4 + 1.8937463 \times 10^4 + 1705.872 = 3.8111807 \times 10^4$$

To calculate the sample size n, we also need the following quantities:

$$\sum_{i=1}^{3} N_i \sigma_i^2 = (3.5 \times 10^4) \times 0.2491 + (4.5 \times 10^4) \times 0.1771 + (10^4) \times 0.0291 = 1.6979 \times 10^4$$

and

$$N^2 \times D = (9 \times 10^4)^2 \times 0.0025 = 2.025 \times 10^7$$

Then, **the sample size required** is:

$$n = \frac{\left(\sum_{k=1}^{3} N_k \sigma_k\right)^2}{N^2 \times D + \sum_{i=1}^{3} N_i \sigma_i^2} = 71.67 \sim 72.$$

This means the sample size required is n=72

(1mk)

Applying, $a_i = \frac{N_i \sigma_i}{\sum_{k=1}^{3} N_k \sigma_k}$, the Neyman allocation fractions are:

$$a_1 = 0.458, \qquad a_2 = 0.497, \qquad \text{and} \quad a_3 = 0.045$$

.

The Neyman sample size allocations are then:

The strata 1 sample size is $n_1 = a_1 \times n = 0.45835 \times 72 = 33.00106 \sim 33$ **(1mk)**

The strata 2 sample size is $n_2 = a_2 \times n = 0.49689 \times 72 = 35.77624 \sim 36$ **(1mk)**

The strata 3 sample size is $n_3 = a_3 \times n = 0.04476 \times 72 = 3.2227 \sim 3$ **(1mk)**

R code

```
# strata 1
N1<-35000
c1<-9
p1<-0.47

# strata 2
N2<-45000
c2<-10
p2<-0.23

# strata 3
N3<-10000
c3<-16
p3<-0.03

N<-c(N1,N2,N3)
c<-c(c1,c2,c3)
p<-c(p1,p2,p3)
q<-1-p
sigma2<-p*q
sigma<-sqrt(p*q)

B<-0.1
D<-B^2/4

# two usefull terms for sample size calculation

Den_term1<-sum(N)^2*D
Den_term2<-sum(N*sigma2)

Den<-Den_term1+Den_term2


# (a) Proportional allocation
ap<-N/sum(N)
Num<-sum( (N^2*sigma2)/ap )
# sample size
np<-Num/Den
# round up
np<-ifelse(round(np)-np>0, round(np), round(np)+1)
# proportion allocation
np1<-round(np*ap[1])
np2<-round(np*ap[2])
np3<-round(np*ap[3])
#c(np,np1,np2,np3)

# (b) Optimal allocation
ao<-( (N*sigma)/sqrt(c) )/sum( ((N*sigma)/sqrt(c)) )
Num<-sum( (N^2*sigma2)/ao )
no<-Num/Den
# sample size
```

```
no<-ifelse(round(no)-no>0, round(no), round(no)+1)
# proportion allocation
no1<-round(no*ao[1])
no2<-round(no*ao[2])
no3<-round(no*ao[3])
#c(no,no1,no2,no3)

# (c) Neyman allocation
an<-N*sigma/sum(N*sigma)
Num<-sum( (N^2*sigma2)/an )
nn<-Num/Den
# sample size
nn<-ifelse(round(nn)-nn>0, round(nn), round(nn)+1)
# proportion allocation
nn1<-round(nn*an[1])
nn2<-round(nn*an[2])
nn3<-round(nn*an[3])
#c(nn,nn1,nn2,nn3)
```