Professional Maters in

Applied Statistics and Data Science

(PM-ASDS)

An Assignment on

Systematic Sampling and Cluster Sampling

Submitted to:

Mohammad Alamgir Kabir, PhD
Professor & Chairman,
Department of Statistics, JU

Submitted by:

Mohammad Saiduzzaman Sayed
ID: 20215063
Batch: 5th, Sec: A
Department of Statistics, JU

Course Name: Sampling Methodology
Course Code: PM-ASDS02

Example of A4_Lecture on systematic sampling

Example: Given below are the daily milk yield (in liters) records of the first lactation of a specified cow belonging to the Tharparkar herd maintained at the Government Cattle Farm.
(The table file name is “systematic_sampling_table.xlsx”. The table is attached to this folder.)
Find the systematic sampling of size 29. Estimate mean yield per day and variance yield.

Solution:

```
import pandas as pd
import numpy as np

dataset = pd.read_excel("systematic_sampling_table.xlsx")
N= len(dataset)
n = 29
k = N/n
print("Interval = ", k)

#Generate systematic sampling
def systematic_sampling(dataset, step):
    indexes = np.arange(0, N, step=step)
    systematic_sample = dataset.iloc[indexes]
    return systematic_sample
systematic_sample = systematic_sampling(dataset, k)

#Sample Table
sample = {"C1":
[10,8,9,10,18,16,15,17,22,19,18,18,20,21,21,20,18,16,16,19,22,18,21,19,14,15,17,12,8],
         "C2":
[11,11,10,11,16,16,14,25,21,19,19,18,18,22,21,21,21,19,14,16,22,20,20,18,13,16,18,16,9],
```

```
        "C3":
[14,10,11,11,13,13,14,22,21,19,21,22,20,22,19,20,20,15,15,16,21,17,20,20
,16,18,17,10,4],
        "C4":
[10,6,11,14,14,16,15,23,23,19,20,22,21,20,20,21,22,15,17,20,22,20,16,19,
16,18,16,13,5],
        "C5":
[14,9,13,15,14,17,17,19,21,19,17,22,21,21,21,20,21,16,16,20,21,20,16,18,
16,15,13,8,6],
        "C6":
[9,8,12,12,15,14,15,18,19,19,16,20,20,21,20,19,21,19,20,18,21,21,15,16,1
8,18,14,8,6],
        "C7":
[10,7,12,17,16,16,16,16,19,19,18,20,18,21,21,21,21,12,15,21,21,21,18,14,
16,16,13,6,4]}

sample = pd.DataFrame(sample)
print(sample)

#population mean
p_mean = dataset["Milk yield"].mean()
print("Population Mean = ","%.2f"% p_mean)

#S_square
S_square = dataset["Milk yield"].var()
print("S_square = ", "%.2f"% S_square)

#S_wsy_square
t_var = sample["C1"].var() + sample["C2"].var() + sample["C3"].var() +
sample["C4"].var() + sample["C5"].var() + sample["C6"].var() +
sample["C7"].var()

S_wsy_2 = t_var / k
print("S_wsy_2 = ", "%.2f"% S_wsy_2)

#var
var = ((N-1)/N * S_square) - (((k*(n-1))/N) * S_wsy_2)
```

```
print("var = ", "%.2f"% var)
```

Output:

```
Interval =  7.0
Population Mean =  16.60
S_square =  18.99
S_wsy_2 =  19.44
var =  0.13
```

Example of A5_Lecture on cluster sampling

Example: A social researcher wishes to estimate the average number of male children in a given community. For this purpose, he prepared a list of N=400 geographical clusters of M=10 households each and a simple random sample of n=4 clusters was selected. The relevant data appear in the accompanying table. Estimate the average number of male children per household for the community and hence obtain an estimate of the variance.

Table: Sample of male children in 10 households.

(The table file name is “cluster_example_table.xlsx”. The table is attached to this folder.)

Solution:

```
import pandas as pd

dataset = pd.read_excel("cluster_example_table.xlsx")

df = pd.DataFrame(dataset)

print(df)

I_mean = df["I"].mean()
print("I_mean = ", I_mean)

II_mean = df["II"].mean()
print("II_mean = ", II_mean)
```

```
III_mean = df["III"].mean()
print("III_mean = ", III_mean)

IV_mean = df["IV"].mean()
print("IV_mean = ", IV_mean)


n = 4
M = 10
N = 400


#Hence an estimate of the average number of male children in the community is
ybar_cl = (I_mean+II_mean+III_mean+IV_mean) / n
print("ybar_cl = ", ybar_cl)


S_b_2 = 1/(n-1) * ((I_mean-ybar_cl)**2 + (II_mean-ybar_cl)**2 + (III_mean-ybar_cl)**2 + (IV_mean-ybar_cl)**2)
print("S_b_2 = ", "%.3f" % S_b_2)


var = (N-n)/(N*n) * S_b_2
print("var = ", "%.4f" % var)
```

Output:

```
I_mean =  1.9
II_mean =  2.0
III_mean =  1.6
IV_mean =  2.0
ybar_cl =  1.875
```

S_b_2 = 0.036

var = 0.0089

Assignment 2:

In a small city there are 5 primary-cum-secondary schools with 10 classes each. The number of students by classes are given in the following Table:

(The table file name is “assignment2_table.xlsx”. The table is attached to this folder.)

Draw a single-stage cluster sample of 3 schools and hence estimate the mean number of students per school and estimate the variance of the mean.

Solution:

```
#Assignment 2

import pandas as pd

dataset = pd.read_excel("assignment2_table.xlsx")

df = pd.DataFrame(dataset)

N=5

n=3

M=10

School_sample = df["School 1"] , df["School 2"], df["School 3"]

School_sample = pd.DataFrame(School_sample)

print(School_sample)

school1_mean = df["School 1"].mean()

print("school1_mean = ", school1_mean)

school2_mean = df["School 2"].mean()
```

```
print("school2_mean = ", school2_mean)

school3_mean = df["School 3"].mean()
print("school3_mean = ", school3_mean)


ybar_cl = (school1_mean + school2_mean + school3_mean) / n
print("ybar_cl = ", ybar_cl)


S_b_2 = 1/(n-1) * ((school1_mean-ybar_cl)**2 + (school2_mean-
ybar_cl)**2 + (school3_mean-ybar_cl)**2)
print("S_b_2 = ", "%.3f" % S_b_2)


var = (N-n)/(N*n) * S_b_2
print("var = ", "%.4f" % var)
```

Output:

```
school1_mean =  47.7
school2_mean =  49.3
school3_mean =  52.7
ybar_cl =  49.9
S_b_2 =  6.520
var =  0.8693
```