**W5 Systematic Selection**

**5.1 Systematic Selection – How it’s done**

Systematic Selection

• A simple method of selecting a sample from a list (e.g. population of transactions)

• Take every so many elements….

• Choose the start – first, any, any one up to the interval

• say taking every 10th, starting with first … (Interval k = 10)

• When do we stop? When we have n selections?

• This list has N elements

• If n = 50, stop with element 501 … and elements 502 to 1,000 have zero chance of selection

• As do elements 2,3,4….,10…and 12, 13, 14…. , 20 and so on

• Two remedies needed:

- spread the selection out over the whole list

- vary the selection start

• For example, for n= 50, don’t take every 10th, but every 20th

• And don’t start with the first, but start from any element from 1 to 20 … at random

• That is, adapt the selection process to the size of sample and size of list

• Calculate an interval of k = N / n (k= 1,000 / 50 = 20)

• And choose to start anywhere from 1 to k = 20 … at random

• Conceptually, this is taking the population, dividing it into k samples, and choosing one of them: it’s kinda like stratified sampling in a way

• This is equivalent to cluster sampling – each possible systematic sample is a cluster of n elements

**5.2 What happens if the interval is not an interval?**

• To repeat the process, first determine the sampling interval k = N / n

• Select a random number (RN) from 1 to k

• Add k repeatedly

• suppose, for example, there were N = 12,000 dwellings in a city and a sample of n = 500 is required

- k = 12,000 / 500 = 24

- take a RN from 01 to 24, say 03

- Take the 3rd dwelling, and every 24th thereafter: 3, 27, 51, etc.

• But what do we do in the more common situation where k is not an integer/

- examples

- N = 9, n = 2 and k = 4.5

- N = 952, n = 200, and k = 4.76

- N = 170,345, n = 1,250, and k = 136.272

• Consider three alternatives…

• First round the fractional interval

- for example, when N = 9, n = 2, take k = 4 or 5

- if k = 4 and RN = 1, the sample is the **three** element 1, 5, 9

- if RN = 2,3, or 4 the sample has **only two** elements

- if k = 5 and RN = 1,2,3 or 4 the sample has **two** elements

- if RN = 5, the sample has **only one** element

- What would happen if N = 952 and n = 200?

- rounding k to 5, RN’s 1,2,3 & 4 select 191, and RN 5 selects 190 – neither sample size is 200!

- what about for N = 170,345 and n = 1,250? 🡺 the sample size can be either 1252 or 1253

- Rounding thus has the problem that the sample size is not fixed, and we don’t get the target sample size!

• Second solution is one some people prefer

• treat the list as circular

- As before, calculate the interval k = N / n, and round up or down, say to k\*

- choose a RN anywhere from 1 to N at random

- then start counting every k\*th thereafter

- keep going until exactly n elements are selected

• But what if you get to the end of the list before you have n elements?

- we do “wrap”

- think about the list like it is a clock

- suppose n = 5 and N = 12, or k = 12/5 = 2.4

- round to [k] = 2, and choose random start 7:

- take every 2 after starting at 7 …

- and then 9… and then 11 …. And then …

- …. And then … 1

- and then 3 … and then … STOP because n = 5

- remember, start anywhere on the list … and wrap

• Here use the **fractional interval**…

- choose a random start from 0.1 to 4.6

- but how do you do that?

• One way is with a table of random numbers

• Since we need a number from 0.1 to 4.6, why not choose a random number from 01 to 46

• Suppose the number is 35

• “Insert” a decimal to make it fractional: 3.5

• Alternatively, generate a UNIFROM random number from zero to 1 in statistical software, say 0.76087

• Multiply by 4.6, and get 3.5

• But then what?

- do systematic counting …

- but “count” every 4.6 ….

- starting with (1) 3.5, we ‘count’ to (2) 3.5 + 4.6 = 8.1

- and again, (3) 8.1 + 4.6 = 12.7 …

- and again, (4) 12.7 + 4.6 = 17.3 …

- and again, (5) 17.3 + 4.6 = 21.9 …

- and just to be sure, one more time gives us, (5) 21.9 + 4.6 = 26.5

- oops! We are off the list

• But before we got off the list, we had n = 5 “selections”

- 3.5, 8.1, 12.7, 17.3, and 21.9

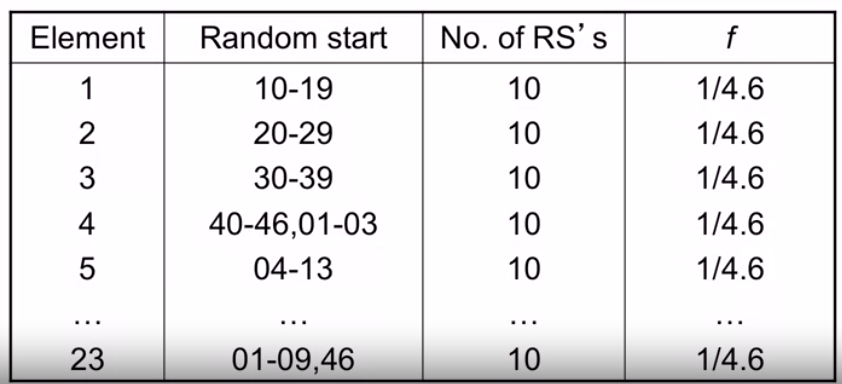
• What do we do with the decimals though?

- truncate to the whole number

- that is our selections are 3, 8, 12, 17, and 21 (truncation: drop decimals at the end)

• What does this all mean?

Simple method that is “epsem”



**5.3 Systematic selection and implicit stratification**

• List order combined with systematic selection can improve the efficiency (in terms of variance) of systematic sample designs

• Arrange the list order in advance

• Determines which kind of samples are selected

- random order: SRS