**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