Department of Statistics Malaysia
www.dosm.gov.my

NEWSLETTER

statsmalaysia
DOSM/BMP/3.2018/Series 18

ADAPTIVE CLUSTER SAMPLING ? (ACS)

What?

ACS refers to a sampling design in which an initial set of objects is selected by some probability procedure, and whenever the variable of interest for selected object satisfies a given criterion, additional objects in the neighbourhood are added to the sample.

Why?

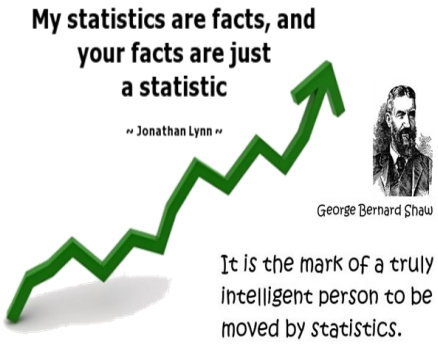
In many surveys, characteristics of interest is sparsely distributed but highly aggregated. Using typical sampling design, the sampling designer will face the situation that only certain area with concentrated objects of interest while other areas with zero objects of interest. In other words, there is a high likelihood that we can not find the target object.

How?

- The simplest solution is to increase sample size in order to increase the probability to encounter the rare and clustered objects.
- If it is known or suspected for some reasons that the rare objects forms clusters or groups within the population of interest, then, one may be interested to establish areas around the sampled areas once a sampled areas contains a rare object; because then, one would expect more objects around that area.
- The sampling strategy, that implies enlarging the area once a target object is found on the initial area, is called *Adaptive Cluster Sampling* (ACS).

When ?

ACS was introduced by Thompson in 1990 as an efficient sampling procedure for estimating totals and means of rare and clustered populations. The selection of objects to include in the sample depends on values of the variable of interest observed during the survey.



An Application
Standard Adaptive Cluster Sampling

Figure A

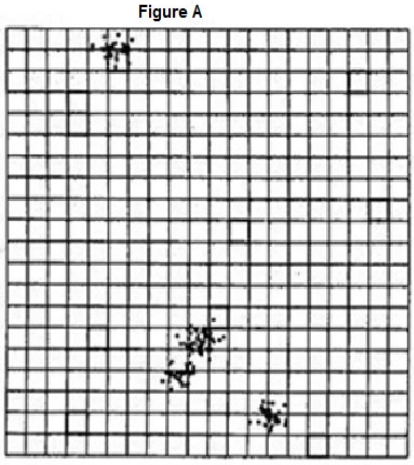
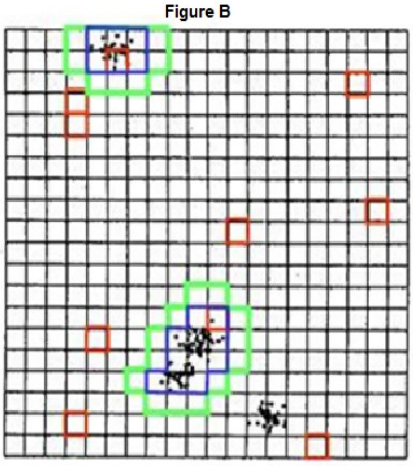


Figure B



- Figure A and B shows the population of clustered and relatively rare objects, thus the population of interest.
- From figure B, the adaptive clustering process is depicted: the red squares are the randomly selected n initial objects. The green depicts the clusters that are eventually expended according to a specific rule, while the blue objects are the sets of objects surrounding the initial sampled area and satisfying the specific rule.

Step 1: A random sample of n area is selected; this is some times called the initial sample or the seed sample.

Step 2: In each initial area, determine whether the target objects is there or not; or in general terms: whether the specified condition is fulfilled or not. If the condition is fulfilled (for example: there is at least one of the target objects on that initial area), then all its neighbouring area are also observed. This procedure is continued until no more areas found at the periphery of the cluster of sub-area that fulfill the condition.

Advantages

- More efficient than the classical design as the estimator variances under ACS will be smaller.
- Can identify areas of high abundance of targeted object.
- More flexible with respect to the choice of initial sample size n, neighbourhood definition, and the adaptive sampling condition.
- Cost saving as the average distances between sampled objects may be smaller and locating quadrats is easier.

Disadvantages

- Total number of sampled objects is random making it difficult to control the total survey cost.
- Flexibility of ACS (such as the choice of specified condition, neighbourhood definition, and initial sample size) complicates the search for a variance optimal design.

Reference:

Steve Thompson (2011). Adaptive Sampling. Simon Fraser University

Mena Patummasut & John J. Borkowski (2017). Stratified Adaptive Cluster Sampling with Spatially Clustered Secondary Units . Thailand Statistician

Thesis: An Analysis of The Adaptive Cluster Sampling Design with Rare plant Point Distributions. Jeremy Tout (2008). State University

J.J.Borkowski, P Yurk (2013) Adaptive Cluster Sampling: An Introduction.

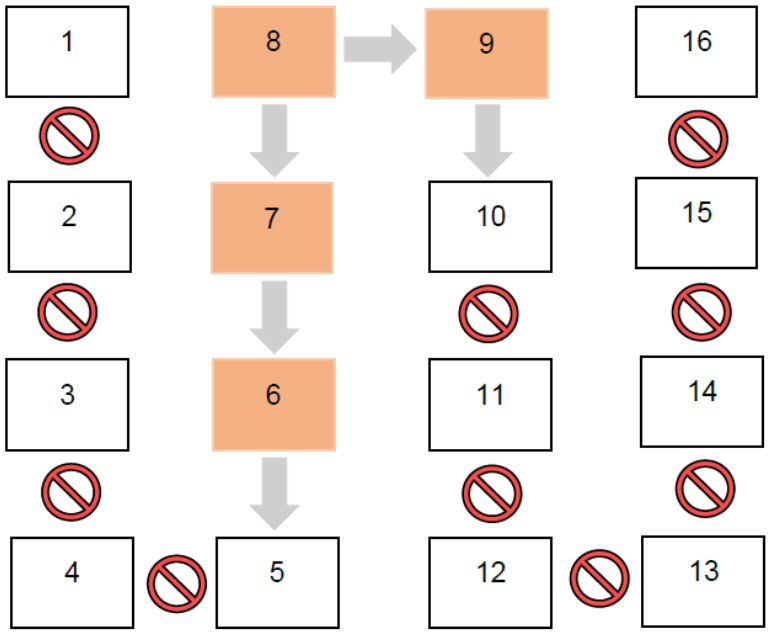
G.Chandra, N.Tiwari, H.Chandra (2011) Adaptive Cluster Sampling Based on Ranked Sets.

Do U Know?

DOSM used ACS in "Employment Survey in Oil Palm Plantations, Malaysia 2018" to select palm oil workers (object of interest). This sampling design is suitable as the palm oil workers living in dwellings close to each other (sparsely distributed but highly aggregated)

A schematic representation of Adaptive Cluster Sampling

A living quarter is initially selected (starting point) with equal probability. Example: living quarter 8 has been chosen as starting point, and screened for the presence of palm oil workers.



An enumeration area with 16 living quarters, 4 with palm oil workers and 12 without palm oil workers, marked in **peach colour**.

Since there is a palm oil worker in living quarter 8, the interview session is conducted and proceeds to the neighbouring living quarters, i.e., living quarters 7 and 9.

Living quarters 7 and 9 have palm oil workers. The interview is carried out. Next, living quarters 6 and 10 is screened for palm oil workers, since there is no palm oil workers in living quarter 10, the process stop here. While for living quarters 6, the interview session is carried out. Next, living quarters 5 is screen for palm oil workers. Since there is none, the process of selecting the sample is stop here.