Cluster sample example survey

Cluster sampling is a sampling plan used when mutually homogeneous but internally heterogeneous groups are visible in the statistical population. It is often used in marketing research. In this sampling plan, the total population is divided into these groups (known as clusters) and a simple random sample of groups is selected. The elements in each cluster are then sampled. Where all the elements in each cluster sampled are sampled, this is called a cluster sampling plan at one stage. If simply randomly subsample elements are selected within each of these groups, this is called a two-stage cluster sampling plan. A common motivation for sampling clusters is to reduce the total number of interviews and costs given the desired accuracy. For a fixed sample size, the expected random error is lower when most variations in the population are present internally within groups rather than between groups. Elements of the cluster The population within the cluster should ideally be as heterogeneous as possible, but there should be homogeneity between clusters. Each cluster should be a small representation of the total population. Clusters should be mutually exclusive and collectively exhaustive. The random sampling technique is then used on all relevant clusters to select which clusters to include in the study. At one stage, all elements from each of the selected clusters are sampled for sampling of clusters. In two-month cluster sampling, an accidental sampling technique is applied to elements from each of the selected clusters. The main difference between cluster sampling and stratified sampling is that when sampling a cluster, it is treated as a sampling unit so that sampling is performed on the cluster population (at least in the first phase). In stratified sampling, sampling is done on elements within each layer. In stratified sampling, a random sample is extracted from each of the layers, while only selected clusters are sampled in cluster sampling. A common motivation for cluster sampling is to reduce costs by increasing sampling efficiency. This is contrary to stratified sampling where the motivation is to increase precision. There is also multi-stage cluster sampling, where at least two phases are taken in the selection of elements from the cluster. When clusters are of different sizes Without altering the estimated parameter, cluster sampling is unbiased when clusters are approximately the same size. In this case, the parameter is calculated by combining all selected clusters. When clusters of different sizes there are several options: One of the methods is to sample clusters, and then inspect all the elements in that cluster. The second method is a two-way method of sampling a fixed unit content (be it 5% or 50%, or another number, depending on the cost) within each of the selected Reliance on a sample drawn from these options will be provided by an impartial appraiser. However, the sample size is no longer predetermined. This leads to a more complex formula for the assessor's standard error, as well as problems with the optics of the study plan (since power analysis and cost estimates often refer to a specific sample size). The third possible solution is to use the probability of proportional sampling size. In this sampling plan, the likelihood of selecting a cluster is proportional to its size, so a large cluster is more likely to be selected than a small cluster. The advantage is that when clusters are selected with a probability proportional in size, the same number of interviews should be conducted in each sampled cluster so that each sampled unit has the same likelihood of selection. Application of cluster sampling An example of cluster sampling is area sampling or sampling of geographical clusters. Each cluster is a geographical area. Since a geographically dispersed population can be expensive to explore, greater economy than simple random sampling can be achieved by grouping several subjects within the local area into a cluster. It is usually necessary to increase the overall sample size to achieve the same precision in assessors, but cost savings can make such an increase in sample size feasible. Cluster sampling is used to estimate high mortality in cases such as wars, famines and natural disasters. [1] The advantage may be cheaper than other sampling plans – e.g. Feasibility: This sampling plan takes into account large populations. Because these groups are so large, introducing any other sampling plan would be very expensive. Economy: The regular two main issues of expenditure, namely travel and listing, have been greatly reduced in this method. For example: Collecting research information about every household in the city would be very expensive, while gathering information on different blocks of the city will be more economical. Here, travel as well as citing efforts will be greatly reduced. Reduced variability: In a rare case of negative intraclassic correlations between subjects within the cluster, assessors produced by cluster sampling will provide more accurate estimates than data obtained from a simple random sample (i.e. the design effect will be less than 1). This is not the usual scenario. Main use: When a sampling framework of all elements is not available, we can resort only by sampling clusters. Disadvantage A larger sampling error, which can be expressed by design effect: the ratio between the appraiser's deviations made from the cluster study samples and the deviation of the appraiser obtained from the sample of subjects in an equally reliable, randomly sampled unsuspect study. [2] The greater the intraclass between the subjects within the cluster the worse the design effect (i.e. greater than 1. Indicating a higher expected increase in the variance of the appraiser). In other words, the more heterogeousness there is between clusters and the more homogeneity between subjects within the cluster, the less accurate our appraisers are. This is because in such cases we are better off sample as many clusters as possible and do so with a small sample of subjects from each cluster (i.e. two-stage cluster sampling). Complexity. Cluster sampling is more sophisticated and requires greater attention to how to plan and how to analyse (i.e.: take into account subjects' weights during parameter assessment, confidence interval, etc.) More on the sampling of cluster Two-stage sampling of cluster Two-stage cluster sampling, a simple case of multi-stage sampling, is obtained by selecting cluster samples in the first phase and then selecting a sample of elements from each sampled group. Consider the total population of the N cluster. In the first phase n clusters are selected using the usual method of sampling clusters. In the second phase, simple random sampling is usually used. [3] It is used separately in each cluster, and the number of elements selected from different clusters is not necessarily the same. The total number of N clusters, the number of n clusters selected, and the number of elements from the selected clusters must be determined in advance by the survey designer. The cluster's two-stad sampling aims to minimise research costs and at the same time control the uncertainty associated with interest assessments. [4] This method can be used in health and social sciences. For example, researchers used two-stage cluster sampling to generate a representative sample of the Iraqi population to conduct mortality research. [5] Sampling in this method may be faster and more reliable than other methods, which is why this method is now commonly used. The conclusion when the number of clusters is a small method of sampling clusters can lead to significant bias when working with a small number of clusters. For example, grouping at the state or city level, units that may be small and fixed in number, may be required. Microeconometric methods for panel data are often used by short panels, which is analogous to few observations by clusters and many clusters. A small cluster problem can be viewed as a random parameter problem. [6] While point estimates can be reasonably accurately estimated, if the number of observations per cluster is high enough, we need the number of $G \rightarrow \infty$ clusters {\displaystyle G\rightarrow \infty } for asymptotics to take effect. If the number of clusters is low, the estimated brake matrix may be biased downwards. [7] A small number of clusters are a risk when there is a serial correlation or when there is an intraclass as in the Moulton context. When we have several clusters, we tend to underestimate serial correlation through observations when there is an accidental shock or intra-class correlation in the Moulton environment. [8] Several studies highlighted the consequences of serial correlation and highlighted the problem of a small cluster. [9] [10] Within the Moulton factor, an intuitive explanation of the problem of a small cluster can be extracted from the formula for the Moulton factor. Let's assume that the number of observations per cluster is fixed to n. Below, V c (β) {\displaystyle V_{c}(\beta)} indicates a grouping-friendly coin matrix, V (β) {\displaystyle V(\beta)} indicates a non-grouping-friendly coin matrix that indicates intraclassical correlations: V c (β ^) V (β ^) = 1 + (n − 1) τ {\displaystyle {\frac {V_ {c}({hat\{\beta }})}{V({\hat {\beta }}}=1+(n-1)\rho } The ratio on the left provides an indicator of how much an unadjusted scenario overestimates precision. Therefore, a large number means a strong downward bias of the estimated kovariance matrix. A small cluster problem can be interpreted as a large n: when data is fixed and the number of clusters is low, the number of data within the cluster can be high. It follows that the conclusion when the number of clusters is small will not have the correct coverage. [8] Several solutions to the small cluster problem have been proposed. It can be used biasedly corrected cluster-robust variable matrix, make T-distribution adjustments, or use bootstrap methods with asymptotic refinement, such as percentil-t or wild bootstrap, which can lead to improved final reasoning of samples. [7] Cameron, Gelbach and Miller (2008) provide microsimulations for different methods and find that wild bootstrap performs well in front of a small number of clusters. [11] See also Multistage Sampling Sampling (Statistics) Simple Random Sampling Stratified Sampling Reference ^ David Brown, Study Claims Iraqi 'Excess' Death Toll Reached 655,000, Washington Post, Wednesday, September 11, 2010. ^ Kerry and Bland (1998); Statistics note: Intraclaster correlation coefficient in cluster randomization. British Medical Journal, 316, 1455/1460 ^ Ahmed, Saifuddin (2009). Methods in sample surveys (PDF). Johns Hopkins University and Saifuddin Ahmed. ^ Daniel Pfeffermann; C. Radhakrishna Rao (film). Statistics Manual Vol.29A Sample Reviews: Theory, Methods and Conclusion. Elsevier B.V. ISBN 978-0-444-53124-7. ^ LP Galway; Nathaniel Bella; Al S SAE; Amy Hagopian; Gilbert Burnham; Abraham Flaxman; Wiliam M. Weiss; Julie Rajaratnam; Team K Takaro (27 April 2012). A two-month method of sampling clusters using gridded population data, GIS and Google EarthTM images in a population-based mortality survey in Iraq. International Health Geographic. 11: 12th doi:10.1186/1476-072X-11-12. PMC 3490933. PMID 22540266. ^ Cameron A.C. and P. K. Trivedi (2005): Microeconometrics: Methods and Applications. Cambridge University Press, New York. ^ a b Cameron, C. and D. L. Miller (2015): Practitioners' Guide to Cluster Robust Conclusion. Journal of Human Resources 50(2), p. 317–372. ^ a b Angrist, J.D. and J.-S. Pischke (2009): Mostly harmless econometrics. The empiricist is a companion. Princeton University Press, New Jersey. ^ Bertrand, M., E. Duflo and S. Mullainathan (2004): How much should we trust estimates of differences in differences? Quarterly Economics Journal 119(1), p. 249–275. ^ Kezdi, G. (2004): Robust standard error assessment in fixed-performance panel models. Hungarian Statistical Review 9, p. 95–116. ^ Cameron, C., J. Gelbach and D. L. Miller (2008): Bootstrap-based improvements to conclude with clustered errors. Economics and Statistics Overview 90, p. 414–427. Retrieved from the