Statistical Checklist 2

SAMPLING

J. N. R. JEFFERS

Institute of Terrestrial Ecology

Sampling

Stating the objectives

1. Have you stated clearly and explicitly the objectives of the investigation and the reasons for undertaking it?
2. Have you translated these objectives into precise questions that sample determinations can be expected to answer?

Defining the population about which inferences are to be made

3. Have you defined carefully the population about which you are seeking to make inferences from the sample?
4. What constraints of space, time or category are to be placed upon the population about which you are seeking to make inferences?
5. Are the individuals to be measured or counted representative of the population?
6. If not, what do you need to do to find representative individuals?
7. Is there a logical framework for the choice of sample individuals from the defined population e.g. a list of all the individuals in the population, a location in one, two or three dimensional space?
8. If not, what do you need to do in order to impose a logical sampling frame from which samples can be taken according to a pre-arranged plan?
9. If there is no logical and practical way of finding samples which are representative of your defined population, is it worth continuing with the investigation at all?

Sample units

10. Are the sample units naturally defined, e.g. as individual organisms, cultures, or objects?
11. If not, how are the sample units to be defined and limited in both space and time?
12. Is the number of sample units in the population finite?
13. If so, is the total number of units in the population sufficiently large to enable you to ignore complications associated with sampling from a finite population, e.g. sampling without replacement, correction factor for the expression of variance, effects of destructive sampling?
14. Is the definition of the sample units appropriate to other objectives of the investigation?
15. Have you considered alternative shapes and sizes of sample units in relation to the objectives of the investigation?
16. Are you satisfied that there is a sufficiently logical definition of the sample units to justify proceeding with the investigation?

Translating the objectives

17. What exactly is to be estimated or tested?
18. Are the estimates that are required proportions, totals, means, totals or means over sub-populations, ratios, productions or other non-linear expressions?
19. Have you constructed blank tables of exactly what you want to estimate?
20. What is the smallest subset of data from which estimates are to be made?
21. What precision is required of the estimates for the various subsets of data?

Preliminary information

22. Do you have any information about the population which may be helpful in designing the sampling scheme?
23. Do you have any estimates of the likely variability of the population for sample units of the size, shape or kind that you intend to use?
24. If not, would some kind of pilot survey be desirable to determine this variability?
25. Are there any known factors which would help you to stratify the population before sampling?
26. If not, would a pilot survey be desirable to determine the possible efficiency of various kinds of stratification?
27. Have you any experience of the kind of sampling you are intending to use in this investigation, including experience in dealing with such matters as sample unit identification, measurement or counting, and the physical conditions in which you will be working?
28. If not, would a pilot sample enable you to test the feasibility of the methods you are intending to use?

Choice of sampling design

Systematic sampling

29. Have you considered the possibility of using systematic sampling?
30. If so, what interval will separate the individual sample units in either time or space, or both?
31. Is there any likelihood that this interval will coincide with some natural periodicity in the population, with the result that the sample estimates, while precise, may be biased?
32. Are you aware of the difficulties of determining the precision of estimates derived from systematic samples?
33. If so, what methods will you use to determine the precision of the estimates?

Simple random sampling

34. Is simple random sampling likely to be effective in deriving population estimates of acceptable precision?
35. If not, have you clearly stated reasons for wanting to use a more complex sample design?

Stratified random sampling

36. Will stratification of the population (i.e. division of the population into strata that will reduce variability) improve the precision of the estimates?

37. If so, what is the basis of the stratification you intend to use?

Multistage sampling

38. Is any advantage likely to be gained by multistage sampling (i.e. by dividing the population into a number of first-stage sampling units, with the selected units being divided into smaller second-stage units)?

39. If so, do you know how to analyse data derived from such a procedure, and how to derive the estimates required from the survey?

Cluster sampling

40. Is any advantage likely to be gained by cluster sampling (i.e. by taking samples as aggregates or "clusters" of the sample units)?

41. If so, do you know how to analyse data derived from such a procedure, and how to derive the estimates required from the survey?

Determining the size of the sample

42. Have you calculated the size of the sample that would be needed to provide standard errors consistent with the desired precision of the investigation?

    e.g. $N = (S/E)^2$ where N = number of samples
    S = standard deviation
    E = standard error

43. If so, is the size of the sample you propose to take adequate, i.e. neither too small nor too large?

44. If you propose to use systematic sampling, is the number of samples needed to cover the population uniformly consistent with the expected precision of the results?

45. If you propose to use stratified random sampling, have you tested the efficiency of the stratification in reducing the size of the sample, or, alternatively, in obtaining the greatest number of samples from that part of the population which is of greatest interest?

46. If you propose to use multistage sampling, have you tested the efficiency of various combinations of the sample units at different stages, using known or estimated variance components for the stages?

47. If you propose to use cluster sampling, have you tested the efficiency of various sizes of clusters?

48. Do the numbers of samples need to be modified to take into account markedly different costs in different parts of the sampling procedure, e.g. considerably greater costs in travelling to a site than in sampling within a site?

49. If so, can multistage or batch sampling be used to improve the cost efficiency of the sampling procedure?

50. Is sequential sampling likely to be of benefit in reducing the field sampling effort, or in helping to revise the allocation of further sampling effort?

51. If preliminary calculations indicate that there is little chance of you obtaining estimates of sufficient precision with the number of sample units you can take, is it worth proceeding with the investigation at all?

Auxiliary variables

52. Have you considered any auxiliary variables or attributes which should be recorded at the time of sampling to aid in the interpretation of the results or to improve the efficiency of the estimates?

53. Will these auxiliary variables be used as ratio or regression estimators?

54. If so, do you know how to do the necessary calculations?

Randomization

55. Have the sample units been selected by an explicit randomizing procedure?

56. Were any constraints on the randomization, e.g. stratification, correctly applied?

57. Were you tempted to re-randomize any part of the selection of sample units because of apparently unfortunate coincidences or clustering of samples?

58. If so, do you have some knowledge of the variation in the defined population which has not been incorporated into the design of the sampling scheme?

59. Does a plan exist, showing the location of the sample units in space and time?

60. Is there likely to be any bias in sampling due to 'non-response' or 'volunteer' effects?

61. If so, have you provided for appropriate 'call-back' or 'follow-up' studies?

Recording of results

62. Does each sample unit have a clear number or designation, linking it unambiguously to the plan of the sampling scheme?

63. Have you defined the exact procedure to be followed in the selection of the sample unit and in recording the results?

64. Have you defined variables or attributes to be measured or counted in each sample unit?

65. If so, are the measurements meaningful and relevant to the objectives of the investigation?

66. Have you designed a record form which will ensure that all assessments are complete and are recorded against the correct sample unit?

67. Have the assessors been trained to measure and count the variables and attributes efficiently and accurately?

68. Is there space on the record forms for observations to be recorded of unexpected occurrences, and have the assessors been encouraged to look for these occurrences?

Analysis of results

69. Do you understand the methods of analysis that will need to be used for this sampling scheme?

70. Have you made arrangements for the computations to be done on a computer, or elsewhere?

71. If the computations are to be done on a computer, does the necessary program exist, and do you understand the constraints that the program places on the data set?

72. If not, have you obtained advice from a statistician on the analysis and interpretation of the results, preferably before starting the sampling?

The final (and most important) question

73. If you are in doubt about the purpose of any of the questions in this checklist, should you not obtain some advice from a statistician with experience in your field of research before continuing with the sampling investigation?

There is usually little that a statistician can do to help you once you have committed yourself to a particular sampling scheme.

Bibliography

If any of the questions in this checklist refer to aspects of statistical theory with which you are unfamiliar, further information can be found in the following texts:–

Campbell, R. C. (1967). *Statistics for biologists*. Cambridge University Press, Cambridge.

Cochran, W. G. (1963). *Sampling techniques*. Wiley, New York and London.

Finney, D. J. (1962). *Introduction to statistical science in agriculture*. 2nd edn. Oliver and Boyd, Edinburgh.

Freese, F. (1962). *Forest sampling*. US Department of Agriculture, Handbook No. 232.

Sampford, M. R. (1962). *Introduction to sampling theory*. Oliver and Boyd, Edinburgh.

Subman, S. (1976). *Applied sampling*. Academic Press, New York.

Sukhatme, P. V. (1954). *Sampling theory of surveys with applications*. Iowa State College Press, Iowa.

Yates, F. (1960). *Sampling methods for censuses and surveys*. 3rd edn. Charles Griffin, London.

The Institute of Terrestrial Ecology

The Institute of Terrestrial Ecology (ITE) was established in 1973, from the former Nature Conservancy's research stations and staff, joined later by the Institute of Tree Biology and the Culture Centre of Algae and Protozoa. ITE contributes to and draws upon the collective knowledge of the fourteen sister Institutes which make up the NATURAL ENVIRONMENT RESEARCH COUNCIL, spanning all the environmental sciences.

The Institute studies the factors determining the structure, composition and processes of land and freshwater systems, and of individual plant species. It is developing a sounder scientific base for predicting and modelling environmental trends arising from natural or man-made change. The results of this research are available to those responsible for the protection, management and wise use of our natural resources.

Nearly half of ITE's work is research commissioned by customers, such as the Nature Conservancy Council and the Department of the Environment. The remainder is fundamental research supported by NERC.

ITE's expertise is widely used by international organizations in overseas projects and programmes of research.

Statistical Checklists

1. This statistical checklist is one of a series currently being developed by the Institute of Terrestrial Ecology (ITE). The aim is to highlight some of the more significant questions to be taken into account in the application of statistical methods to practical research and management. They provide a framework for marshalling thoughts or ideas on the subject, but by their nature cannot expect to cover a subject exhaustively. Wherever possible, references are provided to readily-available textbooks as sources of further information to those wishing to follow up any topic in detail.
2. Single copies of checklists can be obtained from the Publications Officer, Institute of Terrestrial Ecology, 68 Hills Road, Cambridge, at a cost of 30p, if a stamp and addressed envelope is enclosed with your order.
3. Bulk supplies are available for Institutes and organizations wishing to use these checklists for wider internal circulation. Copies are available in multiples of 20 at £3.00 per pack, post paid.
4. The checklists so far issued include: –
   (a) Design of experiments
   (b) Sampling.

   Details of new checklists will be announced in the ITE Annual Report.
5. ITE welcomes comments and suggestions on this series. Please write to the Director, Institute of Terrestrial Ecology, Merlewood Research Station, Grange-over-Sands, Cumbria, LA11 6JU.
6. The help of K. H. Lakhani and M. D. Mountford is gratefully acknowledged.

(5/4) Dd0074426 2M 5/79 HPLtd Su'ton G3313

ISBN 0 904282 29 5