Numeracy for Everyone

4. Numeracy in Research Planning

Anil P Gore and S A Paranjpe

A P Gore is currently Head, Department of Statistics, University of Pune. He initiated teaching and research in ecological statistics at Pune. He has been active for many years in promoting interaction between statisticians and biologists.

S A Paranjpe is Reader in the Department of Statistics, University of Pune. She has long experience of applying statistics in the fields of nutrition, forestry, ecology, agriculture and health.

How to Check if Rice is Cooked

We take a pinch of grains of rice and press to check if the core is soft. If yes, we declare that rice (we mean all grains) is cooked. How do you know? The unstated assumption is that all grains of rice in that pot are alike (in an identical state of being cooked). If this assumption is not true, our conclusion can go wrong. This predicament indeed does occur if the heat is too strong. Then we get multi-tiered performance. The bottom tier is charred and black. The middle layer is edible while the top layer is still raw. Why don't we check every grain to rule out errors? Simply because it is cumbersome, time consuming and usually unnecessary.

All these considerations are equally relevant in many studies. We take water samples, air samples and soil samples. Studies of sample plots in forests are common. By sampling we hope to know about a huge entity at low cost. This is a valid and successful approach. It is only necessary to be careful about assumptions. Is the entire forest area of interest uniform? River banks, eastern and western slopes of hills and rocky outcrops may have very different vegetation. So it is best to divide the forest area into strata or divisions such that each division is relatively homogeneous but there are differences between divisions. Some sampling has to be done in every stratum. If a stratum has a lot of variation in it, more effort has to be spent on studying it. If a stratum is very uniform a little bit of observation will suffice to show what is in it.

How large should a sample be? Or how many individuals should be examined to estimate reliably some feature of the population? Generally, the larger the sample size, the greater is the precision of an estimate. But we cannot go on observing because resources in terms of time, money, personnel, etc. are limited. While this question is very important, it is also very difficult to answer in general. If we want to know the number of mammary glands in different mammals, observing one or two adult females in a species suffices. This is because the trait is almost without variation. The same is true of clutch size in birds. On the other hand if our interest is number of eggs laid by frogs or size of crowns in tree species, a lot more observation would be necessary. If a population is small, like trees in a few hectares of forest, 5% sampling is usually adequate. If the target population (not necessarily human) size is very large, this rule becomes inappropriate. Instead it is the actual sample size that becomes crucial. When election patterns are predicted, though the size of a parliament constituency is of the order of 10 lakhs, a few hundred individuals selected carefully and interviewed skillfully can indicate the result.

Part 1. Why Quantification?, *Resonance*, Vol.4, No.5, 19-30, 1999.

Part 2. Dice of Life, *Resonance*, Vol.4, No.9,14-23, 1999.

Part 3. Just for Ecologists, *Resonance*, Vol.5, No.5, 15-25, 2000.

Ecologists often want to measure diversity in an ecosystem. It appears that here also sampling a few hundred individuals does the job quite well. A good familiarity with the population to be sampled is very essential for effective sampling. If we want to estimate sex ratio in black bucks or langurs or elephants and we examine some breeding groups, we will find that adult males are very few. We have to know that large number of males live separately as bachelor herds or just loners. Otherwise we will get a totally misleading answer even though we use correct statistical procedures.

Having discussed broad outlines of sample survey methods, let us turn to some specifics. In sampling we talk about a 'frame'. It is the list of all individuals in the population of interest. Such a list is available in case of some studies but not always. Electoral roll of a constituency is the frame for sampling of voters. List of houses in municipal records is another example of a frame. In an educational institution, students enrolled are all listed. Given that a frame is available, given that a sample size (number of individuals to be checked) is decided, the important task is to identify in the frame, the individuals to be interviewed or subjected to some measurement. The basic method for this is 'simple random sampling'.

Suppose we have a list of one thousand individual students, houses, shops, families or whatever and a sample of 50 is to be selected randomly. We number the cases from 000 to 999. Now we refer to a table of random numbers and select 50 triplets. Suppose the first triplet in the random number table is 178. Then the individual corresponding to that number is located and relevant observations are recorded. If instead of a printed random number table you have a computer, it can generate random fractions. Multiply the fraction by 1000 and use the integer part. There can be small problems in any such procedure. Suppose our frame contains only 500 cases. Then the above procedure may have a hitch. What if the random number selected happens to be above 500? (It will never exceed 1000 since we are choosing a triplet). Well, one can divide the number by two and ignore any fractional part. The idea is somehow to ensure that each case has the same chance of being selected. We have mentioned earlier that a population should be split into homogeneous subgroups called strata (to ensure higher efficiency). In such 'stratified random sampling', sample size is decided for each stratum and a simple random sample is drawn from it.

Systematic Sampling

One difficulty about simple random sampling is that often the selected individual is hard to locate, or if selection is to be done by field staff they may get confused. Hence sometimes a modified selection procedure is recommended. In our problem of selecting 50 cases out of 1000, we can start by selecting a random number from the set 1 to 20. Say 13. Then we go on adding 20 to this number. Thus 33, 53, 73, 93 become the selected set from the first hundred. 113, 133, etc. are selected from the next hundred and so on. This is systematic sampling with a random start. While there is convenience in it, a word of caution is in order. If there is a rhythmic pattern in the list and we start with a high value, we may get all high values and hence inflated estimates. In American cities, residential areas are divided into blocks of roughly equal size. To choose a sample of houses we choose one from a block and select the corresponding house in each block. But suppose our random start gives us a corner house. Then all cases in the sample may be corner houses. If corner houses have larger plot size or higher market value, we have a sample that is not representative. Eliminating such biases becomes possible with experience of studying the same or similar populations again and again.

Cluster Sampling

Suppose we wish to study agricultural practices of farmers in a district. Suppose there are 1000 villages and about 200 farmers in each village. We plan to interview about 1000 farmers. There are various ways to select these. Select one farmer from each village. Select a simple random sample of 1000 from a frame of two lakh farmers. A drawback of these strategies is that travel time increases immensely. In the first option the interviewer has to go to every village. In the second option travelling is less but still considerable. A strategy to avoid this is to select five villages randomly and interview all farmers in those villages. For this strategy, with very low cost of travelling, to be satisfactory we need that each cluster should be representative of the whole population of interest. Variation from village to village should be low. Variation between farmers in the same village should be high. In the district of Pune, where we live, the assumption is not valid. Western part of the Pune district is hilly and has heavy rainfall while eastern part is in plains and with scanty rainfall. In such a case, sampling should be done separately.

Mail Questionnaire Surveys

One common method of sample survey is to send questionnaires to a random sample of respondents. This saves cost and effort of travel. It allows the respondent to answer the questions leisurely after thinking about them instead of giving impromptu responses under pressure. So far so good. There is one major drawback of these surveys. It is that of nonresponse.

Many people simply ignore the letter requesting them to fill a form and return it (even when postage is prepaid). There are many reasons for this. Lack of time or interest, diffidence about the questions, fear of losing confidential information or even sheer laziness. Someone will argue that we should estimate the degree of nonresponse and compensate for it. If we need a sample of 500, and only half the people are expected to respond, send out 1000 letters. This is correct. However there is one additional problem. Nonrespondents often tend to be different from respondents.

If questions are about one's personal income, people with high incomes are reluctant to answer (or let us assume that it is so). In that case only low income individuals will respond. If we use this information to estimate average income, we may get a serious underestimate. This problem is genuine.

In a democracy, a vociferous minority can create an atmosphere in favour of this or that political decision. But we must never forget the silent majority. A truly democratic decision should take serious note of what the majority believes or feels.

So what can we do about the problem of nonresponse? One pragmatic view is that an attempt should be made to interview at least a few nonrespondents by extra effort. This information should be added to whatever else is available before pronouncing the final estimates. This brings us to the issue of questionnaire construction. It involves expertise of marketing specialists and psychologists. However a few generalities can be clear to anyone. Firstly a questionnaire has to elicit answers from a person over whom you have no (or marginal) control. If the control is significant, the response loses its credibility anyway (Who will (should) believe it if an employee praises his boss?) So it is most important to humour the respondent and to avoid irritating him. Hence the questionnaire should be as brief as possible. Answering the questions should not require any complicated calculation or thinking. Yet the question must not be suggestive. “Do you use toothpaste A?” is not a good question. It is better to ask “which brand of toothpaste do you use?”. Sometimes there is a difference between the intent of the surveyor and the understanding of respondent.

If you ask “what is the cause of ozone depletion over Australia?” and the answer is “I do not know but I closed all gas cylinders,” you are sure of miscommunication. To avoid this confusion, it is customary to conduct a pilot/trial survey and make any corrections in the survey strategy. There is one category of questions that is quite important but difficult to ask. We know that many students cheat in final examinations. We want to estimate the proportion of such students. It is pointless to ask “did you cheat in last year’s final examination?” Everyone will say, “No, I did not”. No one wants to be caught confessing to such misdemeanor. Statisticians have found a way to avoid invasion of privacy and yet to get estimates needed. The method involved is called ‘randomized response’. In this approach we offer two questions with yes or no as possible answers. Here is an example.

Q1. Did you copy or cheat in the last examination?
Q2. Were you completely honest in the last examination?

Everyone knows these questions ahead of time. Then a respondent throws a die. If the score is 1 or 2 he is to answer Q1. Otherwise he answers Q2. No one but the respondent knows which question he has to answer. Hence he can give a candid reply without fear. Suppose out of 100 students 40 say yes and 60 say no. What can we conclude about degree of cheating? The estimate of proportion of cheaters is given by the equation

$$0.40=\frac{1}{3}p+\frac{2}{3}(1\quad p).$$

Here p is the unknown proportion. (Try to see how this equation arose). Solving it we get $p = 0.8$.

Another variant of the method involves asking a different question as Q. 2. For example, Were you born in the first trimester of the year (i.e. Jan Feb March)? We know that roughly the probability of an affirmative answer is a quarter. We will let the readers think about how to write an equation to get p.

Sample surveys are a powerful tool in every walk of life. Experience in sample surveys can be useful if you seek a job in marketing. Even otherwise, you will come across many reports based on sample survey and knowing about the technique will help you interpret the reported material better.

Designing an Experiment

Preliminaries: Experiments are the heart of science. Scientists believe what they observe and what their experiments confirm. A good experiment is one which yields an unequivocal conclusion. Every conceivable objection is raised before a conclusion gets accepted by the scientific community. This skepticism is crucial in scientific enquiry. We need to design an experiment in such a way that there is minimal room for skepticism. It is widely held that overgrazing by cattle affects grassland adversely. Can we design an experiment to check this? Yes. Go to an area where cattle graze. Build a fence around a small portion of the area, to exclude cattle. Come back after say 4 weeks. If condition of grass inside is different from that outside, then cattle must be the reason. If there is no difference, cattle have no impact. Does this sound reasonable?

Wait a minute. What if the experiment was done in May and there were no summer showers? Then all grass would be yellowed and dead even before the experiment begins and no perceptible change may occur. So perhaps it should be done in August. To consider different ways of plugging loop holes, let us discuss a simpler experiment that you can perform at home. There are two companies A and B each selling soap to wash clothes. We wish to know which soap is better. So we design an experiment.

Attempt 1. You wash your shirt with soap A and your friend washes his shirt with soap B. Check which shirt is cleaner. If your shirt is cleaner, declare soap A as better. Remember such a conclusion if widely known and believed, will reduce the sale of soap B. Hence that company will quickly raise objections to discredit the conclusion. Here is one objection. Anything can happen in one case. You should try soap B many times. Only then will the truth come out. This is a valid objection. We need adequate 'replication'.

Attempt 2. You and your friend wash five clothes each and a cleanliness score is given to each piece. If your total is greater, soap A wins. Now defenders of soap B raise another objection. You wash more diligently and for a longer time. Hence your score is better. Not because soap is better. This objection could be wrong. But the experimental design did not take care of it explicitly.

Attempt 3. Each person devotes precisely the same amount of effort. Suppose soap A still fares better. Then there can be yet another objection. Your friend is an athlete while you are a sit-at-home type person. So your clothes are not as dirty as your friend's. Well this objection can be taken care of using only your clothes to test both soaps. This is the principle of 'local control'. Comparisons are valid if experimental units are similar.

Even this precaution may not be enough because the objection may be that soap B is really better for dirty clothes while you tried it only on relatively clean ones. Remember, the side that loses, will leave no stone unturned in the attempt to discredit the experiment. So you will have to run parallel trials with exercise clothes and normal clothes. But the objections will not end. The opponent may attempt hair splitting by arguing that even among your clothes some are cleaner than others and they went to soap A. How do we ensure absolutely 'level playing field'? Perhaps the best thing would be to wash one half of a shirt by soap A and the other half by soap B. Even then there can be an objection that soap A was used to wash left side while soap B to wash right.

And it is the right armpit that gets more sweat. The last resort in designing a fair experiment is to allot sides of shirts to soaps by tossing coins. This is the principle of 'randomization'. Replication, local control and randomization are the three cannons of the art of designing good experiments. In science there are usually no favorite sides and no vested interests but we take good care all the same so as not to be misled.

Let us consider one experiment carried out a few years ago by a Swedish ornithologist named Malt Anderson, to see how he handled such problems. In the theory of evolution, there is a concept called 'sexual selection'. This concept suggests that males develop extraordinary traits (e.g. peacock's plumage) in response to female attraction for those traits. Anderson wanted to verify this proposition in the case of an African bird species called blackbird. In this species, males possess very long tails. If this is due to female preference, Anderson argued, then males with short tails should have lower reproductive success in terms of number of females that lay eggs in the territories of the short tailed males (and number of eggs laid). So, analogous to soap A are the ordinary males with long tails. Analogous to soap B we should have short tailed males. Now in nature all males have long tails. (Presumably short tailed cousins have been wiped out over time). To overcome this problem, Anderson cut off a portion of the tails of a few birds. If the experiment is done in the same area and in the same breeding season local control is taken care of. With effort one can ensure enough number of males of each type (precisely how many is a difficult question and can be answered only with experience).

Is this design adequate? Are all loopholes plugged? If males with long tails get greater number of females in their territories and greater number of eggs laid (presumably fathered by them) is the hypothesis of sexual selection to be treated as confirmed? Anderson anticipated two possible objections.

Firstly, shortening of the tail may affect flying, fighting or foraging abilities of males and females may avoid them for that reason, and not because short tail is ugly. This objection was taken care of, by actually measuring various abilities of the short tailed males. Anderson could confirm that there was no adverse effect on those males.

The second objection is more subtle. Perhaps the act of cutting its tail may have affected the general behaviour of the male in some manner noticeable only by females but not by scientists. If so the attractiveness or otherwise is not in tail length but that behavioral feature. To protect against this possibility, Anderson introduced in the design a third group of male birds. These had their tails shortened but the portion cut out was pasted back. So the group suffered the act of cutting without actually getting a shortened tail. We shall return to this kind of arrangement when we discuss clinical trials and the concept of 'placebo'.

At the end of the experiment Anderson found that short tailed males had lower rate of reproductive success whereas the other two groups had higher rate. So it turned out that the act of cutting does not have any impact but shortened length of tail does create a handicap for a male bird. In this sense, females seem to have promoted long tails. Having basically outlined the approach to design of experiments, let us mention a few basic terms used commonly.

Completely Randomized Design (CRD): Suppose we want to compare several hormone treatments on flowering and fruiting of teak. We identify several trees of the same variety, similar age, in similar soil condition and allocate trees to each hormone treatment randomly. Presumably there are no differences among individual trees and any observed differences in fruiting are attributable to hormones.

Randomized Complete Block Design (RBD): If we have several varieties of teak, we should compare the hormones for each variety. Block is a group of homogeneous experimental units. Here all teak trees of a given variety from one block. Within each block we allocate trees to hormones randomly.

Incomplete Block Design: Here the group of similar individuals is very small. Suppose we wish to check if bird songs are inherited. Our experiment involves rearing in isolation and in different levels (low, medium and high) of exposure to other birds of the same species. Thus we will be comparing four 'treatments'. For this we need littermates which will constitute a natural block. If litter size is only 2, we get into a problem. We cannot have complete blocks in the sense that within a litter we can only try two out of four treatments. This can be handled by trying different treatment pairs in different blocks.

Factorial Experiments: Traditionally scientists adopted a 'change one thing at a time' approach in designing experiments. This means all other conditions were held constant and only one 'factor' was changed to study its effect. Suppose we wish to check if it is economical to use higher doses of fertilizer in growing rice. So we keep all other things constant and use less fertilizer in some fields and more in others. This is quite legitimate. But it has two drawbacks.

Often the number of factors is large. For a biscuit maker the following factors may be important in determining biscuit quality: (a) type of wheat, (b) type of oil and its quantity, (c) oven temperature, (d) amounts of sugar and salt. As the number of factors goes up, the number of trials needed increases and the experiment becomes more and more expensive.

Second problem is that of interaction between factors. Suppose two factors of interest are nitrogenous fertilizer and irrigation in growing wheat. It turns out that the level of fertilizer use that is best depends on the amount of irrigation available. If irrigation is plentiful, high dose of fertilizer may be good. But if irrigation is not available, high dose of fertilizer may not be useful. So we have to study combinations of factors. Sir Ronald Fisher developed the technique of designing experiments in which many factors are changed simultaneously and yet valid inferences about each can be drawn.

Exercises

1. Design a survey strategy to estimate the market share of different daily newspapers in your city/town. How will you obtain the profile of a typical customer of each newspaper?

2. Identify four surveys reported recently in the newspaper you read. Check whether there are (can be) some loopholes in these surveys.

3. How will you generate a triplet of integers 0 to 9 using only a coin (or a six faced die)? Can you show that your procedure ensures equal probabilities for all cases?

4. Aim: to check effect of three treatments on quality of chapati. The three treatments are: after preparing the dough:
   a. immediately make chapatis.
   b. Keep dough for some time and then make chapatis.
   c. Keep dough in refrigerator for a couple of hours and then make chapatis.

Design a suitable experiment.

Suggested Reading

[1] **M Delampady and V R Padmawar, Sampling, Probability Models and Statistical Reasoning, *Resonance*, 1(5) 49-58, 1996.**

[2] **R L Karandikar and A Basu, Opinion Polls and Statistical Surveys: What They Really Tell Us, *Resonance*, 4(7) 49-58, 1999.**

[3] **T Krishnan, Fisher's Contributions to Statistics, *Resonance*, 2(9) 32-37 1997.**

[4] **S Kunte, Statistical Computing 1. Understanding Randomness and Random Numbers, *Resonance*, 4(10) 16-21, 1999.**

Address for Correspondence
A P Gore & S A Paranjpe
Department of Statistics
University of Pune
Pune 411 007, India.

> *The law of gravitation is indisputably and incomparably the greatest scientific discovery ever made.*
>
> W Whewell's
> *History of the Inductive Sciences*
>
> From: *The Mathematical Intelligencer*