# Determining Adequate Sample Sizes

*Antony Fielding*
Department of Economics,
University of Birmingham, England.

## ◆INTRODUCTION◆

PRACTISING statisticians in a consultancy capacity are often asked 'How big a sample should I take?' Of course, the whole subject of an appropriate sample design for a particular inquiry is in itself a skilled task. Yet even if it is accepted that a simple random sample is appropriate, the question is often not an easy one to answer. Investigators unfamiliar with sampling theory and practice are often preoccupied with sampling fractions rather than absolute sample size, although it is the latter (in most cases) rather than the former that is instrumental in determining the precision of sample results. Even when this controversy is resolved it is often difficult to elicit from an investigator the objects of his inquiry, what precision he requires, the extent of his resources, and how to resolve often conflicting requirements of his piece of work. Yet, after some careful questions, and some detailed ways of getting investigators to make their aims more explicit, it is sometimes possible to give some reasonable answers to the question. What follows is a consultancy example in determining whether a particular sample chosen was adequate and illustrates some of the underlying difficulties in answering such questions. It may be useful in introducing some of the essential criteria to be borne in mind when facing students with the above questions.

## ◆BACKGROUND ◆

A large hotel was managed under contract by another chain of hotel owners. It was acquired by a new owner who, upon acquisition, appointed a firm of management consultants to undertake a review of the operational aspects of the hotel. As part of this review a sample of rooms at the hotel was inspected and as a result of this inspection the consultants reached the conclusion that: 'Much of the accommodation seen (bedrooms and bathrooms) is incompatible in quality and standard both to the perceived image of the hotel and to the high luxury market that it sets out to attract'. The number of rooms in the hotel was 289 letting units. The number of rooms visited by the inspectors was 15. These were a 'complete mix between rooms which had been refurbished, partly refurbished and not refurbished at all, and between suites, doubles, and singles'.

The consultants state that: 'Many of the bedrooms seen were dingy, scruffy, and tatty. There was ample evidence of lack of maintenance - not just overdue decoration such as damaged and tired paintwork, but also in daily maintenance and upkeep (e.g. cigarette burns, stains, frayed and torn lampshades). And most of the rooms seen (in service and available to guests) were actually dirty; layers of dust on various surfaces were common as were filthy curtains and nets'. The consultants then detailed a long list of defects which they noted in one or more of the inspected rooms.

Another firm of management consultants was hired by the hotel managers to prepare a response to the report by the owner's consultants. A statistician was asked by this firm to comment on one aspect of this report: the adequacy of the sample size of 15 on which conclusions were based. This presented some difficulty since it was not clear on what basis the 15 had been selected, and also the subjective and unsystematic nature of the observations made on the selected rooms. Yet it proved possible to make some generalisations under various assumptions about the nature of the latter inspections and observations. Of course, the manager's consultants wished to argue that the sample size of 15 was inadequate given the seriously detrimental nature of the consultant's report to the owner.

## ◆ADEQUACY OF SAMPLING ◆ METHOD AND SAMPLE SIZE

The sample size was 15 suites or rooms out of a total number of rooms of 289, although only 217 were available for inspection since contractors were occupying the rest.

There is one facet of the problem that really has no bearing on the matter of sample size - the manner of sample selection. If rooms are selected purposively; either because of convenience, accessibility or whatever, then there can be no scientific basis for making any judgement about likely errors. Such a selection mechanism is open to all sorts of potential subjective biases. Whether this was the case is not clear from the relevant sections of the report by the owner's consultants, and might be the first question asked.

Nevertheless, the question remains of what could be said about likely errors in generalisations if the sample was in fact selected randomly, and in particular by a simple random sampling mechanism. It is possible here to make some general statements that may be useful, but specific statements about sample size are difficult without reference to the sort of information collected and the characteristics of the hotel as a whole, about which it was desired to generalise. On further questioning there appeared to be two broad aspects of this problem, which are possible motivations of the inspectors:

(i) We might suppose that the inspectors had a checklist of possible defective attributes (e.g. torn lampshades; dangerous flex) and the proportion of the 15 rooms which was defective on each attribute was noted as an estimate of the corresponding proportion of rooms in the hotel as a whole which were defective on that attribute.

(ii) There is some acceptable minimum standard for a hotel room (say, for example, 4 or fewer minor defects) and it was noted whether each of the sample of 15 rooms fell below this minimum standard, or not. In this case, interest would focus on the proportion of rooms in the sample which were defective on this minimum standard and this was used as an estimate of corresponding proportion in the hotel as a whole.

What generalisations the owner's consultants made were difficult to say from the sections of their report available to the statistician. However, it appeared that they merely listed all the single defects noted on their tour and made a number of subjective and vaguely phrased generalisations. It might be expected that all rooms might have some minor defect or another, therefore it would seem that (ii) above should be the focus of any generalisation. Further, we might suppose that, apart from a minimum standard for each room which could be observed, there was a minimum standard for the hotel as a whole: i.e. if x% of rooms in the hotel or more were to fall below the minimum room standard, this would be judged unsatisfactory. Certain values of x are possible but in what follows I have assumed a range of say 5 to 20. In certain circumstances (i) above might be judged the appropriate approach if it was desired to focus on specific types of fundamental defect, and similar comments about minimum standards for the hotel as a whole could be made. The tables and comments that follow could apply to either situation so that the hotel proportion, π, could be either, (i) proportion of rooms with a specific defect or (ii) proportion of rooms which fell below the minimum standard. In the following comments, however, it is assumed that (ii) is the focus of attention.

## ◆SAMPLING ERRORS ◆

From standard sampling theory the estimated standard error of the estimate of a proportion in a finite population from a simple random sample is

$$\sqrt{\frac{p(1-p)}{n-1}\left(\frac{N-n}{N}\right)}$$

where p is the sample proportion. This can be used to form confidence intervals for π at say 95% and 99% levels. In layman's terms a 95% confidence interval is a range of values for the hotel proportion which we are reasonably certain covers the true value; or put another way we have only a 1 in 20 chance of being wrong if we assert that the true proportion lies in that range. The estimated standard error for $N$ = 289 ranges from 0.057 when the estimate of π is p = 0.05, to 0.104 when p = 0.20 to 0.1301 when p is 0.50. The latter is the maximum value of this estimated standard error. 95% confidence intervals are approximately two standard errors (s.e.) on either side of an estimate and 99% intervals are approximately 2.5 s.e. on either side. These approximations are based on the usual approximation to the binomial distribution by the normal. Exact intervals could be based on the binomial but the approximation should be sufficient to make the point desired here. Also, in applying the normal approximation for small values of p. the lower limit of π may appear negative. This is clearly a practical impossibility. However, again it is due to the roughness of the normal approximation for such values. The general point about the width of the intervals should again not be invalidated and the lower limits can be taken as practically zero.

Adopting this convention and, for example, supposing the number of 'defectives' in the example was 3, yielding p = 0.20, a 95% interval for the true π would be approximately 0 to 0.408. Similar calculations could be made for other values of p and for 99% or other size intervals. However, it is clear that there is a large chance of a big **margin of error with a** sample of size 15. Results are little affected if $N$ = 217 rather than 289, and this emphasises the point made in the introduction that it is absolute sample size rather than sampling fraction which should be the focus of interest.

## ◆NECESSARY SAMPLE SIZES ◆ TO ACHIEVE DESIRED PRECISION

We might ask what sample size was required to achieve a minimum desirable precision in our estimate of $\pi$. Thus, suppose we desired to achieve a standard error of no more than $d = 0.02$, then what size of $n$ is required? If $d$ was 0.02 this would mean that the half-width of our confidence interval would be 0.04 for 95% confidence and 0.05 for a 99% confidence interval. From standard sampling theory the minimum sample size is

$$n = \frac{N\pi(1-\pi)}{\pi(1-\pi) + (N-1)d^2}$$

This required sample size varies according to the true value of $\pi$ and $d$ in a way indicated in some selected cases in Table 1 below:

**Table 1.** Minimum sample size to achieve desired precision in estimating π

| | Desired standard error (100d%) | | | | | |
|---|---|---|---|---|---|---|
| | **1%** | **2%** | **3%** | **4%** | **5%** | **10%** |
| N=289; π=50% | 260 | 198 | 142 | 102 | 75 | 23 |
| N=217; π=50% | 199 | 161 | 122 | 91 | 69 | 22 |
| N=289; π=l0% | 219 | 126 | 75 | 47 | 32 | 9 |
| N=217;π=10% | 175 | 111 | 69 | 45 | 31 | 9 |
| N=289; π=5% | 179 | 84 | 45 | 27 | 18 | 5 |
| N=217; π=5% | 149 | 77 | 43 | 26 | 17 | 5 |

The general conclusion to emerge from an inspection of the above table is that for practically all reasonable situations a sample size far in excess of $n = 15$ would be required. Thus a sample of at least 30 would be required unless the true $\pi$ was less than 0.10 AND we desired to estimate it with a standard error which is relatively quite large (certainly greater than 0.035). In confidence interval terms we would be willing to accept a sample size of at most 30 if we are willing to accept such a large margin of error that gives us a 95% confidence interval length from our sample of 0.14 for something that in fact is around a value of 0.10. It is doubtful that this would.be a desirable state of affairs if we were to question the investigators.

## ◆THE NULL HYPOTHESIS ◆ THAT HOTEL MEETS STANDARD

We might also approach the problem in a hypothesis testing framework which could lead to an investigation of Type II errors in this situation.

This approaches the problem from the angle of supposing that there is a maximum acceptable figure for $\pi$ of between 5% and 20%. What would be required in the nature of the sample result to reject the null hypothesis that the hotel meets the minimum standard, if the sample was only 15 in a population of 289? Table 2 presents a range of such possibilities.

Table 2. Results of testing hypothesis that minimum standard is met for various values *of p.*

| **Maximum acceptable** | **Sample estimate (l00p%)** | | | | |
|---|---|---|---|---|---|
| **π** | **5%** | **10%** | **20%** | **30%** | **40%** |
| 5% | *NR* | *NR* | + | ++ | ++ |
| 10% | *NR* | *NR* | *NR* | + | ++ |
| 20% | *NR* | *NR* | *NR* | *NR* | + |

*NR* indicates inability to reject at 5% level.
\+ indicates rejection at 5% level.
++ indicates rejection at 1% level.

As will be seen, the general conclusion is that the **sample result** would have to be **considerably greater** than the desirable minimum standard if it was to be claimed that the hotel as a whole fell below this minimum standard. Power could only be increased for much larger sample sizes.

We may now investigate what is the minimum sample size necessary to establish rejection of the above hypothesis at the 5% level for a range of desired minimum standards and a range of sample results $p$ (N=289), i.e. to conclude that results indicate that the hotel as a whole is unsatisfactory? Table 3 presents a range of results.

**Table 3.** Necessary sample size to achieve rejection of hypothesis of satisfactory standards for specific sample results.

| **Desired maximum acceptable** | **Observed sample estimates (l00π%)** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **l00π%** | **7.5%** | **10%** | **15%** | **20%** | **25%** | **30%** | **35%** | **40%** |
| 5% | 120 | 44 | 12 | | | | | |
| 10% | | 289 | 73 | 22 | 10 | | | |
| 15% | | | 289 | 108 | 38 | 15 | | |
| 20% | | | | 289 | 109 | 36 | 18 | |
| 25% | | | | | 289 | 119 | 43 | 20 |

As will be seen, a sample size considerably in excess of the one used would be necessary, unless the sample result was considerably greater than the desirable maximum value of $\pi$. For a rejection at the 1% level the sample sizes would have to be even larger. Results are little affected by population size $N$.

## ◆CONCLUSION◆

We have seen that it is possible to give some advice on the suitability of certain sample sizes in simple random sampling once the objectives of a survey have been clarified. It is clear in this example that, in an examination of the problem from several inferential standpoints, the sample size chosen was far too small. However, it is demonstrated that fairly simple approaches can lead to constructive suggestions about adequate sample sizes. The question is one faced by any statistician whose advice is sought on planning a survey and sometimes not an easy one to answer. This example might serve to illustrate the line of approach that may be chosen in many common situations when faced with a finite population.