



Note from the editor: The editors are engaged in continuing research on the use of electronic distribution lists as a venue for adult incidental learning. Dr. Hill was privately asked by the editors the question that entitles this article. With so many persons now also engaged in survey research on the Internet, we felt - as did the IPCT-J reviewers - his elaborated response would be a useful source of information and reference.

WHAT SAMPLE SIZE is "ENOUGH" in INTERNET SURVEY RESEARCH?

Dr. Robin Hill

The Waikato Polytechnic Hamilton, New Zealand

INTRODUCTION

One of the most frequently asked questions of a research director or mentor is "what size sample should I use?" It is a question pertinent to all forms of research, but a question that creates awkwardness when considering internet based electronic survey (e-survey) methods. When we sample, we are drawing a subgroup of cases from some population of possible cases (say, a subgroup of listserv administrators, worldwide). The sample will deviate from the true nature of the population by a certain amount due to chance variations of drawing few cases from many possible cases. We call this sampling error ([Isaac & Michael, 1995](#)). The issue of determining sample size arises with regard to the investigator's wish to assure that the sample

statistic (usually the sample mean) is as close to, or within a specified distance from, the true statistic (mean) of the entire population under review. There is a good illustration of this notion in [Frankfort-Nachmias & Nachmias](#) (1996, p. 194-195). As [Weisberg & Bowen](#) (1977) point out, once the study is complete it is usually too late to discover that your sample was too small or disproportionately composed.

Calculation of an appropriate sample size generally depends upon the size of the population in question (although [Alreck & Settle](#), 1995, dispute this logic). The suggested sample size for a survey of students enrolled in a first-year course at a university would be a function of the total number of students (population) enrolled in that course. This would be a known, finite number of students. That is where the awkwardness of e-survey arises. Investigators generally cannot determine, nor even guess the size of the population they are interested in; cannot guess the number of subscribers sitting at keyboards exploring the internet. The awkwardness is also compounded by lack of representativeness; e-survey investigators are restricting their studies not just to those with computer equipment but to those of them who have connected their equipment to the outside world. Hence there is a bias in the data gained, since the opinions of those who do not have access to the internet or e-mail have been excluded from the study.

However, take the scenario mentioned briefly above. Suppose an investigator wished to survey those people who "owned," moderated or administered listserv groups. How many would need to be sampled? It is difficult to answer this question, since it is difficult to know how many listserv groups there are and because they're growing in number by the day. If the survey took a month or more to complete, then the population of users may be quite different at the end of the project, compared to when it began.

[Martin and Bateson](#) (1986) indicate that to a point, the more data collected the better, since statistical power is improved by increasing the sample size. However, indefinite collection of data must be weighed up against time, since at some point it becomes more productive to move onto a new study rather than persist with the current one. When sufficient results have been acquired, according to Martin & Bateson, additional results may add little to the conclusions to be drawn.

The problem of sample size may arise in any one of three forms - all under the heading "How many observations do I need to make?" The three forms of problem are:

(a) How many people to use as respondents. If the parent population is 1400 people, how many people should be sampled?

and/or...

(b) Within the sample, what should be the size or proportion of sub-populations within it? If the parent population is 1400 people, then what proportion of the sample should be males, females, other ethnic groups, etc.?

or...

(c) If in a naturalistic setting, or a single-subject case study, how many independent observation sessions are required? For example, if observing a listserv group, and measuring the number of times people from different countries log in or contribute to the discussion, then how many hours of observation are required; on how many days do I need to observe this? [Miles & Huberman](#) (1994) point out that sampling involves not only decisions about which people to observe, but also which settings, events and or social processes.

"How large should the sample be?" [Gay & Diehl](#) (1992) indicate that the correct answer to this question is: "Large enough". While this may seem flippant, they claim that it is indeed the correct answer. Knowing that the sample should be as large as possible helps, but still does not give guidance as to what size sample is "big enough". Usually the researcher does not have access to a large number of people, and in business or management research, and no doubt in e-surveys, obtaining informed consent to participate is not an easy task. Usually the problem is too few subjects, rather than determining where the cut-off should be for "large enough".

As indicated, below, choice of sample size is often as much a budgetary consideration as a statistical one ([Roscoe](#), 1975; [Alreck & Settle](#), 1995), and by budget it is useful to think of all resources (time, space and energy) not just money. [Miles and Huberman](#) (1994) indicate that empirical research is often a matter of progressively lowering your aspirations. You begin by wanting to study all facets of a problem, but soon it becomes apparent that choices need to be made. You have to settle for less. With this knowledge that one cannot study everyone, doing everything (even a single person doing everything), how does one limit the parameters of the study?

The crucial step, according to [Miles & Huberman](#) (1994), is being explicit about what you want to study and why. Otherwise you may suffer the pitfalls of vacuum-cleaner-like collection of every datum. You may suffer accumulation of more data than there is time to analyse and detours into alluring associated questions that waste time, goodwill and analytic opportunity. Between the economy and convenience of small samples and the reliability and representativeness of large samples lies a trade-off point, balancing practical considerations against statistical power and generalisability.

[Alreck & Settle](#), (1995) suggest that surveyors tend to use two strategies to overcome this trade-off problem: Obtain large amounts of data from a smaller sample, or obtain a small amount of data from a large sample.

ROSCOE'S SIMPLE RULES OF THUMB

The formulae for determining sample size tend to have a few unknowns, that rely on the researcher choosing particular levels of confidence, acceptable error and the like. Because of this, some years ago [Roscoe](#) (1975) suggested we approach the problem of sample size with the following rules of thumb believed to be appropriate for most behavioural research. Not all of these are relevant to e-survey, but are worthy of mention all the same.

1 The use of statistical analyses with samples less than 10 is not recommended.

2 In simple experimental research with tight controls (eg. matched-pairs design), according to Roscoe, successful research may be conducted with samples as small as 10 to 20.

3 In most ex post facto and experimental research, samples of 30 or more are recommended. [Experimental research involves the researcher in manipulating the independent variable (IV) and measuring the effect of this on a dependant variable (DV). It is distinguished from ex post facto research where the effect of an independent variable is measured against a dependant variable, but where that independent variable is NOT manipulated. For example, a researcher may be interested in the effect of a specific kind of brain damage (the IV) on computer learning skills (the DV). The researcher is hardly in a position to manipulate or vary

the amount of brain damage in the subjects. The researcher must, instead, use subjects who are already brain damaged, and to classify them for the amount of damage.]

4 When samples are to be broken into sub-samples and generalisations drawn from these, then the rules of thumb for sample size should apply to those sub samples. For example, if comparing the responses of males and females in the sample, then those two sub-samples must comply with the rules of thumb.

5 In multivariate research (eg. multiple regression) sample size should be at least ten times larger than the number of variables being considered.

6 There is seldom justification in behavioural research for sample sizes of less than 30 or larger than 500. Samples larger than 30 ensure the researcher the benefits of central limit theorem (see for example, [Roscoe](#), 1975, p.163 or [Abranovic](#), 1997, p. 307-308). A sample of 500 assures that sample error will not exceed 10% of standard deviation, about 98% of the time.

Within these limits (30 to 500), the use of a sample about 10% size of parent population is recommended. [Alreck & Settle](#) (1995) state that it is seldom necessary to sample more than 10%. Hence if the parent population is 1400, then sample size should be about 140.

While Roscoe advocates a lower limit of 30, [Chassan](#) (1979) states that 20 to 25 subjects per IV group would appear to be an absolute minimum for a reasonable probability of detecting a difference in treatment effects. Chassan continues, that some methodologists will insist upon a minimum of 50 to 100 subjects. Also, in contrast to Roscoe, [Alreck & Settle](#) (1995) suggest 1,000 as the upper limit.

7 Generally choice of sample size is as much a function of budgetary considerations as it is statistical considerations. When they can be afforded, large samples are usually preferred over smaller ones.

DETERMINING SUFFICIENCY OF THE SAMPLE SIZE

A crude method for checking sufficiency of data is described by [Martin & Bateson](#) (1986), as "split-half analysis of consistency." This seems a useful tool for e-survey investigators. Here the data is divided randomly into two halves which are then analysed separately. If both sets of data clearly generate the same conclusions, then sufficient data is claimed to have been collected. If the two conclusions differ, then more data is required.

True split-half analysis involves calculating the correlation between the two data sets. If the correlation coefficient is sufficiently high ([Martin & Bateson](#), 1986, advocate greater than 0.7) then the data can be said to be reliable. Split-half analysis provides the opportunity to carry out your e-survey in an ongoing fashion, in small manageable chunks, until such time as an acceptable correlation coefficient arises.

As stated earlier [Alreck and Settle](#) (1995) dispute the logic that sample size is necessarily dependant upon population size. They provide the following analogy. Suppose you were warming a bowl of soup and wished to know if it was hot enough to serve. You would probably taste a spoonful. A sample size of one spoonful. Now suppose you increased the population of soup, and you were heating a large urn of soup for a large crowd. The supposed population of soup has increased, but you still only require a sample size of one spoonful to determine

whether the soup is hot enough to serve.

A number of authors have provided formulae for determining sample size. These are too many and varied to reproduce here, and are readily available in statistics and research methods publications.

A formula for determining sample size can be derived provided the investigator is prepared to specify how much error is acceptable ([Roscoe](#) 1975, [Weisberg & Bowen](#) 1977, [Alreck & Settle](#) 1995) and how much confidence is required ([Roscoe](#) 1975, [Alreck & Settle](#) 1995). Readers are advised to see [Frankfort-Nachmias & Nachmias](#) (1996) for a more detailed account of the role of standard error and confidence levels for determining sample size. A probability (or significance) level of 0.05 has been established as a generally acceptable level of confidence in most behavioural sciences. There is more debate about the acceptable level of error, and just a hint in the literature, that it is what ever the Investigator decides as acceptable.

[Roscoe](#) seems to use 10% as a "rule of thumb" acceptable level. [Weisberg & Bowen](#) (1977) cite 3% to 4% as the acceptable level in survey research for forecasting election results, and state that it is rarely worth the compromise in time and money to try to attain an error rate as low as 1%. Beyond 3% to 4%, an extra 1% precision is not considered worth the effort or money required to increase the sample size sufficiently.

[Weisberg & Bowen](#) (1977, p. 41), in a book dedicated to survey research, provide a table of maximum sampling error related to sample size for simple randomly selected samples. This table, reproduced below as Table One, insinuates that if you are prepared to accept an error level of 5% in your e-survey, then you require a sample size of 400 observations. If 10% is acceptable then the a sample of 100 is acceptable, provided the sampling procedure is simple random.

Table One: Maximum Sampling Error For Samples Of Varying Sizes

| Sample Size | Error |
|-------------|-------|
| 2,000 | 2.2 |
| 1,500 | 2.6 |
| 1,000 | 3.2 |
| 750 | 3.6 |
| 700 | 3.8 |
| 600 | 4.1 |
| 500 | 4.5 |
| 400 | 5.0 |
| 300 | 5.8 |
| 200 | 7.2 |
| 100 | 10.3 |

(Based on Weisberg & Bowen,1977, p. 41)

[Krejcie & Morgan](#) (1970) have produced a table for determining sample size. They did this in response to an article called "Small Sample Techniques" issued by the research division of the National Education Association. In this article a formula was provided for the purpose, but, according to Krejcie & Morgan, regrettably an easy reference table had not been provided.

They therefore produced such a table based on the formula. No calculations are required to use the table which is also reproduced below, as Table Two. According to Krejcie & Morgan, if one wished to know the sample size required to be representative of the opinions of 9,000 specified electronic users, then one enters the table at N=9,000. The sample size in this example is 368. The table is applicable to any population of a defined (finite) size.

Table Two

Required Sample Size, Given A Finite Population, Where N = Population Size and n = Sample Size

| N - n | N - n | N - n | N - n | N - n |
|---------|-----------|-----------|------------|--------------|
| 10 - 10 | 100 - 80 | 280 - 162 | 800 - 260 | 2800 - 338 |
| 15 - 14 | 110 - 86 | 290 - 165 | 850 - 265 | 3000 - 341 |
| 20 - 19 | 120 - 92 | 300 - 169 | 900 - 269 | 3500 - 346 |
| 25 - 24 | 130 - 97 | 320 - 175 | 950 - 274 | 4000 - 351 |
| 30 - 28 | 140 - 103 | 340 - 181 | 1000 - 278 | 4500 - 354 |
| 35 - 32 | 150 - 108 | 360 - 186 | 1100 - 285 | 5000 - 357 |
| 40 - 36 | 160 - 113 | 380 - 191 | 1200 - 291 | 6000 - 361 |
| 45 - 40 | 170 - 118 | 400 - 196 | 1300 - 297 | 7000 - 364 |
| 50 - 44 | 180 - 123 | 420 - 201 | 1400 - 302 | 8000 - 367 |
| 55 - 48 | 190 - 127 | 440 - 205 | 1500 - 306 | 9000 - 368 |
| 60 - 52 | 200 - 132 | 460 - 210 | 1600 - 310 | 10000 - 370 |
| 65 - 56 | 210 - 136 | 480 - 214 | 1700 - 313 | 15000 - 375 |
| 70 - 59 | 220 - 140 | 500 - 217 | 1800 - 317 | 20000 - 377 |
| 75 - 63 | 230 - 144 | 550 - 226 | 1900 - 320 | 30000 - 379 |
| 80 - 66 | 240 - 148 | 600 - 234 | 2000 - 322 | 40000 - 380 |
| 85 - 70 | 250 - 152 | 650 - 242 | 2200 - 327 | 50000 - 381 |
| 90 - 73 | 260 - 155 | 700 - 248 | 2400 - 331 | 75000 - 382 |
| 95 - 76 | 270 - 159 | 750 - 254 | 2600 - 335 | 100000 - 384 |

(Adapted from Krejcie & Morgan, 1970, p.608)

Krejcie and Morgan state that, using this calculation, as the population increases the sample size increases at a diminishing rate (plateau) and remains, eventually constant at slightly more than 380 cases. There is little to be gained to warrant the expense and energy to sample beyond about 380 cases. [Alreck and Settle](#) (1995) provide similar evidence.

According to [Gay & Diehl](#), (1992), generally the number of respondents acceptable for a study depends upon the type of research involved - descriptive, correlational or experimental.

For **descriptive research** the sample should be 10% of population. But if the population is small then 20% may be required.

In **correlational research** at least 30 subjects are required to establish a relationship.

For **experimental research**, 30 subjects per group is often cited as the minimum.

LARGE OR SMALL SAMPLE SIZES?

[Isaac and Michael](#) (1995) provide the following conditions where research with large samples is essential and also where small samples are justifiable:

1 Large Samples are essential:

- (a) When a large number of uncontrolled variables are interacting unpredictably and it is desirable to minimise their separate effects; to mix the effects randomly and hence cancel out imbalances.
- (b) When the total sample is to be sub-divided into several sub-samples to be compared with one another.
- (c) When the parent population consists of a wide range of variables and characteristics, and there is a risk therefore of missing or misrepresenting those differences - a potential with e-survey, considering that the parent population is now global.
- (d) When differences in the results are expected to be small.

2 Small sample sizes are justifiable:

- (a) In cases of small sample economy. That is when it is not economically feasible to collect a large sample.
- (b) When computer monitoring. This may take two forms: (i) Where the input of huge amounts of data may itself introduce a source of error - namely, key punch mistakes. (ii) Where, as an additional check on the reliability of the computer program, a small sample is selected from the main data and analysed by hand. The purpose of this is to compare the small sample data, and the large sample data for similar results.
- (c) In cases of exploratory research and pilot studies. Sample sizes of 10 to 30 are sufficient in these cases. They are large enough to test the null hypothesis and small enough to overlook weak treatment effects. Statistical significance is unlikely to be obtained on this size sample however.
- (d) When the research involves in-depth case study. That is, when the study requires methodology such as interview and where enormous amounts of qualitative data are forthcoming from each, individual respondent.

Presumably, to these may be added the converse of those listed under 1, above. Namely: when control is extremely tight and interacting variables are neither large in number nor unpredictable; when the population is homogeneous; when differences in the results are expected to be very large.

[Gay & Diehl](#) (1992) state that in one way the typically smaller sample sizes used in applied or practical research have a redeeming feature. Their argument states that large sample sizes enhance the likelihood of yielding statistically significant results. Thus with very large sample sizes, a very small difference between means may yield a significant result, and yet be of little practical use. It results in research that [Bannister](#) (1981) described as specifically and precisely irrelevant. On the other hand, if you obtain a statistically significant result from a small sample size, then the impact of the difference is probably more obvious and useful - but this is

admittedly tenuous argument and Gay & Diehl advise care in interpretation of results.

SUMMARY

It appears that determining sample size for an e-survey is not a cut-and-dried procedure. Despite a large amount of literature on the topic, seemingly in all cases there is an element of arbitrary judgement and personal choice involved. Maybe the terms "arbitrary" and "personal choice" are too harsh ... "informed judgement" may come closer to the mark.

It is obvious that the nature of the methodology used is a major consideration in selecting sample size. For instance, if the methodology attracts large amounts of qualitative information, as is the case with ideographic techniques such as interview, case study or repertory test, then practical constraints may mean that the researcher needs to settle for a small sample size. In these circumstances the argument goes that it is better to have collected some data, to have gained some information and to have done some research, than to have collected no data, gained no information, and to have conducted no research. A good deal of important information would be missed if we insisted on large sample sizes always. The analysis of the content of messages in a listserv group's archives would fall within this category.

In other cases [Roscoe's](#) rules of thumb might be kept in mind.

If the researcher is in a position to keep collecting data and to assess the sufficiency of the sample size as the research progresses, then the split-half method of [Martin & Bateson](#) (1986) may be useful. This seems to be a specific advantage of some e-survey methods. For example where the e-survey can remain available on the internet for a sustained period of time.

Because of the problem of getting "enough" respondents and the problems that raises regarding generalisability, according to [Gay & Diehl](#) (1992) there is therefore, a great deal to be said for replication of findings. The current author takes this to mean replication of research (a) to increase the subject pool and (b) to create greater validity for generalisability.

When preparing this document, the author created a standard scenario (finite population with a known statistical mean and standard deviation). This related to use of a specific IQ test where the population mean is known to be 100 and the population standard deviation is known to be 15. This scenario was subjected to 7 different formulae found in the literature for establishing sample size, including [Roscoe's](#) rules of thumb. The result produced 7 different "required" sample sizes, with enormous spread (from a sample size of 35 through to 400 for the same research scenario). This outcome reinforces the view that there is no one accepted method of determining necessary sample size.

[Gay & Diehl](#) (1992) refer the reader to [Cohen](#) (1988) for more precise statistical techniques for estimating sample size.

REFERENCES

Abranovic, W.A. (1997) *Statistical Thinking and Data Analysis for Managers*. Reading, MA: Addison-Wesley.

Alreck, P.L. & Settle, R.B. (1995) *The Survey Research Handbook*, 2nd edition. Chicago: Irwin.

Bannister, D. (1981). Personal Construct Theory and Research Method. In P. Reason & J. Rowan (Eds.). *Human Inquiry*. Chichester: John Wiley & Sons.

- Chassan, J.B. (1979). *Research Design in Clinical Psychology and Psychiatry*. New York: Irvington Publishers Inc.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioural Sciences*, 2nd edition. Hillsdale, N.J.: Lawrence Erlbaum.
- Frankfort-Nachmias, C. & Nachmias, D. (1996). *Research Methods in the Social Sciences*. Fifth Edition. London: Arnold.
- Gay, L.R. & Diehl, P.L. (1992). *Research Methods for Business and Management*. New York: Macmillan.
- Isaac, S. & Michael, W.B. (1995). *Handbook in Research and Evaluation*. San Diego: EdITS.
- Krejcie, R.V. & Morgan, D.W. (1970). Determining sample size for research activities. *Educational & Psychological Measurement*, 30, 607-610.
- Martin, P. & Bateson, P. (1986). *Measuring Behaviour: An Introductory Guide*. Cambridge: Cambridge University Press.
- Miles, M.B. & Huberman, A.M. (1994). *Qualitative Data Analysis*. Beverly Hills: Sage.
- Roscoe, J.T. (1975) *Fundamental Research Statistics for the Behavioural Sciences*, 2nd edition. New York: Holt Rinehart & Winston.
- Weisberg, H.F. & Bowen, B.D. (1977). *An Introduction to Survey Research and Data Analysis*. San Francisco : W. H. Freeman.

BIOGRAPHICAL NOTES

Robin Hill has a PhD in organisational psychology with a particular interest in Personal Construct Psychology. Following time spent lecturing in psychology and working in industry as a Human Resource Manager, Robin has completed 6 years at The Waikato Polytechnic, where he is a Principal Lecturer in organisational behaviour. He is also the Research Leader for the department of business studies: a role in which he is a mentor for colleagues, facilitates their research activities and teaches research methods.

Address for correspondence:

bsrah@TWP.AC.NZ

Copyright Statement

Interpersonal Computing and Technology: An Electronic Journal for the
21st Century

© 1998 The Association for Educational Communications and Technology. Copyright of individual articles in this publication is retained by the individual authors. Copyright of the compilation as a whole is held by AECT. It is asked that any republication of this article state that the article was first published in IPCT-J.

Contributions to IPCT-J can be submitted by electronic mail in APA style to:

Susan Barnes, Editor

SBBARNES@PIPELINE.COM or
BARNES@MURRAY.FORDHAM.EDU