MESO-00-A003-06

An Introduction To Environmental Sampling Planning

15 September 2000

Prepared By:

C. Kurtz, CSC
Marine Environmental Support Office
SPAWARSYSCEN D3621
53475 Strothe Road
San Diego, CA 92152-6326

EXECUTIVE SUMMARY

One of the most difficult challenges scientists/environmental managers must address when conducting a site assessment or survey is determining the number of samples to be taken. Taking too many samples will waste time and resources, both in collecting and analyzing the data. On the other hand, taking too few samples can make the whole study meaningless or lead to errors in interpretation. This paper offers some guidelines on how to take available data (or to use pilot study data) and design a sampling plan to suit the needs of a specific project in the most time-efficient and cost effective way.

ACKNOWLEDGEMENTS

The author and the Marine Environmental Support Office are grateful to Ms. Vikki Kirtay, Environmental Chemistry and Biotechnology Branch, and Dr. Ken Richter, Marine Environmental Quality Branch, Space and Naval Warfare Systems Center, San Diego, for their invaluable assistance in the preparation of this document.

TABLE OF CONTENTS

1.0 Introduction

One of the most difficult challenges scientists/environmental managers must address when conducting a site assessment or survey is determining the number of samples to be taken. Too many samples will waste time and resources, both in collection and analysis of the data. On the other hand, too few samples can make the whole study scientifically indefensible, or even worse, lead to errors in interpretation (Eckblad, 1991). Consideration must be given to cost of the sampling itself and to the availability of the resources needed to complete the analysis. In some cases, time of year is also a factor and will influence the design of a sampling plan and also the statistics used to analyze the data. Statistics can be defined as a theory of information. Such information is obtained by experimentation or, equivalently, by sampling; it is employed to make an inference about a larger set of measurements, called a population (Ott, 1988).

The aim of this paper is to offer some guidelines on how to take available data (or to use pilot study data) and design a sampling plan to suit the needs of a specific project in the most time-efficient and cost effective way.

2.0 Hypothesis Testing

The cornerstone of scientific analysis is hypothesis testing. The idea is that you formulate a hypothesis (H_1) into a statement, collect appropriate data and then use statistics to determine whether the hypothesis is true or not. However, the statistical tests do not provide a simple answer of true or not. For every hypothesis there will be an associated null hypothesis (H_0), which is the parameter that most statistical procedures test.

The null hypothesis is simply a statement of "no difference" between the actual value (derived from an experiment) and an expected value. A statistical test determines the probability that the null hypothesis is true (p-value). If the probability is low then the null hypothesis is rejected and the original hypothesis (H_1) accepted.

Wrong inferences, however, can be drawn for testing the null hypothesis. These are called Type I (α) and Type II (β) errors. In a Type I error, the null hypothesis is really true but the statistical test leads to the assumption that it is false. This type of error thought of as a "*false positive*". In a Type II error, the null hypothesis is really false but the test has not conclusively established this difference. Small sample sizes can often lead to a Type II error. Although Type II errors are less serious than Type I errors, a Type II error should still be avoided, if possible. Usual convention is to use a critical p-value of 0.05, which states that the probability of a null hypothesis being true is 5%. (Dytham, 1999). Depending on the study objectives and subject matter, the wrong interpretation of Type I and Type II errors can provide misleading results.

3.0 Sample Design

Several things must be considered before the sample design and analysis can occur. Study objectives (hypotheses) must be clearly stated, as well as conceptually defined in temporal and spatial terms (*i.e.* determination of time frame and physical distances necessary to define a population). Collection of background information and historic data applicable to a given site is crucial in picking the optimal sampling method. Potential trends and variability in historic data can be examined for usefulness in a current study, assuming there is confidence in the original study design. Part of the pre-planning process is to determine sample types required and other field parameters to collect. Above all else, a solid quality assurance program must be implemented. (Gilbert, 1987). If a proposed sampling site has large-scale environmental patterns of some kind, break the area up into homogenous subsections and take samples from those areas that are proportional to the size of the subsection. Additionally, pick a sample size that will give the precision desired, yet be representative of the population as a whole (Green, 1979). Determine what statistics to use for sample analysis, and assess the uncertainties associated with estimated quantities like arithmetic means, trends, average maximums, *etc.* (Gilbert, 1987).

This paper focuses on the two concepts presented by Green (1979) because they are the key to effective sampling design and data gathering: 1) breaking up sampling sites with large-scale environmental patterns into homogenous subsections and taking subsamples from those areas that are proportional to the size of the subsection, and 2) picking a sample size that will give you the precision desired and yet be representative of the population as a whole.

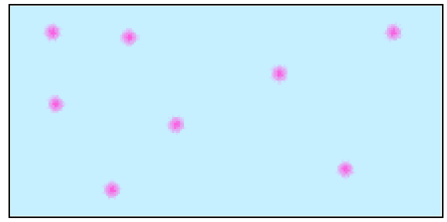
Depending on site specific conditions and program objectives, multiple sampling techniques may be used to best characterize each different regions within an area. Also, certain analytical techniques require sampling be conducted in certain ways. The four basic methods for sample collection are:

- haphazard sampling;
- judgement sampling;
- probability (statistical) sampling; and
- search sampling.

Haphazard sampling consists of sample collection from anywhere within the study area the investigator wishes, which can obviously lead to severely biased results. Judgement sampling involves the selection of specific sampling locations by professional judgement. If the person or group is knowledgeable, this method may actually result in accurate data collection, but the degree of accuracy can be difficult to quantify. Hot spots and area of pollution sources, for instance, may be found through the use of judgment sampling. Be advised, however, this technique requires a lot of background data to support station selection and can possibly require more high tech equipment to complete accurately. Probability sampling uses a specific method of random sampling and allows for determination of spatially distributed variables, as well as allowing sampling along time and space scales. There are six probability sampling methods: simple random sampling, stratified random sampling, two-stage sampling, cluster sampling, systematic sampling, and double sampling (Gilbert, 1987).

3.1 Simple Random Sampling

An unbiased estimate of a population is only possible if the sample units are representative of the total population. The easiest way of achieving this is for each sample unit to contain a random sample of the population under investigation (Dytham, 1999). A random sample may be defined as a sample drawn in such a way as to ensure that every member of the population has an equal chance of being included. Random sampling procedures increase the chances that a sample will be unbiased and one of their major purposes is to minimize possible bias on the part of the experimenter (Schefler, 1979).



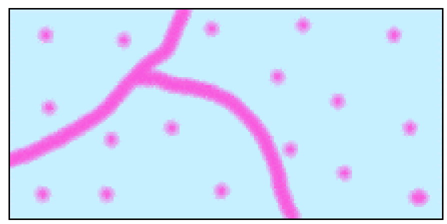
Simple Random Sampling

Simple random sampling is an efficient way for estimating means and totals if the population does not contain major trends, cycles, or patterns of contamination (Gilbert, 1987). There are several techniques which can be used for random sampling. Random sampling gives an unbiased estimate of population parameters, allows for the calculation of confidence intervals and allows hypothesis testing to be carried out. It is very time consuming, however, to implement. Basic assumptions of random sampling include:

- individuals are selected in one stage;
- the chance of being selected is equal for each sample unit in the population; and
- selection of one sample does not effect the chances that another sample would be selected. (Portier & Arvanitis, 1998).

3.2 Stratified Random Sampling

Stratified Random Sampling involves the division of the target population into subdivision or strata in order to get a better estimate of the mean or total of the entire population. Sampling locations within each strata are then selected using the simple random sampling technique. This is useful to break down a heterogeneous population into groups that are themselves homogeneous (Gilbert, 1987). Strata must be constructed so that strata averages are as different as possible and that strata variances are as small as possible.



Stratified Random Sampling

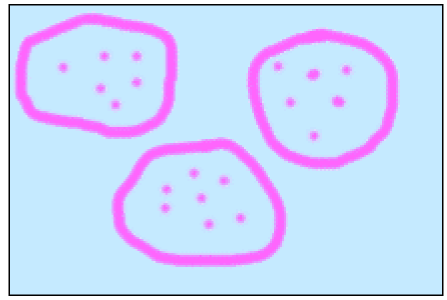
Stratified random sampling requires prior knowledge of the population being sampled and is very time consuming. It is, however, likely to give a more representative sample of the population. The main assumption of this sampling method is that all variables of interest are variable within the population (Portier & Arvanitis, 1998).

3.3 Two-Stage Sampling

Two-Stage Sampling involves taking samples (primary units) using the simple random sampling technique and then taking an aliquot of those samples (Gilbert, 1987). Two-stage sampling allows for a lower error than cluster sampling and costs less than simple random sampling for larger populations. It does, however, have a higher error than simple random sampling.

3.4 Cluster Sampling

Cluster Sampling is very selective in how it can be employed. It is useful in situations where population units cluster together naturally (like schools of fish, rooted vascular plants) and each unit in each cluster can be measured independently of one another (Gilbert, 1987). Cluster sampling, in general, is easier to carry out and in most cases less expensive, but it may not be representative of the whole population because samples would not be random, which increases sample error and causes statistical calculations to be complex.

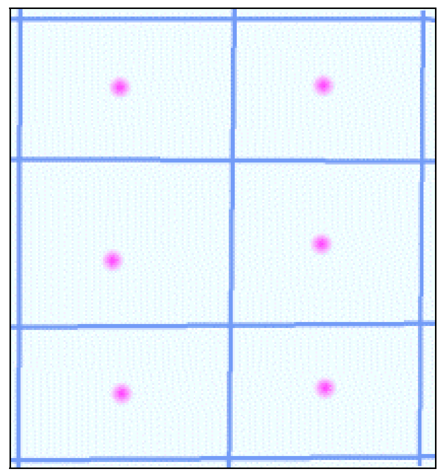


Cluster Sampling

3.5 Systematic Sampling

Systematic Sampling is good when the intention is to estimate trends or patterns over space. Samples are taken at even intervals after a random start is chosen. It can be a dangerous method to use if the sampling pattern corresponds to an unsuspected pattern over space or time (Gilbert, 1987). Cochran (1977) suggests using this sampling method when:

- the ordering of the population is essentially random or it contains at most a mild stratification;
- stratification with numerous strata is employed and an independent systematic sample is drawn from each stratum;
- subsampling cluster units; and
- sampling populations with variation of a continuous type, provided that an estimate of the sampling error is not regularly required.



Systematic Sampling

Systematic sampling may contain unsuspected bias, but is easy to do and quick to carry out. It can be expensive to carry out when dealing with large populations, however. It can be unbiased if the population is arranged randomly in respect to the parameter(s) being investigated. It assumes that the sample frame is randomly ordered with respect to values of the random variables of interest (Portier & Arvanitis, 1998).

3.6 Double Sampling

Double Sampling involves having samples collected in two different ways. (For example, using a portable monitoring device in the field along with traditional analysis from a certified environmental laboratory. It is useful when data using one measurement technique are nearly linear to data obtained from a less resource-intensive technique.

Double sampling determines the most cost effective way of both sampling and measuring samples. It enables a minimum sample to be taken for clear-cut cases of accept/reject hypotheses. This design assumes that there is a high degree of correlation between the supplemental and target measurement random variables, and that this degree of correlation can be expressed as a linear relationship between measurements (Portier & Arvanitis, 1998).

3.7 Search Sampling (Hot Spot Sampling)

Search Sampling is good for determining whether or not local areas of contamination, a.k.a. "hot spots" are present in a region. In order to use this method, the following assumptions are required:

1. the definition of a hot spot is precisely stated;
2. the hot spot is of the circular nature;
3. samples are taken on a triangular, square or rectangular grid;
4. only a very small proportion of the area being studied can actually be measured; and
5. there is no confusion in deciding when a hot spot was hit.

There are specific equations used with this kind of sampling design to calculate:

- grid spacing;
- the maximum size of a hot spot that can be located at a given risk and cost;
- the probability of that a hot spot exists even though it may not be found; and
- the probability of not finding a hot spot at all.

This type of sampling is most beneficially used for examining, in much greater detail, areas already known to be of concern, ones where prior samples have already been taken, and can pinpoint study areas more accurately. It is not good for estimating average concentrations over large areas, nor for evaluating large areas that have no prior background data available (Gilbert, 1987).

4.0 How Many Samples Are Enough?

The optimum number of sample to collect is nearly always limited by the amount of resources available. However, it is possible to calculate the number of samples required to estimate population size with a particular degree of accuracy. The best sample number is the largest sample number, keeping in mind that no sample number will compensate for poor sampling design. In other words, quantity should not be increased at the expense of quality. Poor quality data will have more inherent error and, therefore, make the statistics less powerful. (Dytham, 1999). Precision of estimated mean values will increase with increasing sample number, but there is a law of diminishing returns. Standard error is defined as an estimate of the standard deviation, or spread, of the sampling distribution of means, based on the data from one or more random samples. (Dytham, 1999). The standard error of the mean decreases in proportion to the square root of the sample number. An increase in sample number from 4 to 9 reduces the standard error by a third. To achieve another reduction by 1/3 a sample number of 21 would be required, and to achieve another, 46. (Green, 1979).

Elliott (1977) suggests a simple way, although limited in its applications, to estimate suitable sample size when dealing with benthic invertebrate samples. Elliott suggests taking samples in 5 sample-increments (5, 10, 15, 20…) and calculating the means of every 5 samples until the point is reached where sample means do not vary much. The sample number used to reach that point can be considered a suitable sample size for the study. This method is not useful when in the field, but is a quick approach if a small pilot study is to be conducted.

Each of the six types of probability sampling methods mentioned in the previous section have their own ways of calculating sample number given predetermined levels of variance, correlation and degree of error. See Gilbert (1987) & Elliott (1977) for specific equations and assumptions. An example is provided below for illustrative purposes. The equations utilized are not the only equations available. For further details, see references listed at the end of this paper.

5.0 Example

A local group of scientists has agreed to assist in collecting data on the abundance of an economically and recreationally important species of flatfish (legal catch-size) in a relatively large bay. A proper random sampling design must be developed in order to obtain useful data of high confidence and yet try to keep costs down. The bay area consists of commercial boatyards, numerous residential uses, marinas, oil tanks, agricultural uses (farms), a wetland, an airport, a recreational (beach) area, and several military operations. It is a good mix of military, industrial and commercial land/water use.

Pre-study sampling was used to estimate the sample mean and the sample variance by applying the following formula to calculate the needed sample size:

$$\text{Sample size} \cong \frac{(\text{t - value})^2\,(\text{sample variance})}{(\text{accuracy} \times \text{sample mean})^2}$$

where:

$$\text{t - value} = \frac{(\text{sample mean - population mean})}{\sqrt{\dfrac{\text{sample variance}}{\text{sample size}}}}$$

Alternatively, this can be obtained from a t-table at (n-1) degrees of freedom at p=0.05:

$$\text{sample variance} = \frac{\sum(x^2) - \dfrac{(\sum x)^2}{n}}{n - 1}$$

$$\text{accuracy} \approx \frac{(\text{t} - \text{value})\left(\sqrt{\dfrac{\text{sample variance}}{\text{sample size}}}\right)}{\text{sample mean}}$$

(Eckblad, 1991).

From extensive literature searches, it was found that a fish abundance study of the back-bay area had been conducted a few months earlier. The study derived the following results concerning the size lengths of the flatfish of concern.

The results were:

9, 11, 19, 11, 17, 8, 17, 7, 9, 5, 8, 11, 6, 8, and 11.

Using the formulas above, we calculate a few simple sample statistics

$$\text{Sample Mean} = \frac{9+11+19+11+17+8+17+7+9+5+8+11+6+8+11}{15} = 10.47$$

$$\text{Standard Error of the Mean} = \sqrt{\frac{\text{sample variance}}{\text{number of samples}}} = \sqrt{\frac{108.82}{15}} = 2.69$$

Sample Variance = [$(9)^2$+$(11)^2$+$(19)^2$+ $(11)^2$+$(17)^{2+}$ $(7)^2$+$(9)^2$+ $(5)^2$+$(8)^2$+ $(11)^2$+$(6)^2$+$(8)^2$+ $(11)^2$] – [$(9+11+19+11+17+8+17+7+9+5+8+11+6+8+11)^2$ ÷15] ÷14 = 108.82.

t-value = obtained from the *t*-table at (15-1) or 14 degrees of freedom at a 0.05 probability (p) level; equals 2.145.

$$\text{Accuracy} \cong \frac{(2.145)(\sqrt{\frac{108.82}{15}})}{8.47} = 0.682$$

Accuracy is defined as the closeness of a measure to its true value. It differs from precision in that precision deals with the closeness of repeated measures to the same value. (Dytham, 1999). Accuracy ties into estimation of sample size because the more accurate you are in your sampling, the less samples you have to take. Accuracy itself is a usually a practical and scientific consideration, rather than a statistical question.

$$\text{Sample Size} \cong \frac{(2.145)^2(108.82)}{(0.10 \times 8.47)^2} = 698 \text{ samples}$$

(Eckblad, 1991).

Based on the historical data, 698 samples would need to be taken during the main sampling in order to be within 10% of the true mean, at a 0.05% level of significance. Now that the amount of samples needed has been determined, dividing the total area of the section on the grid by the number of samples needed will tell you approximately what the spacing between each sample should be. You can then point out, roughly on the map, where each individual sample will be taken.

For areas where no background data has been found, it is best to use the equation set forth by Sokal & Rohlf (1969). It defines how to calculate the number of samples needed to detect a given "true" difference between two means as:

$$n \geq 2\left(\frac{\sigma}{\delta}\right)^2 \{t_{\alpha[v]} + t_{2(1-p)[v]}\}^2$$

where:

n = number of samples

σ = true standard deviation (or coefficient of variation)

δ = the smallest difference that is desired to detect.

ν = the degrees of freedom of the sample standard deviation [= (# groups)(n-1)].

α = the significance level

p =desired probability that a difference will be found to be significant

$t_{\alpha[\nu]}$ and $t_{2(1-p)[\nu]}$ = values from the two-tailed t-table with νdegrees of freedom and corresponding probabilities α and $2(1-p)$,, respectively.

From the equation you can see that any refinement of experimental technique that reduces the standard error of the sample results in smaller required sample sizes or the possibility of detecting smaller differences.

We can apply this equation in the above example. Hypothetically, we want to be 90% certain of detecting a 5% difference between the size lengths of five samples of each of four different flatfish families in the central bay area, knowing that a similar area had a standard deviation of 6% at a 1% level of significance. We assume we are taking 5 samples. (These are all arbitrary values - any values can be used). Plugging these values into the above equation yields:

$$\nu = 4(5-1) = 16$$

$$n \geq 2\left(\frac{6}{5}\right)^2 \{t_{.01[16]} + t_{2(1-0.90)[16]}\}^2$$

$$n \geq 2\,(1.2)^2\,\{2.921+1.337\}^2$$

$$n \geq 2\,(1.44)\,\{4.258\}^2$$

$$n \geq 2.88\,\{18.13\}$$

$$n \geq 52.2 \text{ or } 53 \text{ samples}$$

These formulas can be used for any area. The size of each sample to be taken will depend on the specific laboratory method which will be used for each parameter analyzed. Keep in mind that without an estimate of the variability of the items, no answers can be given at all (Sokal & Rohlf, 1969).

6.0 Summary

This paper outlines the major facets of scientific analysis-from basic statistics and hypothesis testing to sampling design. It offers some guidelines on how to take available data (or to use pilot study data of your own) and design a sampling plan to suit the needs of any project in the most time-efficient and cost effective way. A few examples were given to illustrate the two main focuses of the article-those of sample design and calculating ample sample size. A list of highly suggested references to consult for further details and information on the application of the concepts mentioned here is listed in the Bibliography.

7.0 Bibliography

Cochran, W.G. 1977. Sampling Techniques. 3rd ed. John Wiley & Sons, Inc., New York.

Dytham, Calvin,. 1999. *Choosing and Using Statistics: A Biologist's Guide*. Blackwell Science Ltd., Malden, Massachussetts.

Eckblad, James W. 1991. How Many Samples Should be Taken?. *BioScience*, Volume 41, Number 5, pp. 346-348.

Elliott, J.M. Some Methods for the Statistical Analysis of Samples of Benthic Invertebrates. Freshwater Biological Association Scientific Publication No. 25, 2nd Edition, 1977.

Gilbert, Richard O. 1987. Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Reinhold, New York.

Green, Roger H. 1979. Sampling Design and Statistical Methods for Environmental Biologists. John Wiley & Sons, New York.

Ott, Lyman. 1988. An Introduction to Statistical Methods and Data Analysis. PWS-Kent Publishing Company, Massachusetts.

Portier, Kenneth M. and Arvanitis, C. 1998. "Natural Resource Sampling Notes". University of Florida Department of Statistics Institute of Food and Agricultural Sciences. http://ifasstat.ufl.edu/nrs/nrshome.htm

Schefler, William C. 1979. Statistics for the Biological Science. Addison-Wesley Publishing Company, Massachusetts.

Sokal, R.R. and Rohlf, F.J. 1995. Biometry: The Principles and Practice of Statistics in Biological Research, 3rd Edition. W.H. Freeman & Company, New York.