

Manual of Fisheries Survey Methods II: with periodic updates

Chapter 6: Sample Size for Biological Studies

Roger N. Lockwood and Daniel B. Hayes

Suggested citation:

Lockwood, Roger N. and D. B. Hayes. 2000. Sample Size for Biological Studies. Chapter 6 *in* Schneider, James C. (ed.) 2000. Manual of fisheries survey methods II: with periodic updates. Michigan Department of Natural Resources, Fisheries Special Report 25, Ann Arbor.

Chapter 6: Sample Size for Biological Studies

Roger N. Lockwood and Daniel B. Hayes

The goals of this chapter are to provide fisheries personnel with a better understanding of sampling issues related to data collection and to present methods for determining appropriate sample sizes. Much of this chapter is devoted to a discussion of variability in data since, essential to any data collection, is an appreciation of variability. This variability greatly affects our ability to show differences between groups (e.g., years, lakes, etc.). For example, if we consistently catch 10 fish per net lift every year it would be relatively easy to see an increase to 15 fish per net lift. However, if our catch varied from 1 to 50 per lift (still averaging 10 fish), it would be much more difficult to detect an increase to 15 fish per lift.

Data may be collected to characterize population parameters for descriptive purposes, or as part of a research or management study in which some type of “treatment” is administered and some level of change (either an increase or decrease) is anticipated. The variables measured may be continuous (e.g., length of fish at a chosen age) or discrete (e.g., fin ray count).

This chapter presents various methods for estimating sample size corresponding to desired precision or power, given some prior measure or estimate of variability. These methods presume that the data at hand are approximately normally distributed for continuous variables, or binomially distributed for discrete variables. As data deviate from these assumptions, transformation of the data or different statistical methods become necessary.

Also included in this chapter is a discussion of bias associated with estimation of average fish length from average weight using weight-length regression. Appropriate methods to compensate for this bias are presented.

6.1 Precision

Precision is a measure of variability associated with a sample from a population compared to a descriptive statistic, such as the mean, of that sample. Variability is usually displayed in terms of standard deviations or standard errors. Standard deviation (s) is the variability of individual measures within a sample compared to the mean \bar{x} of that sample and is calculated as:

$$s = \sqrt{\frac{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}}{n-1}}, \quad (1)$$

where, n is the sample size and x_i are the individual measures. From a normal distribution, plus or minus one standard deviation encompasses about 68% of all values and two standard deviations about 95%.

The standard error ($s_{\bar{x}}$) is calculated as:

$$s_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}}{n(n-1)}} = \frac{s}{\sqrt{n}}, \quad (2)$$

and measures the variability of the mean through repeated samples. That is, if we randomly sampled a population again and again, we would expect about 68% of the means to fall within one standard error of the grand mean (average of all the means) and about 95% within two standard errors. This is one way to express confidence intervals for a mean. Another way is as a percent error or relative error. For example, we may wish to calculate a mean with two standard errors being 25% of that mean. Expressed this way, the relative error L (error relative to the mean) is the product of the sample mean and precision p :

$$L = \bar{x}p, \quad (3)$$

where, the precision p is expressed as a fraction, say 0.25 or 25%.

To estimate an appropriate sample size (n), we need level of confidence (e.g., 95%, 99%), relative error, and a measure of variability of the population in question. Typically, a sample is collected and the mean (\bar{x}) and standard deviation (s) are calculated. Estimates of \bar{x} and s from populations thought to be similar may also be used. However, if that assumption is incorrect, many more or far fewer samples than necessary may be collected. Sample size then is (Snedecor and Cochran 1989:52):

$$n = t^2 \frac{s^2}{L^2}. \quad (4)$$

Values for t are:

	Confidence level				
	70%	80%	90%	95%	99%
t	1.036	1.282	1.645	1.960	2.576

Example 6.1—A sample of 10 age-3 fish has the following lengths (inches): 2.3, 4.1, 3.9, 3.7, 3.0, 2.5, 3.0, 2.7, 2.9, 3.0. The mean length from the sample is 3.1 and the standard deviation is 2.4. We wish to sample enough fish to achieve precision $\pm 25\%$ of the mean with 95% certainty.

The first variable estimated is L , our relative error:

$$L = 3.1 * 0.25 = 0.775$$

Sample size then is:

$$n = 1.96^2 \frac{2.4^2}{0.775^2} = 36.8$$

Thus, a minimum of 37 age-3 fish should be measured.

An investigator may also wish to determine what fraction of a sample falls into some classification (e.g., percent males). This follows a binomial distribution where P is the fraction of a sample in one category and Q the fraction in the alternative category. Thus,

$$P = \frac{n_1}{n}, \quad (5)$$

and

$$Q = 1 - P, \quad (6)$$

where, n_1 is the count of individuals in the first category. As with continuous variables, measures of variability can be determined for P and Q . The standard deviation (s):

$$s = \sqrt{PQ}, \quad (7)$$

and standard error ($s_{\bar{x}}$):

$$s_{\bar{x}} = \sqrt{\frac{PQ}{n}}, \quad (8)$$

may both be calculated (Ferguson 1976:156). Similarly, confidence intervals can also be computed for P and Q . Sample size (n) and P control confidence interval coverage. That is, the standard error gets smaller as sample size increases or P deviates from 0.5. The reduction in standard error due to P is greatest when $P > 0.70$ or $P < 0.30$, with minimal change occurring for P between 0.30 and 0.70. Examples of confidence interval coverage for several different values of P and n are given in the following table:

n	P	Q	1 Standard error	2 Standard errors
20	0.20	0.80	±8.9%	±17.8%
20	0.50	0.50	±11.2%	±22.4%
50	0.20	0.80	±5.7%	±11.4%
50	0.50	0.50	±7.1%	±14.2%
100	0.20	0.80	±4.0%	±8.0%
100	0.50	0.50	±5.0%	±10.0%
500	0.20	0.80	±1.8%	±3.6%
500	0.50	0.50	±2.2%	±4.4%

Estimating sample size to determine if some fraction of a population is at a chosen level with given certainty follows (Cochran 1977:72):

$$n = \frac{(4)(P * 100)(Q * 100)}{(\alpha * 100)^2}, \quad (9)$$

where, α is 1 – the chosen level of certainty.

Example 6.2—Trout are fin clipped and placed in a raceway prior to stocking. The hatchery manager wants to be 95% certain that at least 90% of all trout are properly fin clipped. How many fish should be randomly sampled?

$$\begin{aligned}n &= \frac{(4)(0.90 * 100)(0.10 * 100)}{[(1 - 0.95) * 100]^2} \\ &= \frac{(4)(90)(10)}{[5]^2} \\ &= 144\end{aligned}$$

Thus, a minimum of 144 fish must be randomly selected to determine if at least 90% of the fish in the raceway are properly clipped.

6.2 Power

Another measure of data precision and reliability is power. Precision and power measure the probability of correctly concluding a hypothesis given a data set. Commonly, we express the null hypothesis as the sample mean of population A is not different than the sample mean of population B. Conversely, we express the alternate hypothesis as the sample means of populations A and B do in fact differ. Two types of statistical error apply to these outcomes, Type I and Type II. A Type I error is the probability of falsely concluding a significant difference exists when in fact there is no difference. Type I error is denoted as α (alpha). For example, we compare two samples using a t-test. Our results indicate that we are 95% certain that these samples, and the populations from which they were taken, are different; or there is only a 5% chance that they are not different. The probability that we have made a Type I error is 5% ($\alpha=0.05$). If a Type I error is made, a treatment, method, etc. will be accepted and “wrongly” used. We may, for example, accept that some type of stream improvement increases survival of trout. While continuing to use this treatment may not have a negative effect, time and money are being spent on a treatment that has no effect. The cost of making a Type I error in this example is measured in time and money.

But if we show no difference between samples, what is the probability we are wrong? Type II error is measured by β (beta) and is the probability of concluding there is no difference when a difference does indeed exist. Power is $1-\beta$. Power analysis then, provides an investigator with an estimate of sample size necessary to detect some anticipated outcome. If a Type II error is made, a treatment, method, etc. will not be accepted and may not be used. Using our stream improvement example, if no increase in survival of trout is shown the method will be rejected. The result of making a Type II error in this situation is an opportunity cost. That is, the opportunity to use a method that has benefits. While statistical confidence of differences (α) is routinely reported, power is rarely reported.

For decisions to be useful, an appropriate sample size must be taken. When sample sizes are too small, significant differences are rarely found and a treatment or method may be discarded when it may in fact be useful (Type II error occurs). Conversely, sample sizes which are larger than necessary, are not cost effective.

Numerous statistical methods for power analysis exist and presenting a complete review is beyond the scope of this chapter. (Commercial software packages are available for power analysis and

suggestions are given at the end of this section.) However, methods for comparing samples collected from different periods or locations using a t-test are covered in this chapter.

To perform an *a priori* power analysis on data that will be evaluated using a t-test, some preliminary estimates of population characteristics (\bar{x} and s) are necessary, as well as an estimate of anticipated change to the population. We may, for example, anticipate an increase in mean length at age of some fish species following some management technique (e.g., introduction of a predator to thin a bluegill population that is believed to be stunted due to overcrowding).

Several assumptions must be met to accurately measure power using the t-test:

1. Data are approximately normal;
2. Variances of both populations (before and after treatment periods, etc.) are equal;
3. The two populations are independent (that is one group, time period, etc. is not influencing the other);
4. Samples have been randomly selected and accurately represent their respective populations.

Numerous texts discuss power analysis methods, such as Dixon and Massey (1957) and Cohen (1988). Equations presented below were derived from Cohen (1988) and presented in Hintz (1996:59). The methods presented are for directional change (one-tailed test), because we typically expect to see a benefit due to a management action. This benefit could be an increase or decrease in the variable depending on the situation. For example, benefits might be increased abundance of trout, or decrease abundance of sea lamprey. One-tailed tests have greater power than two-tailed tests with equivalent sample sizes because we are focusing on only positive changes.

Power analysis for a one-tailed t-test is done in the following manner. First, calculate standard deviation (s) from a preliminary sample s_1 following equation (1). Next, calculate pooled standard error of \bar{x} using an estimate of before (n_1) and after (n_2) sample size:

$$s_{\bar{x}} = s_1 \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}. \quad (10)$$

Note that s_1 is assumed to be equivalent to standard deviation s_2 during the second period (or location). The measure of variability ($s_{\bar{x}}$) and the anticipated change (d) are then standardized following a normal distribution:

$$Z_p = \frac{Z_{\alpha} s_{\bar{x}} - d}{s_{\bar{x}}}, \quad (11)$$

where Z_{α} is the Z value from Table 1 at the chosen confidence level α (Area column). Power then is estimated as:

$$power = 1 - \beta, \quad (12)$$

where β is the area value opposite the Z value (substituting Z_p for Z) from Table 1. Several iterations using different values for n_1 and n_2 may be necessary to establish acceptable sample sizes.

There are no set values for α or β . Traditionally, $\alpha=0.05$ has been used for natural resource data because we generally feel that falsely concluding one method is better than another would be a substantial problem. β equal to 0.20 or less may be a reasonable level of certainty for protection against Type II errors where rejection of a treatment or method does not have severe consequences.

Example 6.3—Walleye are to be introduced in a lake having a stunted bluegill population. A trap net sample of 32 bluegill yielded a mean length of 6.4 inches with a standard deviation of 1.01. The goal of the project is to increase the average length of bluegill, sampled by trap net, to 7.0 inches. The manager wishes to be 95% certain that a Type I error has not occurred ($\alpha=0.05$). To be certain that an effective treatment opportunity is not missed, 95% power that a Type II error ($\beta=0.05$) has not occurred is chosen. How many samples should be collected?

Starting with 32 fish before and 32 fish after:

$$s_{\bar{x}} = 1.01 \sqrt{\frac{1}{32} + \frac{1}{32}} = 0.2525,$$

$$Z_p = \frac{(1.645)(0.2525) - (0.6)}{(0.2525)} = -0.7312.$$

Using Table 6.1, read under Z column to find a value as close to -0.7312 as possible, use -0.7388 , then across to the right in the “Area” column find value 0.23 (β):

$$power = 1 - 0.23 = 0.77$$

Since we selected a power of 95%, more samples are required. Next try 50 before and 50 after samples:

$$S_{\bar{x}} = 1.01 \sqrt{\frac{1}{50} + \frac{1}{50}} = 0.2020,$$

$$Z_p = \frac{(1.645)(0.2020) - (0.6)}{(0.2020)} = -1.3253,$$

$$\beta = 0.09 \text{ and } power = 0.91$$

Next try 65 before and 65 after samples:

$$S_{\bar{x}} = 1.01 \sqrt{\frac{1}{65} + \frac{1}{65}} = 0.1772,$$

$$Z_p = \frac{(1.645)(0.1772) - (1.3)}{(0.1772)} = -1.7410,$$

$$\beta = 0.04 \text{ and } power = 0.96$$

By sampling 32 fish before and 32 fish after, we may not be able to show a difference in mean length when in fact a difference may exist. By roughly doubling our sample to 65 fish before and 65 fish after, power=0.96 and the risk of making a Type II error is minimal.

Example 6.4—Sample sizes (e.g., before and after) may not always be equal. Once a sample has been collected, additional data may not be available for collection during a given time period or at a location. Using the data from Example 3, suppose the collection of 32 fish during our before period cannot be changed because no additional fish can be measured. How many fish must be collected following the introduction of walleye to detect an increase in the average length of bluegill to 7.0 inches? Again, the before period mean is 6.4 inches with standard deviation 1.01, $\alpha=0.05$ and $\beta=0.05$.

This time start with 32 fish before and 300 fish after:

$$S_{\bar{x}} = 1.01 \sqrt{\frac{1}{32} + \frac{1}{300}} = 0.1878,$$

$$Z_p = \frac{(1.645)(0.1878) - (0.6)}{(0.1878)} = -1.5499,$$

$$\beta = 0.06 \text{ and } power = 0.94.$$

Substantially more fish need to be sampled during the after period to detect our anticipated 0.6 inch increase in average length. Approximately 600 bluegill need to be measured during the after period:

$$S_{\bar{x}} = 1.01 \sqrt{\frac{1}{32} + \frac{1}{600}} = 0.1832,$$

$$Z_p = \frac{(1.645)(0.1832) - (0.6)}{(0.1832)} = -1.6301,$$

$$\beta = 0.05 \text{ and } power = 0.95.$$

Currently, the State does not have power analysis software on contract. However, Pass 6.0 is being used at the Institute for Fisheries Research, and is quite easy to use and does not require a systems administrator to install. Power estimates using Pass 6.0 may be done for a variety of parametric and nonparametric statistical tests. These include: t-test on one mean, t-test on two means, one-way ANOVA, fixed effects ANOVA, randomized block ANOVA, repeated measures ANOVA, bioequivalence-means, log-rank survival test, one proportion, two proportions, Fishers exact test, bioequivalence-proportions, matched case/control, one correlation, two correlations, multiple regression, and logistic regression. Trial copies of Pass are currently available on their web site (<http://www.ncss.com/>) and are good for 30 days.

6.3 Length-weight bias

Length-weight regressions are often used to estimate the weight of fish from their length. Similarly, length of fish may be estimated from their weight. These regression equations generally follow the form:

$$\log L = \log a + b(\log W), \quad (13)$$

for estimation of length L from weight W and:

$$\log W = \log a + b(\log L), \quad (14)$$

for estimation of weight from length. For either equation, a is the intercept and b the slope of the line. Equations (13) and (14) may also be written as:

$$L = aW^b, \quad (15)$$

and

$$W = aL^b. \quad (16)$$

When large numbers of fish are processed, individual lengths are seldom taken and mean length \bar{L} is estimated from bulk weights using mean weight \bar{W} . Substitution of \bar{L} for L and \bar{W} for W in equation (15) produces a biased estimate of \bar{L} (Nielsen and Schoch 1980).

For estimation of \bar{L} using equation (15), b is usually about 1/3 and length is overestimated by about 2.5%, but may be greater depending on the range of fish lengths in the sample. Predicted \bar{W} from equation (16) is underestimated by about 10% when b is around 3, but may also be greater depending on the sample. This under- or overestimation is due to the non-linear relationship between length and weight. As b deviates from 1 (when the relationship is linear), under- or overestimation occurs. While this current discussion relates to length and weight only, bias will occur for any power relationships – such as mean length and mean fecundity of female fish. (Note: Do not confuse power relationship with statistical power in previous section – they are entirely different.) The high bias associated with equation (15) is troublesome when the calculated mean length of hatchery-reared fish at the raceway differ from observed mean lengths at the stocking site.

To compensate for bias associated with prediction of mean length from mean weight in a hatchery setting, the following procedure is suggested.

1. Prior to removing fish from a raceway for planting, obtain a sample of 150-200 fish, and measure their individual lengths and weights. From this sample, compute mean length, mean weight, and a weight-length regression specific to that raceway.
2. Using that raceway-specific regression, calculate the predicted mean length based on the mean weight determined in #1. The predicted mean length will be slightly larger than the actual mean length calculated in #1. This difference (expressed as a percentage) will be used to adjust predicted mean lengths planted.
3. During the planting operation, for each group of fish cut off from the rest of the raceway, obtain 5 subsamples (each about 100 fish) for weighing and counting. Subsample each time a portion of the raceway is segregated. (Based on samples from the Harrietta Hatchery, this should give you an average number of fish per kilogram ± 1.5 fish [95% confidence level]).
4. To compute the mean length of fish being planted, insert the mean weight, calculated in #3, into the weight-length regression. The predicted length is then reduced by the percentage determined in #2.

Similar procedures could be used to estimate mean weights from mean lengths of fish.

Example 6.5—A sample of 20 hatchery fish yield the following length (inches) and weight (grams) values (note this is less than the recommended sample of 150-200 and is done so for demonstration purposes):

<u>Length</u>	<u>Weight</u>
3.0	15.0
3.1	29.7
4.6	99.3
4.7	113.8
5.7	200.0
5.9	205.3
6.0	206.0
6.1	236.9
6.9	348.5
3.6	46.6
3.8	54.8
3.9	59.3
4.5	91.1
4.7	103.8
4.8	110.5
4.9	117.6
5.0	125.0
6.0	216.0
6.8	314.4
6.9	328.5

Step 1 – Enter length weight data into SPSS (9.0.0) and calculate average length and weight using “Analyze Descriptive statistics Descriptives”. Average length is 5.0450 and average weight 151.1400. Output will appear similar to:

Descriptive Statistics		
	N	Mean
LENGTH	20	5.0450
WEIGHT	20	151.1400
Valid N (listwise)	20	

Next calculate length weight regression using “Analyze Regression Curve estimation”. Length is the dependent variable and weight the independent. For models check “Power”. Output similar to the following should appear:

```

MODEL:  MOD_9.
-

Dependent variable.. LENGTH          Method.. POWER
Listwise Deletion of Missing Data

Multiple R          .98928
R Square           .97867
Adjusted R Square  .97748
Standard Error     .03782

Analysis of Variance:

              DF   Sum of Squares      Mean Square
Regression    1     1.1809908        1.1809908
Residuals    18     .0257447          .0014303

F =      825.71570      Signif F = .0000

----- Variables in the Equation -----
Variable          B          SE B          Beta          T          Sig T
WEIGHT            .300109    .010444    .989275    28.735    .0000
(Constant)       1.177706    .059277          19.868    .0000

```

Example 6.5—continued.

Reading under “Variables in the Equation” find 0.300109 and 1.177706 (both are in bold type). Substituting in our length-weight regression, $a=1.177706$, $b=0.300109$ and mean weight from above descriptive statistics:

$$L = 1.177706 \cdot (151.1400)^{0.300109} = 5.3099$$

We now have:

Observed mean length = 5.0450
Observed mean weight = 151.1400
Predicted mean length = 5.3099

Step 2 – Calculate correction factor:

$$correction = \frac{5.3099}{5.0450} = 1.0525$$

Step 3 – Each time fish are to be removed from a raceway and planted, collect 5 samples of 100 fish. Calculate the average weight from these 5 samples. This step is repeated each time that a raceway is segregated for removal of fish.

Step 4 – Estimate mean length using regression equation from Step 1 and correction factor from Step 2:

$$corrected\ mean\ length = \frac{1.177706 \cdot W^{0.300109}}{1.0525}$$

Table 6.1.—Areas of the normal distribution. Table values were taken from Remington and Schork (1970).

<i>z</i>	Area	<i>z</i>	Area	<i>z</i>	Area
-2.3290	0.01	-0.4125	0.34	0.4399	0.67
-2.0540	0.02	-0.3853	0.35	0.4677	0.68
-1.8810	0.03	-0.3585	0.36	0.4959	0.69
-1.7510	0.04	-0.3319	0.37	0.5244	0.70
-1.6450	0.05	-0.3055	0.38	0.5534	0.71
-1.5550	0.06	-0.2793	0.39	0.5828	0.72
-1.4760	0.07	-0.2533	0.40	0.6128	0.73
-1.4050	0.08	-0.2275	0.41	0.6433	0.74
-1.3410	0.09	-0.2019	0.42	0.6745	0.75
-1.2820	0.10	-0.1764	0.43	0.7063	0.76
-1.2270	0.11	-0.1510	0.44	0.7388	0.77
-1.1750	0.12	-0.1257	0.45	0.7722	0.78
-1.1264	0.13	-0.1004	0.46	0.8064	0.79
-1.0800	0.14	-0.0753	0.47	0.8416	0.80
-1.0360	0.15	-0.0502	0.48	0.8779	0.81
-0.9945	0.16	-0.0251	0.49	0.9154	0.82
-0.9542	0.17	0.0000	0.50	0.9542	0.83
-0.9154	0.18	0.0251	0.51	0.9945	0.84
-0.8779	0.19	0.0502	0.52	1.0360	0.85
-0.8416	0.20	0.0753	0.53	1.0800	0.86
-0.8064	0.21	0.1004	0.54	1.1264	0.87
-0.7722	0.22	0.1257	0.55	1.1750	0.88
-0.7388	0.23	0.1510	0.56	1.2270	0.89
-0.7063	0.24	0.1764	0.57	1.2820	0.90
-0.6745	0.25	0.2019	0.58	1.3410	0.91
-0.6433	0.26	0.2275	0.59	1.4050	0.92
-0.6128	0.27	0.2533	0.60	1.4760	0.93
-0.5828	0.28	0.2793	0.61	1.5550	0.94
-0.5534	0.29	0.3055	0.62	1.6450	0.95
-0.5244	0.30	0.3319	0.63	1.7510	0.96
-0.4959	0.31	0.3585	0.64	1.8810	0.97
-0.4677	0.32	0.3853	0.65	2.0540	0.98
-0.4399	0.33	0.4125	0.66	2.3290	0.99

6.4 References

- Cochran, W. G. 1977. Sampling techniques. John Wiley & Sons, New York, New York.
- Cohen, J. 1988. Statistical power analysis for the behavioral sciences. Lawrence Erlbaum Associates, Hillsdale, New Jersey.
- Dixon, W. J., and F. J. Massey, Jr. 1957. Introduction to statistical analysis. McGraw-Hill Book Company Inc., New York, New York.
- Ferguson, G. A. 1976. Statistical analysis in psychology and education. McGraw-Hill Book Company Inc., New York, New York.
- Hintze, J. L. 1996. PASS user's guide. NCSS, Kaysville, Utah.
- Nielsen, L. A., and W. F. Schoch. 1980. Errors in estimating mean weight and other statistics from mean length. Transactions of the American Fisheries Society 109:319-322.
- Remington, R. D., and M. A. Schork. 1970. Statistics with applications to the biological and health sciences. Prentice-Hall Inc., Englewood Cliffs, New Jersey.
- Snedecor, G. W., and W. G. Cochran. 1989. Statistical methods. Iowa State University Press, Ames, Iowa.

Written 3/1976 by G. P. Cooper and J. R. Ryckman
Extensively revised 1/2000 by R. N. Lockwood and D. B. Hayes