Sample size determination

Ecological Stats

- The detailed objectives of a research program drive the amount of work that has to be done
- You must decide the level of inference you need
- Your question + difficulty + level of confidence = the number of samples you have to take

Ecological Stats

- Some questions require more samples than are feasible to take, and it can be discouraging
- Requires a different question, or may not be worth doing at all
- Don't ignore the sample size issues!
- Ultimately, it's all about resource allocation (dollars)

Ecological Stats

- What would you do if you had unlimited funds and could acquire perfect information?
- Given that funds are limited, how can you optimally allocate your efforts to measure the same things?

Sample size determination

- Types of questions:
- How many animals to weigh so that 95% CI are within +/- 5 g?
- How many animals to weigh to test for differences in mean weight between two lakes with an alpha = 0.1?

Sample size for continuous variables

- Estimate average length of fish in a lake
  - Assume normal distribution
  - CI for normal

$$\bar{x} \pm t_{\alpha} s_{\bar{x}}$$

Sample size for continuous variables

- Do you want your estimate of the mean to be accurate within 10% of true mean? 1%?
  - This is an ecological or management question, not a statistical one
  - What are the consequences of being wrong?

Sample size for continuous variables

- Types of deviations from the mean
- Absolute error
  - 4*ug* difference around a mean of 10
- Relative error (r)
  - 40% difference from the average
  - Simply the absolute error divided by the mean

Sample size for continuous variables

- n= sample size needed to estimate mean
- t_α= student's t-value for alpha and n
- s = standard deviation of variable
- d = desired absolute error

$$n = \left(\frac{t_\alpha s}{d}\right)^2$$

Sample size for continuous variables

- n= what we want to know
- t_α= usually about 2
- s = where do we get this?
- d = we choose

$$n = \left(\frac{t_\alpha s}{d}\right)^2$$

Sample size for continuous variables

- Ways to get *s*
  - *Previous sampling of a similar population*
  - *Pilot study*
  - *Guess*
  - *Two-stage sampling*

$$n = \left(\frac{t_\alpha s}{d}\right)^2$$

Example

- Measure whitefish to obtain 95% CI of ± 2.8 mm
  - From sampling in another lake you know that whitefish SE is about 9.4 mm
    - n= ????
    - t_α= ~2
    - S= 9.4
    - d= 2.8

$$n \approx \left(\frac{(2)(9.4)}{2.8}\right)^2 = 45.1$$

Example

- To the spreadsheet

Other tricks from Krebs

- Example on how to estimate sample size using relative error (r) of precision like CV

$$n \approx \left(\frac{(100)(CV)(t_\alpha)}{r} \right)^2$$

Other tricks from Krebs

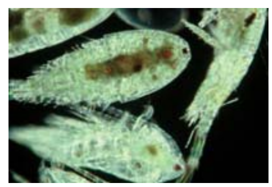
- Example on how to estimate sample size using relative measures (r) of precision like CV

$$n \approx \left(\frac{(100)(CV)(t_\alpha)}{r} \right)^2$$

Other tricks from Krebs

- Example CV for plankton sampling is about 0.70. You want 25% relative precision of your mean estimate. How many samples do you have to take?

$$n \approx \left(\frac{(100)(0.70)(2)}{25} \right)^2 = 31.4 = 32$$



Comparing two means

- Collect size data from two locations and you wish to compare the means between the two locations
- How big a sample do you need?
- This is driven by
  - Size of difference you wish to detect
  - Alpha and Beta
  - Type I, Type II error, and statistical power

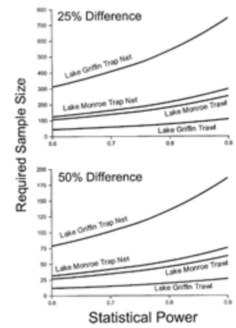
Comparing two means

- Remember alpha, beta and power?
- How do you decrease alpha and beta, increase power?

Comparing two means

- We will examine how to estimate the number of samples required to detect a difference between two means in Excel and in R
- R can also easily calculate power

Examples….



Discrete variables – proportions and percentages

- Proportional data like age classes, sex ratio, fraction of juveniles in the population are described by binomial and multinomial distributions.
- Simplest case has two classes and the distribution has two parameters
  - *p* = Proportion of *x* types in the population
  - q = 1-p = Proportion of *y* types in the population

Discrete variables – proportions and percentages

- Similar process as with the normal distribution
  - Specify margin of error (d) that is acceptable to the estimate of p
  - Specify alpha of not achieving this margin of error
  - If sample size is above 20, normal approximation can be used

Discrete variables – proportions and percentages

- n= what we want to know
- tα= usually about 2
- est. p = estimate of proportion of group 1
- est d.= 1-(est. p)
- d = desired margin of error

$$n = \left(\frac{t_\alpha^{\ 2} \hat{p} \hat{d}}{d^2} \right)$$

Discrete variables – proportions and percentages

- What if you don't have an estimate of the proportion?
  - It can't be any worse than 0.5!
  - Use literature values
  - Pilot sampling

$$n = \left(\frac{t_\alpha^{\ 2} \hat{p} \hat{d}}{d^2} \right)$$

Back to the spreadsheet

Summary

- Estimating required sample sizes forces you to decide on an acceptable level of precision
- Can be bad news, but best to get bad news early, not late
- Can save lots of time and dollars

Summary

- Sample size estimation essentially simulates the sampling process
- Build a model of your system, then collect "data" and see how your sampling design or sampling effort performs
- Does it give you what you need?
- A very useful thing to do that is very common