Module 2: Environmental Sampling

2.6 Choosing Sample Sizes

Choosing Sample Sizes

- Choosing an appropriate sample size is a function of:
  - Study goals
  - Degree of precision required
  - Design type
  - Budget
  - and other factors

Choosing Sample Sizes

- Often a high degree of precision is specified and calculations yield large sample sizes
- Iterative calculations are done until sample sizes come into a “reasonable” range
- This is okay since it ultimately leads to either
  - a balance between competing study requirements of cost and precision
  - a realignment of study requirements or resources

Module 2.6

Choosing Sample Sizes

- Need to know, estimate, or specify the amount of variability expected in the data (σ) and the precision desired (δ)
- δ is half of the width of a 95% confidence interval on the mean

Module 2.6

Choosing Sample Sizes

- For simple random sampling with data from an approximately normal distribution (at least not highly skewed):
- A 95% confidence interval on the mean with width 2δ will result from

$$n = \frac{4\sigma^2}{\delta^2}$$

Module 2.6

Choosing Sample Sizes

- A 95% confidence interval on the difference between two means will result from

$$n = \frac{8\sigma^2}{\delta^2}$$

Module 2.6

Choosing Sample Sizes

- For stratified random sampling, the n observations must be allocated to strata.
- Proportional allocation gives $n_i/n = N_i/N$
- So $n_i = (N_i/N)*n = w_i*n$

Module 2.6

Choosing Sample Sizes

- If σ_i are not expected to be equivalent across strata, and/or sampling costs differ, then set

$$n_i = \frac{N_i \sigma_i}{\sqrt{c_i}}$$

- where c_i is the cost of sampling for strata i

Module 2.6

Example Sample Size Calculation

- You are asked to help with designing a study
- The budget is limited to $15,000
- There is some preliminary data that was taken to see if a problem exists (and it looks like it does). From that data, you estimate s = 20.2.

Example Sample Size Calculation

- After discussion with the team, you would like the confidence interval on the true mean to be no wider than 5 units.

$$n = \frac{4\sigma^2}{\delta^2}$$

- So, n=(4*20.2*20.2)/(5/2)*(5/2)
- Note: I had to square s to get s^2 and half the width of the CI to get delta, then square it

Example Sample Size Calculation

So, n = (4*20.2*20.2)/(5/2)*(5/2)

n = 1632.16/6.25

N = 261.1 = 261

The problem is that collecting the samples will cost about $10 per sample and having them analyzed will be $100 per sample for a total cost of $110 per sample or $28,710 or about $13,710 more than the budget

Module 2.6

11

Example Sample Size Calculation

What can you do?

Refine the estimate of s but that might not help, it might even go up and it will cost money to collect that data

Use composite sampling?

Allow the CI to be wider?

Find more money for the study?

Give up?

Find another lab that's cheaper?

Other ideas?

Module 2.6

12