

A photograph of a pond with large green lily pads and a purple water lily flower in bloom. The lily pads are large and have a prominent vein pattern. The water lily flower is purple with a yellow center. The background shows more lily pads and a reflection of the sky.

Optimizing Sample Size for Environmental Studies

Kenneth E. Osborn
Laboratory Services Division
East Bay Municipal Utility District
Oakland, California
kosborn@ebmud.com

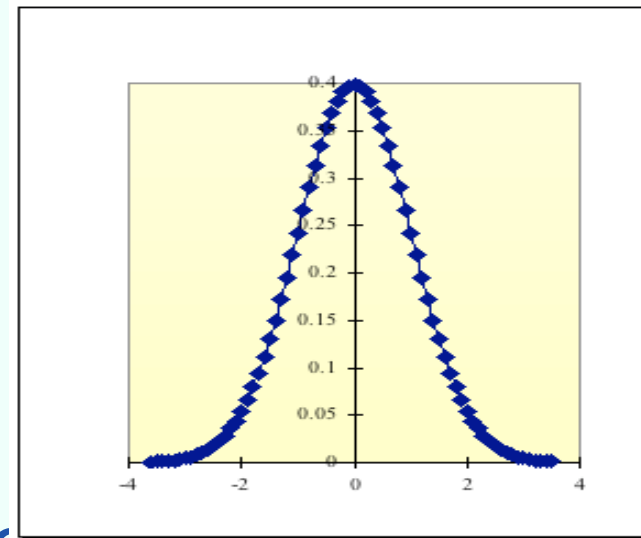
Overview

- Question to statistician “How many samples do we need?”
- Question from statistician “How close do you want to know?”

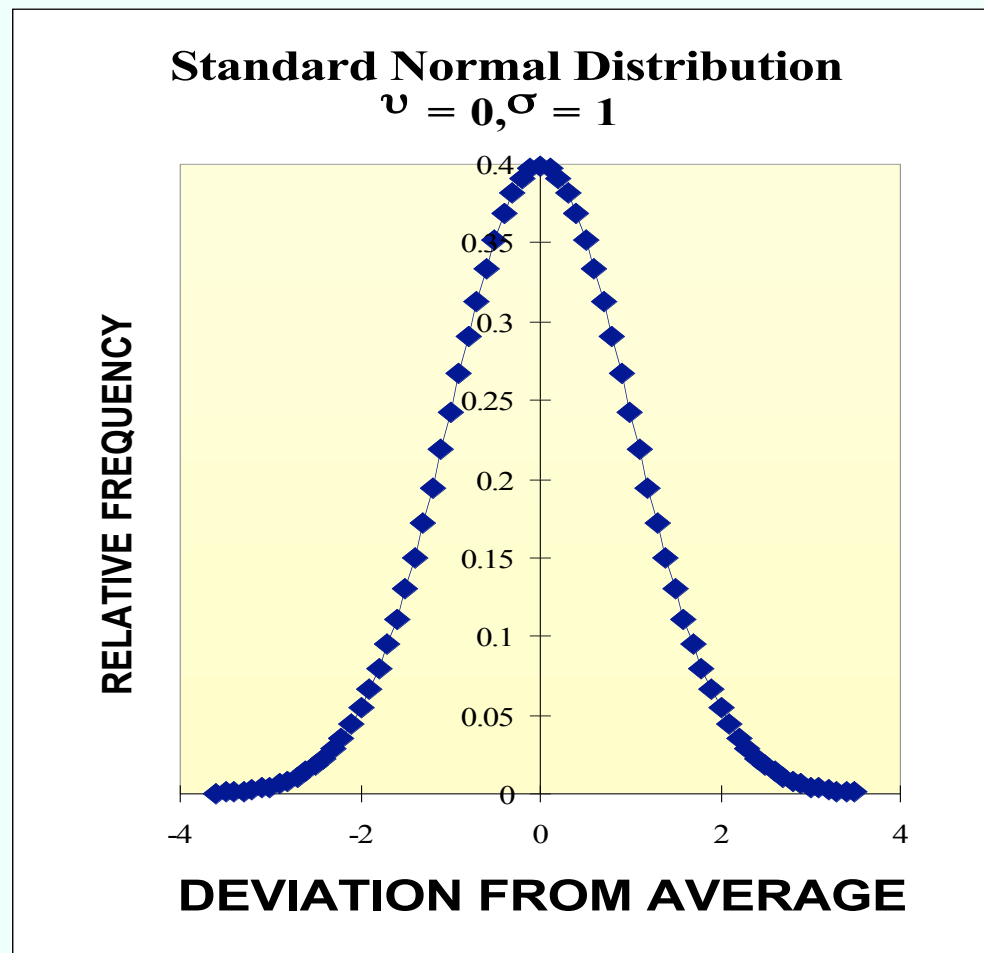
A Bit Of Sadistics

Gaussian Distribution Properties

- Center
- Spread
- Symmetry
- Mean = median
- Large deviations from center are less frequent than small deviations



You've Seen This Before ...



Using the Z-Statistic

$$Z = (\bar{x} - u) / \sqrt{\sigma^2 / N}$$

Where:

\bar{x}

= sample average

u

= population mean

σ

= population standard deviation

N

= sample size

What We Don't Know

- Don't know true population mean
- Don't know N
- Don't know sample mean
- Don't know sample standard deviation

What we do know

The probability that Z is between -2 and + 2 equals 95%

$$P(-2 < Z < 2) = 0.95$$

Zapping Z



$$N = \{z\sigma / (\bar{x} - u)\}^2$$

$$N = 4\{\sigma / (\bar{x} - u)\}^2$$

Deviant Derivations

- Express deviation of sampling mean from population mean as a function of sampling mean

$$D = (\bar{x} - u) / \bar{x}$$

$$u = \bar{x}(1 - D)$$

A bit of algebra...

- Combine sample mean and standard deviation into relative standard deviation to give ...

$$N = 4\{rsd / D\}^2$$

What is D?

- Whatever you want ---

What is rsd?

- A good guess or some preliminary data



An Example

Example: you want the final result to be within 20% of the true value and the expected *rsd* of future samples is 50%

Example continued

- Solving with the equation for N:

$$N = 4\{rsd / D\}^2$$

$$N = 4\{0.5 / 0.2\}^2$$

$$N = 25$$