Design IV

E232 Spring 07

Class 14

Bruce McNair
bmcnair@stevens.edu

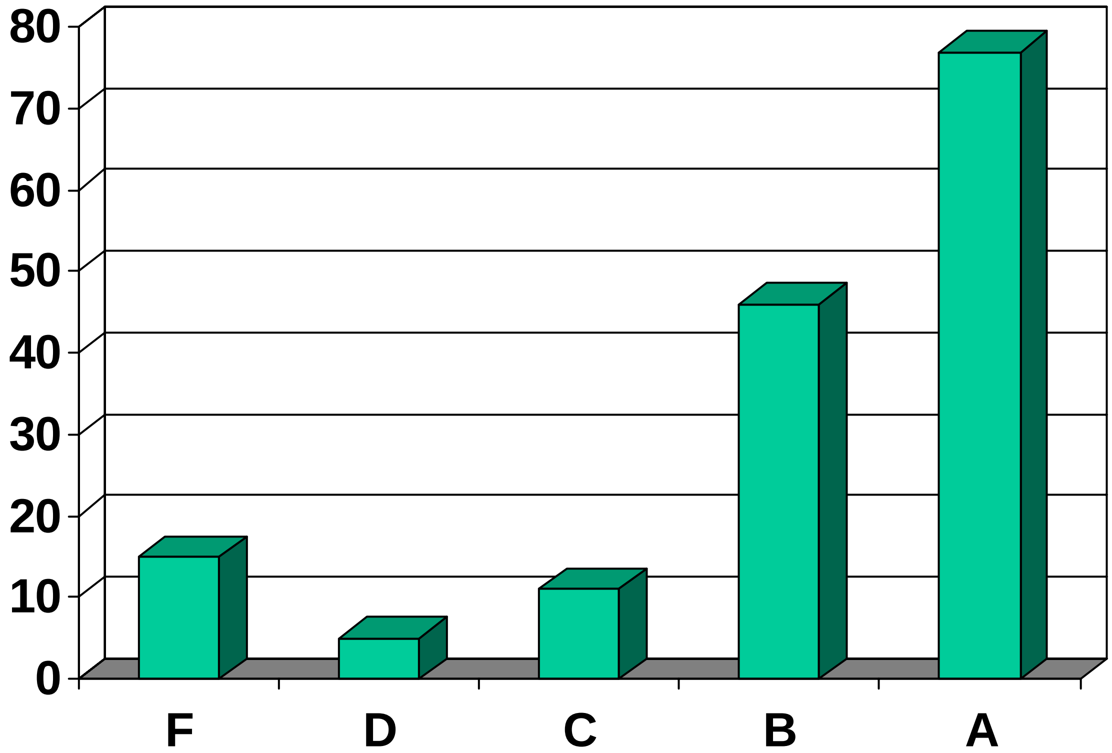
Notes on Midterm Grades

- All 100 and 200 level course instructors are required to assign student midterm grades
- Midterm grades are used to alert students to areas they need to address, they are not on your final transcript
- E232 midterm grades were whole letter grades only, no + or – grades were assigned. The standard letter grade levels were used without adjustment.
- For the E232 midterm grade, there wasn't a lot of feedback from the lab instructors. It was of the form: "do not modulate the lecture component" or "modulate the overall lecture score (upwards | downwards) by 10%"

Notes on Midterm Grades

- In E232, the final grade will be an equal balance between the lab component and lecture component of the course (as noted in Class 1)
- I always grade my courses in the same manner (for final grades):
  - Calculate the overall score
  - Lookup the standard letter grade corresponding to that score
  - Examine the letter grade distribution and boundaries between letter grades
  - Adjust the mapping of numeric grades to letter grades to fit actual population data and special cases, adjusting the mapping downwards only, (e.g. while the standard distribution requires a 95 to get A, I might decide that 93.5 is sufficient, but will never increase the level above 95)
  - The final assignment of numeric scores to letter grades will ***always*** be monotonic.

Notes on Midterm Grades



Regarding less than desired grades:

- Four students did not turn in Q1
- One turned Q1 in with nothing answered.
- Four students have turned in no weekly assignments

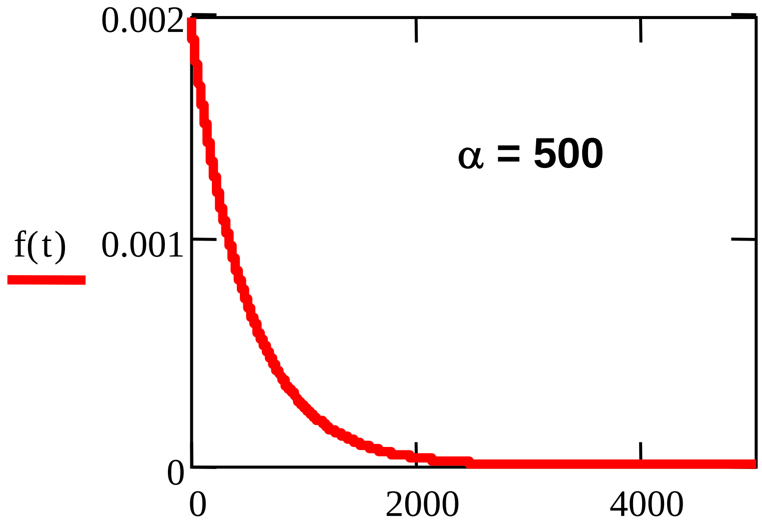
Exponential Distribution

- Probability Density Function

$$P(t \le T) = \int_0^T f(t)dt$$

$$f(t) = \begin{cases} 0 & t < 0 \\ \frac{1}{\alpha} e^{-\frac{t}{\alpha}} & t \ge 0 \end{cases}$$

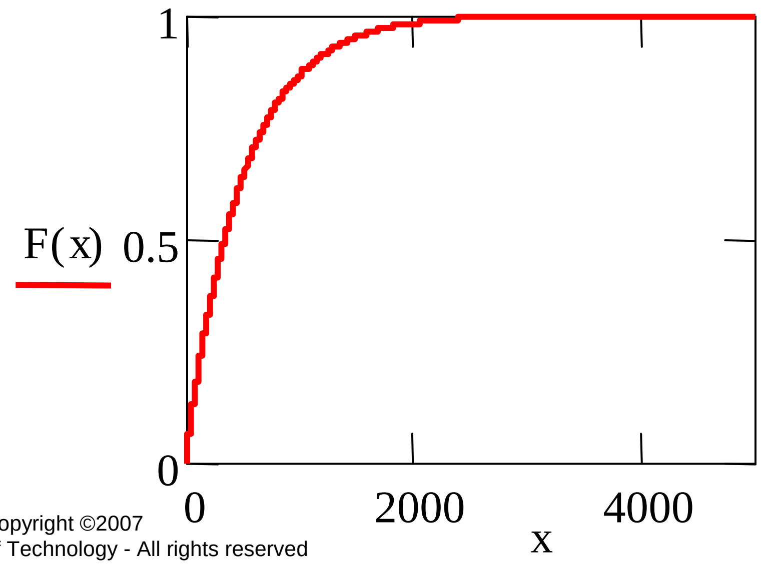
$$P(t \le T) = \int_0^T \frac{1}{\alpha} e^{-\frac{t}{\alpha}} dt$$



- Cumulative Distribution Function

$$P(rv \le x) = \int_{-\infty}^x f(t)dt \doteq F(x)$$

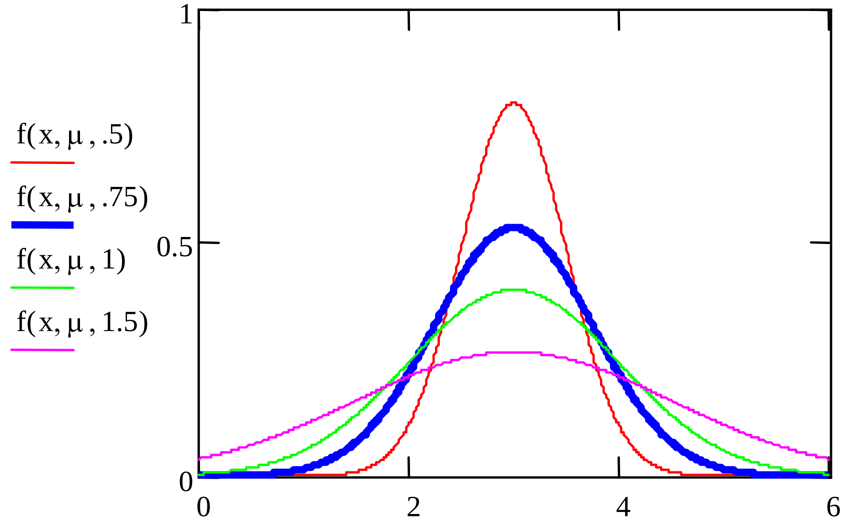
$$F(x) = \int_0^x \frac{1}{\alpha} e^{-\frac{t}{\alpha}} dt = 1 - e^{-\frac{x}{\alpha}}$$



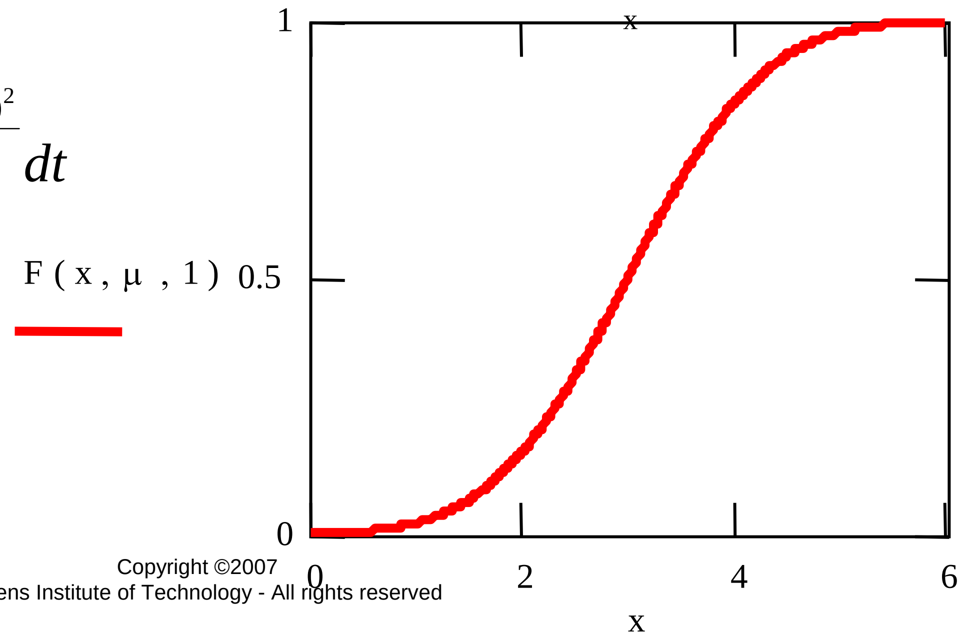
Gaussian Distribution

- Probability density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



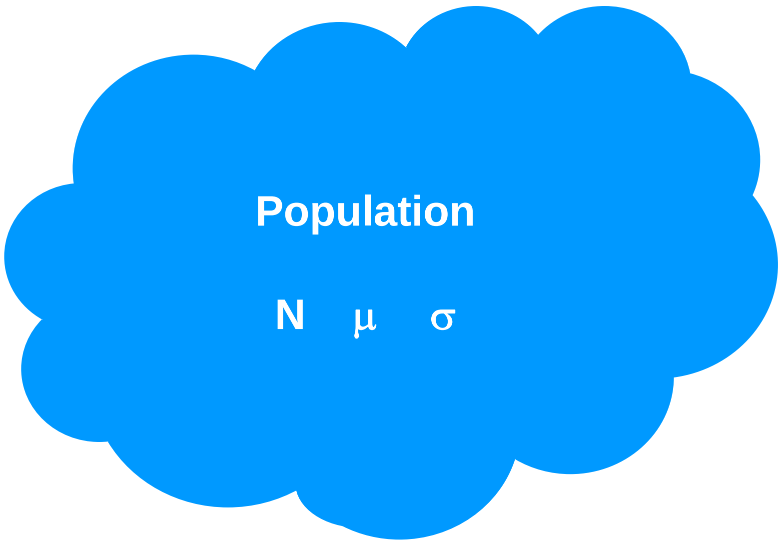
$$F(x) = \int_{-\infty}^{t} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$



Today's topics

- More On Statistical Analysis of Experimental Data
  - Parameter estimation

Parameter Estimation

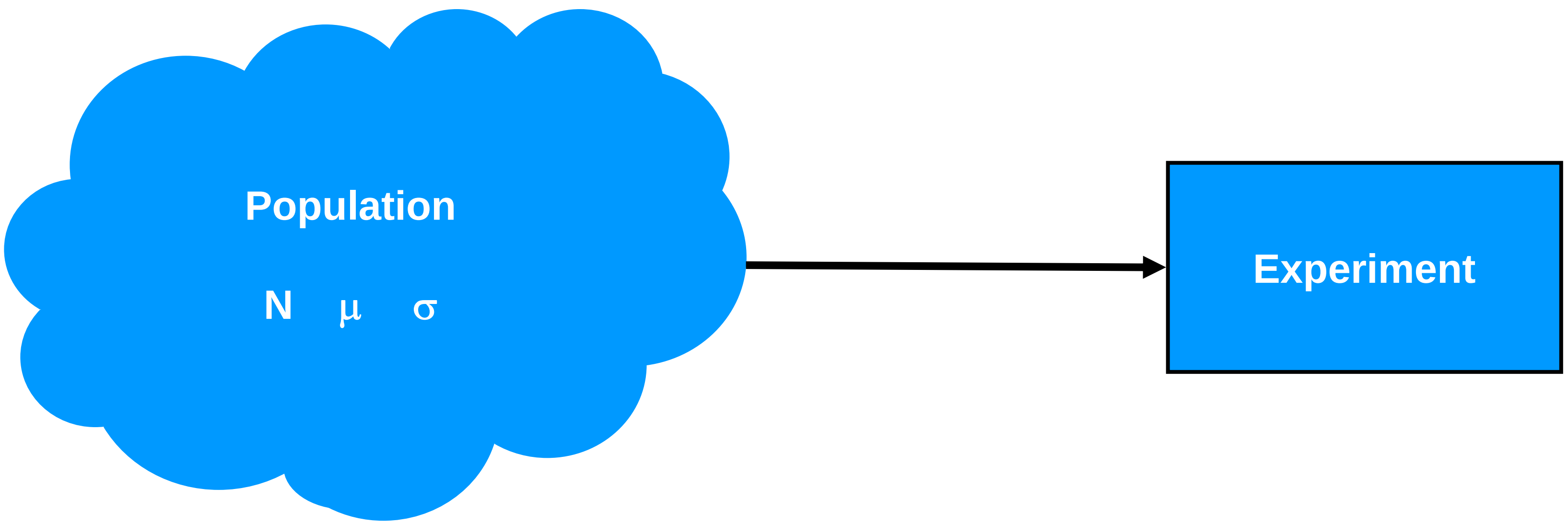


N potential outcomes

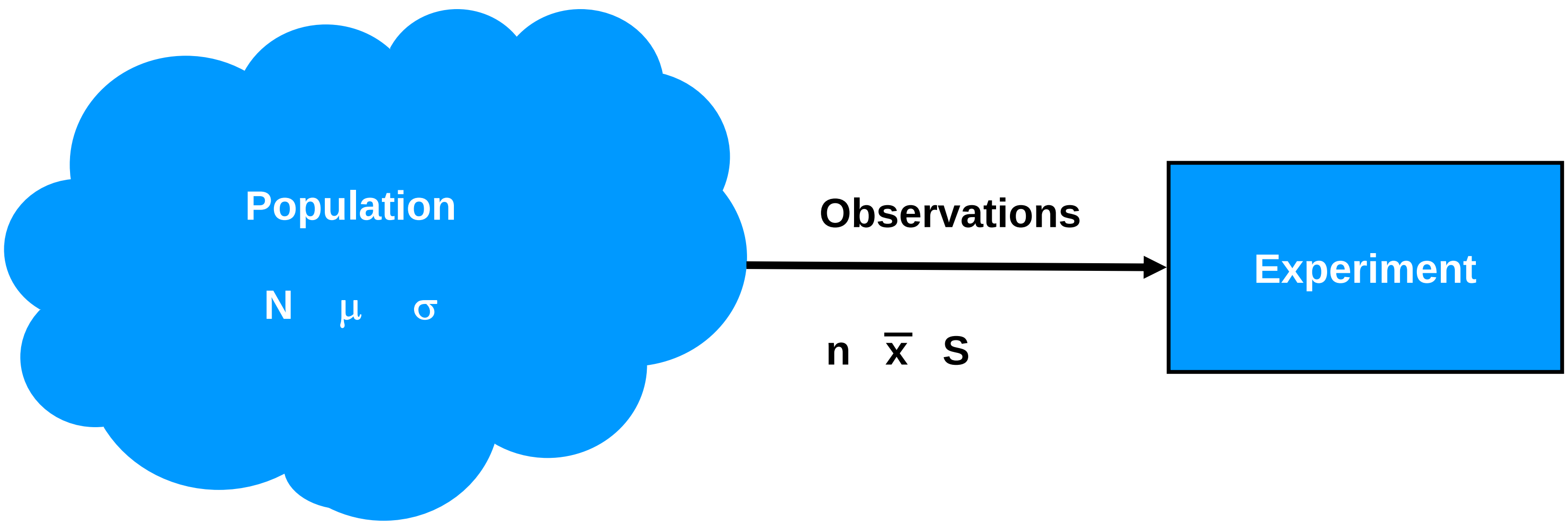
Mean value μ

Standard deviation σ

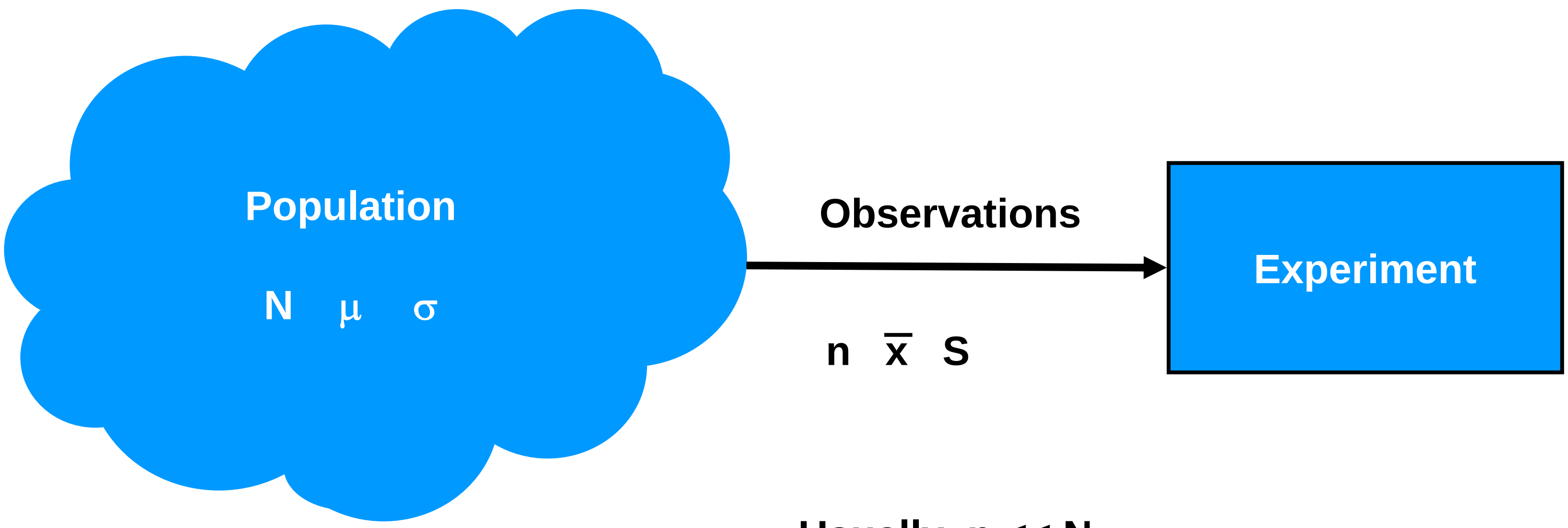
Parameter Estimation



Parameter Estimation

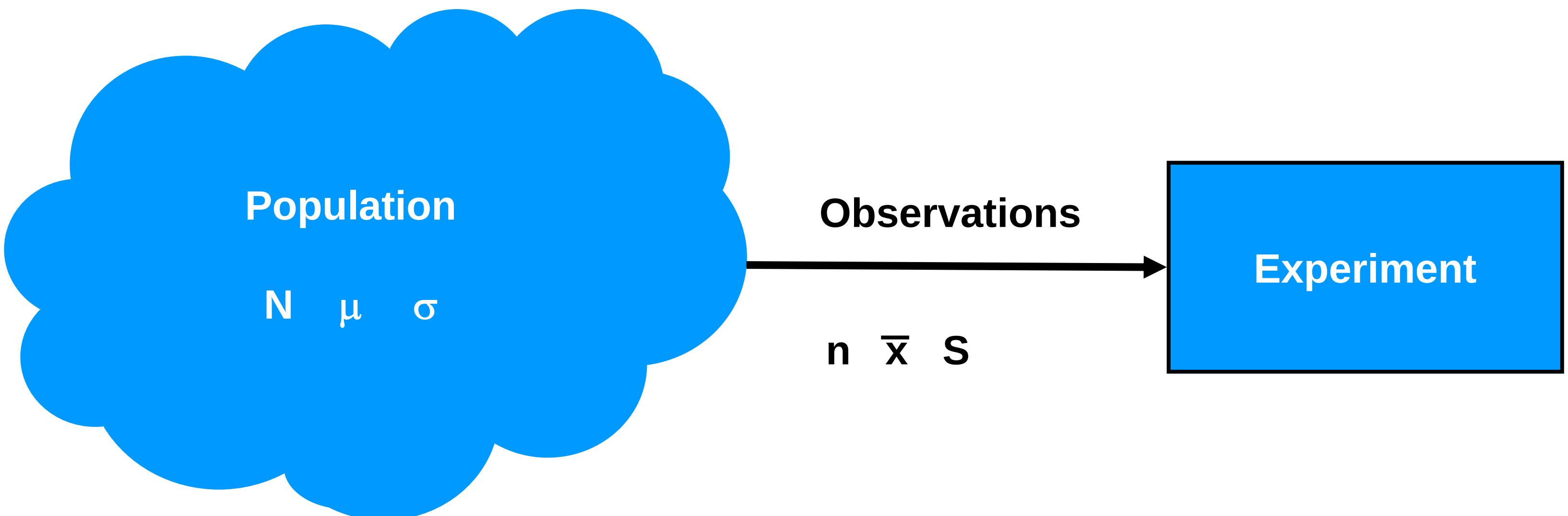


Parameter Estimation



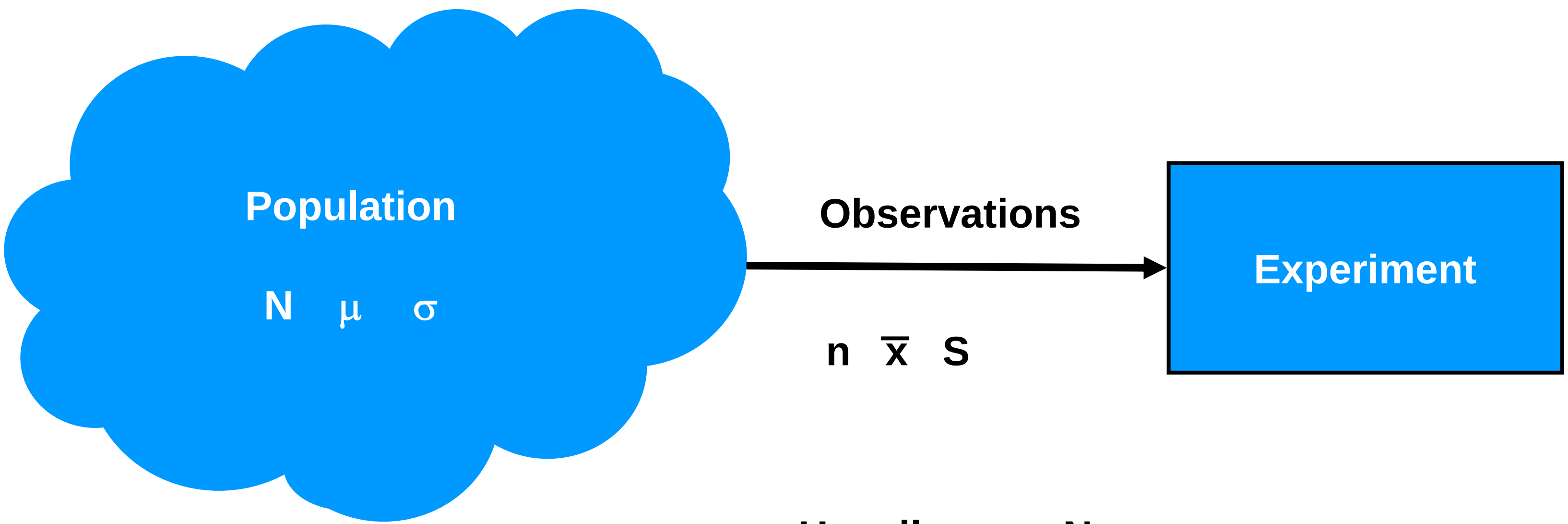
Usually, n << N

Parameter Estimation



Usually, $n \ll N$
so $\bar{x} \sim \mu$
$S \sim \sigma$

Parameter Estimation



Usually, n << N
so $\bar{x}$ ~ μ
S ~ σ

But if δ = |$\bar{x}$-μ|,
how large is δ likely to be?

Confidence Interval and Confidence Level

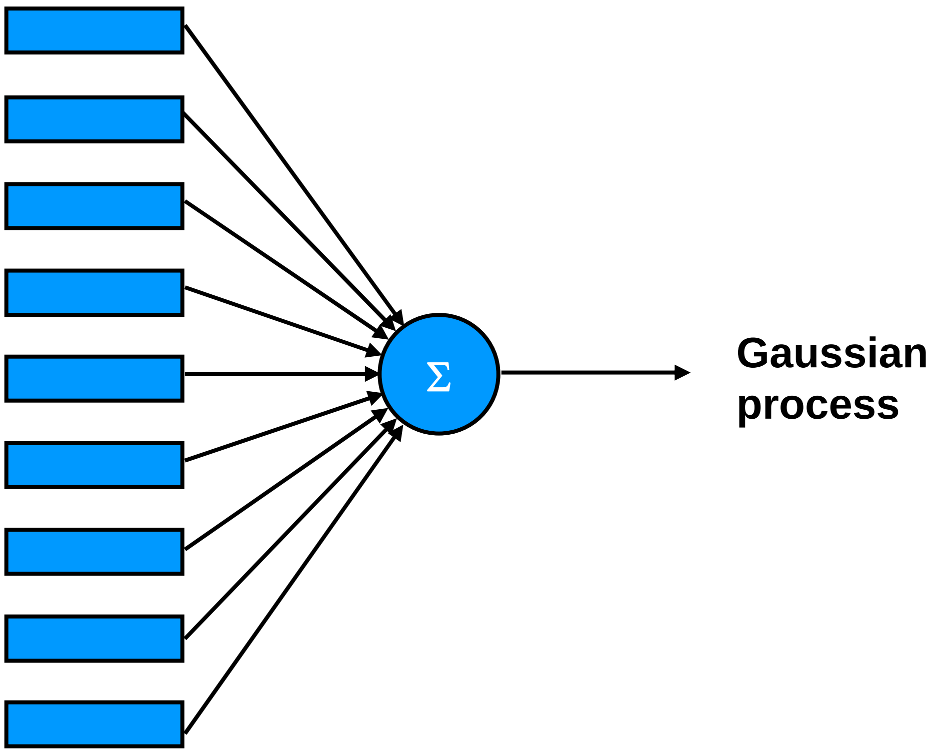
- For a population having a mean μ, the observed mean of n samples measured in one experiment is $\overline{x}$. The confidence interval, i.e., the region within δ of μ is:

$$\overline{x} - \delta \le \mu \le \overline{x} + \delta$$

- If α is the probability that the observed mean will not be within δ of μ, the confidence level is:

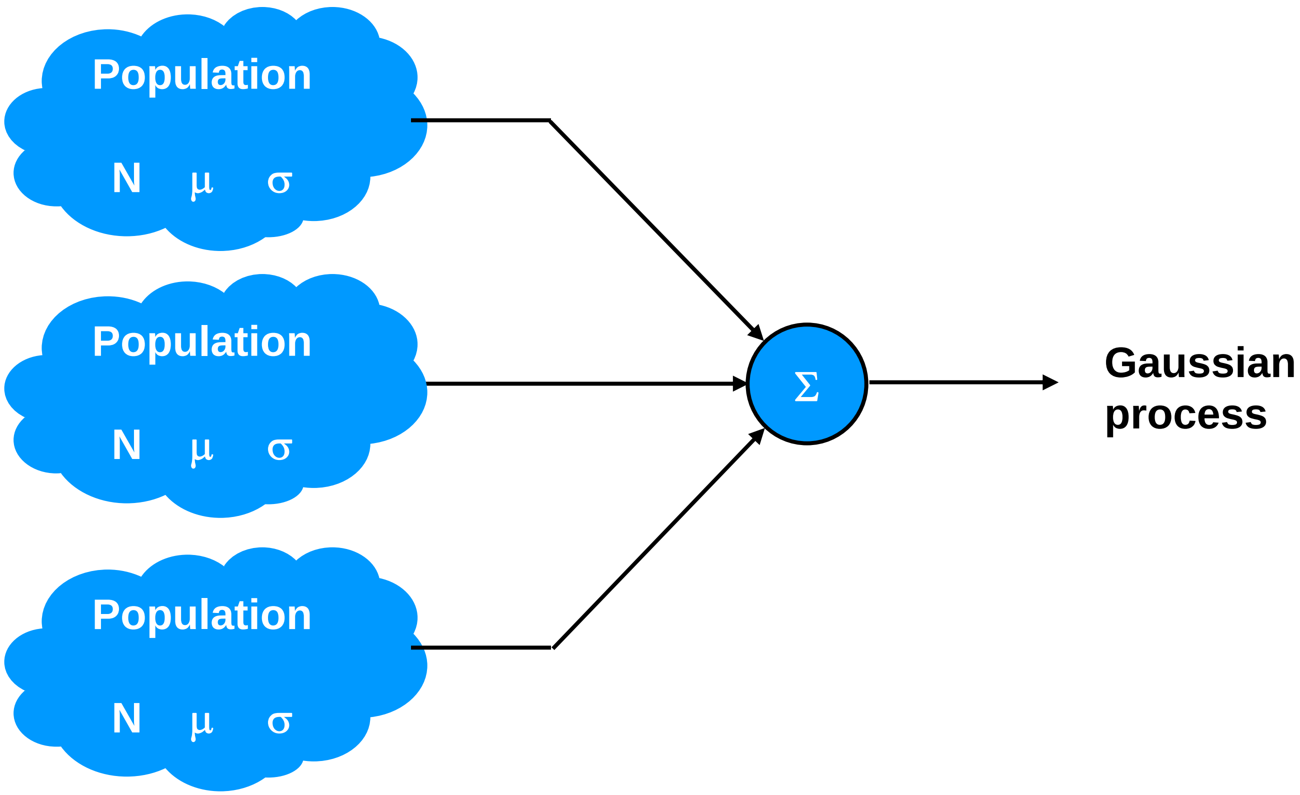
$$1 - \alpha = P(\overline{x} - \delta \le \mu \le \overline{x} + \delta)$$

Source of Gaussian Distribution



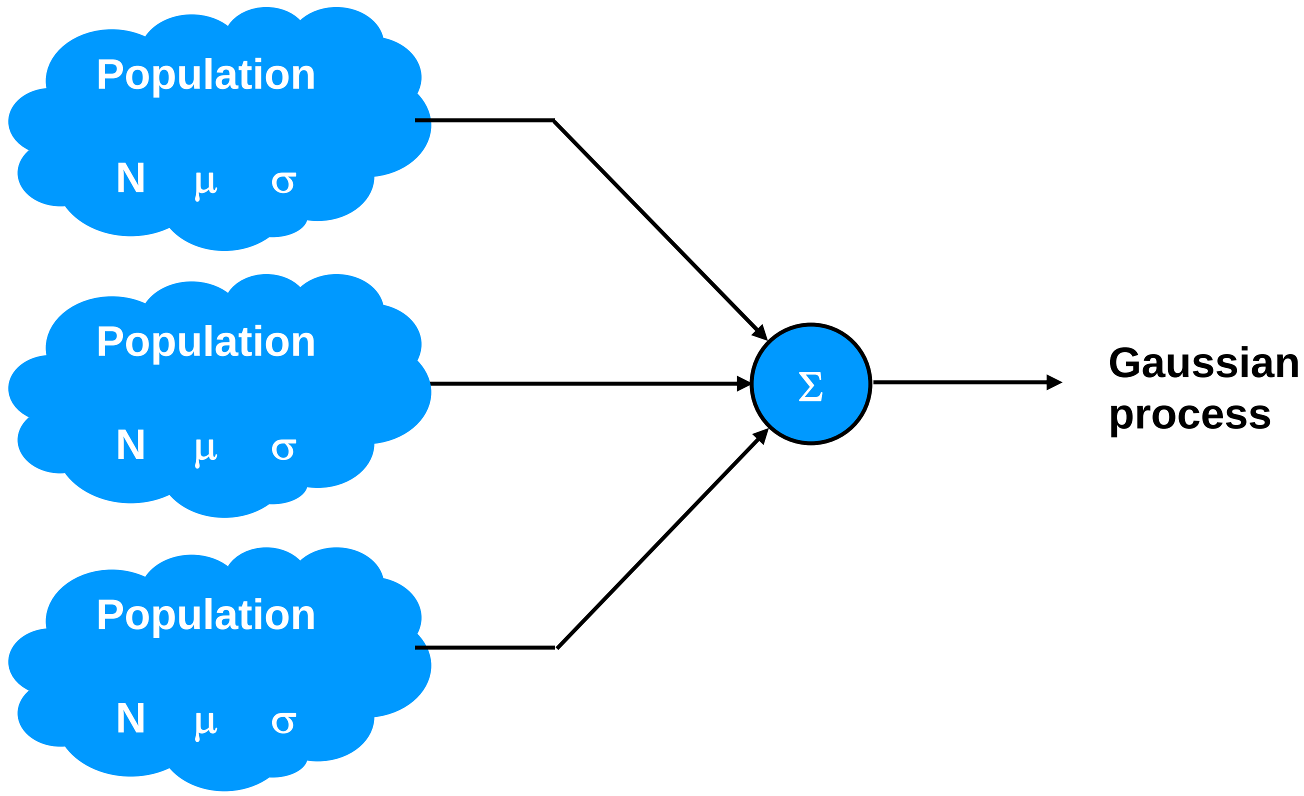
Large number of independent, identically distributed r.v.s

From last time: Source of Gaussian Distribution



Assume experimental samples are picked randomly from population
Assume number of samples is "large",
Assume identically distributed samples

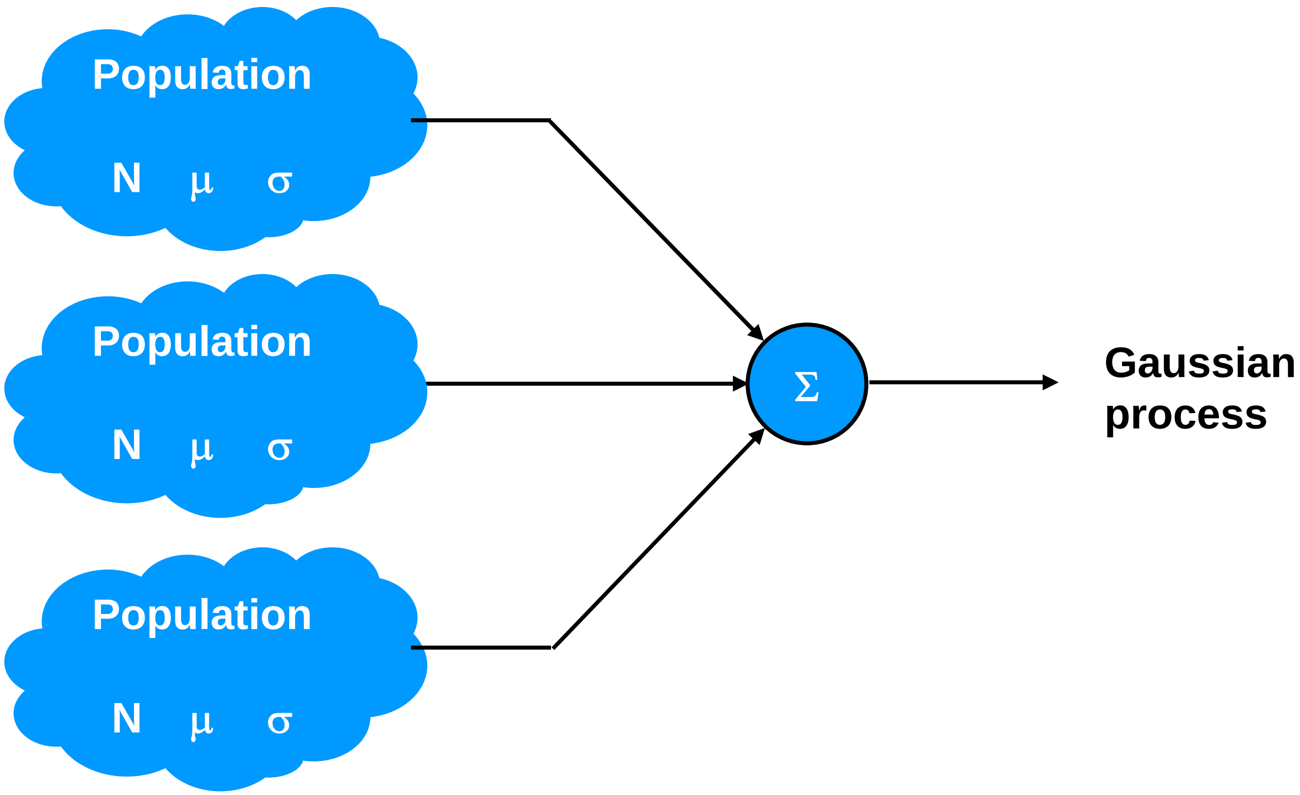
From last time: Source of Gaussian Distribution



Assume experimental samples are picked randomly from population
Assume number of samples is "large",
Assume identically distributed samples

If population is gaussian, n>0 is gaussian

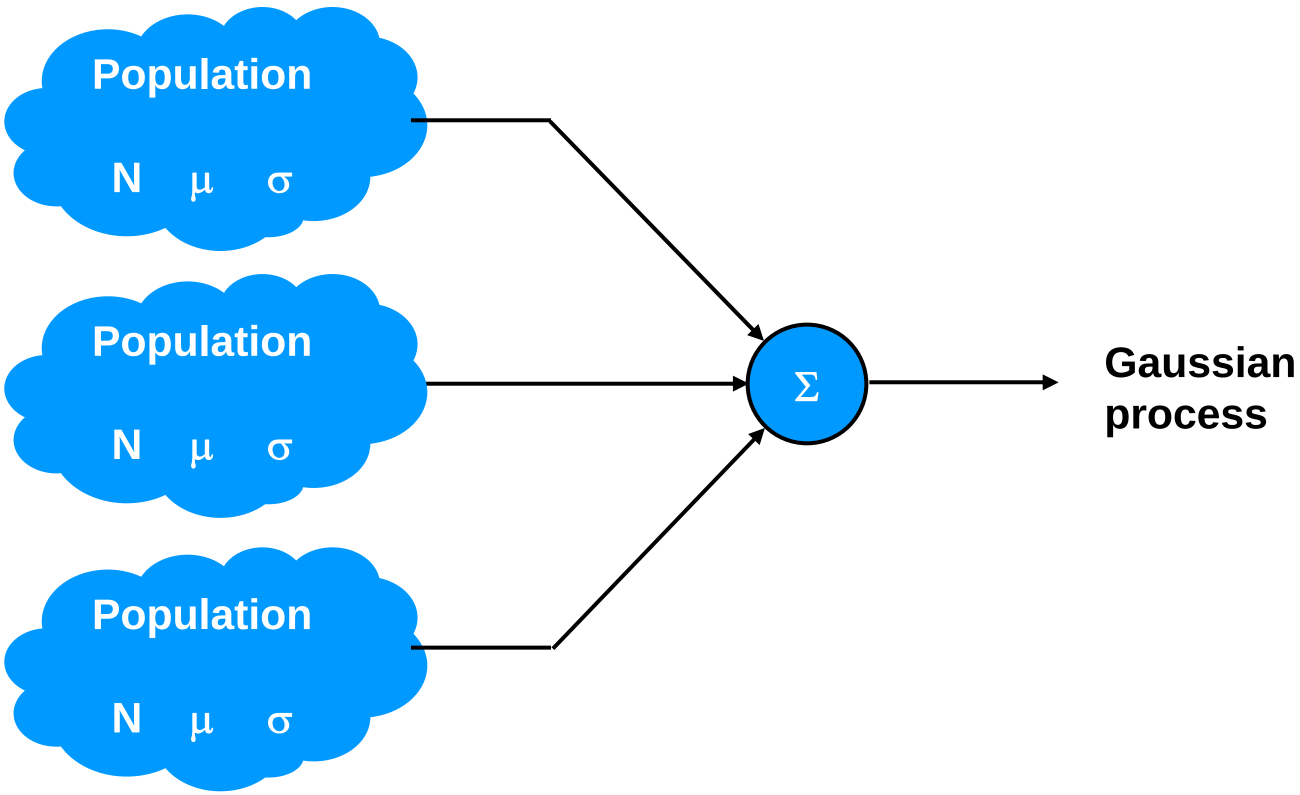
From last time: Source of Gaussian Distribution



Assume experimental samples are picked randomly from population
Assume number of samples is "large",
Assume identically distributed samples

If population is non gaussian, but n>30, result is very close to gaussian

From last time: Source of Gaussian Distribution

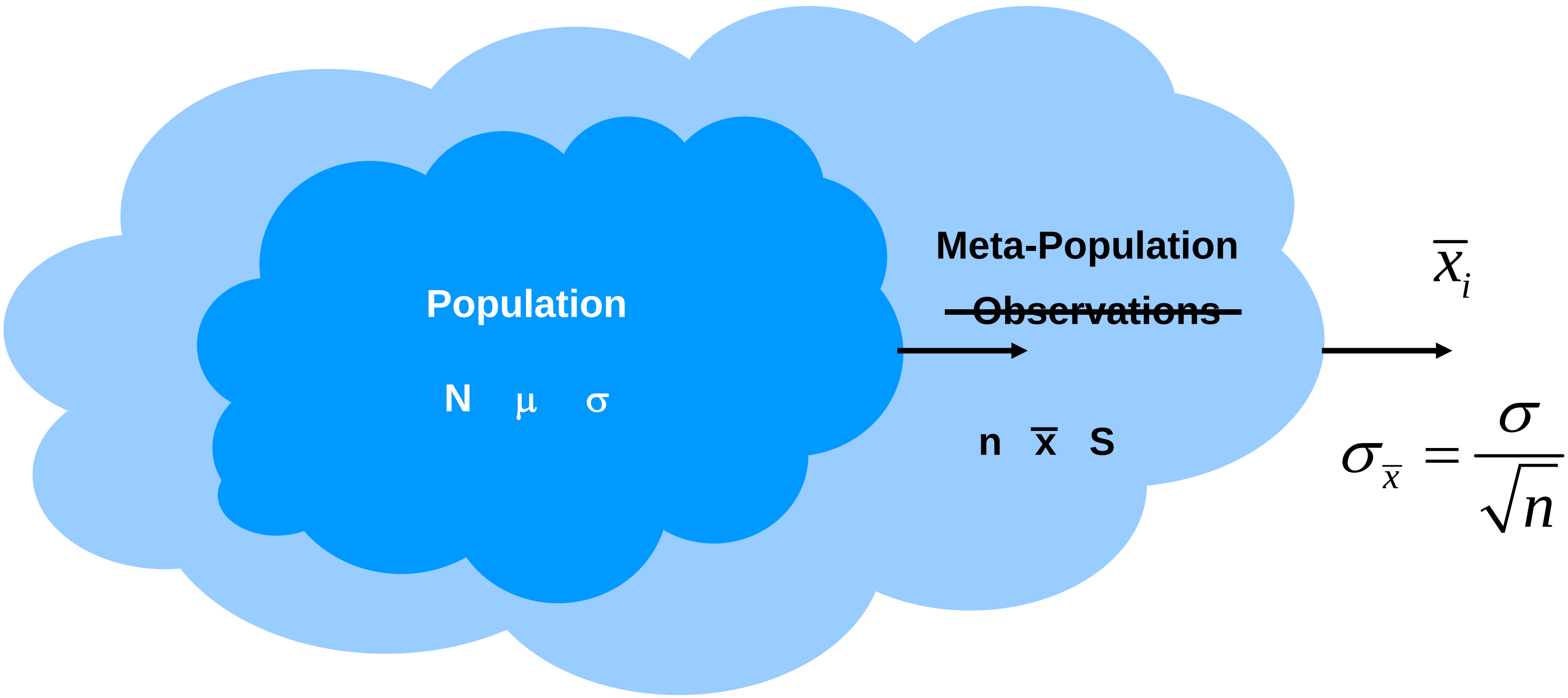


Assume experimental samples are picked randomly from population
Assume number of samples is "large",
Assume identically distributed samples

If population is non gaussian,
and n<30 result is not gaussian

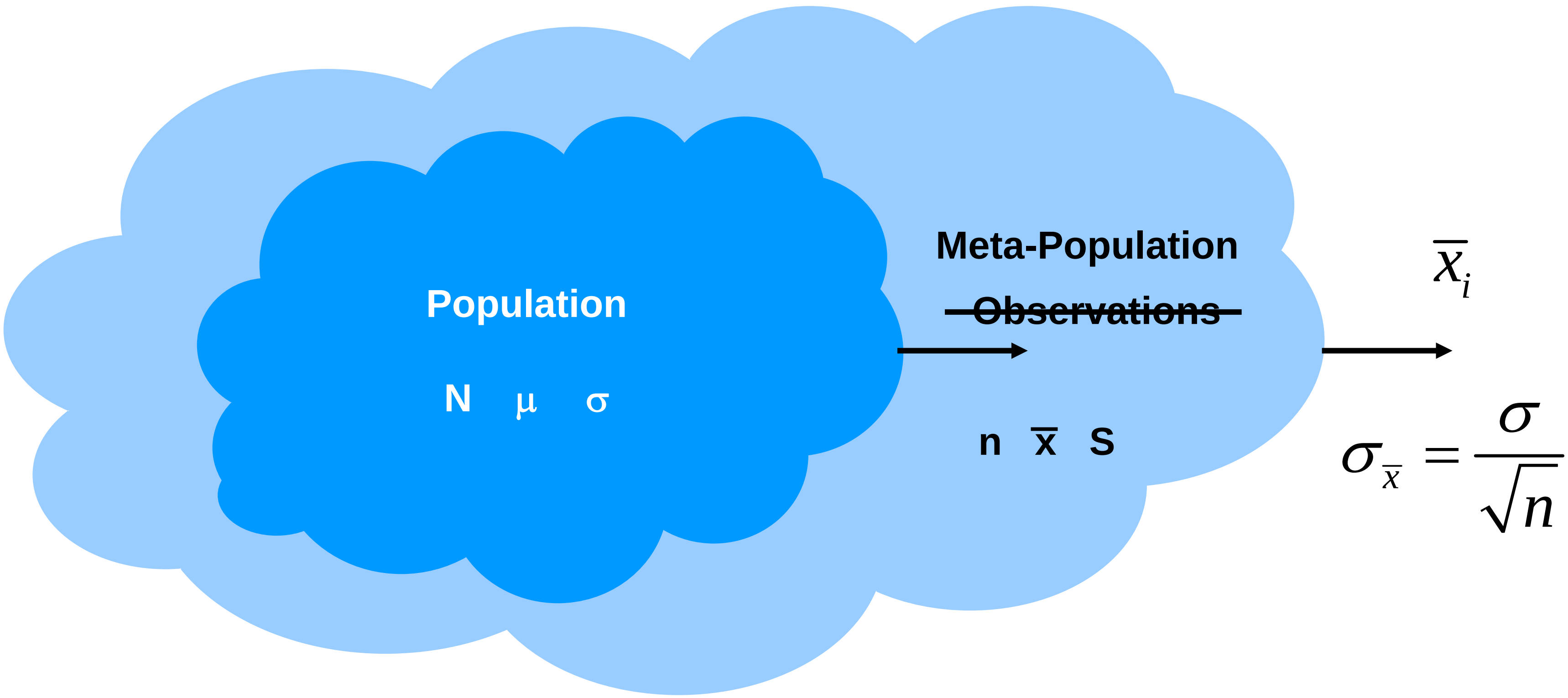
Estimating Confidence Interval

- Assume a large enough sample size, use Central Limit Theorem



Estimating Confidence Interval

- Assume a large enough sample size, use Central Limit Theorem



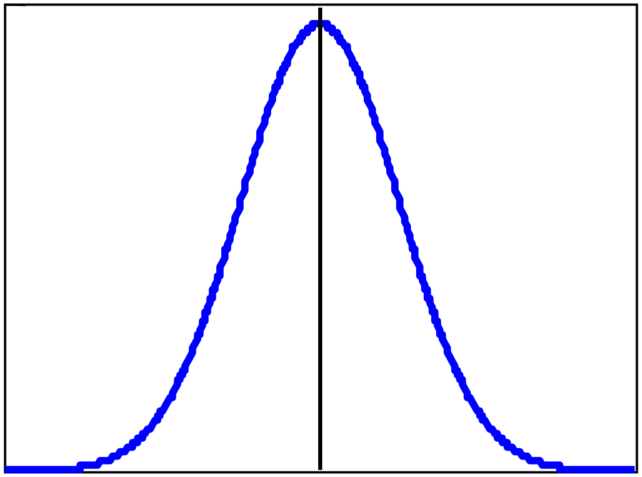
What can be said, statistically, about confidence interval?

Computing Confidence Interval

- Define statistic z:

$$z = \frac{\overline{x} - \mu}{\sigma_{\overline{x}}}$$

z is normally distributed with <u>zero mean</u>



z=0

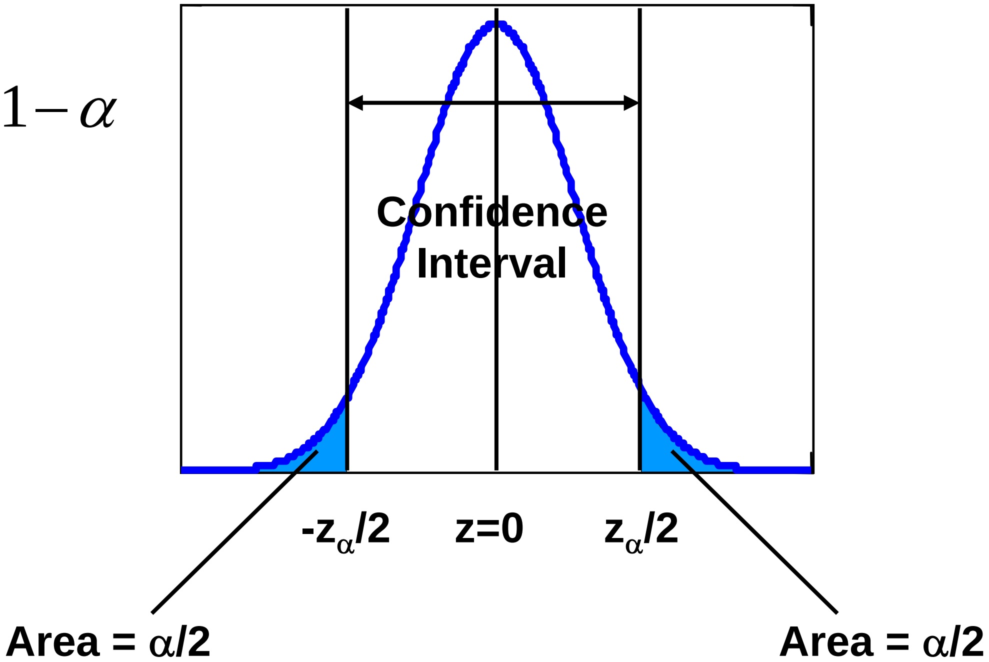
Computing Confidence Interval

- Define statistic z:

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

z is normally distributed with <u>zero mean</u>

$$P(-z_{\alpha/2} \leq \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \leq z_{\alpha/2}) = 1 - \alpha$$



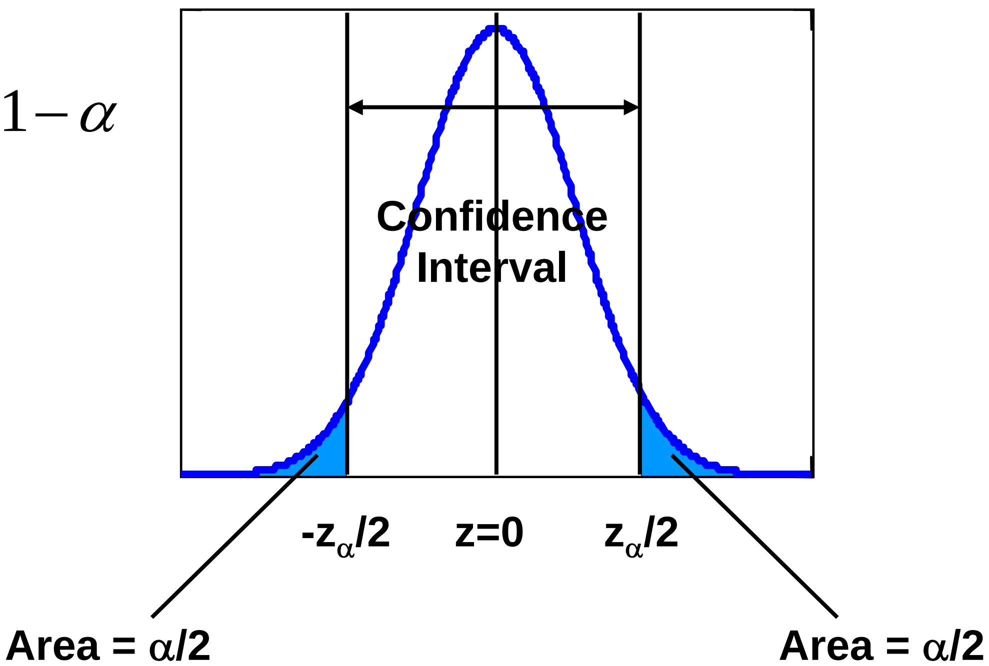
Computing Confidence Interval

- Define statistic z:

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

z is normally distributed with <u>zero mean</u>

$$P(-z_{\alpha/2} \le \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \le z_{\alpha/2}) = 1 - \alpha$$



Or,

$$\mu = \bar{x} \pm \left(z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

with confidence 1-α

Confidence Interval Example

- We need to know the melting point of an organic compound being manufactured by a chemical process, but the results depend on the specific composition, which varies randomly. 50 samples are tested, and the average melting point is found to be 80 ^oC with a standard deviation of 3 ^oC. What is the 98% confidence interval for the average melting temperature?

Confidence Interval Example

- We need to know the melting point of an organic compound being manufactured by a chemical process, but the results depend on the specific composition, which varies randomly. 50 samples are tested, and the average melting point is found to be 80 ^{o}C with a standard deviation of 3 ^{o}C. What is the 98% confidence interval for the average melting temperature?
    - 50 samples are enough to assume a Gaussian distribution of results.

Confidence Interval Example

- We need to know the melting point of an organic compound being manufactured by a chemical process, but the results depend on the specific composition, which varies randomly. 50 samples are tested, and the average melting point is found to be 80 ^oC with a standard deviation of 3 ^oC. What is the 98% confidence interval for the average melting temperature?
    - 50 samples are enough to assume a Gaussian distribution of results.
    - 98% confidence implies $\alpha = .02$

Confidence Interval Example

- We need to know the melting point of an organic compound being manufactured by a chemical process, but the results depend on the specific composition, which varies randomly. 50 samples are tested, and the average melting point is found to be 80 ^{o}C with a standard deviation of 3 ^{o}C. What is the 98% confidence interval for the average melting temperature?

  - 50 samples are enough to assume a Gaussian distribution of results.
  - 98% confidence implies α = .02
  - We need to find $z_\alpha/2$ such that

$$\int_0^{z_\alpha/2} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} = .5 - \frac{\alpha}{2} = .49$$

Confidence Interval Example

- We need to know the melting point of an organic compound being manufactured by a chemical process, but the results depend on the specific composition, which varies randomly. 50 samples are tested, and the average melting point is found to be 80 ^{o}C with a standard deviation of 3 ^{o}C. What is the 98% confidence interval for the average melting temperature?

  - 50 samples are enough to assume a Gaussian distribution of results.
  - 98% confidence implies α = .02
  - We need to find $z_\alpha/2$ such that

$$\int_0^{z_\alpha/2} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} = .5 - \frac{\alpha}{2} = .49$$

  - Use Table 6.3 with a value .49

    $z_\alpha/2 = 2.326$

Confidence Interval Example

- We need to know the melting point of an organic compound being manufactured by a chemical process, but the results depend on the specific composition, which varies randomly. 50 samples are tested, and the average melting point is found to be 80 ^oC with a standard deviation of 3 ^oC. What is the 98% confidence interval for the average melting temperature?

  - 50 samples are enough to assume a Gaussian distribution of results.
  - 98% confidence implies α = .02
  - We need to find $z_\alpha/2$ such that

$$\int_0^{z_\alpha/2} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} = .5 - \frac{\alpha}{2} = .49$$

  - Use Table 6.3 with a value .49

    $z_\alpha/2 = 2.326$
  - Find confidence interval:

$$\overline{x} - \frac{z_{\alpha/2}S}{\sqrt{n}} \le \mu \le \overline{x} + \frac{z_{\alpha/2}S}{\sqrt{n}} = 80 - \frac{2.326 \cdot 3}{\sqrt{50}} \le \mu \le 80 + \frac{2.326 \cdot 3}{\sqrt{50}}$$

$$79.013 \le \mu \le 80.987$$

Next time

- More On Statistical Analysis of Experimental Data
  - Correlation

Homework 6

- Problems 6.17, 6.26, 6.29, 6.34, 6.44