IE 340/440

PROCESS IMPROVEMENT THROUGH PLANNED EXPERIMENTATION

Confidence Interval Estimation

Dr. Xueping Li
University of Tennessee

Chapter Topics

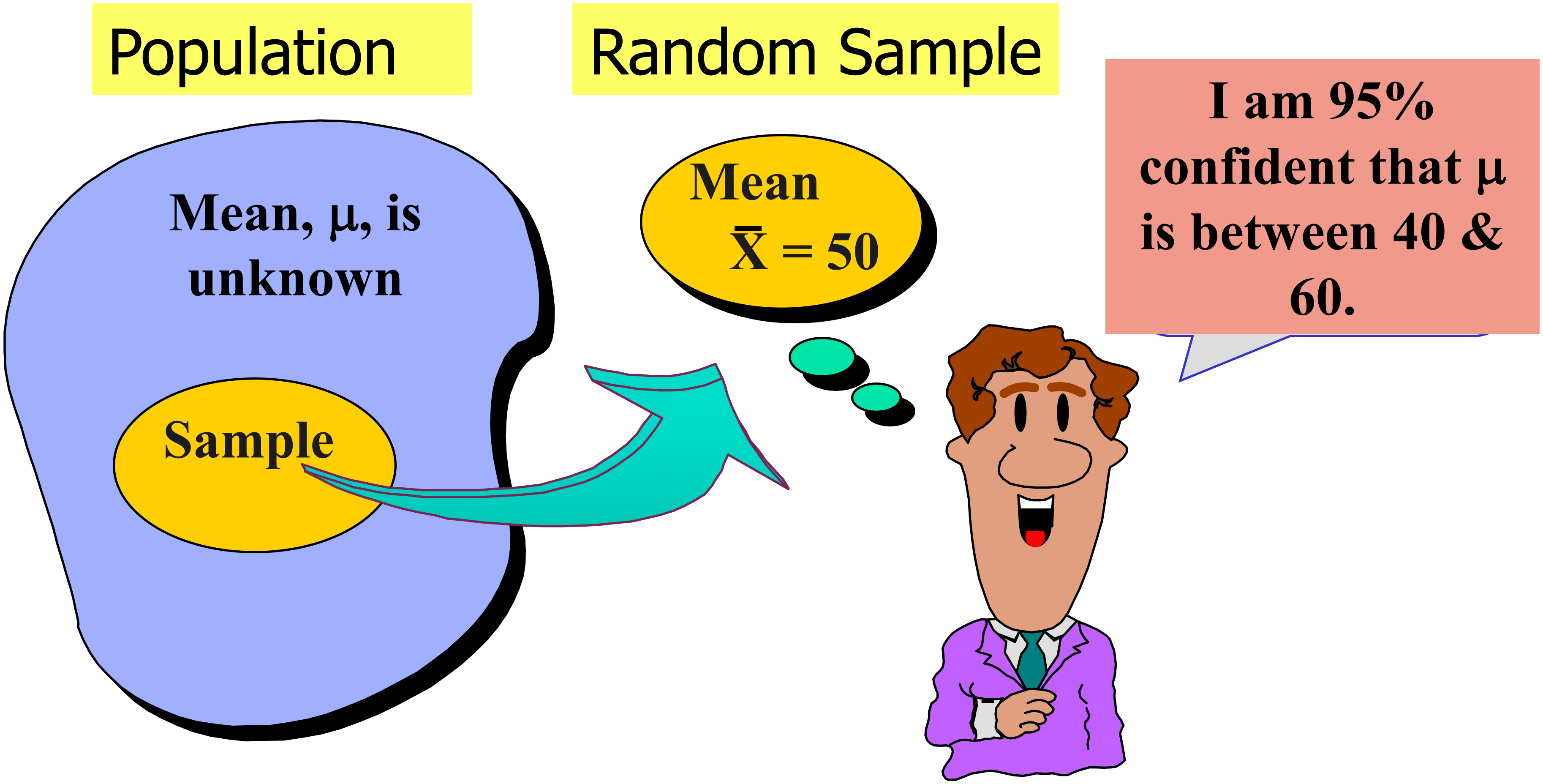
- Estimation Process
- Point Estimates
- Interval Estimates
- Confidence Interval Estimation for the Mean (σ Known)
- Determining Sample Size
- Confidence Interval Estimation for the Mean (σ Unknown)

Chapter Topics

(continued)

- Confidence Interval Estimation for the Proportion
- Confidence Interval Estimation for Population Total
- Confidence Interval Estimation for Total Difference in the Population
- Estimation and Sample Size Determination for Finite Population
- Confidence Interval Estimation and Ethical Issues

Estimation Process



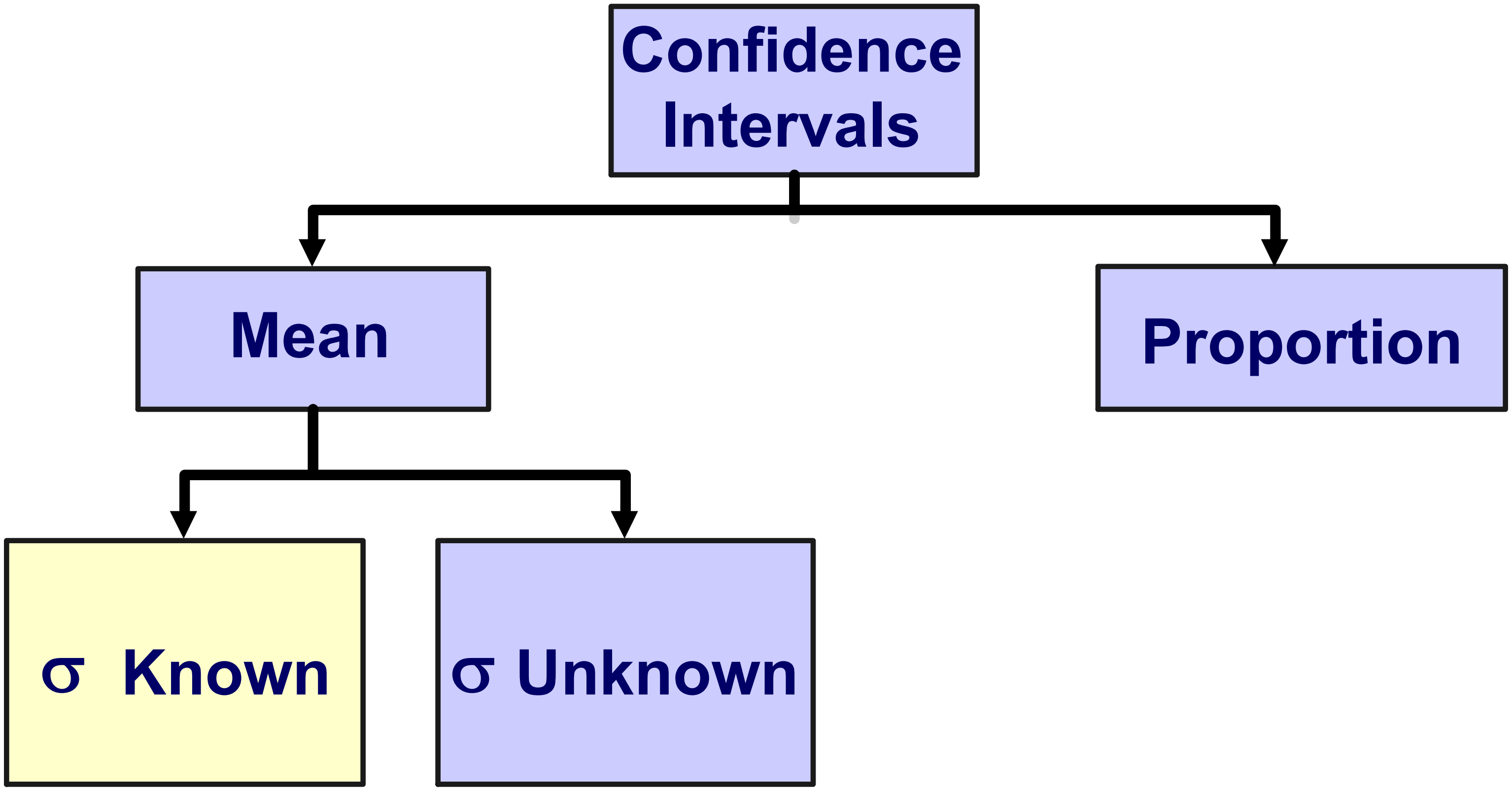
Point Estimates

Estimate Population Parameters …		with Sample Statistics
Mean	μ	$\bar{X}$
Proportion	p	P_S
Variance	σ^2	S^2
Difference	$\mu_1 - \mu_2$	$\bar{X}_1 - \bar{X}_2$

Interval Estimates

- Provide Range of Values
  - Take into consideration variation in sample statistics from sample to sample
  - Based on observation from 1 sample
  - Give information about closeness to unknown population parameters
  - Stated in terms of level of confidence
    - Never 100% sure

Confidence Interval Estimates



Confidence Interval for μ (σ Known)

Critical Value

Standard Error

- Assumptions
  - Population standard deviation is known
  - Population is normally distributed
  - If population is not normal, use large sample
- Confidence Interval Estimate
  - $$\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$
  - $e = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ is called the **sampling error** or **margin of error**

Elements of Confidence Interval Estimation

- Level of Confidence
  - Confidence that the interval will contain the unknown population parameter
- Precision (Range)
  - Closeness to the unknown parameter
- Cost
  - Cost required to obtain a sample of size n

Level of Confidence

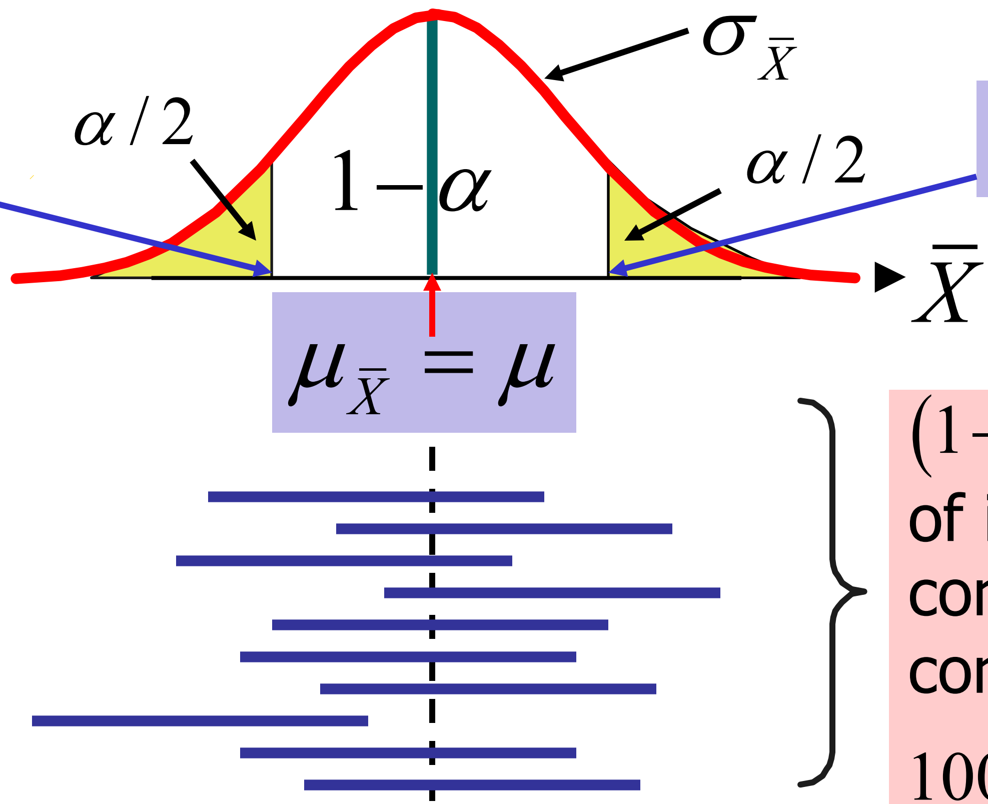
- Denoted by $100(1-\alpha)\%$
- A Relative Frequency Interpretation
  - In the long run, $100(1-\alpha)\%$ of all the confidence intervals that can be constructed will contain (bracket) the unknown parameter
- A Specific Interval Will Either Contain or Not Contain the Parameter
  - No probability involved in a specific interval

Interval and Level of Confidence

Sampling Distribution of the Mean

$\mu - Z_{\alpha/2}\sigma_{\bar{X}}$

$\mu + Z_{\alpha/2}\sigma_{\bar{X}}$



$\mu_{\bar{X}} = \mu$

Intervals extend from

$\bar{X} - Z\sigma_{\bar{X}}$

to

$\bar{X} + Z\sigma_{\bar{X}}$

$(1-\alpha)100\%$ of intervals constructed contain μ ;

$100\alpha\%$ do not.

Confidence Intervals

Example

A random sample of 15 stocks traded on the NASDAQ showed an average shares traded to be 215000. From past experience, it is believed that the population standard deviation of shares traded is 195000 and the shares traded are very close to a normal distribution. Construct a 99% confidence interval for the average shares traded on the NASDAQ. Interpret your result.

PHStat output

Confidence Interval Estimate for the Mean	
Population Standard Deviation	**195000**
Sample Mean	**215000**
Sample Size	**15**
Confidence Level	**99%**
Standard Error of the Mean	50348.7835
Z Value	-2.57583451
Interval Half Width	129690.1343
Interval Lower Limit	**85309.86569**
Interval Upper Limit	**344690.1343**

The 99% CI for the population mean:

$$85309 < \mu < 344690$$

Example: Interpretation

(continued)

If all possible samples of size 15 are taken and the corresponding 99% confidence intervals are constructed, 99% of the confidence intervals that are constructed will contain the true unknown population mean.

We are 99% confident that the population average number of shares traded on the NASDAQ is between 85309 and 344690.

For this particular confidence interval [85309, 344690], the unknown population mean can either be in the interval or not in the interval. It is, therefore, **incorrect** to state that the probability is 99% that the unknown population mean will be in the interval [85309, 344690].

Example: Interpretation

(continued)

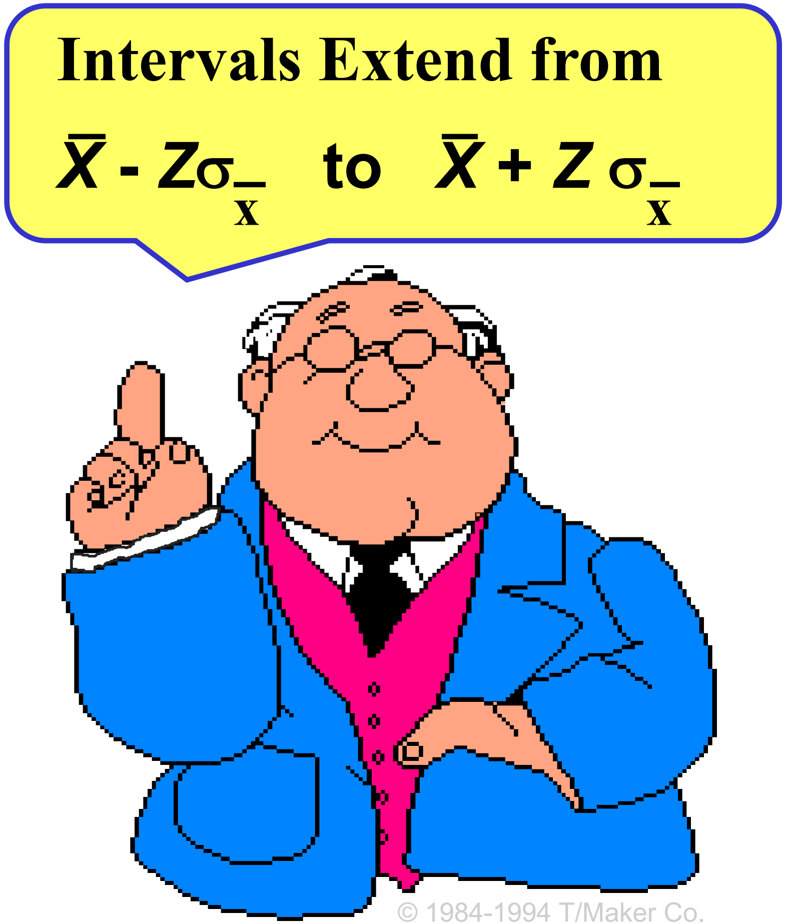
Using the confidence interval method on repeated sampling, the probability that we will have constructed a confidence interval that will contain the unknown population mean is 99%.

Obtaining Confidence Interval in PHStat

- PHStat | Confidence Interval | Estimates for the Mean, Sigma Known

Factors Affecting Interval Width (Precision)

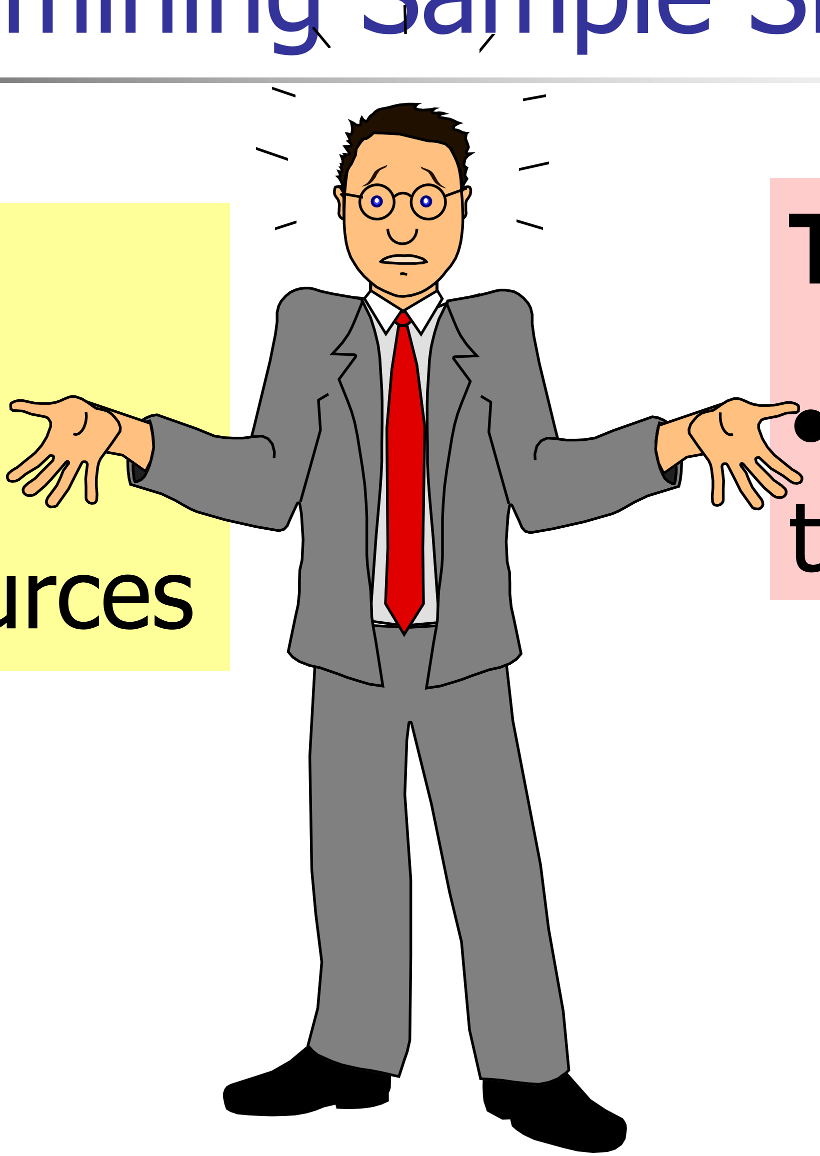
- Data Variation
  - Measured by σ
- Sample Size
  - $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}$
- Level of Confidence
  - $100(1-\alpha)\%$



Determining Sample Size (Cost)

Too Big:

• Requires more resources



Too small:

• Won’t do the job

Determining Sample Size for Mean

What sample size is needed to be 90% confident of being correct within ± 5? A pilot study suggested that the standard deviation is 45.

$$n = \frac{Z^2\sigma^2}{\text{Error}^2} = \frac{1.645^2\left(45^2\right)}{5^2} = 219.2 \cong 220$$

Round Up

Determining Sample Size for Mean in PHStat

- PHStat | Sample Size | Determination for the Mean …
- Example in Excel Spreadsheet

Microsoft Excel Worksheet

Sample Size Determination	
Data	
Population Standard Deviation	**45**
Sampling Error	**5**
Confidence Level	**90%**
Intemediate Calculations	
Z Value	-1.644853
Calculated Sample Size	219.1488528
Result	
Sample Size Needed	**220**

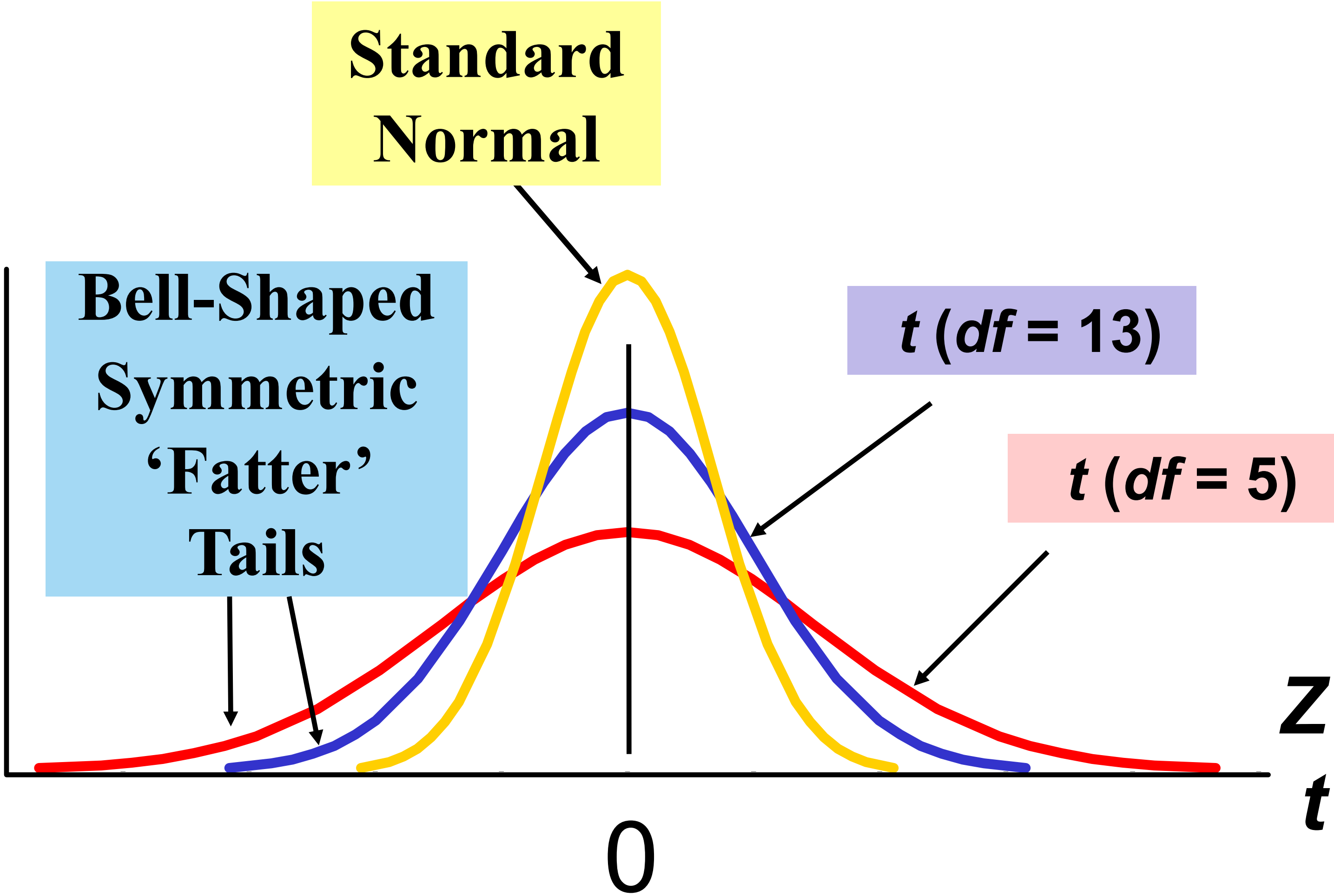
Confidence Interval for μ (σ Unknown)

Standard Error

Margin of Error

- Assumptions
  - Population standard deviation is unknown
  - Population is normally distributed
  - If population is not normal, use large sample
- Use Student's t Distribution
- Confidence Interval Estimate
  - $$\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} \le \mu \le \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}$$

Student's *t* Distribution



Student's t Table

	Upper Tail Area		
df	.25	.10	**.05**
1	1.000	3.078	6.314
2	0.817	1.886	**2.920**
3	0.765	1.638	2.353

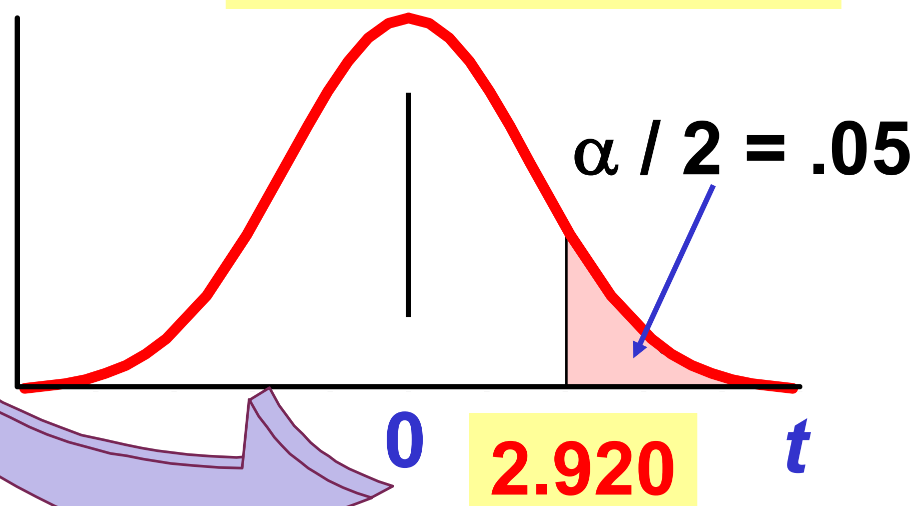
t Values

Let: n = 3

df = n - 1 = 2

$\alpha = .10$

$\alpha/2 = .05$



Example

A random sample of $n = 25$ has $\bar{X} = 50$ and $S = 8$. Set up a 95% confidence interval estimate for μ.

$$\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} \le \mu \le \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}$$

$$50 - 2.0639\frac{8}{\sqrt{25}} \le \mu \le 50 + 2.0639\frac{8}{\sqrt{25}}$$

$$46.69 \le \mu \le 53.30$$

We are 95% confident that the unknown true population mean is somewhere between 46.69 and 53.30.

Confidence Interval for μ (σ Unknown) in PHStat

- PHStat | Confidence Interval | Estimate for the Mean, Sigma Unknown
- Example in Excel Spreadsheet

Microsoft Excel Worksheet

Confidence Interval Estimate for the Mean	
Data	
Sample Standard Deviation	**8**
Sample Mean	**50**
Sample Size	**25**
Confidence Level	**95%**
Intermediate Calculations	
Standard Error of the Mean	1.6
Degrees of Freedom	24
t Value	2.063898137
Interval Half Width	3.302237019
Confidence Interval	
Interval Lower Limit	**46.70**
Interval Upper Limit	**53.30**

Confidence Interval Estimate for Proportion

- Assumptions
  - Two categorical outcomes
  - Population follows binomial distribution
  - Normal approximation can be used if $np \geq 5$ and $n(1-p) \geq 5$
  - Confidence Interval Estimate
  - $$p_S - Z_{\alpha/2}\sqrt{\frac{p_S(1-p_S)}{n}} \leq p \leq p_S + \underbrace{Z_{\alpha/2}\underbrace{\sqrt{\frac{p_S(1-p_S)}{n}}}_{\text{Standard Error}}}_{\text{Margin of Error}}$$

Example

A random sample of 400 voters showed that 32 preferred Candidate A. Set up a 95% confidence interval estimate for *p*.

$$p_s - Z_{\alpha/2}\sqrt{\frac{p_s(1-p_s)}{n}} \le p \le p_s + Z_{\alpha/2}\sqrt{\frac{p_s(1-p_s)}{n}}$$

$$.08 - 1.96\sqrt{\frac{.08(1-.08)}{400}} \le p \le .08 + 1.96\sqrt{\frac{.08(1-.08)}{400}}$$

$$.053 \le p \le .107$$

We are 95% confident that the proportion of voters who prefer Candidate A is somewhere between 0.053 and 0.107.

Confidence Interval Estimate for Proportion in PHStat

- PHStat | Confidence Interval | Estimate for the Proportion …
- Example in Excel Spreadsheet

Microsoft Excel Worksheet

Confidence Interval Estimate for the Mean	
Data	
Sample Size	**400**
Number of Successes	**32**
Confidence Level	**95%**
Intermediate Calculations	
Sample Proportion	0.08
Z Value	-1.95996108
Standard Error of the Proportion	0.01356466
Interval Half Width	0.026586206
Confidence Interval	
Interval Lower Limit	**0.053413794**
Interval Upper Limit	**0.106586206**

Determining Sample Size for Proportion

Out of a population of 1,000, we randomly selected 100, of which 30 were defective. What sample size is needed to be within ± 5% with 90% confidence?

$$n = \frac{Z^2 p(1-p)}{\text{Error}^2} = \frac{1.645^2 (0.3)(0.7)}{0.05^2}$$

$$= 227.3 \cong 228$$

Round Up

Determining Sample Size for Proportion in PHStat

- PHStat | Sample Size | Determination for the Proportion …
- Example in Excel Spreadsheet

Microsoft Excel Worksheet

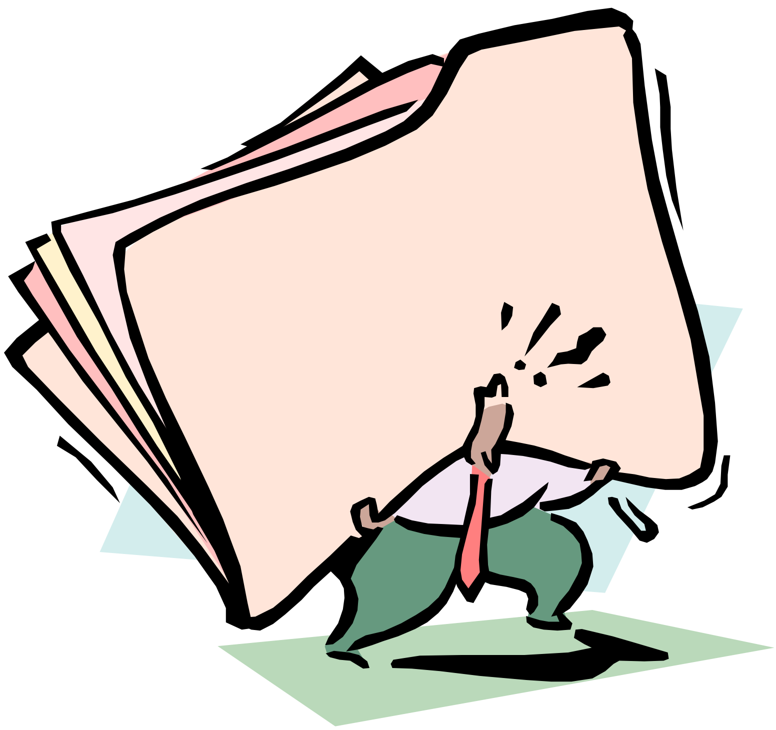
Sample Size Determination	
Data	
Estimate of True Proportion	**0.3**
Sampling Error	**0.05**
Confidence Level	**90%**
Intermediate Calculations	
Z Value	-1.644853
Calculated Sample Size	227.265477
Result	
Sample Size Needed	**228**

Confidence Interval for Population Total Amount

- Point Estimate
  - $N\bar{X}$
- Confidence Interval Estimate
  - $N\bar{X} \pm N\left(t_{\alpha/2,n-1}\right)\dfrac{S}{\sqrt{n}}\sqrt{\dfrac{(N-n)}{(N-1)}}$

Confidence Interval for Population Total: Example

An auditor is faced with a population of 1000 vouchers and wishes to estimate the total value of the population of vouchers. A sample of 50 vouchers is selected with the average voucher amount of $1076.39, standard deviation of $273.62. Set up the 95% confidence interval estimate of the total amount for the population of vouchers.



Example Solution

$$N = 1000 \qquad n = 50 \qquad \bar{X} = \$1076.39 \qquad S = \$273.62$$

$$N\bar{X} \pm N\left(t_{\alpha/2,n-1}\right)\frac{S}{\sqrt{n}}\sqrt{\frac{(N-n)}{(N-1)}}$$

$$= (1000)(1076.39) \pm (1000)(2.0096)\frac{273.62}{\sqrt{100}}\sqrt{\frac{1000-50}{1000-1}}$$

$$= 1,076,390 \pm 75,830.85$$

The 95% confidence interval for the population total amount of the vouchers is between 1,000,559.15 and 1,152,220.85.

Example Solution in PHStat

- PHStat | Confidence Intervals | Estimate for the Population Total
- Excel Spreadsheet for the Voucher Example

Microsoft Excel Worksheet

Confidence Interval for Total Difference in the Population

- Point Estimate
  - $N\bar{D}$ where $\bar{D} = \frac{\sum_{i=1}^{n} D_i}{n}$ is the sample average difference
- Confidence Interval Estimate
  - $$N\bar{D} \pm N\left(t_{\alpha/2,n-1}\right)\frac{S_D}{\sqrt{n}}\sqrt{\frac{(N-n)}{(N-1)}}$$
  - where $$S_D = \sqrt{\frac{\sum_{i=1}^{n}\left(D_i - \bar{D}\right)^2}{n-1}}$$

Estimation for Finite Population

- Samples are Selected Without Replacement
  - Confidence interval for the mean (σ unknown)
    - $$\bar{X} \pm t_{\alpha/2,n-1} \frac{S}{\sqrt{n}} \sqrt{\frac{(N-n)}{(N-1)}}$$
  - Confidence interval for proportion
    - $$p_S \pm Z_{\alpha/2} \sqrt{\frac{p_S(1-p_S)}{n}} \sqrt{\frac{(N-n)}{(N-1)}}$$

Sample Size (n) Determination for Finite Population

- Samples are Selected Without Replacement
  - $$n = \frac{n_0 N}{n_0 + (N-1)}$$
  - When estimating the mean
    - $$n_0 = \frac{Z_{\alpha/2}^2 \sigma^2}{e^2}$$
  - When estimating the proportion
    - $$n_0 = \frac{Z_{\alpha/2}^2 p(1-p)}{e^2}$$

Ethical Issues

- Confidence Interval (Reflects Sampling Error) Should Always Be Reported Along with the Point Estimate
- The Level of Confidence Should Always Be Reported
- The Sample Size Should Be Reported
- An Interpretation of the Confidence Interval Estimate Should Also Be Provided

Chapter Summary

- Illustrated Estimation Process
- Discussed Point Estimates
- Addressed Interval Estimates
- Discussed Confidence Interval Estimation for the Mean (σ Known)
- Addressed Determining Sample Size
- Discussed Confidence Interval Estimation for the Mean (σ Unknown)

Chapter Summary

(continued)

- Discussed Confidence Interval Estimation for the Proportion
- Addressed Confidence Interval Estimation for Population Total
- Discussed Confidence Interval Estimation for Total Difference in the Population
- Addressed Estimation and Sample Size Determination for Finite Population
- Addressed Confidence Interval Estimation and Ethical Issues