Interval Estimation

- Interval Estimation of a Population Mean:
  Large-Sample Case
- Interval Estimation of a Population Mean:
  Small-Sample Case
- Determining the Sample Size

Inferential Statistics

- Based on a sample, inferential statistics is all about making some type of statement concerning the possible value of the population parameter
    - Statements are made in a probabilistic sense, i.e., we can never say “I am absolutely sure that the true value of the population parameter is ….”

Types of statements

- **Point Estimate:** you best guess as to the value of the population parameter
- **Confidence interval**
  - Based on a sample, make a statement like “I am 90% sure that the true value of the population mean is BETWEEN 65 and 72 (here 65 is a lower bound and 72 is an upper bound)
- **Hypothesis testing**
  - Assume some value for the population parameter, eg, I think the true mean is at least 85. Then, take a sample and see if the evidence supports or refutes this claim

Two Population Parameters

- We are generally making statements concerning 2 population parameters, the population mean and the population proportion, and we are going to use the sample mean and the sample proportion respectively to estimate the parameters

Parameter	Estimate	When to use
μ	$\bar{X}$	When the outcome of each individual trial has many different possible outcomes
π	p	When the outcome of each individual trial only has 2 possible outcomes

Point Estimate

- This is the “Best Guess” of the value of the population parameter, given your sample information
  - Recall, the sample mean is normally distributed with a mean of μ which means that, on average, the sample mean will be equal to the true population mean.
  - Therefore, given a sample mean of $\bar{X}$ obtained from the sample – this is your point estimate of μ.
  - Likewise for π : the sample proportion p is the point estimate for π

Confidence Interval

- A "confidence interval" consists of a **range** in which population parameter may fall and a **confidence level**
  - The **range** is a lower and upper bound between which you think the population parameter lies
  - The **confidence level** is how sure you are that the parameter in within this range
- Interpretation: a 95% confidence interval means that 95% of similarly constructed intervals will contain the population parameter

Confidence Interval (or Interval Estimate)

- A range (or an interval) of values used to estimate the true value of the population parameter

Lower # < population parameter < Upper #

- Probability Statements about the Sampling Error

Degree of Confidence

(level of confidence or confidence coefficient)

- The probability 1 - α (often expressed as the equivalent percentage value) that is the relative frequency of times the confidence interval actually does contain the population parameter, assuming that the estimation process is repeated a large number of times

usually 90%, 95%, or 99%

(α = 10%), (α = 5%), (α = 1%)

Interpreting a Confidence Interval

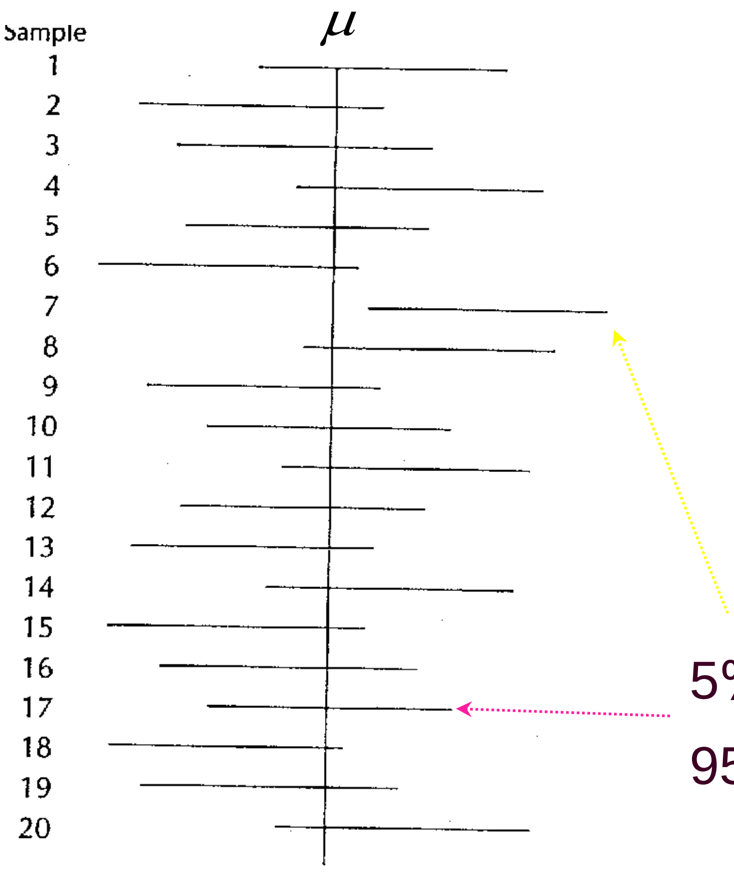
$98.08 < \mu < 98.32$

- **Correct Use:** We are 95% confident that the interval from 98.08 to 98.32 actually does contain the true value of μ. This means that if we were to select many different samples of size 100 and construct the confidence intervals, 95% of them would actually contain the value of the population mean μ.

- **Wrong Use:** There is a 95% chance that the true value of μ will fall between 98.08 and 98.32.

Meaning of a 95% Confidence Interval

In repeated sampling, the interval would encompass the true parameter value 95% of the time.



5% of the intervals don't include the true mean.

95% of the intervals include the true mean.

Sampling Error

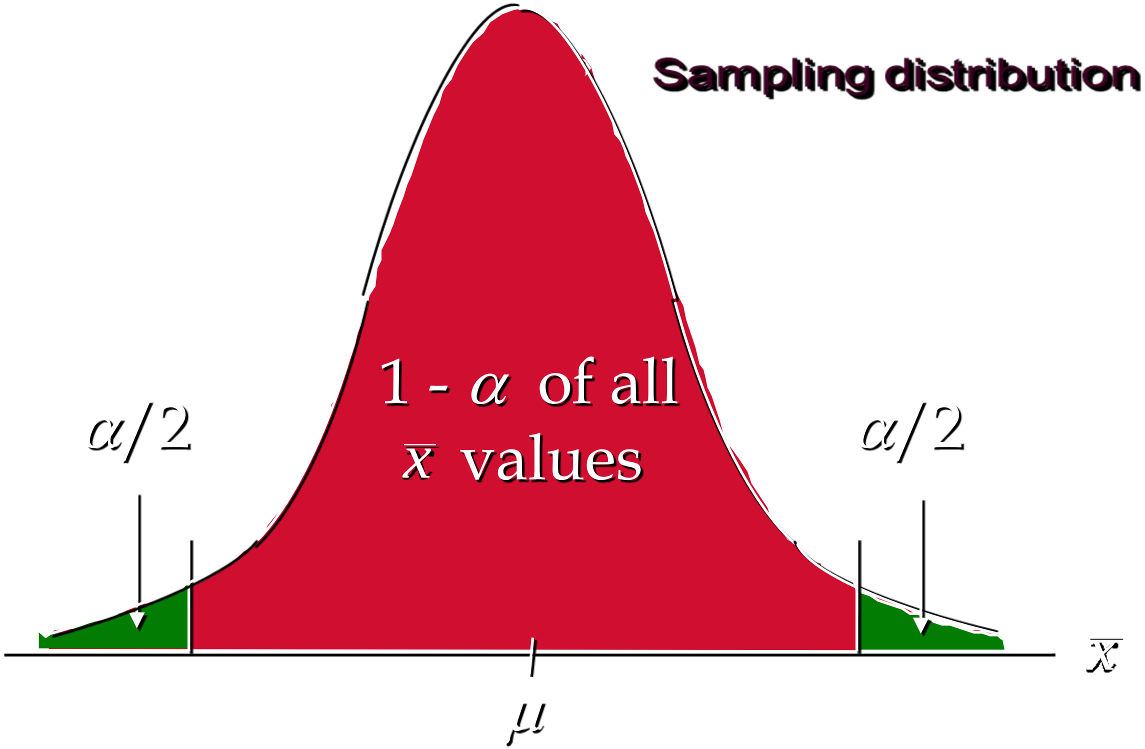
- The absolute value of the difference between an unbiased point estimate and the population parameter it estimates is called the <u>sampling error</u>.
- For the case of a sample mean estimating a population mean, the sampling error is

$$\text{Sampling Error} = |\bar{x} - \mu|$$

Probability Statements About the Sampling Error

- **Precision Statement**

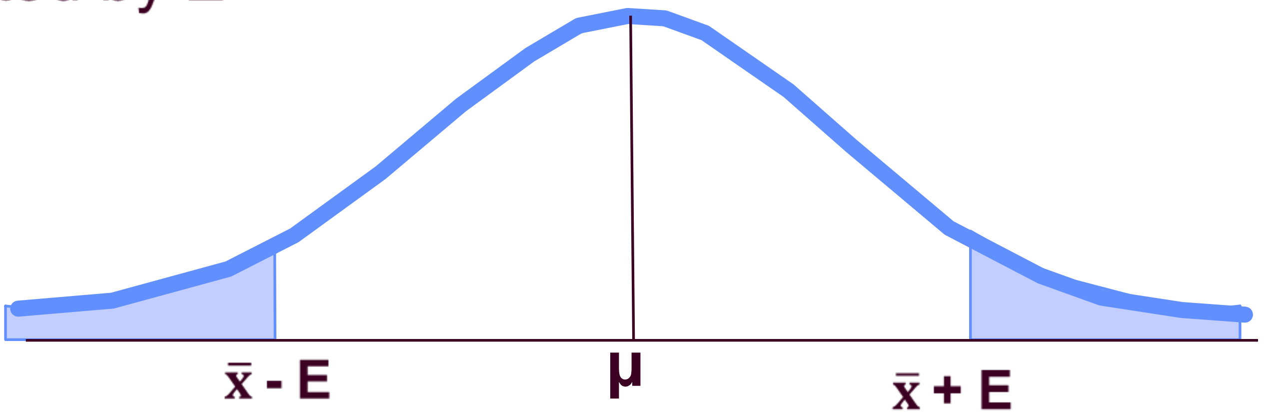
  There is a 1 - α probability that the value of a sample mean will provide a sampling error of $z_{\alpha/2}\sigma_{\bar{x}}$ or less.



Margin of Error

is the maximum likely difference observed between sample mean $\bar{x}$ and true population mean μ.

denoted by E



$$\bar{x} - E < \mu < \bar{x} + E$$

lower limit upper limit

Interval Estimate of a Population Mean: Large-Sample Case ($n \geq 30$)

- **With σ Known** $\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

where 1 - α is the confidence coefficient

$z_{\alpha/2}$ is the z value providing an area of α/2 in the upper tail of the standard normal probability distribution

- **With σ Unknown** $\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$

where s is the sample standard deviation

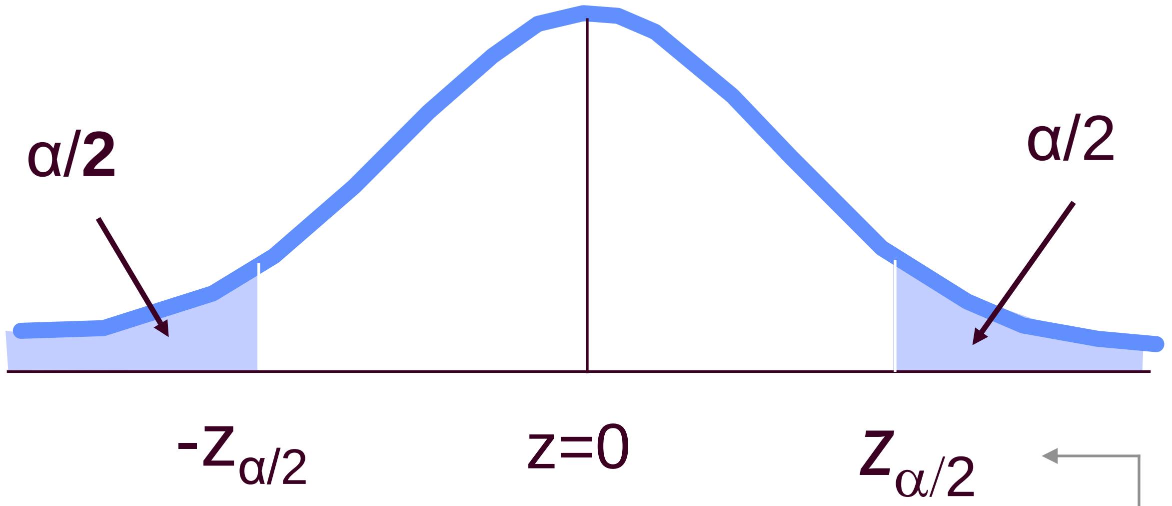
Confidence Coefficient

- If your confidence level is $(1-\alpha)100$ % then to calculate the appropriate confidence coefficient
  - Take $1- \alpha/2$
  - Look this number up as a probability in the standard normal table (meaning, try to find the number as close to this in the body of the table because recall that the numbers in the body of the table are probabilities whereas the numbers on the left and top are Z's)
  - Find the Z that this probability corresponds to
  - This Z is your confidence coefficient

Confidence Coefficient

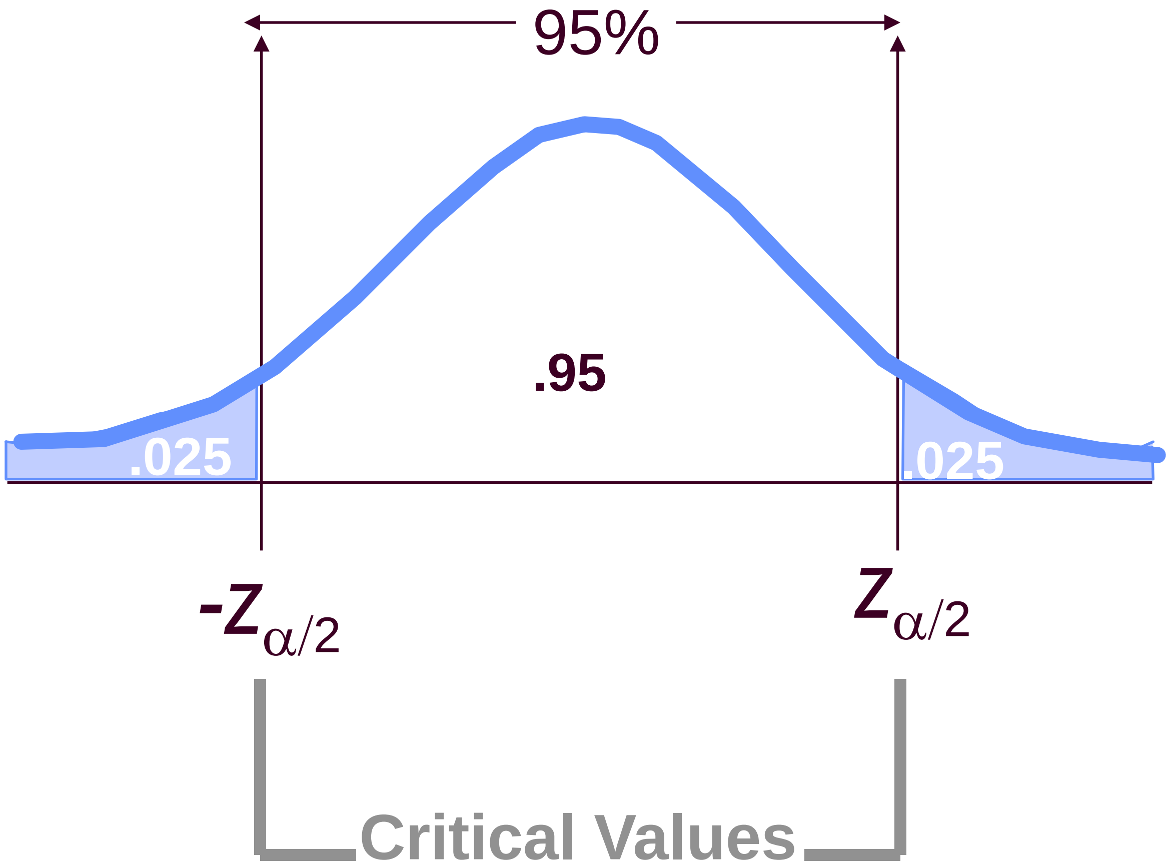
- Suppose you wanted to find the confidence coefficient for confidence level of 90% ($1 - \alpha = .90$)
  - Take $1 - \alpha/2 = 0.95$
  - Try to find the number in the body as close to 0.95 as you can
  - Note that you see a .9495 and a .9505 and these are as close to 0.95 as you can get (it doesn't matter which of these two you choose, but we will go with the .9495 number)
  - The .4495 number corresponds to a Z = 1.64 so the confidence coefficient for a 90% confidence interval is 1.64
  - If the confidence level is 95% then the confidence coefficient is 1.96
  - If the confidence level is 99% then the confidence coefficient is 2.57 (or 2.58)

The Critical Value $z_{\alpha/2}$



Found from standard normal table (corresponds to area of 1.0 – α/2)

Finding $z_{\alpha/2}$ for 95% Degree of Confidence



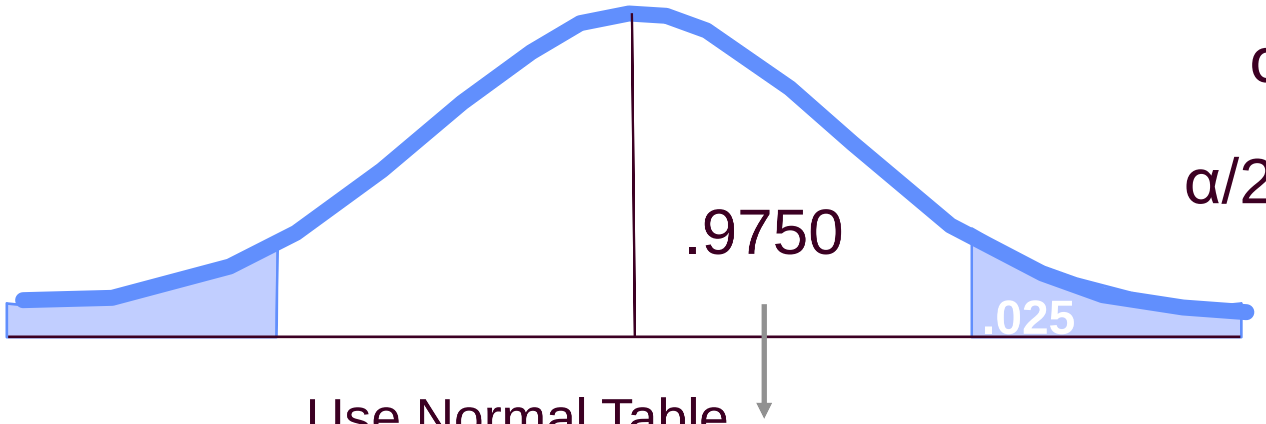
α = 5%

$\alpha/2$ = 2.5% = .025

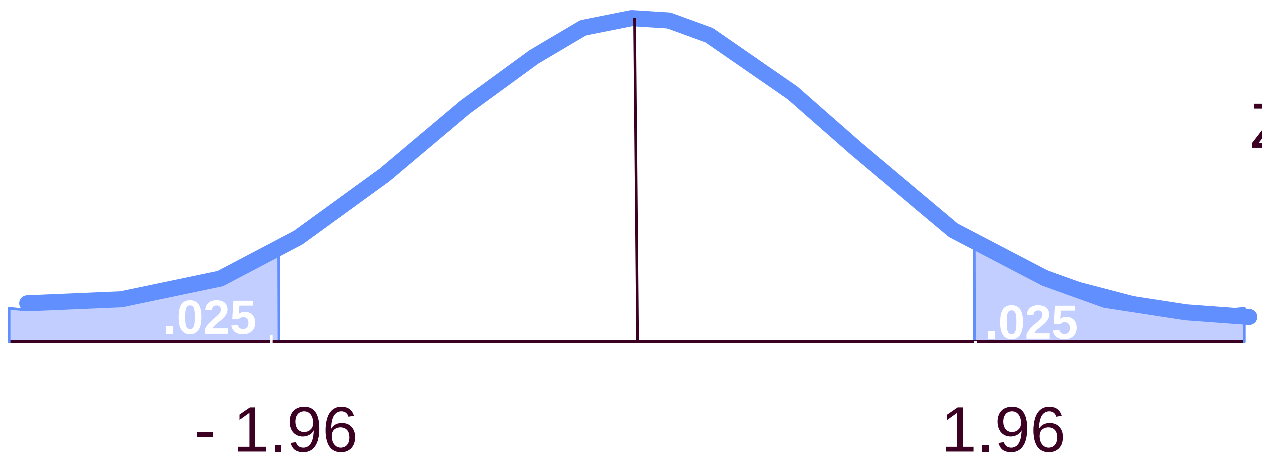
Finding $z_{\alpha/2}$ for 95% Degree of Confidence

$\alpha = 0.05$

$\alpha/2 = 0.025$



Use Normal Table
to find a z score of 1.96



$Z_{\alpha/2} = 1.96$

Example

- **Interval Estimate of the Population Mean: $n \geq 30$**

  Dell Publishing samples 106 shipments to estimate the mean postal cost. The sample mean is \$98.2 with a standard deviation of \$0.62. Calculate the margin of error E and the 95% confidence interval for the mean postal cost.

$n = 106$

$\bar{x} = 98.20$

$s = 0.62$

$\alpha = 0.05$

$\alpha/2 = 0.025$

$z_{\alpha/2} = 1.96$

$$\mathbf{E} = z_{\alpha/2} \bullet \frac{s}{\sqrt{n}} = 1.96 \bullet \frac{0.62}{\sqrt{106}} = 0.12$$

$$\bar{x} - E < \mu < \bar{x} + E$$

$$98.08 < \mu < 98.32$$

Interval Estimate of μ is 98.2 ± 0.12
or 98.08 to 98.32

Example

- **Interval Estimate of the Population Mean: $n \geq 30$**

  Based on the sample provided, the confidence interval for the population mean is $98.08 < \mu < 98.32$. If we were to select many different samples of the same size, 95% of the confidence intervals would actually contain the population mean μ.

Example

- Dell Publishing samples 48 shipments to estimate the mean postal cost. The sample mean is $25.36 with a standard deviation of $4.80. Calculate the 98% confidence interval for the mean postal cost.
  - Note that the sample size is > 30 so the Central Limit Theorem applies and the sample mean is normally distributed
  - $\overline{X} = 25.36$, $s = 4.80$, $n = 48$, $Z_{.98} = 2.33$
  - 98% Confidence interval for the mean postal cost is

$$\overline{X} - (Z_{1-\alpha/2})\left(\frac{s}{\sqrt{n}}\right), \quad \overline{X} + (Z_{1-\alpha/2})\left(\frac{s}{\sqrt{n}}\right)$$

$$25.36 - (2.33)\left(\frac{4.80}{\sqrt{48}}\right), \quad 25.36 + (2.33)\left(\frac{4.80}{\sqrt{48}}\right)$$

$$= 23.75, \quad 26.97$$

Small Sample Confidence Intervals

- The previous section on constructing confidence intervals is valid if the sample size is **> 30**
  - The Central Limit Theorem only applies when the sample size is > 30.
  - If the sample size is < 30 then the sample mean is not approximately normally distributed, but instead has a STUDENT-T distribution
    - The Student-t distribution looks like the normal distribution, but it has more area in the tails of the distribution
    - The ONLY difference when constructing confidence intervals for small samples versus constructing them for large samples is that for small samples, use a "t" number instead of a Z number.

Interval Estimation of a Population Mean: Small-Sample Case (n < 30) with σ Unknown

- **Interval Estimate**

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

where $1 - \alpha$ = the confidence coefficient

$t_{\alpha/2}$ = the t value providing an area of $\alpha/2$ in the upper tail of a t distribution with $n - 1$ degrees of freedom

s = the sample standard deviation

Z interval in MINITAB

- **Stat > Basic Statistics > 1-Sample Z**

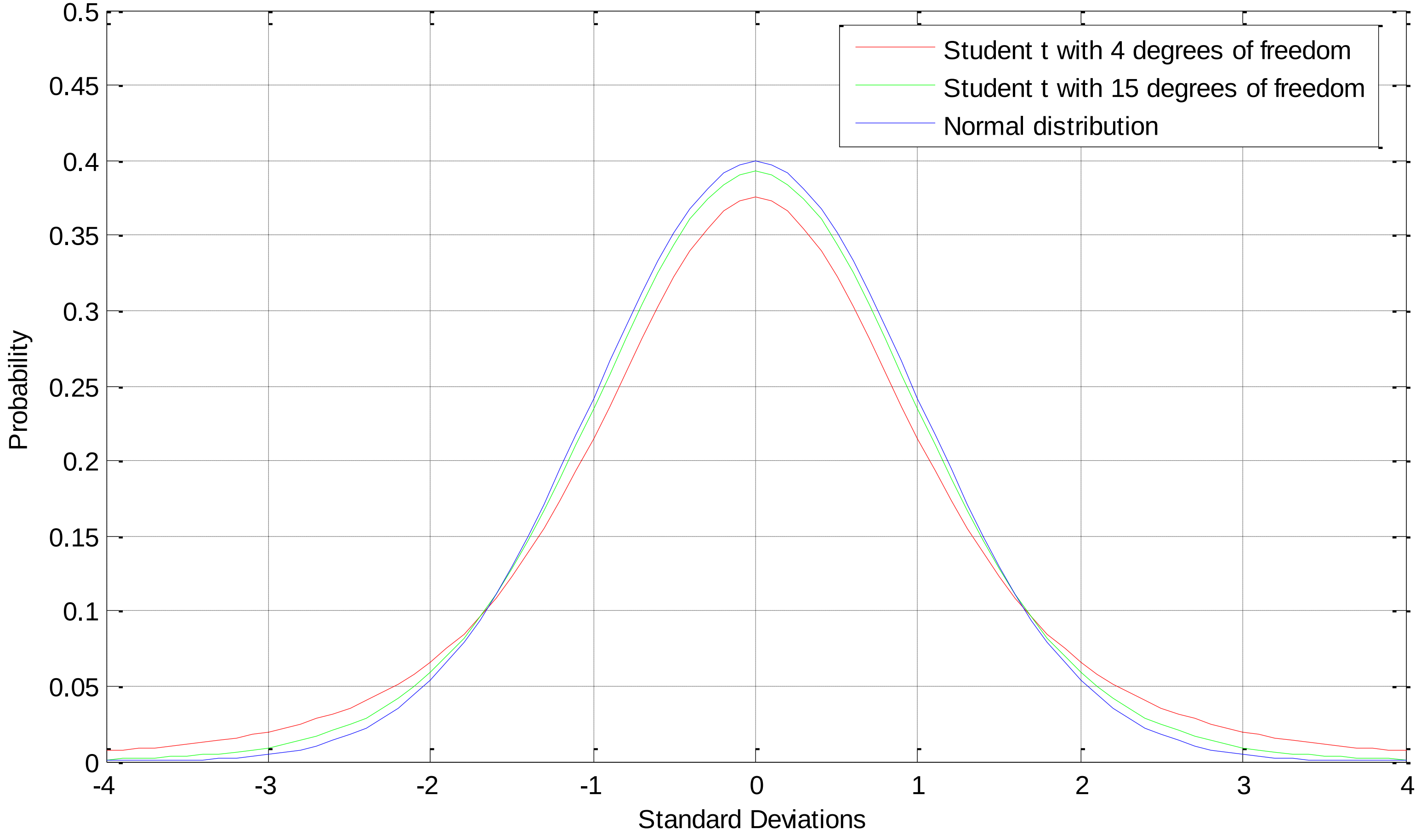
  Dialog box items

  - **Samples in columns:** Choose if you have entered raw data in columns. Enter the columns containing the sample data
  - **Summarized data:** Choose if you have summary values for the sample size and mean.
    - **Sample size:** Enter the value for the sample size
    - **Mean:** Enter the value for the sample mean.
    - **Standard deviation:** Enter the value for the population standard deviation.

t Distribution

- The t distribution is a family of similar probability distributions.
- A specific t distribution depends on a parameter known as the degrees of freedom.
- As the number of degrees of freedom increases, the difference between the t distribution and the standard normal probability distribution becomes smaller and smaller.
- A t distribution with more degrees of freedom has less dispersion.
- The mean of the t distribution is zero.

Comparing the Student-t distribution to the Normal distribution



Choosing a "t"

- Which t number should you use?
  - There is a different "t-distribution" for every "degree of freedom"
    - DEGREES OF FREEDOM = n-1
    - The numbers in the body of the t-distribution table are STANDARD DEVIATIONS – not probabilities
    - If your sample size is 28 then you have 28-1=27 degrees of freedom.
- As the degrees of freedom get larger, the t-distribution starts to look EXACTLY like the normal distribution

Choosing a "t"

- If you want to form a 90% confidence level (and you have 27 degrees of freedom) you want to choose the column that is headed by ".05" and the 27 degrees of freedom row
  - You should find a number of 1.703
    - t numbers are subscripted by the degrees of freedom and how much area is beyond a certain point
      - In this example, we would have $t_{27,.05} = 1.703$
        - » We will also start doing this with our Z numbers as well – subscripting the Z's with the amount of area BEYOND a certain point
        - » If we had a LARGE sample and wanted to form a 95% confidence interval, we would use a $Z_{.05} = 1.96$
    - Notice that the column headings are the areas in the TAILS of the t-distribution
    - Therefore, there is .05-"area in the upper tail" between 0 and a given number of standard deviations

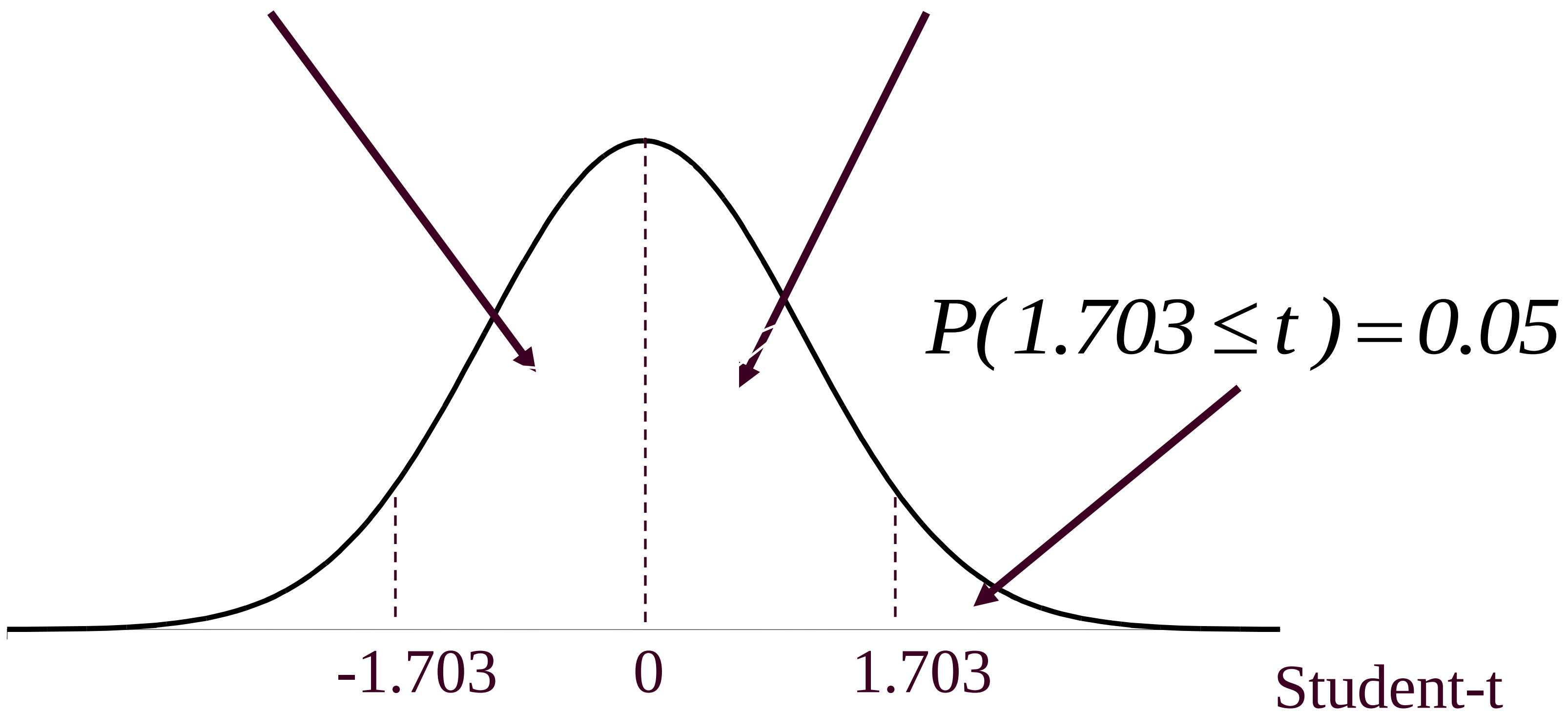
Choosing a "t"

- **Interpretation:** for a t-distribution with 27 degrees of freedom, you need to move 1.703 standard deviations from the mean to have .45 between 0 and 1.703
  - And therefore, .90 between –1.703 and 1.703
  - See the next slide for a graph
  - Note: If you have a small sample, you cannot form any general confidence intervals of any given confidence level – You can only look up confidence levels of 60%, 80%, 90%, 95%, 98%, and 99%

90% Confidence Interval with Small Samples

$$P(-1.703 \le t \le 0) = 0.45 \qquad P(0 \le t \le 1.703) = 0.45$$

$$P(1.703 \le t) = 0.05$$



Finding t-Numbers

- Suppose you have 21 observations and you want to form a 95% confidence interval
  - $t_{20,.025}$ = 2.086
- Suppose you have 30 observations and you want to form a 80% confidence interval
  - $t_{29,.10}$ 1.311
- Suppose you have 16 observations and you want to form a 99% confidence interval
  - $t_{15,.005}$ = 2.947

Example: Apartment Rents

- **Interval Estimation of a Population Mean: Small-Sample Case (n < 30) with σ Unknown**

A reporter for a student newspaper is writing an article on the cost of off-campus housing. A sample of 10 one-bedroom units within a half-mile of campus resulted in a sample mean of $350 per month and a sample deviation of $30.

Let us provide a 95% confidence interval estimate of the mean rent per month for the population of one-bedroom units within a half-mile of campus. We'll assume this population to be normally distributed.

Example: Apartment Rents

- **t Value**

At 95% confidence, $1 - \alpha = .95$, $\alpha = .05$, and $\alpha/2 = .025$.

$t_{.025}$ is based on $n - 1 = 10 - 1 = 9$ degrees of freedom.

In the t distribution table we see that $t_{.025} = 2.262$.

Degrees of Freedom	Area in Upper Tail .10	.05	.025	.01	.005
.	.	.	.	.	.
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
.	.	.	.	.	.

Example: Apartment Rents

- **Interval Estimation of a Population Mean: Small-Sample Case (n < 30) with σ Unknown**

$$\bar{x} \pm t_{.025} \frac{s}{\sqrt{n}}$$

$$350 \pm 2.262 \frac{30}{\sqrt{10}}$$

$$350 \pm 21.46$$

or $328.54 to $371.46

We are 95% confident that the mean rent per month for the population of one-bedroom units within a half-mile of campus is between $328.54 and $371.46.

Interval Estimation of a Population Mean: Small-Sample Case ($n < 30$)

- **Population is Not Normally Distributed**

  Use nonparametrics or increase the sample size to $n > 30$ and use the large-sample interval-estimation procedures.

- **Population is Normally Distributed and σ is Known**

  The large-sample interval-estimation procedure can be used.

- **Population is Normally Distributed and σ is Unknown**

  The appropriate interval estimate is based on a probability distribution known as the t distribution.

Choosing Between Normal and t Distributions

START

Is n > 30?

YES

By the Central Limit Theorem, we can use the normal distribution with:

$$E = z_{\alpha/2} \frac{s}{\sqrt{n}}$$

NO

Does the population have a normal distribution?

NO

You must use nonparametric or bootstrap methods

YES

Use the t distribution with:

$$E = t_{\alpha/2} \frac{s}{\sqrt{n}}$$

and n-1 degrees of freedom.

Sample Size for an Interval Estimate of a Population Mean

- Let E = the maximum sampling error
- E is the amount added to and subtracted from the point estimate to obtain an interval estimate.
- E is often referred to as the margin of error.
- We have

$$E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

- Solving for n we have

$$n = \frac{(z_{\alpha/2})^2 \sigma^2}{E^2}$$

Example:

- **Sample Size for an Interval Estimate of a Population Mean**

    As a restaurant owner, you need to decide how much food to prepare each night. In a sample of 100 nights, the mean number of customers is 85 with a standard deviation of 38

    How large must your sample be if you want to be 99% sure that your sample error is no larger than 4?

Example:

- **Sample Size for Interval Estimate of a Population Mean**

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 4$$

At 99% confidence, $z_{.005}$ = 2.57. Recall that s = 38.

Solving for n we have

$$n = \frac{(2.57)^2(38.0)^2}{(4)^2} = 569$$

Interpretation: You have to take a sample of 569 in order for you to be 99% sure that your maximum margin of error is no larger than 4 customers.

We need to sample 569 to reach a desired precision of ± 4 at 99% confidence.

t interval in Minitab

- Ask for t interval by Stat → Basic Statistics → 1 - α Sample t….
- Select desired variable.
- Specify desired confidence level.
- Say OK.

t interval in MINITAB

- **Stat > Basic Statistics > 1-Sample t**

  Dialog box items

- **Samples in columns:** Choose if you have entered raw data in columns. Enter the columns containing the sample data
- **Summarized data:** Choose if you have summary values for the sample size, mean, and standard deviation.
  - **Sample size:** Enter the value for the sample size
  - **Mean:** Enter the value for the sample mean.
  - **Standard deviation:** Enter the value for the population standard deviation.