JOHNS HOPKINS
CAREY
BUSINESS SCHOOL

Statistical Analysis

BU.510.610

Confidence Intervals

1. z-Based Confidence Intervals for a Population Mean: σ Known
2. t-Based Confidence Intervals for a Population Mean: σ Unknown
3. Sample Size Determination
4. Confidence Intervals for a Population Proportion
5. Confidence Intervals for Parameters of Finite Populations

z-Based Confidence Intervals for a Mean: σ Known

- Confidence interval for a population mean is an interval constructed around the sample mean so we are reasonable sure that it contains the population mean
- Any confidence interval is based on a confidence level

The Car Mileage Case

- Automaker conducted mileage tests on n=50 cars
- Sample mean is 31.56

  This is a point estimate of the population mean
- Do not know how good this estimate is
- Will use a confidence interval

The Car Mileage Case *continued*

- There were many samples of 50 cars

    Each would give different means

- Consider the probability distribution of all the sample means

    Called the sampling distribution

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

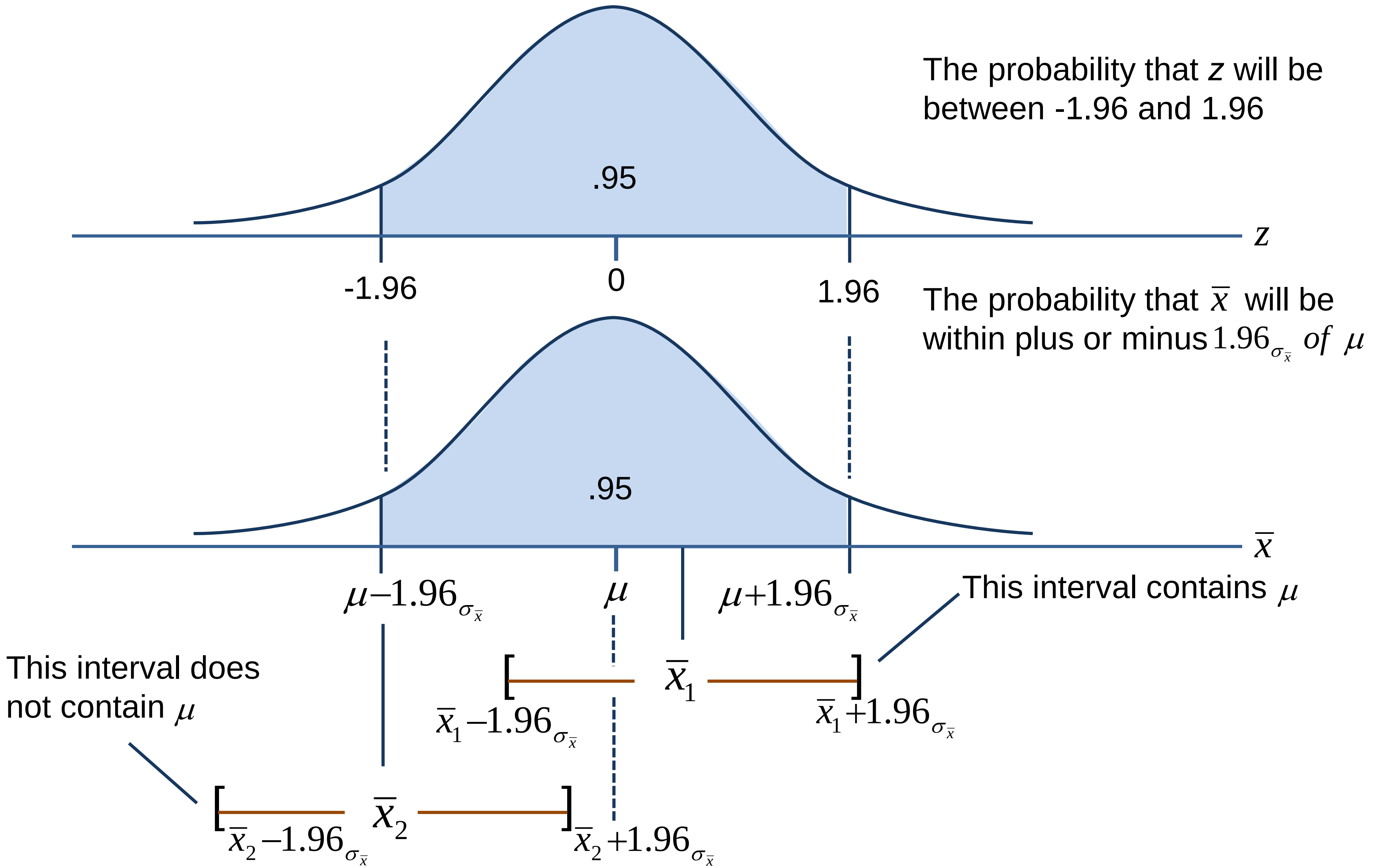
$$\sigma_{\bar{x}} = \sigma / \sqrt{n}$$

The Car Mileage

1. Because the sampling distribution of the sample mean is a normal distribution, we can use the normal distribution to compute probabilities about the sample mean
2. Saying $\bar{x}$ is within $\pm 1.96\sigma_{\bar{x}}$ *of* μ is the same as saying the interval $[\bar{x} \pm 1.96\sigma_{\bar{x}}]$ *contains* μ
3. The 95 percent confidence interval is

$$\left[\bar{x} \pm 1.96\sigma_{\bar{x}}\right] = \left[\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}\right]$$

A Confidence Interval for the Population Mean



Generalizing

- In the example, we found the probability that μ is contained in an interval of integer multiples of $\sigma_{\bar{x}}$
- Usually we specify the (integer) probability and find the corresponding number of $\sigma_{\bar{x}}$
- The probability that the confidence interval will ***not*** contain the population mean μ is denoted by α

Generalizing *continued*

The probability that the confidence interval will contain the population mean μ is denoted by $1-\alpha$

- $1-\alpha$ is referred to as the confidence coefficient
- $(1-\alpha)\times 100\%$is called the confidence level

Two decimal point probabilities for $1-\alpha$ are used

- Here, focus on $1-\alpha = 0.95 \; or \; 0.99$

General Confidence Interval

In general, the probability is $1-\alpha$ that the population mean μ is contained in the interval

$$\left[\bar{x} \pm z_{\alpha/2}\sigma_{\bar{x}}\right] = \left[\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

- The normal point $z_{\alpha/2}$ gives a right hand tail area under the standard normal curve equal to $\alpha/2$
- The normal point $-z_{\alpha/2}$ gives a left hand tail area under the standard normal curve equal to $\alpha/2$
- The area under the standard normal curve between $-z_{\alpha/2}$ and $z_{\alpha/2}$ is $1-\alpha$

General Confidence Interval *continued*

If a population has standard deviation σ (known),

and if the population is normal or if sample size is large ($n \geq 30$), then …

… a **(1-α)100% confidence interval for μ** is

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = \left[\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

95% Confidence Level

For a 95% confidence level, $1 - \alpha = 0.95$, so $\alpha = 0.05$, and $\alpha/2 = 0.025$

Need the normal point $z_{0.025}$

- The area under the standard normal curve between $-z_{0.025}$ and $z_{0.025}$ is 0.95
- Then the area under the standard normal curve between 0 and $z_{0.025}$ is 0.475
- From the standard normal table, the area is 0.475 for $z = 1.96$
- Then $z_{0.025} = 1.96$

95% Confidence Interval

The 95% confidence interval is

$$\left[\bar{x} \pm z_{0.025}\sigma_{\bar{x}}\right] = \left[\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}\right]$$

$$= \left[\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right]$$

99% Confidence Interval

For 99% confidence, need the normal point $z_{0.005}$

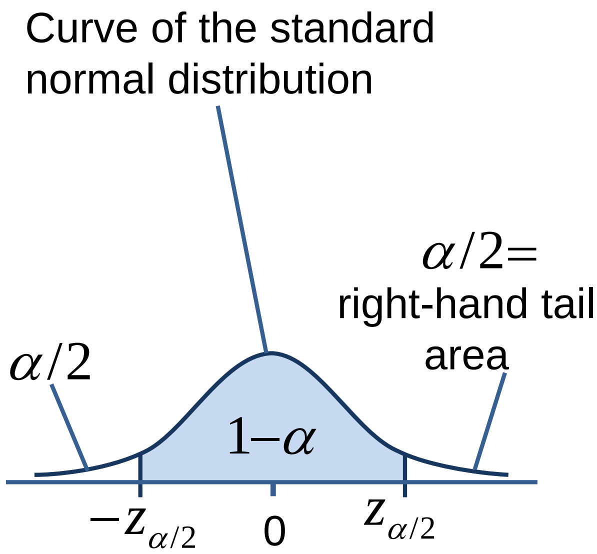
- Reading between table entries in the standard normal table, the area is 0.495 for $z_{0.005} = 2.575$

The 99% confidence interval is

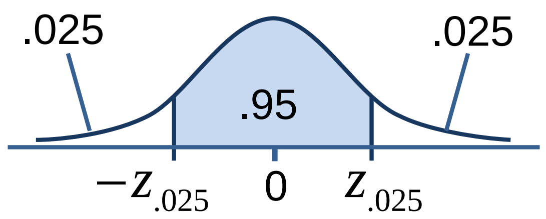
$$\left[\bar{x} \pm z_{0.025}\sigma_{\bar{x}}\right] = \left[\bar{x} \pm 2.575\frac{\sigma}{\sqrt{n}}\right]$$

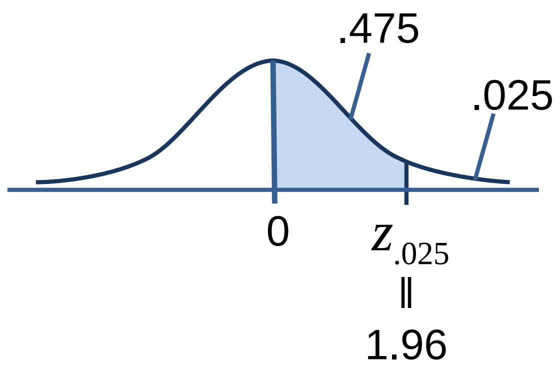
$$= \left[\bar{x} - 2.575\frac{\sigma}{\sqrt{n}}, \bar{x} + 2.575\frac{\sigma}{\sqrt{n}}\right]$$

The Effect of a on Confidence Interval Width

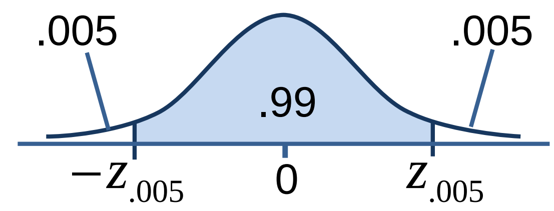


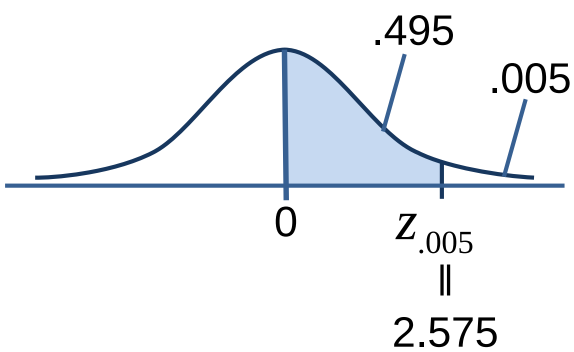
$$z_{\alpha/2} = z_{0.025} = 1.96$$





$$z_{\alpha/2} = z_{0.025} = 2.575$$





t-Based Confidence Intervals for a Mean: σ Unknown

- If σ is unknown (which is usually the case), we can construct a confidence interval for μ based on the sampling distribution of

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

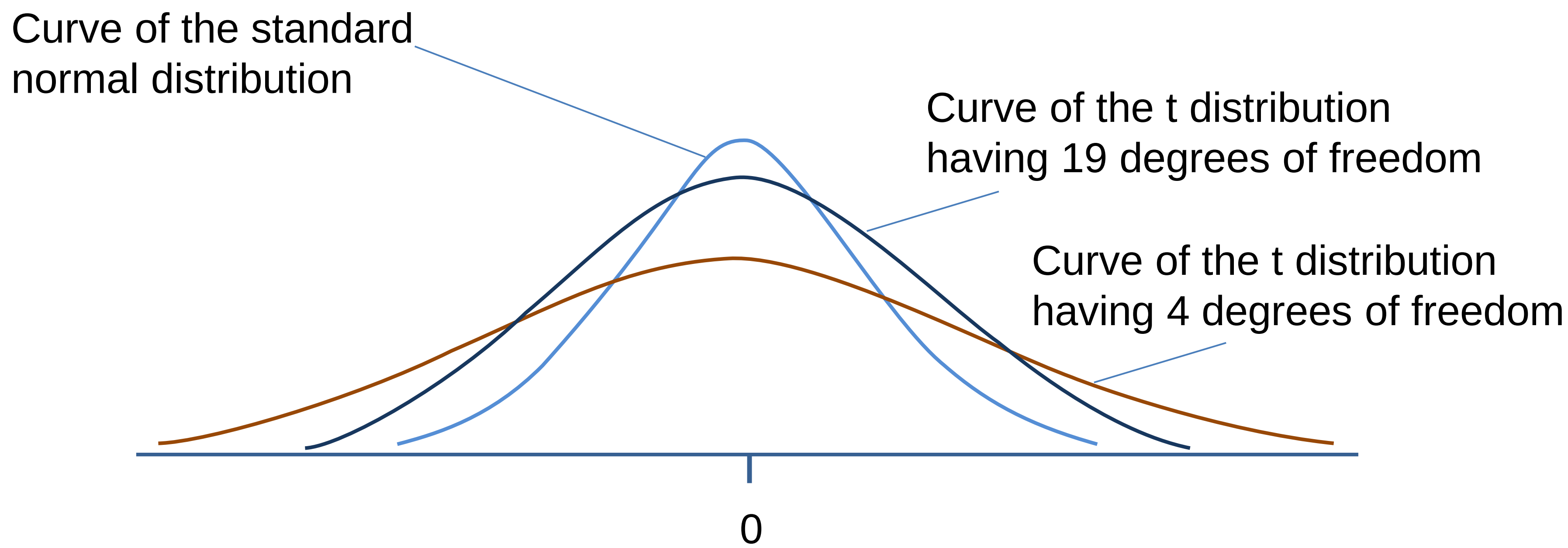
- If the population is normal, then for any sample size *n*, this sampling distribution is called the ***t* distribution**

The *t* Distribution

The curve of the *t* distribution is similar to that of the standard normal curve

- Symmetrical and bell-shaped
- The *t* distribution is more spread out than the standard normal distribution
- The spread of the *t* is given by the **number of degrees of freedom**
    a. Denoted by *df*
    b. For a sample of size n, there are one fewer degrees of freedom, that is, df = n – 1

Degrees of Freedom and the t-Distribution



As the number of degrees of freedom increases, the spread of the *t* distribution decreases and the *t* curve approaches the standard normal curve

The t Distribution and Degrees of Freedom

- As the sample size n increases, the degrees of freedom also increases

- As the degrees of freedom increase, the spread of the t curve decreases

- As the degrees of freedom increases indefinitely, the t curve approaches the standard normal curve

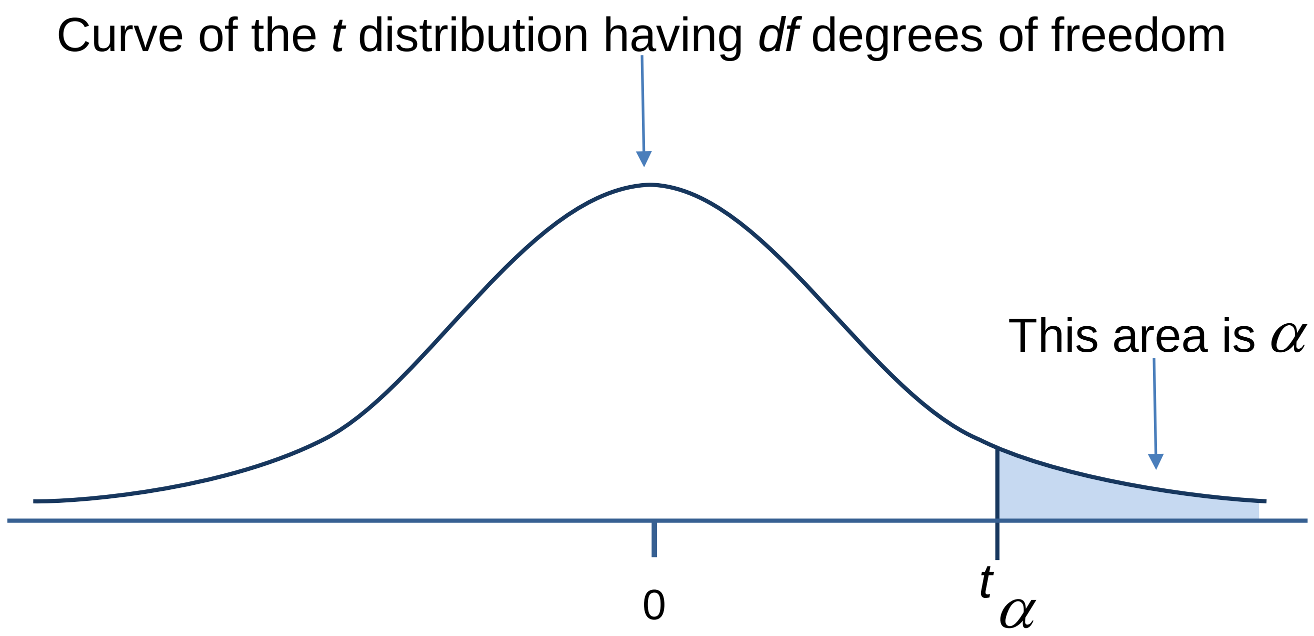
  If $n \geq 30$, so $df = n - 1 \geq 29$, the t curve is very similar to the standard normal curve

t and Right Hand Tail Areas

Use a t point denoted by t_α

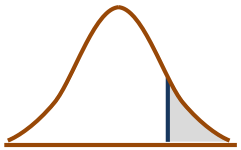
1. t_α is the point on the horizontal axis under the t curve that gives a right hand tail equal to α
2. So the value of t_α in a particular situation depends on the right hand tail area and the number of degrees of freedom
   - df = n – 1
   - α in one tail

t and Right Hand Tail Areas



Using the t Distribution Table

t table with right tail probabilities



t(p.df)

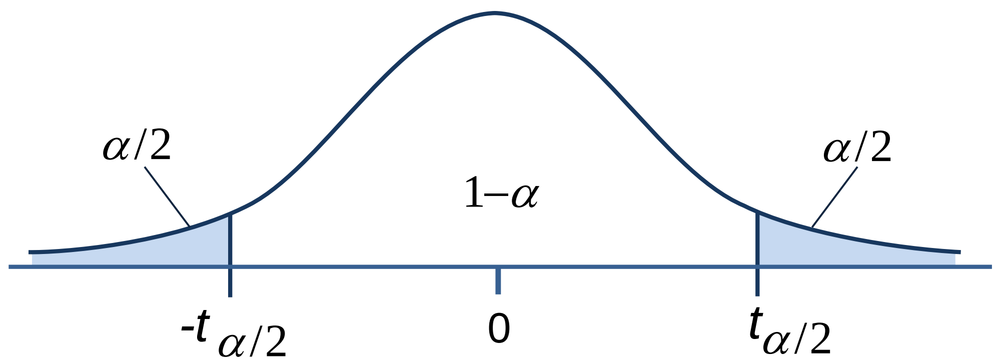
df\p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091	12.9240
4	0.270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6.8688
6	0.264835	0.717558	1.439756	1.943180	2.44691	3.14267	3.70743	5.9588
7	0.263167	0.711142	1.414924	1.894579	2.36462	2.99795	3.49948	5.4079
8	0.261921	0.706387	1.396815	1.859548	2.30600	2.89646	3.35539	5.0413
9	0.260955	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984	4.7809
10	0.260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869

t-Based Confidence Intervals for a Mean: σ Unknown

- If the sampled population is normally distributed with mean μ, then a $(1\alpha)100\%$ confidence interval for μ is

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

Curve of the t distribution having *n-1* degrees of freedom



- $t_{\alpha/2}$ is the t point giving a right-hand tail area of $\alpha/2$ under the t curve having n1 degrees of freedom

Confidence Intervals for a Population Proportion

If the sample size n is large, then a (1α)100% confidence interval for p is

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}$$

Here, n should be considered large if both

- $n \cdot \hat{p} \geq 5$
- $n \cdot (1\text{-}\hat{p}) \geq 5$

Confidence Intervals for Parameters of Finite Populations

- For a large (n ≥ 30) random sample of measurements selected without replacement from a population of size N, a (1- α)100% confidence interval for μ is

$$\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

- A (1- α)100% confidence interval for the population total is found by multiplying the lower and upper limits of the corresponding interval for μ by N

Confidence Intervals for Proportion and Total for a Finite Population

- For a large random sample of measurements selected without replacement from a population of size N, a $(1-\alpha)$ 100% confidence interval for p is

$$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}\left(\frac{N-n}{N-1}\right)}$$

- A $(1-\alpha)$100% confidence interval for the total number of units in a category is found by multiplying the lower and upper limits of the corresponding interval for p by N

A Comparison of Confidence Intervals and Tolerance Intervals

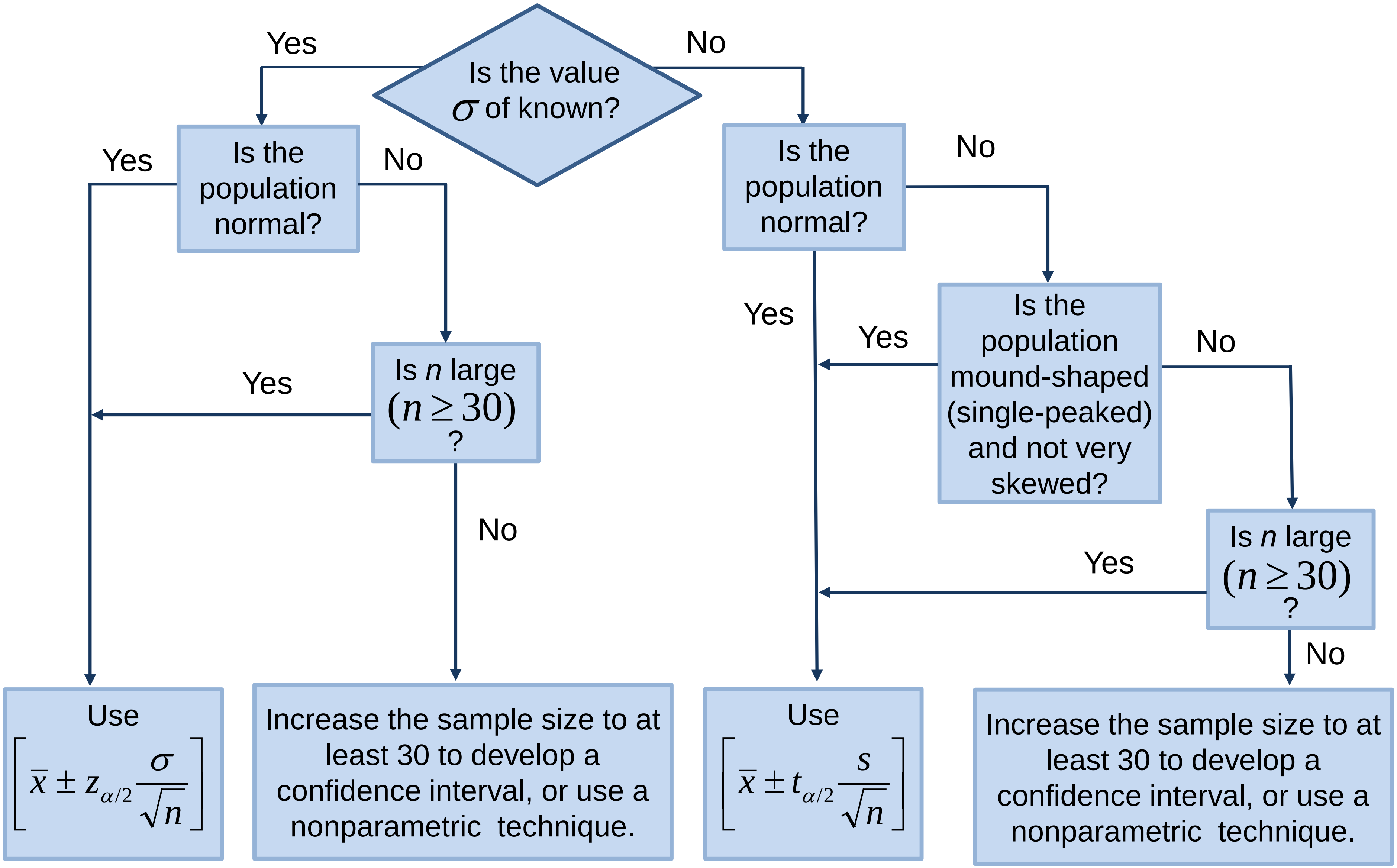
Tolerance interval: contains specified percentage of individual population measurements

- Often 68.26%, 95.44%, 99.73%

Confidence interval: interval containing the population mean μ, and the confidence level expresses how sure we are that this interval contains μ

- Often level is set high (e.g., 95% or 99%)
- Such a level is considered high enough to provide convincing evidence about the value of μ

Selecting an Appropriate Confidence Interval for a Population Mean



Sample Testing *example*

A sample of 100 Johns Hopkins Carey Business School take the standard IQ test. The sample mean score is 108. Assuming the IQ is a normal distribution and the population standard deviation, σ is 15, determine a 95% confidence interval?

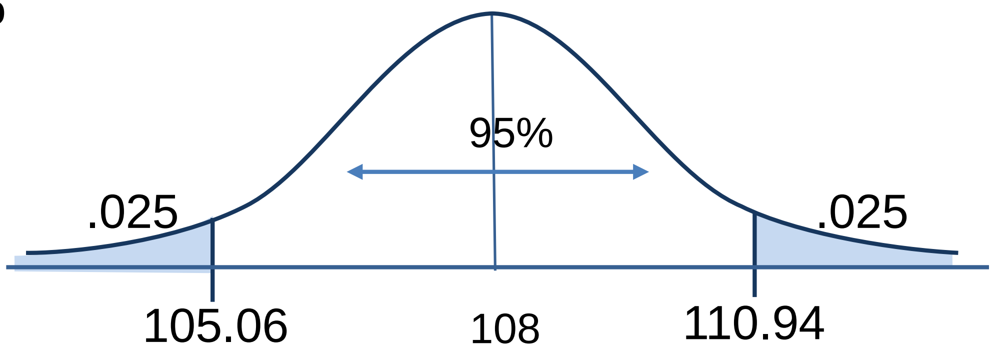
$$\bar{x}=108, \sigma=15, \mathrm{n}=100, \mathrm{CI}=95\%$$

$$\mathrm{CI} = \bar{x} \pm z\sigma_{\bar{\mathrm{x}}}$$

$$\mathrm{CI} = 108 \pm 1.96\,(15/\sqrt{100})$$

$$\mathrm{CI} = 108 \pm 2.94$$

$$105.06 \le \mu \le 110.94$$

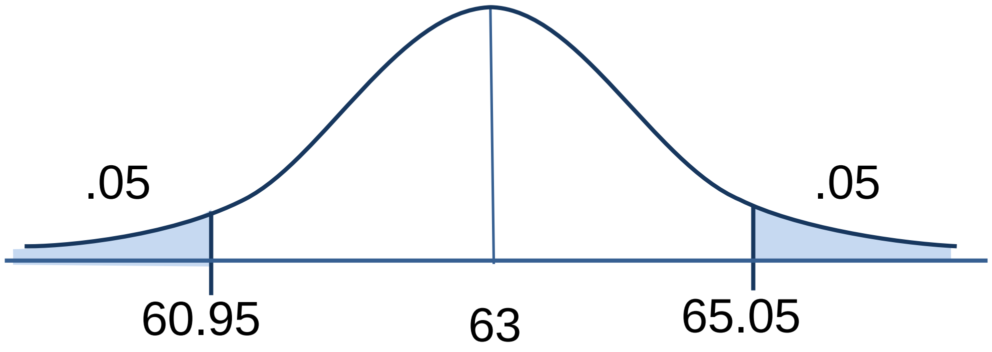


Sample Testing *example*

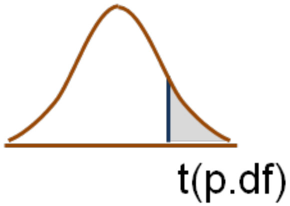
A sample of 25 technicians are used to measure daily production output for one week. The average output was 63 with a sample standard deviation s = 6. Determine the 90% confidence interval.

$$n = 25$$
$$\bar{x} = 63$$
$$s = 6$$
$$CI = 90\%$$

$$CI = \bar{x} \pm t \, {}^{s}\!/\!{}_{\sqrt{n}}$$
$$CI = 63 \pm 1.711 ({}^{6}\!/\!{}_{\sqrt{25}})$$
$$CI = 63 \pm 2.05$$
$$60.95 \le \mu \le 65.05$$



t table with right tail probabilities



Sample Testing *example*

A sample of 200 customers were asked if they would purchase the new personal electronic device. 160 indicated a positive response. Determine the 88% Confidence Interval.

$$n = 200$$

$$r = 160$$

$$CI = 88\%$$

$$CI = \hat{p} \pm z\,\sigma_{\hat{p}}$$

$$\hat{p} = \frac{160}{200} = .80$$

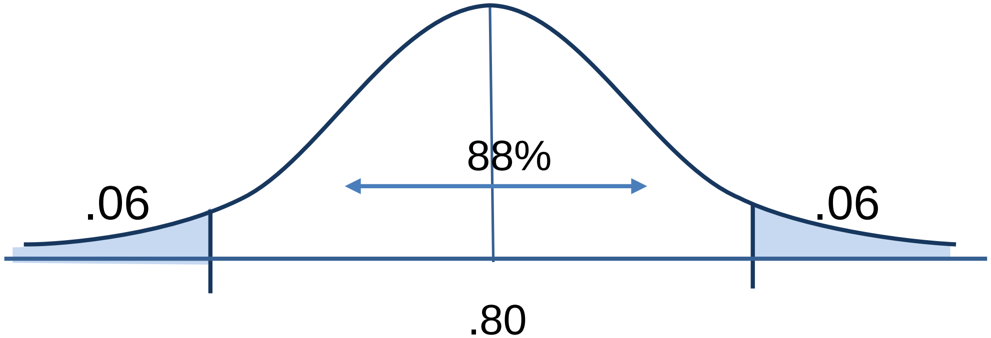
$$\sigma_{\hat{p}} = \sqrt{\frac{(\hat{p})(1-\hat{p})}{n}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{(.80)(.20)}{200}} = .028$$

$$CI = .80 \pm 1.56(.028)$$

$$.80 \pm .0437$$

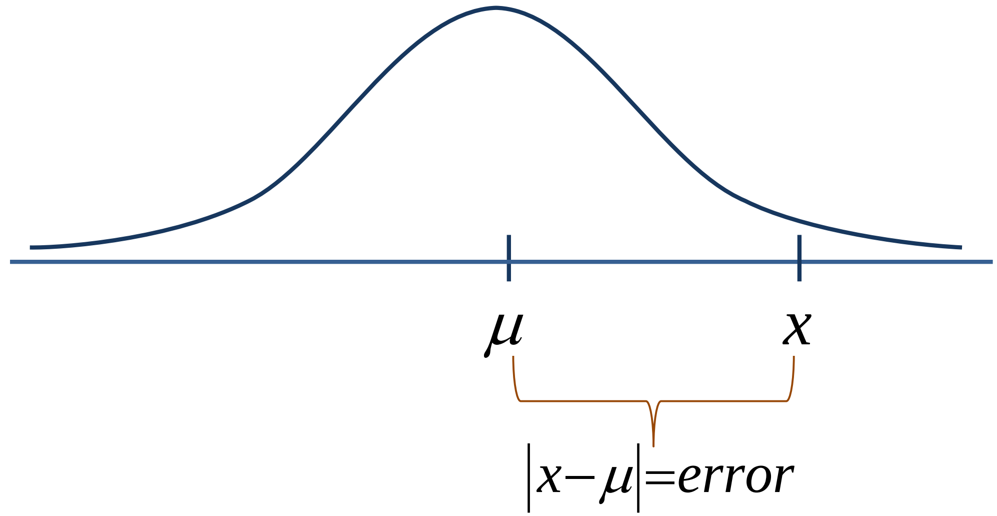
$$.7563 \le \pi \le 84.37$$



Sample Testing *example*

Calculating sample size n.

$$z=\frac{x-\mu}{\sigma/\sqrt{n}}$$



$$\frac{z\sigma}{\sqrt{n}}=x-\mu$$

$$\frac{z^z\sigma^2}{n}=(x-\mu)^2$$

$$n=\frac{z^2\sigma^2}{(x-\mu)^2}$$

Sample Testing *example*

You are assigned the task to determine the gas mileage for the new Chinese auto. How large should your sample be so that the error is 2 miles. You want a 99% CI. The σ is 8.5

$$N = \frac{z^2 \sigma^2}{(x-\mu)^2} = \frac{(2.56)^2 (8.5)^2}{2^2} = 118.37$$