# Lecture 6
# Sampling Distributions and Confidence Intervals

# Topics

- The Central Limit Theorem.
- Large sample confidence intervals for μ.
- Small sample confidence intervals for μ.

# Section 6.5

# Distribution of the Sample Mean: The Central Limit Theorem

# Definitions

- A *point estimate* is a single number calculated from sample data. It is used to estimate a parameter of the population.
- A *point estimator* is the formula or rule that is used to calculate the point estimate for a particular set of data.

# Definitions (cont.)

- The probability distribution of a point estimator or sample statistic is called a *sampling distribution*.
- The *standard error* is the standard deviation of the sampling distribution of a point estimator.

# Goals

- To introduce point estimators.
- To introduce sampling distributions for point estimators.
- To use the Central Limit Theorem to compute sampling distributions of the sample mean for large samples.

# Example – Home Sales

- The population of interest consists of all houses that are for sale in this area.
- The variable of interest is the number of days the house sat on the market until sold.
- The parameter of interest is:

  μ = mean number of days on the market

# Example – Home Sales

- Based on a sample of five homes that sold recently consider the following point estimator for μ:

$$\bar{X} = \frac{1}{5}\sum_{i=1}^{5} x_i$$

- Suppose that a sample of five homes produced the following data: 76, 48, 42, 9 and 27 days.

# Example – Home Sales

- Then the point estimate for the mean number of days on the market would be:

$$\bar{x} = \frac{1}{5}(76 + 48 + 42 + 9 + 27) = 40.4$$

- Suppose a second sample of five homes produced the following data: 140, 10, 75, 24 and 4 days.

# Example – Home Sales

- Then the point estimate for the mean number of days on the market would be:

$$\bar{x} = \frac{1}{5}(140 + 10 + 75 + 24 + 4) = 50.6$$

- Same point estimator but different point estimates.
- So the point estimator, $\bar{X}$, is a random variable with a probability distribution, a mean and a variance.

# Central Limit Theorem (CLT)

- In random sampling from a population with mean μ and standard deviation σ, when *n* is large enough (≥ 30), the distribution of $\bar{X}$ is:
  - approximately normal
  - $\mu_{\bar{X}}$ = mean of the sampling distribution of $\bar{X} = \mu$
  - $\sigma_{\bar{X}}$ = standard error of $\bar{X} = \dfrac{\sigma}{\sqrt{n}}$

# Special Exception

- If the population we are sampling from has a normal distribution (or at is least approximately normal) then the CLT applies for all sample sizes.

# Section 6.6

# The Central Limit Theorem – A More Detailed Look

# Goal

- A better appreciation of the CLT.

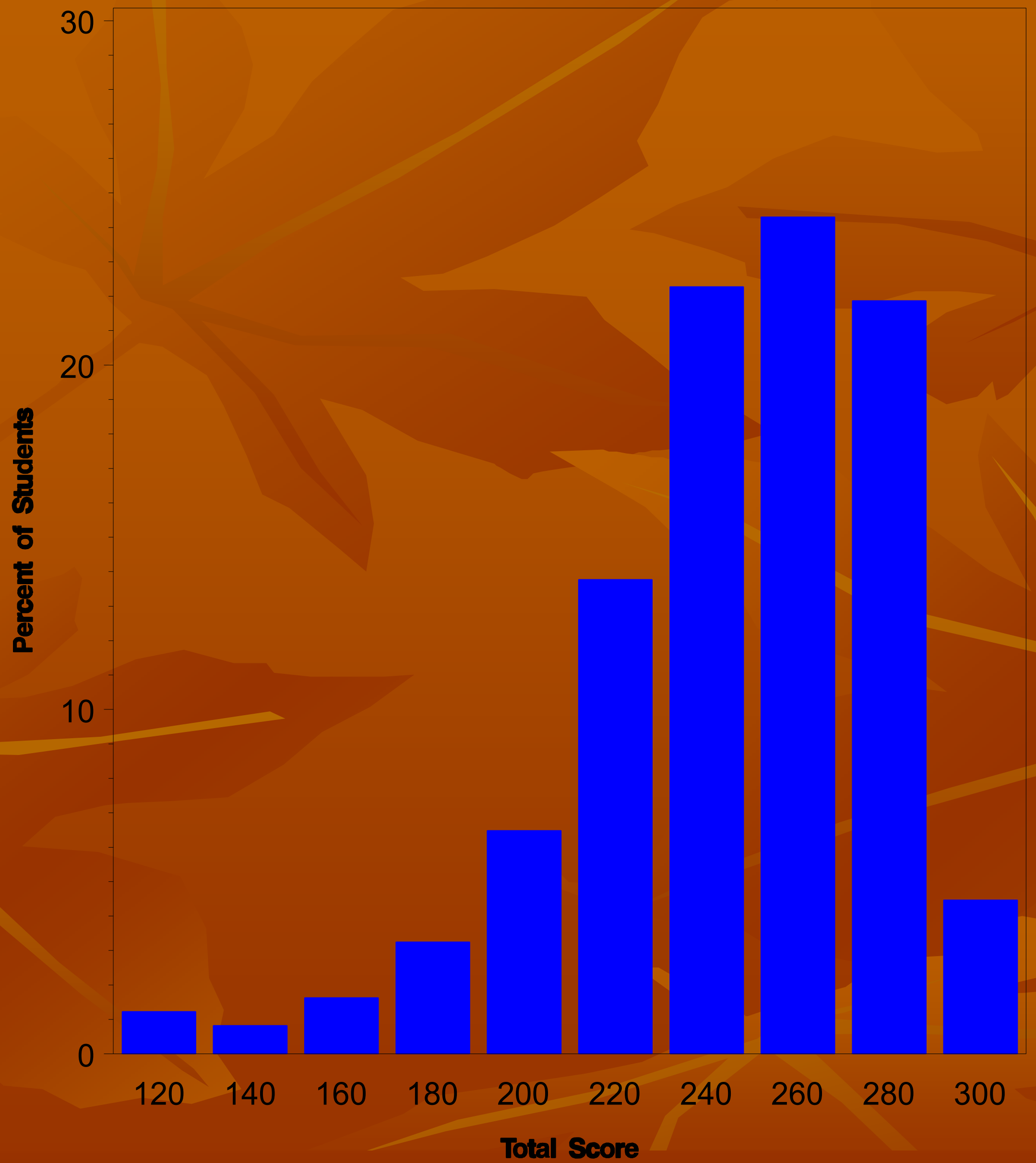
# Shape of the Sampling Distribution of $\bar{X}$

- No matter what the shape of the distribution of the population, if $n \geq 30$, the shape of the distribution of $\bar{X}$ for repeated sampling is normal.

# Example – Total Scores

- Suppose the population of interest is the 247 total scores to date (Exams 1-2, Quizzes 1-4 and Minute Papers 1-3).

# Total Score

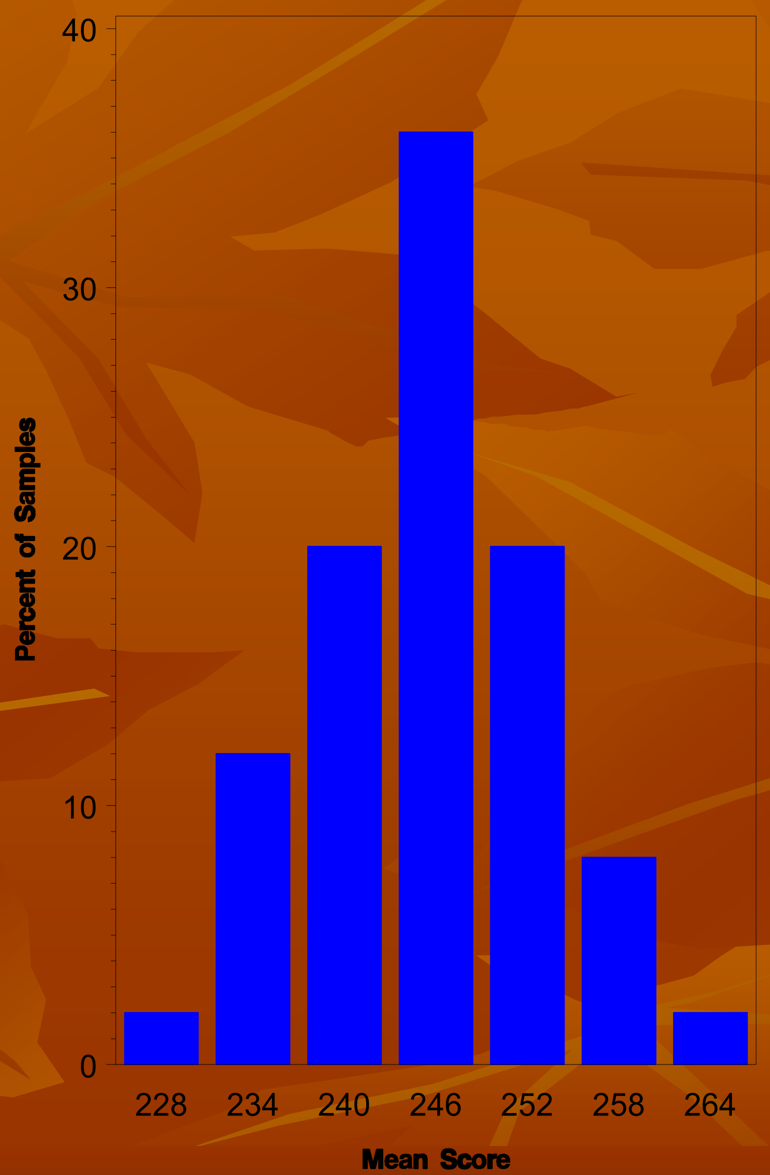


30
20
10
0
Percent of Students
120
140
160
180
200
220
240
260
280
300
Total Score

# Example – Total Scores

- Clearly the distribution of total scores is not normally distributed.
- Suppose you take a sample of size $n = 30$ scores and compute the average total score.
- Repeat this process for a total of 50 times.
- This would yield 50 values for $\bar{X}$.

# Mean Score – n = 30



# Mean of Sampling Distribution $\bar{X}$ of

- If $n \geq 30$, then the mean of the sampling distribution of $\bar{X}$ is the mean of the population.

# Example – Total Scores

- The mean total score for the 247 students in the population was

  $\mu = 245.7$

- The mean of the sample mean total scores for the 50 samples of size $n = 30$ was

  $\mu_{\bar{x}} = 245.6$

# Standard Error of the Sampling Distribution of $\bar{X}$

- If $n \geq 30$, the standard error of $\bar{X}$ is the standard deviation of the population divided by the square root of $n$.

# Example – Total Scores

- The standard deviation of the total scores for the 247 students in the population was

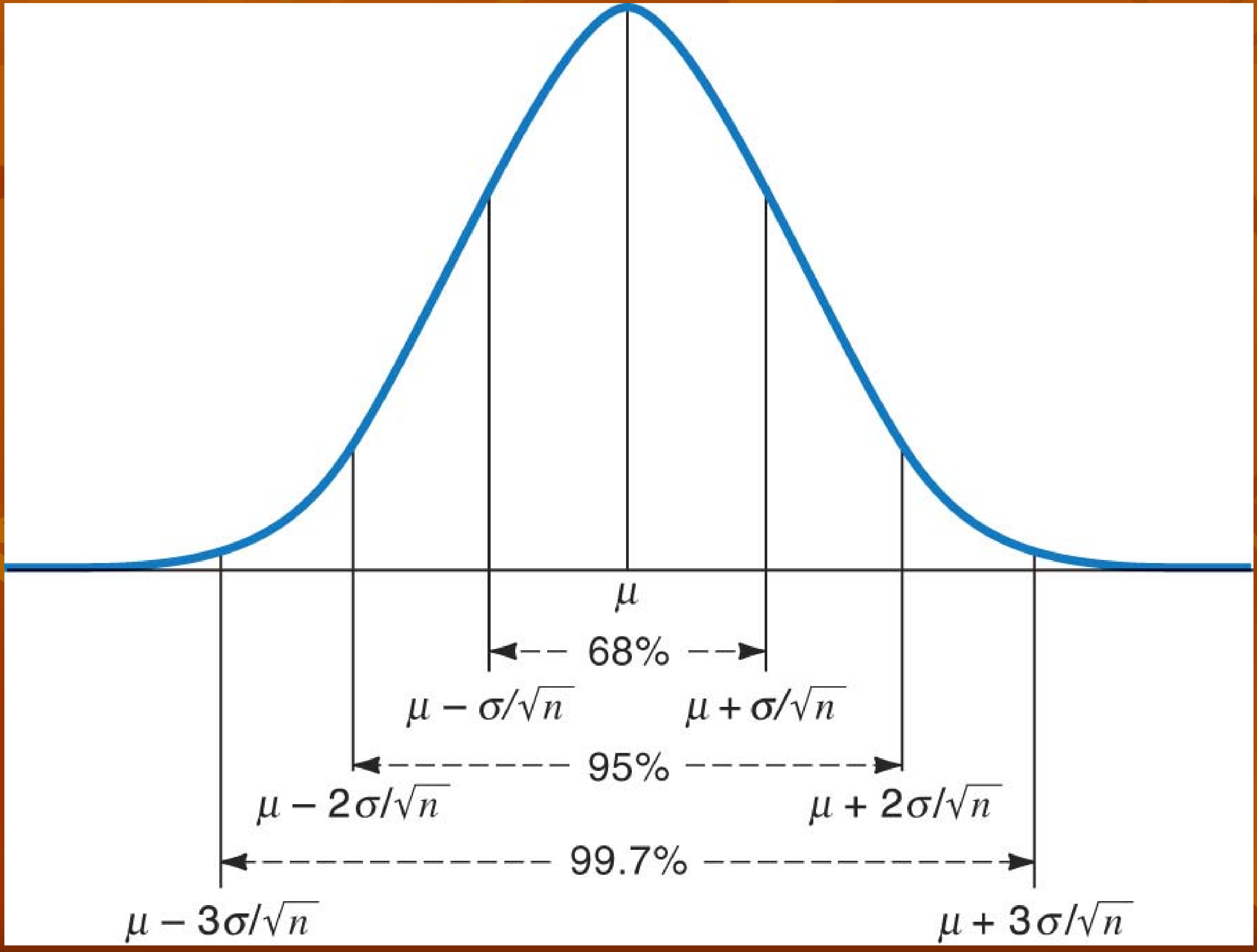
  $\sigma = 34.8$

- The standard deviation of the sample total mean scores for the 50 samples was

  $\sigma_{\bar{X}} = 7.4$

- $\dfrac{34.8}{\sqrt{30}} = 6.4 \cong 7.4$

# Summary of CLT



# Special Bonus

- As $n$ increases, the standard error of $\bar{X}$ decreases.
- As $n$ increases, $\bar{X}$ becomes a better (more precise) point estimator of μ.
- More is better.

# Section 6.7

# Drawing Inferences by Using the Central Limit Theorem

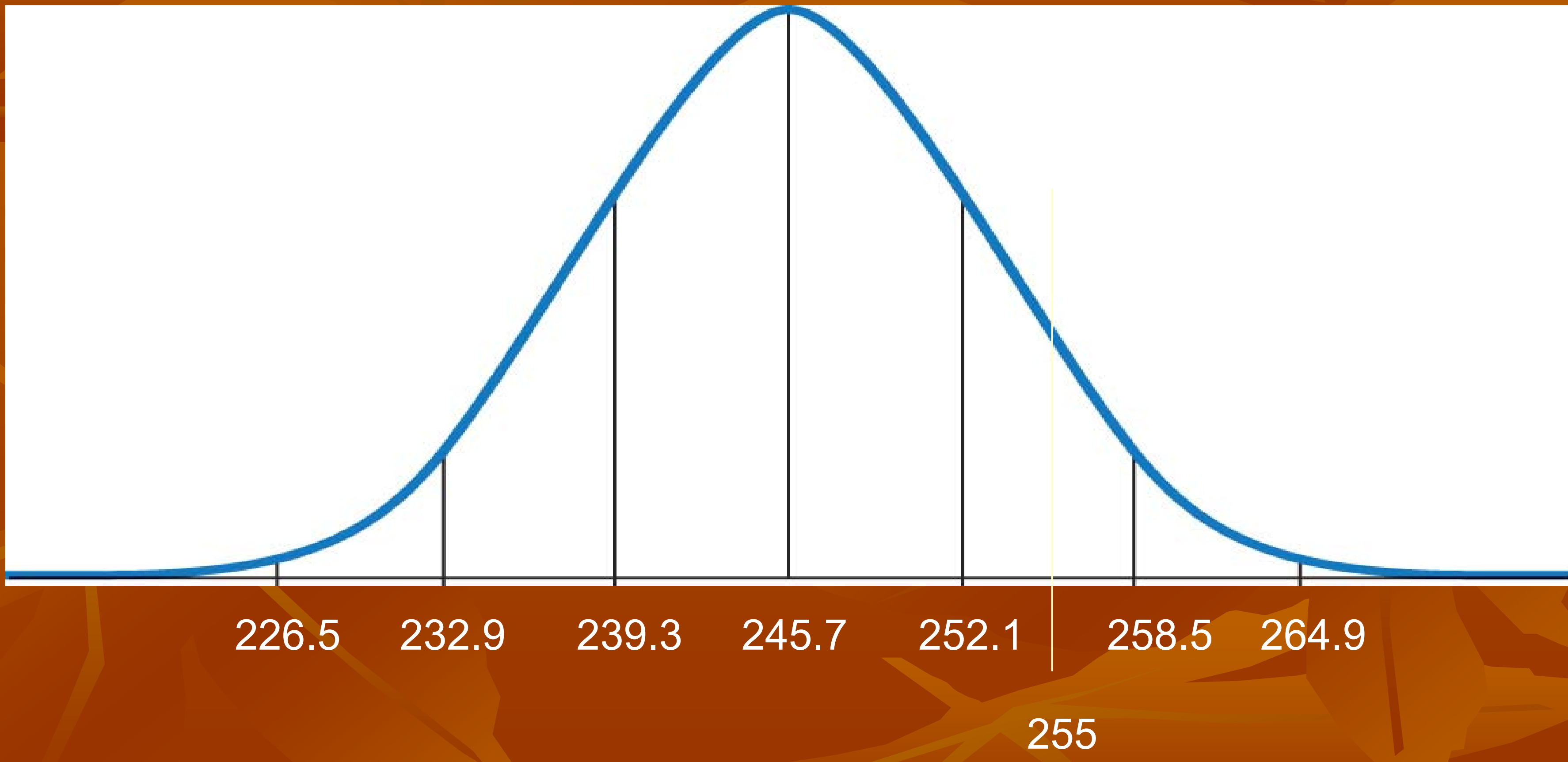
# Goal

- To put the CLT to good use.

# Example – Total Scores

- Suppose that you and 29 other students in the this class are members of the Nihilists for a Better America (NBA). You believe you are pretty hot and as a group your average total score (255 points) should be well above the class average.
- Assume $\bar{X} \sim N(245.7, 6.4)$

# Example – Total Scores



# Example – Total Scores

- $z = \dfrac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \dfrac{255 - 245.7}{6.4} = 1.45$

- $P(\bar{X} > 255) = P(Z > 1.45)$

  $= 1 – P(Z \leq 1.45)$

  $= 1 – 0.9265$

  $= 0.0735$

- So you’re hot, but not white hot.

# σ Unknown

- What if μ is <u>known</u> but σ is <u>unknown</u>?
- Use *s* in place of σ.
- Replace $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ with $s_{\bar{x}} = \frac{s}{\sqrt{n}}$.

# Example – Car Battery Warranty

- A automobile battery manufacturer guarantees them for 36 months.
- The warranty states that if the battery fails before 36 months it will be replaced with a new one whose cost will be discounted by the fraction of time left on the old one.
- You, as a member of Skeptics for a Better America (SBA), believe it's a scam to keep you hooked on their batteries, i.e., you believe $\mu < 36$ months.

# Example – Car Battery Warranty

- Suppose that the life lengths of 9 of these batteries were 35.38, 35.87, 35.49, 34.38, 34.55, 39.99, 36.18, 35.09 and 34.57 months.
- $\bar{X} = 35.72$ months
- $s = 1.71$ months
- Assume that the population of all life lengths is normally distributed.

# Example – Car Battery Warranty

- Assume that $\mu = 36$ months.
- If $P(\bar{X} < 35.72)$ is small, assuming that $\mu = 36$, then the sample would suggest $\mu < 36$.

- $$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu_{\bar{x}}}{s_{\bar{x}}} = \frac{35.72 - 36}{1.71/\sqrt{9}} = -0.49$$

- $$P(\bar{X} < 35.72) \cong P(Z < -0.49) = 0.3121$$

# Example – Car Battery Warranty

- The sample does not suggest that $\mu < 36$ months.

# Section 6.8

# Large-Sample Confidence Intervals for the Mean

# Definitions

- A *confidence interval* or an *interval estimate* is a range of values with an associated probability or confidence level $1 - \alpha$.
- The *confidence level* quantifies the chance that, in repeated sampling, the interval contains the true population parameter.

# Goal

- To construct a range of values that will contain $\mu$ with high probability.

# Bits & Pieces of a Confidence Interval

- interval = $L$ to $U$
  - $L$ = lower bound
  - $U$ = upper bound
- $1 - \alpha$ = confidence level
- $P(L \leq \mu \leq U) = 1 - \alpha$
  - Read this as, "The probability that the interval contains $\mu$ is $1 - \alpha$."

# CI Construction: σ known

- Start with a good point estimator of μ: $\bar{X}$
- If $n \geq 30$ then $\bar{X} \sim N\left(\mu, \sigma/\sqrt{n}\right)$
- $P\left(\mu - 2\sigma/\sqrt{n} \leq \bar{X} \leq \mu + 2\sigma/\sqrt{n}\right) \cong 0.95$
- $P\left(\bar{X} - 2\sigma/\sqrt{n} \leq \mu \leq \bar{X} + 2\sigma/\sqrt{n}\right) \cong 0.95$

# CI Construction: σ known

- Therefore, a confidence interval for μ with confidence level 0.95 (if $n \geq 30$ and σ known) is given by:
  - $L = \bar{X} - 2\sigma / \sqrt{n}$
  - $U = \bar{X} + 2\sigma / \sqrt{n}$
- This is also referred to as a 95% confidence interval for μ.

# New & Improved 95% CI

- $$\frac{\left(\mu - 2\sigma/\sqrt{n}\right) - \mu}{\sigma/\sqrt{n}} = -2$$
- $$\frac{\left(\mu + 2\sigma/\sqrt{n}\right) - \mu}{\sigma/\sqrt{n}} = 2$$
- $$P\left(\mu - 2\sigma/\sqrt{n} \le \bar{X} \le \mu + 2\sigma/\sqrt{n}\right)$$

$$= P\left(-2 \le Z \le 2\right) = 0.9772 - 0.0228 = 0.9544$$

# New & Improved 95% CI

- Need to find *z* such that

$$P\left(\mu - z\sigma/\sqrt{n} \le \bar{X} \le \mu + z\sigma/\sqrt{n}\right)$$

$$= P\left(-z \le Z \le z\right)$$

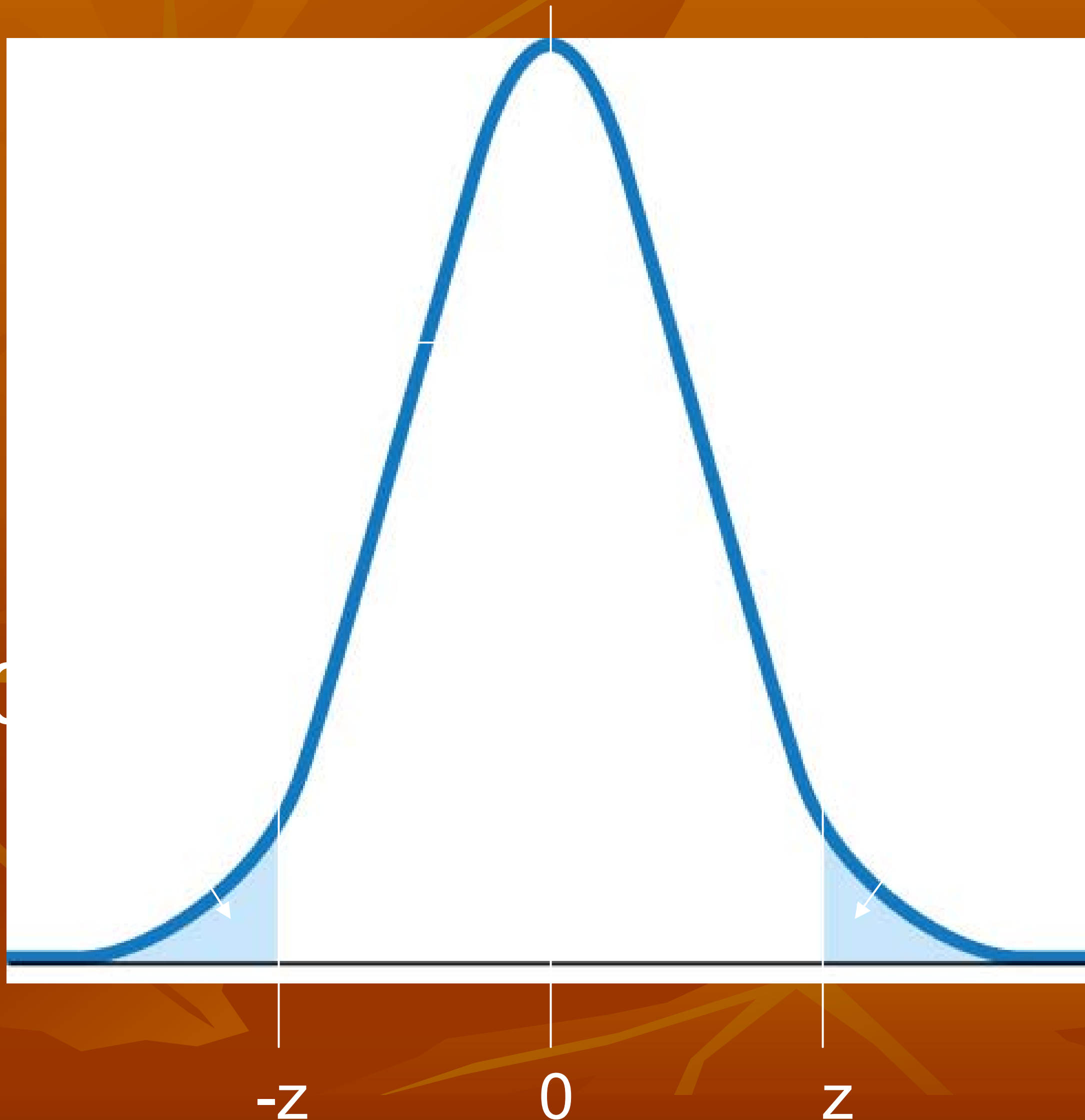
$$= 0.95$$

# New & Improved 95% CI

= 0.975

0.0

25



# New & Improved 95% CI

- Since P($Z < 1.96$) = 0.975 then $z = 1.96$
- Therefore, a confidence interval for μ with confidence level 0.95 is given by:
  - $L = \bar{X} - 1.96\sigma / \sqrt{n}$
  - $U = \bar{X} + 1.96\sigma / \sqrt{n}$

# Notation

- For a 95% CI (confidence level = 0.95) we needed $z_{.025} = 1.96$
- For a $(1 - \alpha)100\%$ CI (confidence level = $1 - \alpha$) we require $z_{\alpha/2}$

# Common values of $\alpha$ and $z_{\alpha/2}$

| Confidence level $(1 - \alpha)$ | | | | | |
|---|---|---|---|---|---|
| | **0.90** | **0.95** | **0.98** | **0.99** | **0.995** |
| $\alpha$ | 0.10 | 0.05 | 0.02 | 0.01 | 0.005 |
| $z_{\alpha/2}$ | 1.645 | 1.96 | 2.33 | 2.58 | 2.81 |

# Example – Total Scores

- Suppose you wish to estimate the mean total score for the 247 students in this class with 95% confidence.
- Assume that you know only that $\sigma = 34.8$ points.
- Suppose that from a sample of $n = 30$ students you obtained $\bar{X} = 250.7$

# Example – Total Scores

- Confidence level = 0.95
- $\alpha = 0.05$
- $z_{\alpha/2} = z_{.025} = 1.96$
- $1.96\dfrac{\sigma}{\sqrt{n}} = 1.96\dfrac{34.8}{\sqrt{30}} = 12.5$
- $L = 250.7 - 12.5 = 238.2$
- $U = 250.7 + 12.5 = 263.2$

# Example – Total Scores

- We are 95% confident that the mean total score for all 247 students is between 238.2 and 263.2 points.

# CI Construction: σ Unknown

- If σ is unknown and $n \geq 30$ then replace σ by $s$ in the CI calculations.
- If σ is unknown and $n < 30$ then . . .

# $n \geq 30$

| | σ | |
|---|---|---|
| Population | known | unknown |
| normal | $\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ | $\bar{X} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$ |
| nonnormal | Same as above | Same as above or Chapter 16 |

# $n < 30$

| | σ | |
|---|---|---|
| Population | known | unknown |
| normal | $\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ | Section 7.9 |
| nonnormal | Chapter 16 | Chapter 16 |

# “I am 95% Confident . . .” ???

- In the previous example we said we were 95% confident that the mean total score for all 247 students ($\mu$) was between 238.2 to 263.2 points.
- Does this mean that there is a 95% chance that $\mu$ is between 238.2 and 263.2 points?
- No! Negative!

# "I am 95% Confident . . ." ???

- In the previous example we knew that $\mu$ = 245.7 points.
- 245.7 is between 238.2 and 263.2
- So our 95% confidence interval contains $\mu$ with probability 1.
- In general we do not know $\mu$.
- Hence, in general, we do not know whether or not the particular interval we've calculated contains $\mu$.

# "I am 95% Confident . . ." ???

- What we do know is that in repeated sampling, 95% of the intervals would contain μ.

# Twenty 95% CIs for Mean Total Score

| $\bar{X}$ | $L$ | $U$ | Covers? | $\bar{X}$ | $L$ | $U$ | Covers? |
|---|---|---|---|---|---|---|---|
| 238.2 | 225.7 | 250.6 | Yes | 250.7 | 238.2 | 263.1 | Yes |
| 263.3 | 250.8 | 275.7 | **No** | 234.4 | 221.9 | 246.8 | Yes |
| 247.8 | 235.4 | 260.3 | Yes | 253.0 | 240.6 | 265.5 | Yes |
| 246.3 | 233.8 | 258.7 | Yes | 248.9 | 236.4 | 261.3 | Yes |
| 246.9 | 234.4 | 259.3 | Yes | 233.8 | 221.3 | 246.2 | Yes |
| 241.7 | 229.2 | 254.2 | Yes | 256.7 | 244.2 | 269.1 | Yes |
| 248.5 | 236.1 | 261.0 | Yes | 252.8 | 240.4 | 265.3 | Yes |
| 237.8 | 225.3 | 250.2 | Yes | 257.4 | 244.9 | 269.8 | Yes |
| 241.3 | 228.8 | 253.7 | Yes | 246.6 | 234.1 | 259.0 | Yes |
| 252.0 | 239.5 | 264.5 | Yes | 249.4 | 236.9 | 261.8 | Yes |

# Twenty 95% CIs for Mean Total Score

- Of the 20 CIs for $\mu$, 19 (19/20 = 0.95) contained 245.7, the true mean total score.

# Section 6.9

# Distribution of the Sample Mean: Small Sample and Unknown σ

# Goal

- To become familiar with Student's *t* distribution.

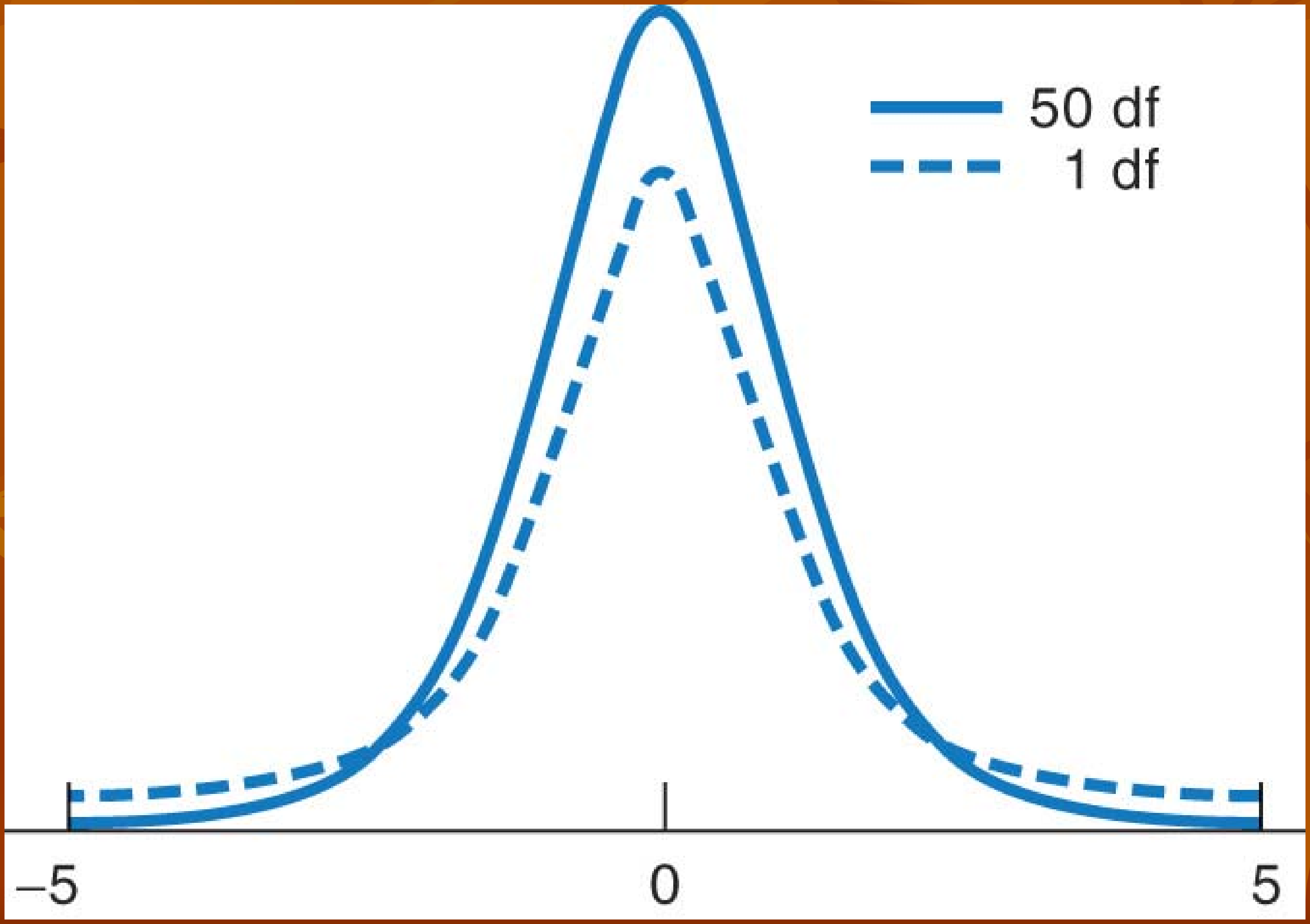
# Student's *t* Distribution

- By the CLT and Slutsky's Theorem, if $n \geq 30$ and σ unknown then:
  - $$Z = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim N(0,1)$$
- If n < 30, σ unknown and the population is normally distributed then:
  - $$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

# Student's *t* Distribution

- $t_{n-1}$ is Student's *t* distribution with ($n$ – 1) degrees of freedom (*df*).
- Due to William S. Gosset of the Guinness Breweries in Ireland.
- Published under the pseudonym *I. Student*.
- Similar to standard normal distribution but more variable (i.e., fatter tails).
- Variability decreases with increasing *df*.

# *t* distributions with *df* = 1 and 50



# Section 6.10

# Small-Sample Confidence Intervals for the Mean

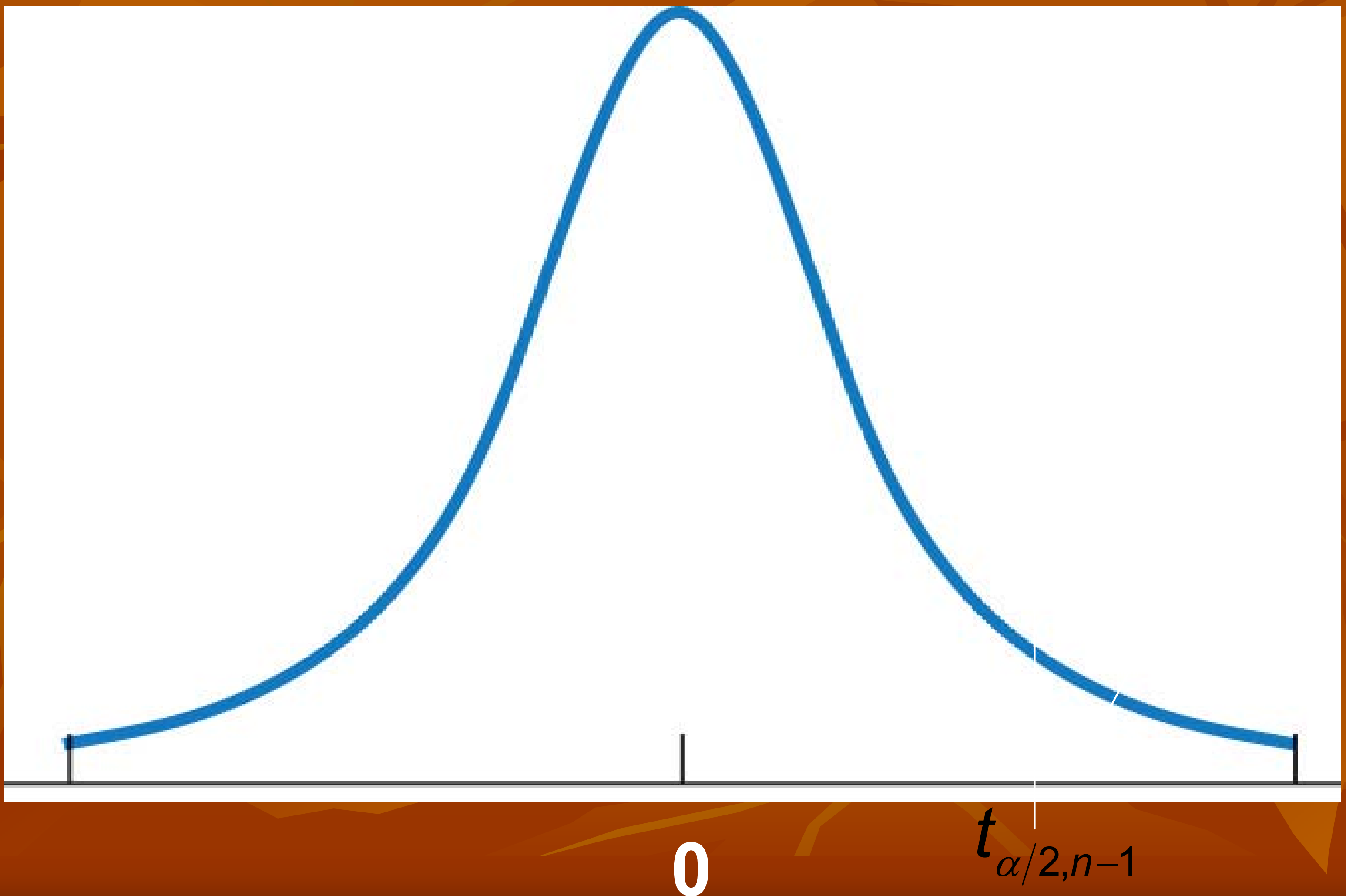
# Goal

- To construct a range of values that will contain $\mu$ with high probability for $n < 30$.

# (1 – α)100% CI for μ: n < 30 & σ Unknown

- Assume the population we are sampling from is normally distributed.
- $L = \bar{X} - t_{\alpha/2,n-1} s/\sqrt{n}$
- $U = \bar{X} + t_{\alpha/2,n-1} s/\sqrt{n}$
- $t_{\alpha/2,n\text{-}1}$ = (1-α/2)100$^{th}$ percentile of the t distribution with ($n$ – 1) *df*.

# Percentiles of the *t* Distribution



# Example – Gas Mileage

- You are interested in estimating the average mpg for a Honda Civic with 95% confidence.
- You have available to you 25 measurements on the mpg from 25 fillups.
- The mean of the 25 measurements was 28.99 mpg with a standard deviation of 3.22 mpg.
- Assume the population of mpg is normally distributed.

# Example – Gas Mileage

- $\alpha = 0.05$
- $(n - 1) = 25 - 1 = 24$
- $t_{\alpha/2,n\text{-}1} = t_{0.025,24} = 2.064$

# *t* CRITICAL VALUES

| *df* | 0.15 | 0.10 | 0.05 | 0.025 | 0.015 | 0.01 | 0.005 | 0.001 | 0.0005 |
|---|---|---|---|---|---|---|---|---|---|
| 4 | 1.190 | 1.533 | 2.132 | 2.776 | 3.298 | 3.747 | 4.604 | 7.173 | 8.610 |
| 5 | 1.156 | 1.476 | 2.015 | 2.571 | 3.003 | 3.365 | 4.032 | 5.893 | 6.869 |
| 6 | 1.134 | 1.440 | 1.943 | 2.447 | 2.829 | 3.143 | 3.707 | 5.208 | 5.959 |
| 7 | 1.119 | 1.415 | 1.895 | 2.365 | 2.715 | 2.998 | 3.499 | 4.785 | 5.408 |
| 8 | 1.108 | 1.397 | 1.860 | 2.306 | 2.634 | 2.896 | 3.355 | 4.501 | 5.041 |
| 9 | 1.100 | 1.383 | 1.833 | 2.262 | 2.574 | 2.821 | 3.250 | 4.297 | 4.781 |
| 10 | 1.093 | 1.372 | 1.812 | 2.228 | 2.527 | 2.764 | 3.169 | 4.144 | 4.587 |
| 11 | 1.088 | 1.363 | 1.796 | 2.201 | 2.491 | 2.718 | 3.106 | 4.025 | 4.437 |
| 12 | 1.083 | 1.356 | 1.782 | 2.179 | 2.461 | 2.681 | 3.055 | 3.930 | 4.318 |
| 13 | 1.079 | 1.350 | 1.771 | 2.160 | 2.436 | 2.650 | 3.012 | 3.852 | 4.221 |
| 14 | 1.076 | 1.345 | 1.761 | 2.145 | 2.415 | 2.624 | 2.977 | 3.787 | 4.140 |
| 15 | 1.074 | 1.341 | 1.753 | 2.131 | 2.397 | 2.602 | 2.947 | 3.733 | 4.073 |
| 16 | 1.071 | 1.337 | 1.746 | 2.120 | 2.382 | 2.583 | 2.921 | 3.686 | 4.015 |
| 17 | 1.069 | 1.333 | 1.740 | 2.110 | 2.368 | 2.567 | 2.898 | 3.646 | 3.965 |
| 18 | 1.067 | 1.330 | 1.734 | 2.101 | 2.356 | 2.552 | 2.878 | 3.610 | 3.922 |
| 19 | 1.066 | 1.328 | 1.729 | 2.093 | 2.346 | 2.539 | 2.861 | 3.579 | 3.883 |
| 20 | 1.064 | 1.325 | 1.725 | 2.086 | 2.336 | 2.528 | 2.845 | 3.552 | 3.850 |
| 21 | 1.063 | 1.323 | 1.721 | 2.080 | 2.328 | 2.518 | 2.831 | 3.527 | 3.819 |
| 22 | 1.061 | 1.321 | 1.717 | 2.074 | 2.320 | 2.508 | 2.819 | 3.505 | 3.792 |
| 23 | 1.060 | 1.319 | 1.714 | 2.069 | 2.313 | 2.500 | 2.807 | 3.485 | 3.768 |
| 24 | 1.059 | 1.318 | 1.711 | 2.064 | 2.307 | 2.492 | 2.797 | 3.467 | 3.745 |
| 25 | 1.058 | 1.316 | 1.708 | 2.060 | 2.301 | 2.485 | 2.787 | 3.450 | 3.725 |
| 26 | 1.058 | 1.315 | 1.706 | 2.056 | 2.296 | 2.479 | 2.779 | 3.435 | 3.707 |
| 27 | 1.057 | 1.314 | 1.703 | 2.052 | 2.291 | 2.473 | 2.771 | 3.421 | 3.690 |
| 28 | 1.056 | 1.313 | 1.701 | 2.048 | 2.286 | 2.467 | 2.763 | 3.408 | 3.674 |
| 29 | 1.055 | 1.311 | 1.699 | 2.045 | 2.282 | 2.462 | 2.756 | 3.396 | 3.659 |
| 30 | 1.055 | 1.310 | 1.697 | 2.042 | 2.278 | 2.457 | 2.750 | 3.385 | 3.646 |
| 40 | 1.050 | 1.303 | 1.684 | 2.021 | 2.250 | 2.423 | 2.704 | 3.307 | 3.551 |
| 50 | 1.047 | 1.299 | 1.676 | 2.009 | 2.234 | 2.403 | 2.678 | 3.261 | 3.496 |
| 60 | 1.045 | 1.296 | 1.671 | 2.000 | 2.223 | 2.390 | 2.660 | 3.232 | 3.460 |
| 120 | 1.041 | 1.289 | 1.658 | 1.980 | 2.196 | 2.358 | 2.617 | 3.160 | 3.373 |

# Example – Gas Mileage

- $t_{\alpha/2,n-1}\dfrac{s}{\sqrt{n}} = 2.064\dfrac{3.22}{\sqrt{25}} = 1.33$

- $L = 28.99 - 1.33 = 27.66$
- $U = 28.99 + 1.33 = 30.32$

# Example – Gas Mileage

- With 95% confidence, the true mean mpg for a Honda Civic is between 27.66 and 30.32 mpg.