Lesson 8
Part 1

Estimation of the Mean from One group

Lesson 8 Part 1 Overview

- One common statistical inference is to estimate a population parameter. Lesson 8 Part 1 covers estimation of the population mean (μ).
- The sample mean is a statistic that estimates the population mean.
- Confidence intervals are constructed to provide information about the precision of the estimate.

Lesson 8 Part 1 Outline

- Review
  - Sampling Distributions
  - CLT
- Overview of Estimation and Confidence Intervals
  - Confidence Intervals of the Mean when population standard deviation is known
  - Confidence Intervals of the Mean when standard deviation of the population is estimated by the sample standard deviation.

Review: Sampling Distributions

- Why is it important that the sample is random?
  - Statistics from random samples are random variables
  - Probability theory describes the sampling distribution of random variables (statistics).
  - The Central Limit Theorem describes the sampling distribution of the statistic.

Central Limit Theorem

- The Central Limit Theorem (CLT) is based on the sampling distribution of sample means from all possible samples of size n drawn from ANY population regardless of it’s distribution.
- However, to apply the CLT it isn’t necessary to generate all possible samples of size n and calculate the mean for each sample to determine the sampling distribution.
- If the population mean and standard deviation are known and the sample is sufficiently large (n > 40) sampling distribution of the sample mean is known:

$$\bar{x} \sim N\left(\mu, \ \frac{\sigma}{\sqrt{n}} \right)$$

Review: Parameters and Statistics

- One goal of statistical inference is to estimate a population parameter from a sample statistic

Statistic	Parameter	Pronounced:
$\overline{X}$	μ	mu
p	π	pi
SD or s	σ	sigma
s^2	σ^2	sigma-squared
r	ρ	rho

Parameters vs. Statistics

- Parameters are
  - Numerical characteristic of a population
  - Constant (fixed) at any one moment
  - Usually unknown
- Statistics are
  - Numerical summary of a sample
  - Calculated from sample data (not constant)
  - Used to *estimate* a parameter

Estimating Parameters

- The statistic calculated from the sample is a point estimate of the corresponding population parameter. For example:
  - The sample average is a point estimate of the true population mean
  - The sample proportion is a point estimate of the population proportion
- The SE of the statistic provides a measure of the precision of the estimate
  - A larger SE indicates a less precise point estimate
  - A smaller SE indicates a more precise point estimate

Definition of Precision in Statistics

- The term ‘precision’ has a specific definition in statistics usage.
- Precision refers to the standard error of the point estimate (the statistic).

Confidence Interval

General form of confidence interval:

Point Estimate $\pm$ Margin of Error

Point Estimate $\pm$ Confidence Coefficient * SE

Confidence Interval of the Mean

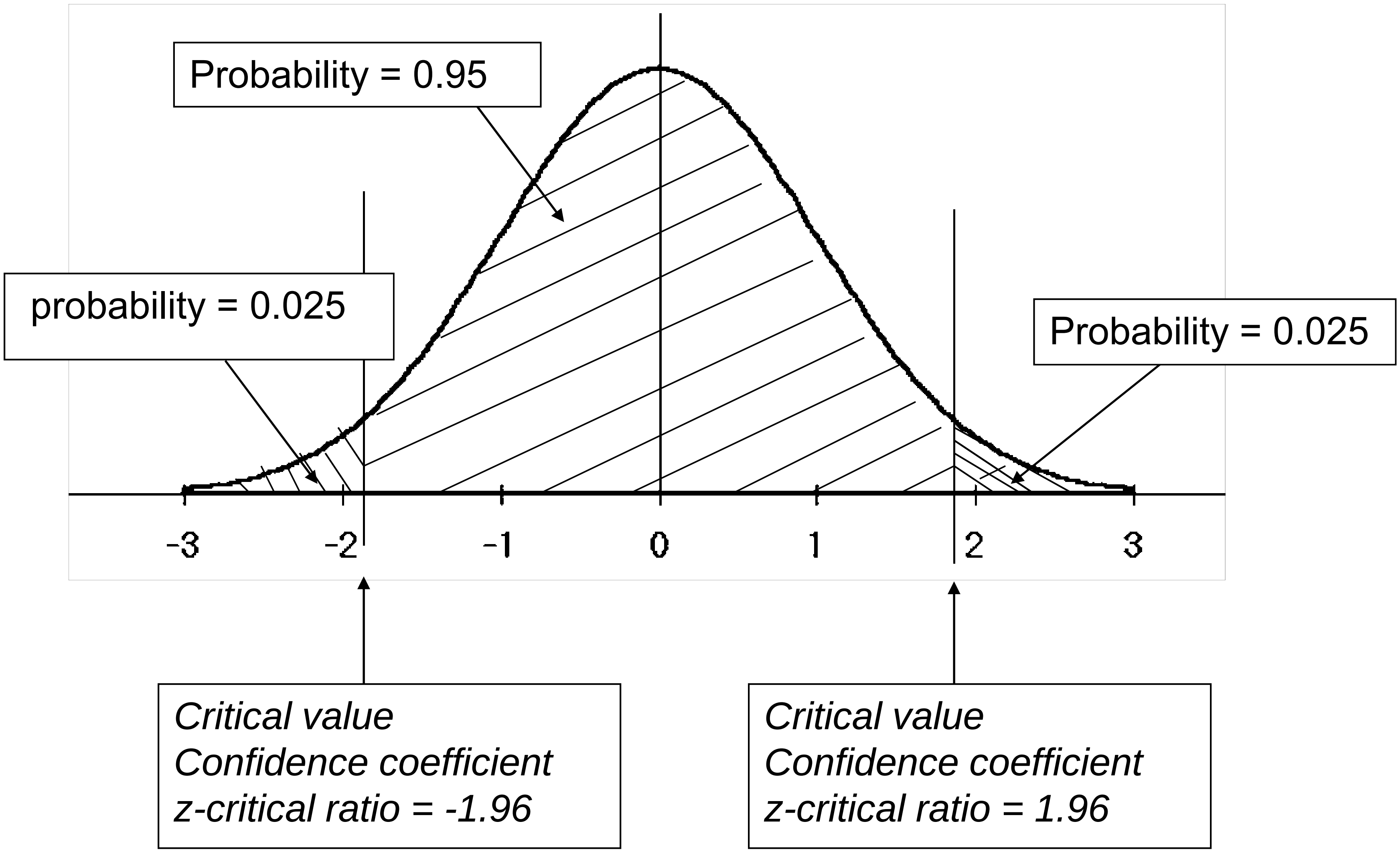
- The sample statistic is the sample mean ($\bar{x}$) which is a *point estimate* of the true population mean (μ).
- Because the mean can vary from sample to sample, this estimate from one sample does not necessarily equal the true mean.
- Confidence Intervals provide a measure of the precision of the estimate of the mean from one sample.

Calculating the Confidence Coefficients (σ known)

- For a 95% confidence level: the z critical ratios that divide the central 95% of the standard normal distribution from the 5% in the two tails (2.5% in each tail)

- =NORMSINV(-0.025) = -1.96
- =NORMSINV(0.975) = 1.96

Confidence Coefficients for 95% Confidence Interval From standard normal distribution



Example: 95% confidence interval of the mean

- Select a sample of 100 from a population with *unknown* mean and *known* variance:

  $\sigma = 4$

- The sample mean for this sample = 18
- SE of the mean:

$$\text{SE} = \frac{4}{\sqrt{100}} = 0.40$$

- The confidence coefficient for a 95% confidence interval is the z-score for probability of 0.95 around the mean: ± 1.96

95% Confidence Interval for the Mean

- General Formula:

$$\text{Sample Mean} \pm \text{Confidence Coefficient} * \text{SE}$$

- 95% Confidence Interval for this example

$$18 \pm 1.96 * 0.40 = (17.2,\ 18.8)$$

Interpretation of Confidence Intervals

- Correct interpretation
  - If we were to select 100 samples from the population and use these samples to calculate 100 different confidence intervals for μ, approximately 95 of the intervals (95%) would include the true population mean.
- Incorrect interpretation
  - The probability that the *true mean* (μ) is in the interval = 0.95

Confidence Interval Illustration

μ

Confidence Intervals Interactive Website

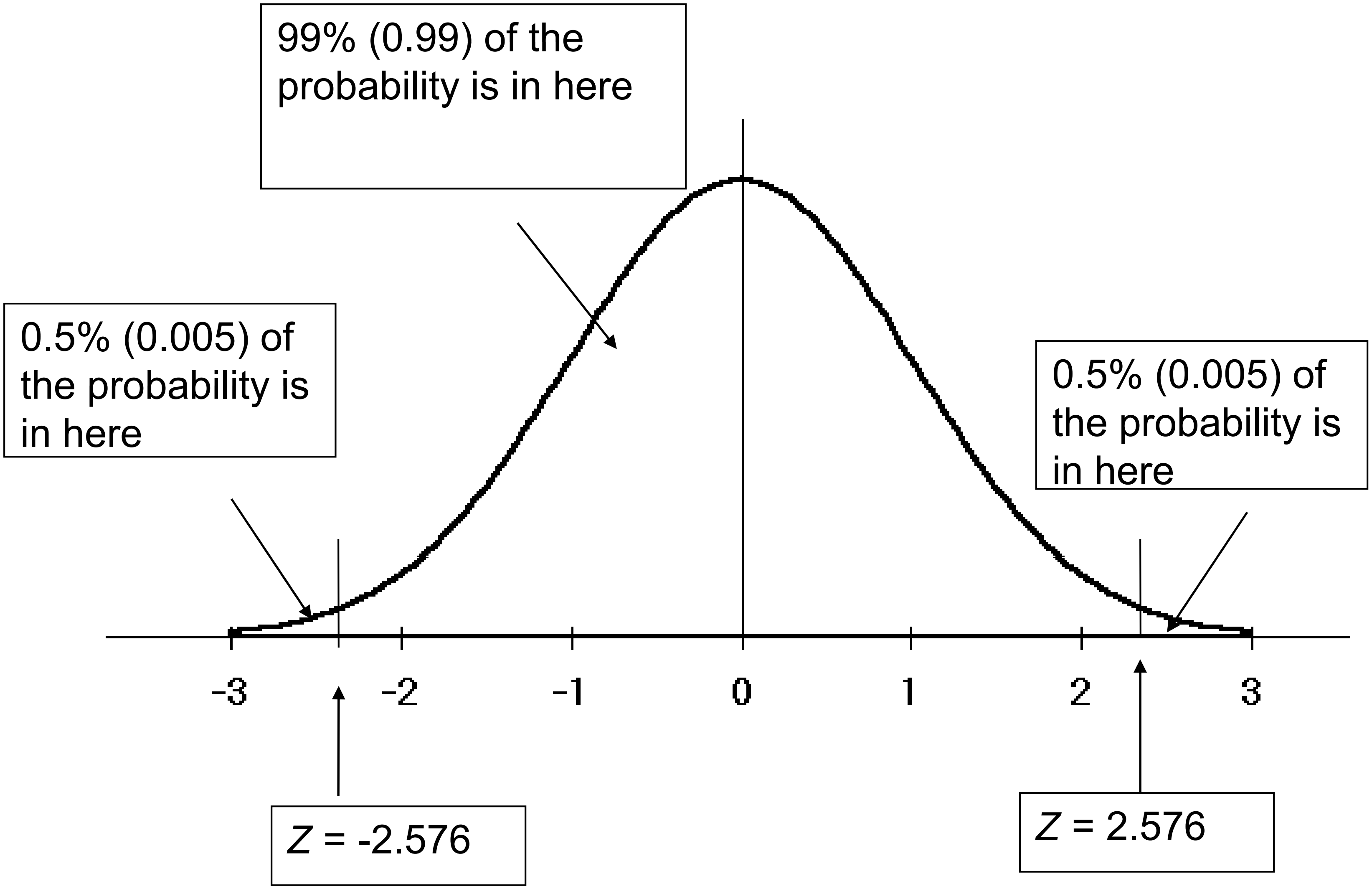
- This website has an applet that illustrates the concept of Confidence Intervals:

http://www.ruf.rice.edu/~lane/stat_sim/conf_interval/

Confidence Coefficient for 99% Confidence Interval

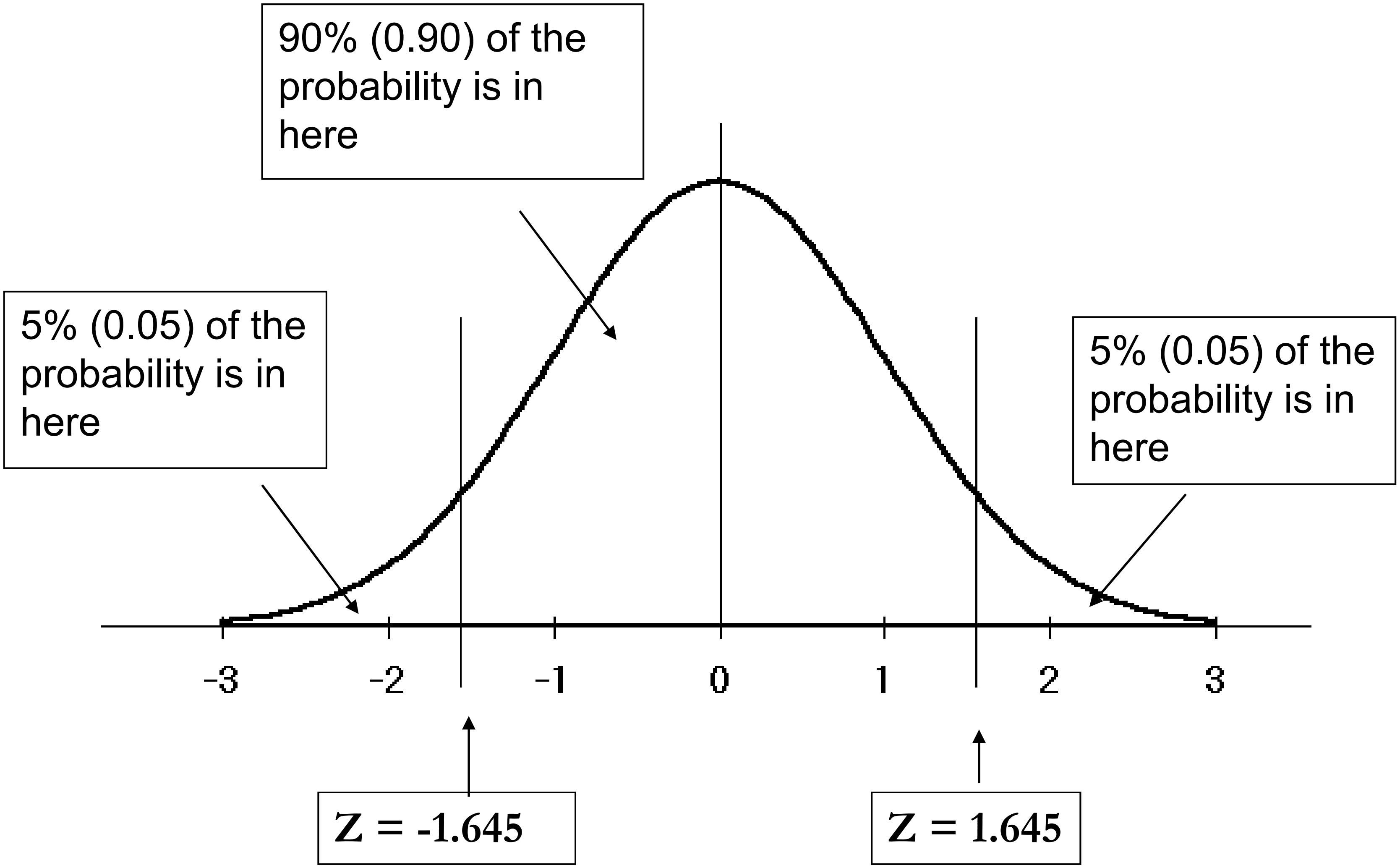
- If the population standard deviation is known, use the standard normal distribution to find the confidence coefficients for a confidence interval of the mean
- 99% confidence coefficients are the z critical ratios that divide the middle 99% of the distribution from the 1% in the tails
  - = NORMSINV(0.005) = -2.576
  - = NORMSINV(0.995) = 2.576

- Will the 99% confidence interval be wider or narrower than a 95% confidence interval with the same estimate and SE?


99% (0.99) of the probability is in here
0.5% (0.005) of the probability is in here
0.5% (0.005) of the probability is in here
-3
-2
-1
0
1
2
3
Z = -2.576
Z = 2.576

Confidence Coefficient for 90% Confidence Interval

- If the population standard deviation is known, use the standard normal distribution to find the confidence coefficients for a confidence interval of the mean
- 90% confidence coefficients are the z critical ratios that divide the middle 90% of the distribution from the 10% in the tails
  - = NORMSINV(0.05) = -1.645
  - = NORMSINV(0.95) = 1.645
- Will the 90% confidence interval be wider or narrower than a 95% confidence interval with the same estimate and SE?


90% (0.90) of the probability is in here
5% (0.05) of the probability is in here
5% (0.05) of the probability is in here
-3
-2
-1
0
1
2
3
Z = -1.645
Z = 1.645

Confidence Intervals of the Mean Using Z-scores

- Z-scores are appropriate confidence coefficients for a confidence interval of the mean when the population standard deviation (σ) is KNOWN

- *However, most of the time when the population mean is being estimated from sample data the population variance is unknown and must also be estimated from sample data*

- The sample standard deviation (s) provides an estimate of the population standard deviation (σ)

Confidence Intervals when the Population Variance is Unknown

- The sampling distribution of the sample mean is a t-distribution with n-1 degrees of freedom instead of a normal distribution

- Instead of using the standard normal distribution to find confidence coefficients, the t-distribution with n-1 df is used to find the confidence coefficients which are called t-coefficients or t-critical values

CLT and the t-distribution

- If the population mean and standard deviation are unknown and the sample is sufficiently large

  - n > 30 Underlying population is unimodal

  OR

  - N> 15 and underlying population approximately normal

  OR

  - N any size, but underlying population is normal

THEN,

$$\bar{x} \sim t_{n-1}\left(\mu,\ \frac{s}{\sqrt{n}}\right)$$

Confidence Interval of the Mean Using t-distribution

$$\text{Sample Mean} \pm \text{Confidence Coefficient} * \text{SE}$$

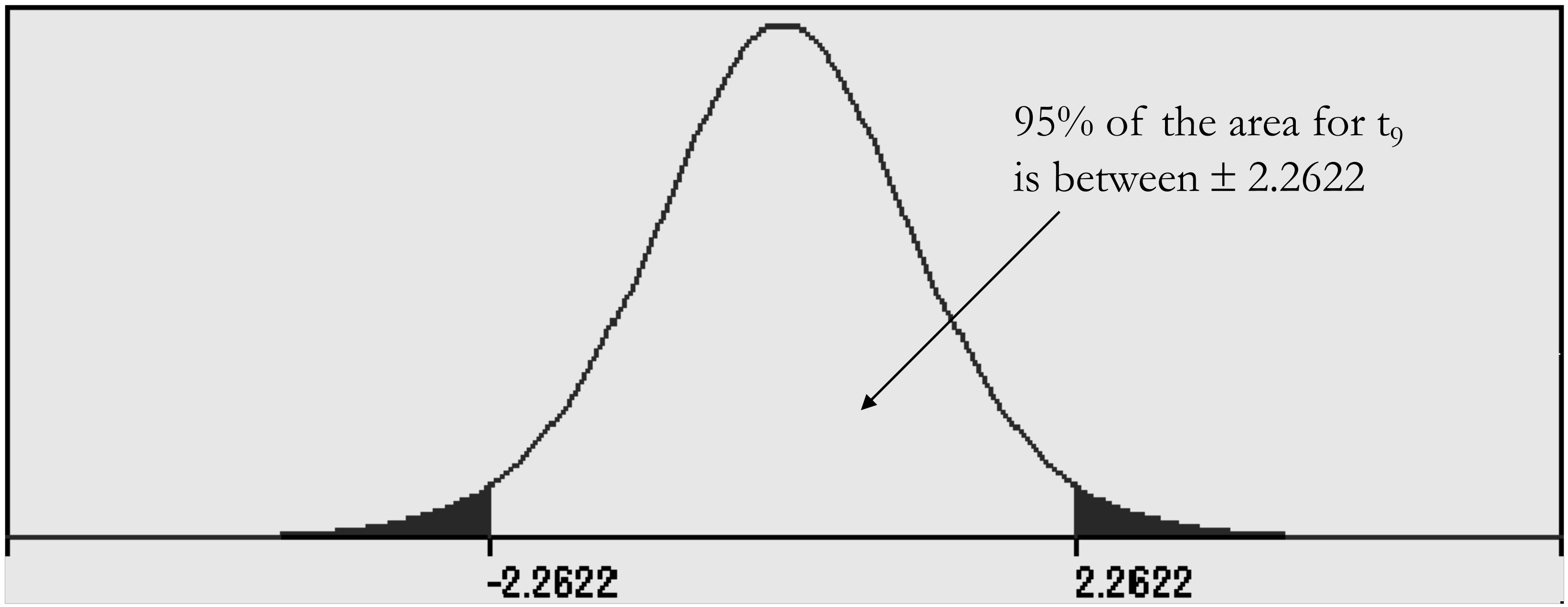
- The confidence coefficient is a t-coefficient from the t-distribution with n-1 df instead of a z-coefficient from the standard normal distribution.
- For a 95% confidence interval the t-coefficient is the value on the t-distribution with n-1 df such that 95% of the area is between ± t-coefficient.
- The value of the t-coefficient will be different for different sample sizes.

Find the t-coefficient with the TINV function

- The TINV function in Excel returns the t-coefficient associated with the specified area in the two tails of the t-distribution.
  - For a 95% confidence interval, 5% of the area is in the tails

- The t-coefficient for a 95% confidence interval of the mean from a sample of size 10 is found using: TINV(0.05,9) = 2.262
- 95% of the area under the t_9 curve is between –2.262 and + 2.262

t-distribution with 9 df



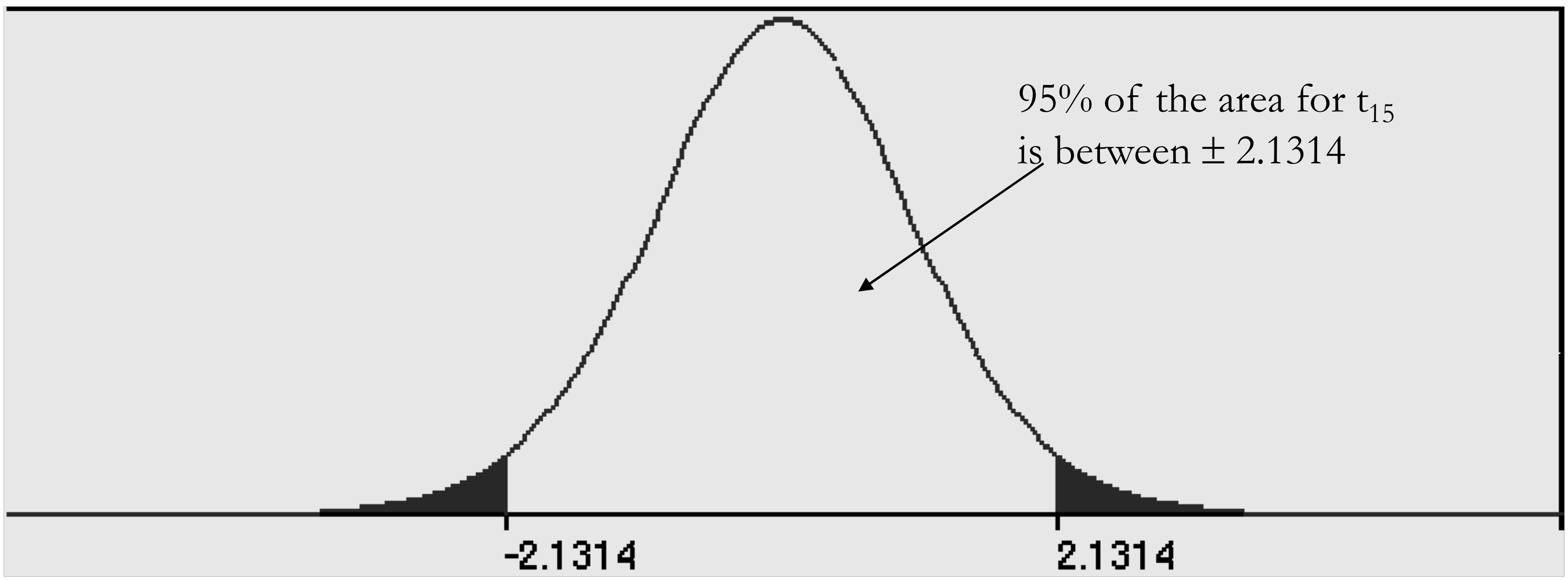
http://noppa5.pc.helsinki.fi/koe/tdistr/tj.html

t-coefficient for 15 df

- The t-coefficient for a 95% confidence interval of the mean from a sample of size 16 is found using: TINV(0.05,15) = 2.134

- 95% of the area under the t_{15} curve is between –2.134 and + 2.134
- To construct the 95% confidence interval, multiply the t-coefficient times the SEM. Add and subtract this from the sample mean.

t-distribution with 15 df



http://noppa5.pc.helsinki.fi/koe/tdistr/tj.html

95% Confidence Interval for Mean Distance Walked

- Data on the distance walked (in ft) in 6 minutes was collected for 16 COPD (Chronic Obstructive Pulmonary Disease) patients
  - Sample mean = 804.4
  - Sample standard deviation = 205.9

- Assume the population from which the sample is drawn is approximately normal.
- Use the sample data to construct a 95% Confidence interval of the mean (μ) distance walked for the population of COPD patients from which this sample was drawn.

95% Confidence Interval for Mean Distance Walked

- Point estimate = 804.4
- SE = 205.9 / 4 = 51.475

- Since we don't know the variance of 6 minute walk data for this population, the confidence coefficients will be from the t- distribution with 15 df.

95% Confidence Interval for Mean Distance Walked

- The Excel function to find the t-Coefficients for a 95% Confidence Interval from the t-distribution with 15 df
  - = TINV (0.05, 15) = 2.134

- Calculate the lower and upper limits of the 95% confidence interval
  - Lower limit: 804.4 – 2.134*51.475 = 694.55
  - Upper limit: 804.4 + 2.134*51.475 = 914.25

Reporting the Confidence Interval

The mean distance walked in 6 minutes for COPD patients is estimated to be 804.4 ft with 95% CI: 694.55 ft to 914.25 ft.

Reducing the Margin of Error of a Confidence Interval

- The margin of error of a confidence interval decreases as
  - The sample size increases
  - The standard deviation decreases
  - The level of confidence decreases

Readings and Assignments

- Reading
  - Chapter 5 pgs 95 – 102: Confidence intervals of the Mean from one group
- Work through the Lesson 8 Part 1 Practice Exercises
- Work through Excel Module 8 Examples
- Start Homework 5