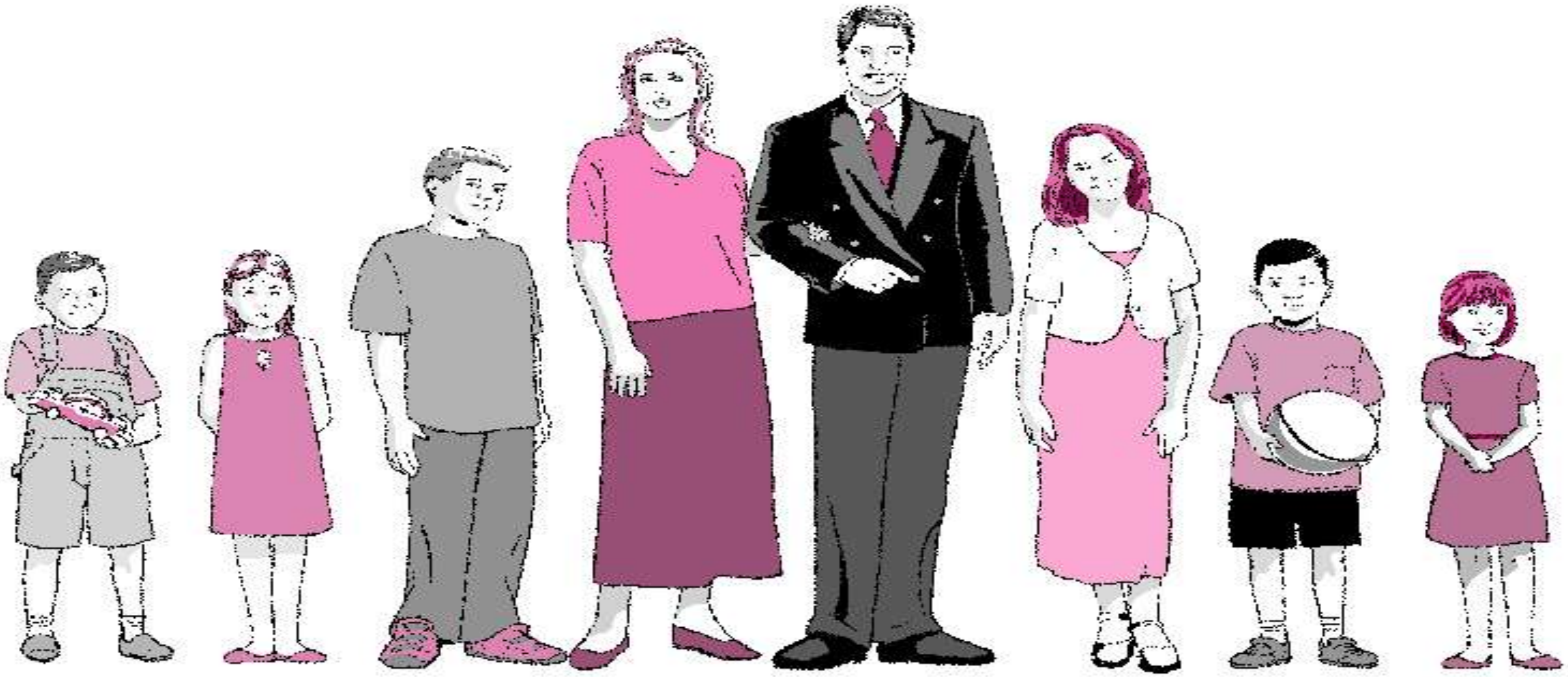
Descriptive Statistics



A Normal Distribution?

Statistics vs. Parameters

- A parameter is a characteristic of a population.
  - It is a numerical or graphic way to summarize data obtained from the population
- A statistic is a characteristic of a sample.
  - It is a numerical or graphic way to summarize data obtained from a sample

Types of Numerical Data

- There are two fundamental types of numerical data:
  1) Categorical data: obtained by determining the frequency of occurrences in each of several categories
  2) Quantitative data: obtained by determining placement on a scale that indicates amount or degree

Techniques for Summarizing Quantitative Data

- Frequency Distributions
- Histograms/Stem and Leaf Plots
- Distribution curves
- Averages/Spread
- Variability/Correlations

Frequency Polygons

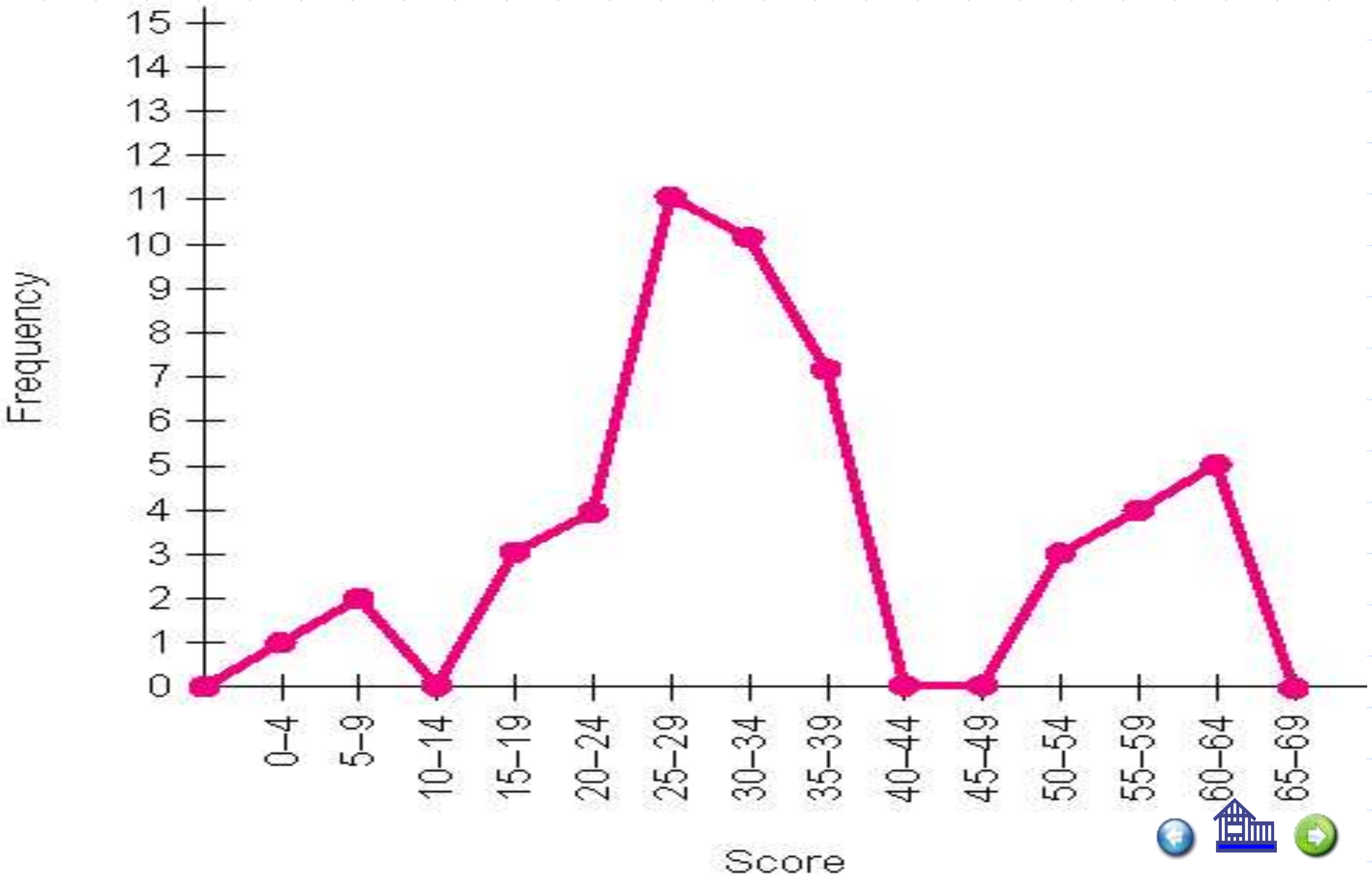
- Places data in some sort of order (ascending or desending).
- A frequency distribution lists scores from high to low (Table 10.1)
- This results in a grouped frequency distribution (Table 10.2)
- Since the information is not very visual, a graphical display called a frequency polygon can help with this (Figure 10.1)
  - Frequency polygons can be negatively or positively skewed (Figure 10.2)
  - They can be useful in comparing two or more groups

Example of a Frequency Distribution

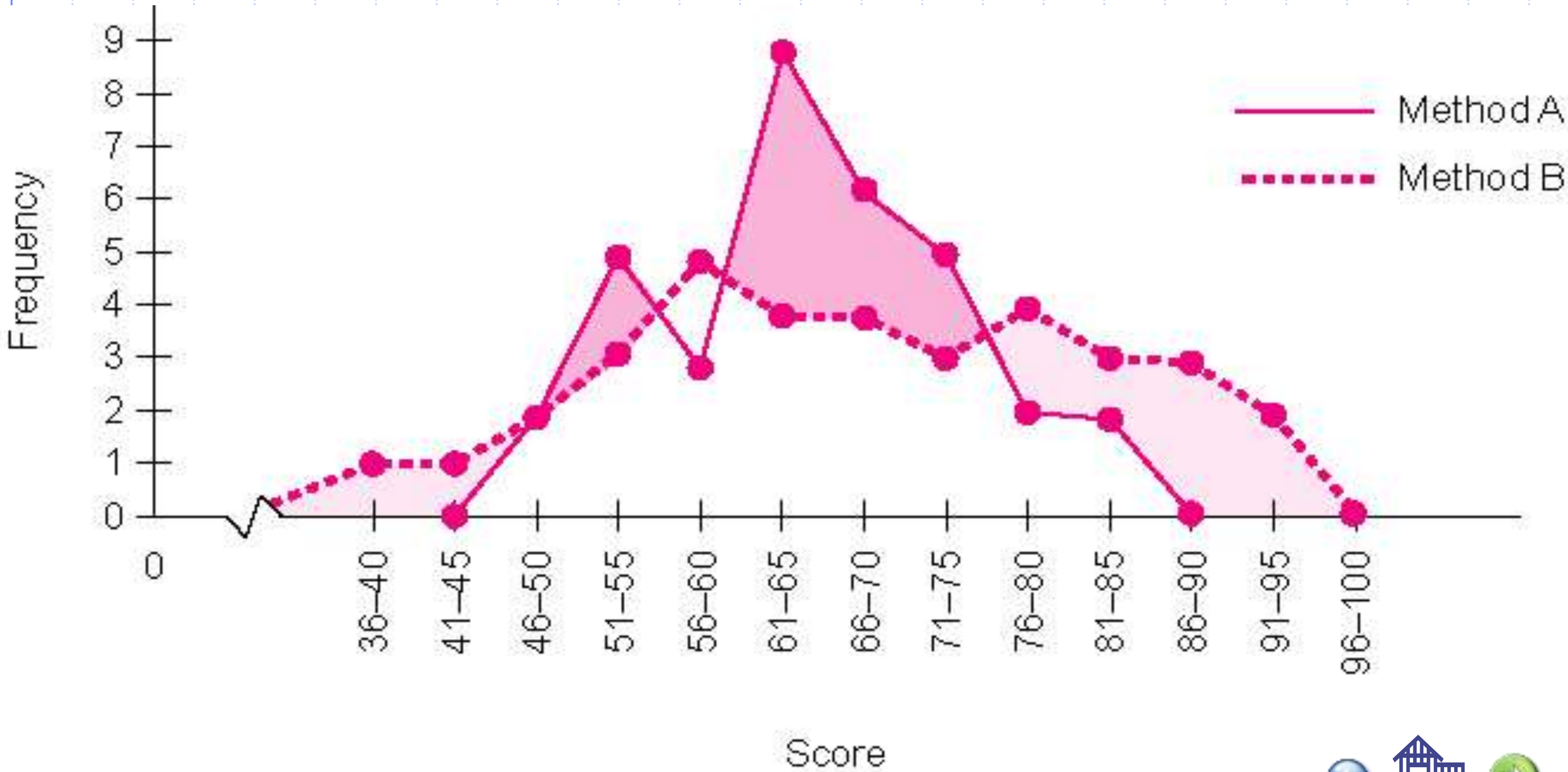
Raw Score	**Frequency**
64	2
63	1
61	2
59	2
56	2
52	1
51	2
38	4
36	3
34	5
31	5
29	5
27	5
25	1
24	2
21	2
17	2
15	1
6	2
3	1
	$n = 50$

Technically, the table should include all scores, including those for which there are zero frequencies. We have eliminated those to simplify the presentation.

Example of a Frequency Polygon



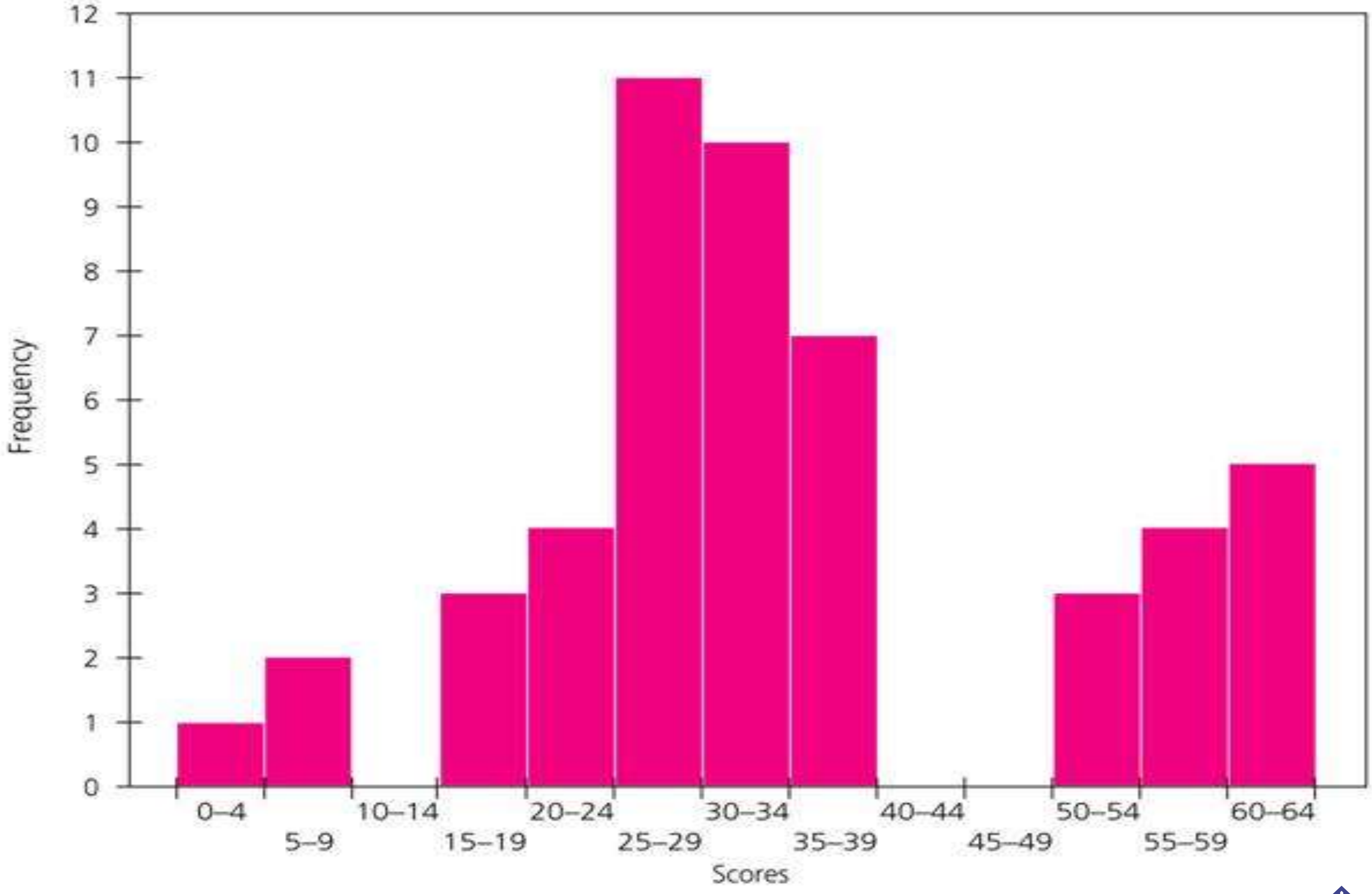
Two Frequency Polygons Compared



Histograms and Stem-and-Leaf Plots

- A histogram is a bar graph used to display quantitative data at the interval or ratio level of measurement (Table 10.2)
- A Stem-leaf Plot (stem plot) looks like a histogram, except instead of bars, it shows values for each category
  - They are helpful for comparing and contrasting two distributions (Table 10.1)

Histogram of Data in Table 10.2

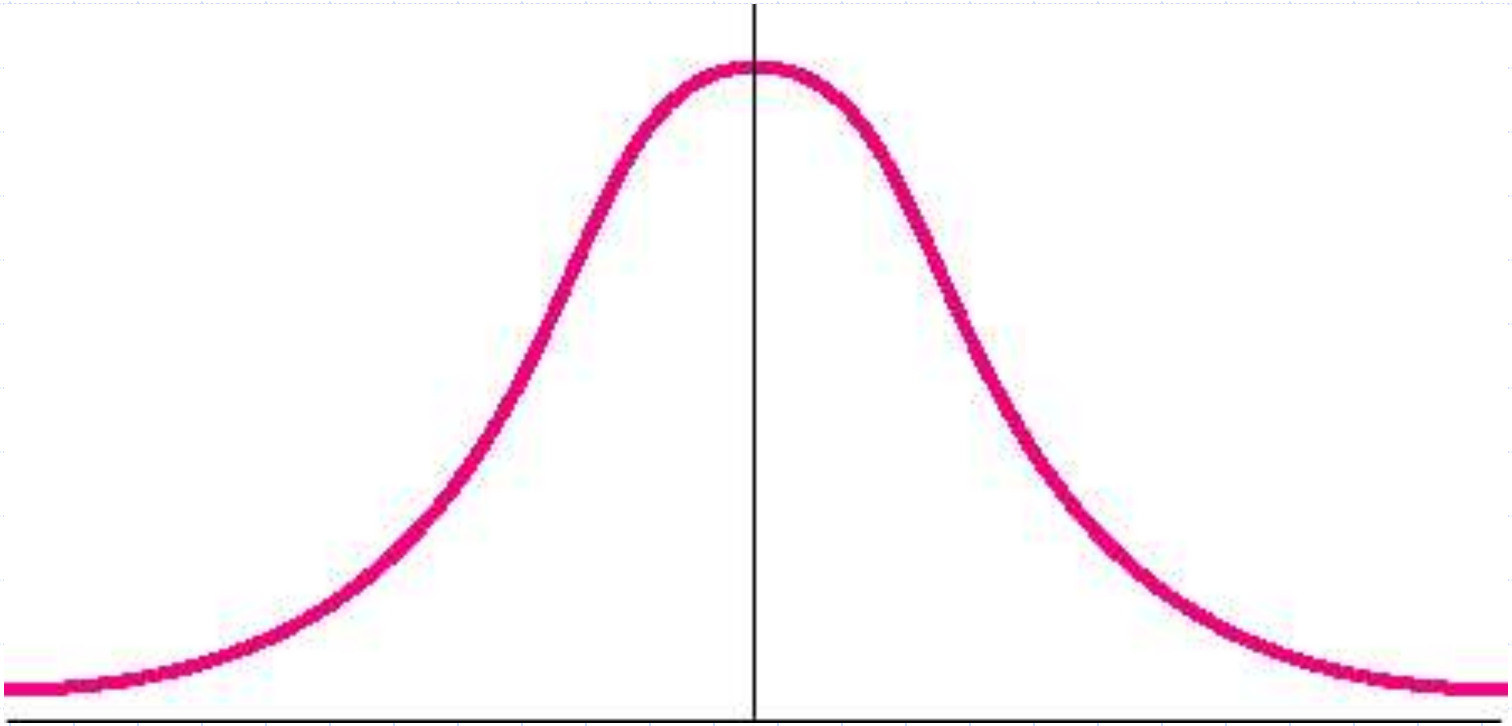


The Normal Curve

- This distribution curve shows a generalized distribution of scores vs. straight lines (frequency polygon)
- Distribution of data tends to follow a specific shape called a normal distribution (see Figure 10.6)
- This distribution is considered 'bell shaped' and allows the plotting of the following averages:
  - Mean
  - Medium
  - Mode

*These measures of central tendencies enable one to summarize the data in a frequency distribution with a single number

The Normal Curve

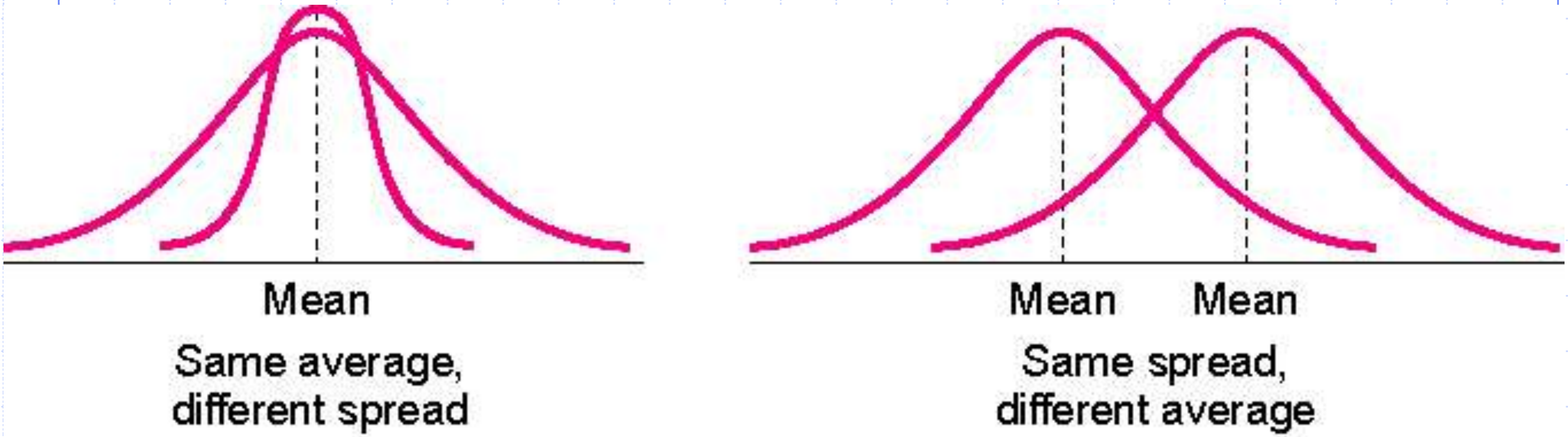


Example of the Mode, Median and Mean in a Distribution

Raw Score	Frequency
98	1
97	1
91	2
85	1
80	5
77	7
72	5
65	3
64	7
62	10
58	3
45	2
33	1
11	1
5	1
	$n = 50$

Mode = 62; median = 64.5; mean = 66.7

Different Distributions Compared



Variability

- Refers to the extent to which the scores on a quantitative variable in a distribution are spread out.
- The range represents the difference between the highest and lowest scores in a distribution.
- A five number summary reports the lowest, the first quartile, the median, the third quartile, and highest score.
- Five number summaries are often portrayed graphically by the use of box plots.

Standard Deviation

- Considered the most useful index of variability.
- It is a single number that represents the spread of a distribution.
- See p. 348 to calculate the mean of the distribution.
- Table 10.5 will illustrate the calculation of the SD of a distribution.
- If a distribution is normal, then the mean plus or minus 3 SD will encompass about 99% of all scores in the distribution.

Calculation of the Standard Deviation of a Distribution

Raw Score (X)	Mean $(\bar{X})$	$X - \bar{X}$	$(X - \bar{X})^2$
85	54	31	961
80	54	26	676
70	54	16	256
60	54	6	36
55	54	1	1
50	54	-4	16
45	54	-9	81
40	54	-14	196
30	54	-24	576
25	54	-29	841

$$\text{Variance } (SD^2) = \frac{\Sigma(X - \bar{X})^2}{n}$$

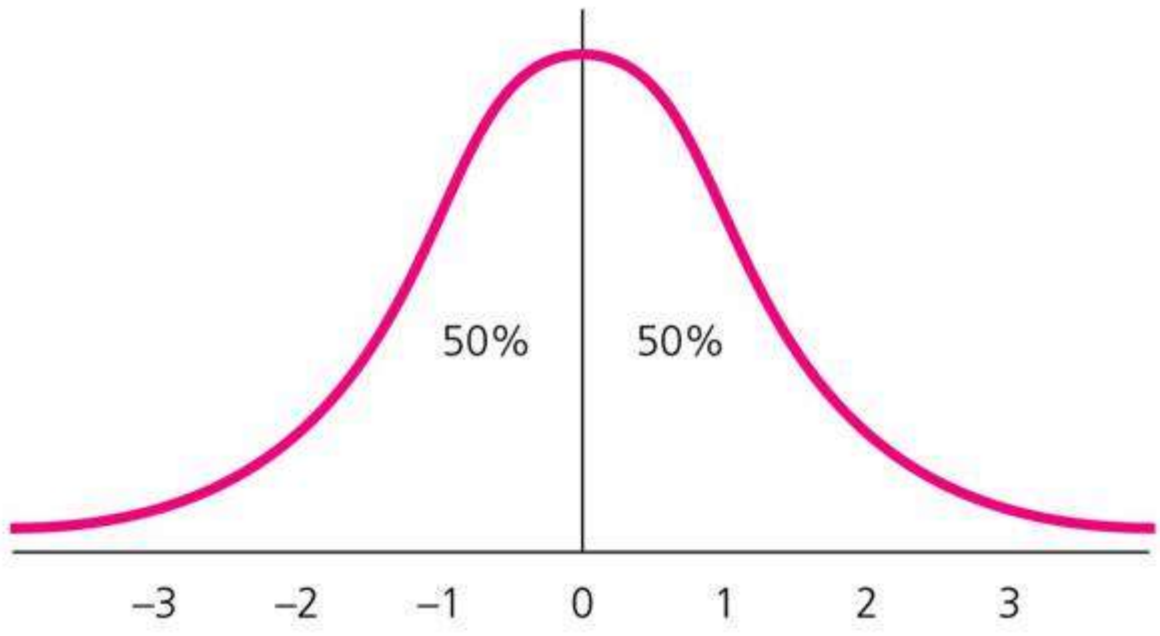
$$= \frac{3640}{10} = 364$$

$$\text{Standard deviation (SD)} = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n}}$$

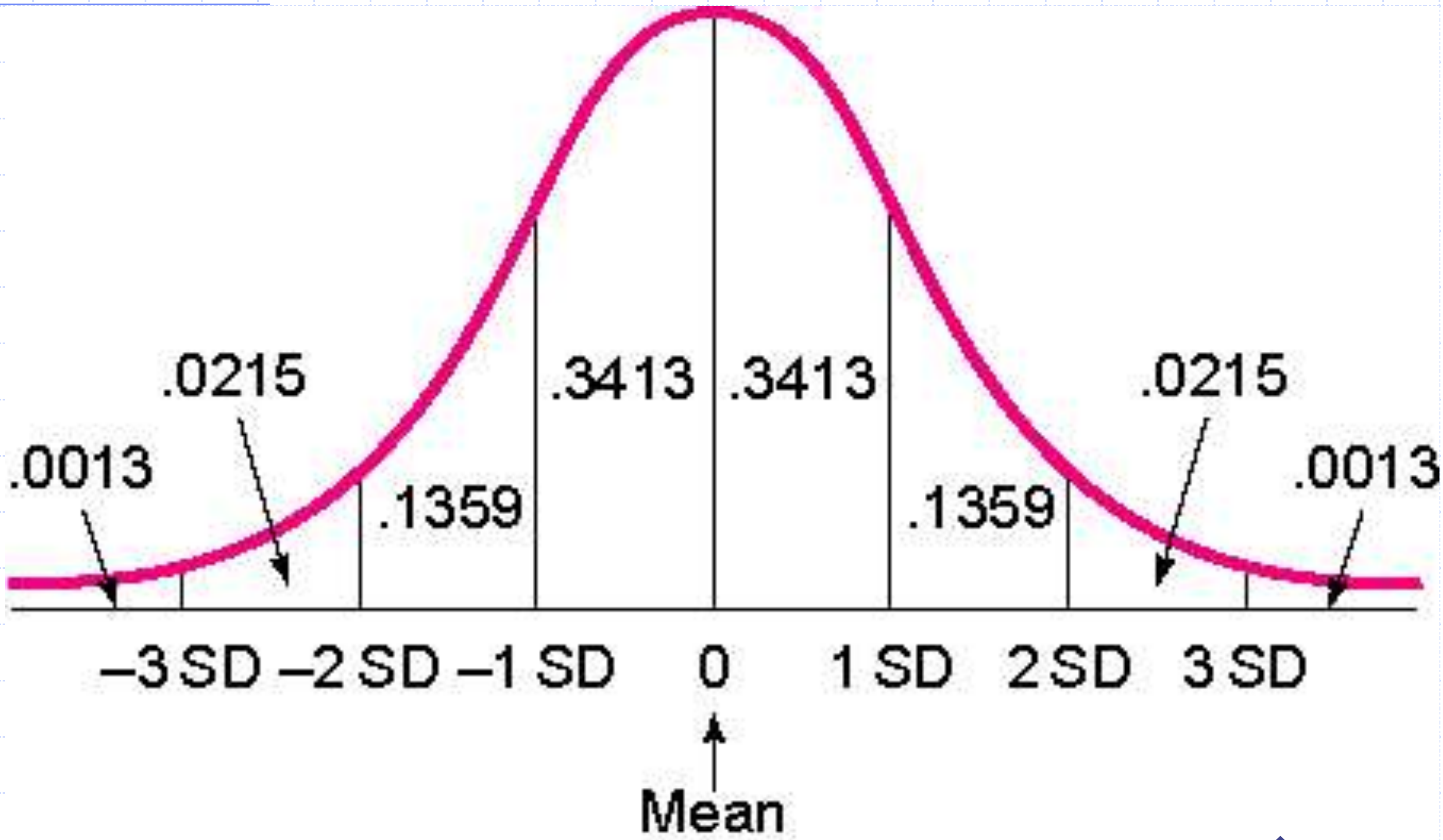
Facts about the Normal Distribution

- 50% of all the observations fall on each side of the mean. (Figure 10.11)
- 68% of scores fall within 1 SD of the mean in a normal distribution.
- 27% of the observations fall between 1 and 2 SD from the mean.
- 99.7% of all scores fall within 3 SD of the mean. (Figure 10.12)
- This is often referred to as the **68-95-99.7 rule**

Fifty Percent of All Scores in a Normal Curve Fall on Each Side of the Mean



Probabilities Under the Normal Curve



Techniques for Summarizing Categorical Data

- The Frequency Table
- Bar Graphs and Pie Charts
- The Cross break Table

Frequency and Percentage of Responses to Questionnaire

Response	Frequency	Percentage of Total (%)
Lecture	15	30
Class discussions	10	20
Demonstrations	8	16
Audiovisual presentations	6	12
Seatwork	5	10
Oral reports	4	8
Library research	2	4
Total	50	100

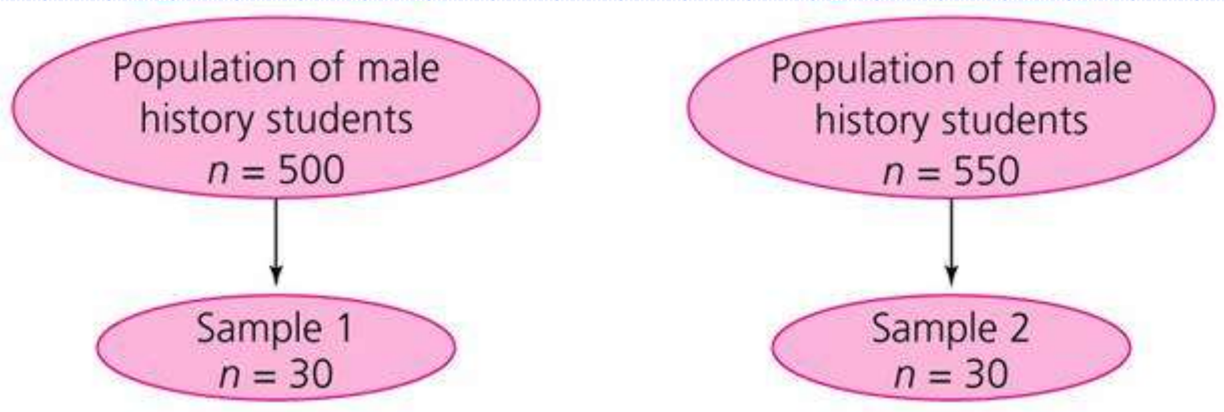
Inferential Statistics

Chapter Eleven

What are Inferential Statistics?

- Inferential Statistics refer to certain procedures that allow researchers to make inferences about a population based on data obtained from a sample.
- Obtaining a random sample is desirable since it ensures that this sample is representative of a larger population.
- The better a sample represents a population, the more researchers will be able to make inferences.
- Making inferences about populations is what Inferential Statistics are all about.

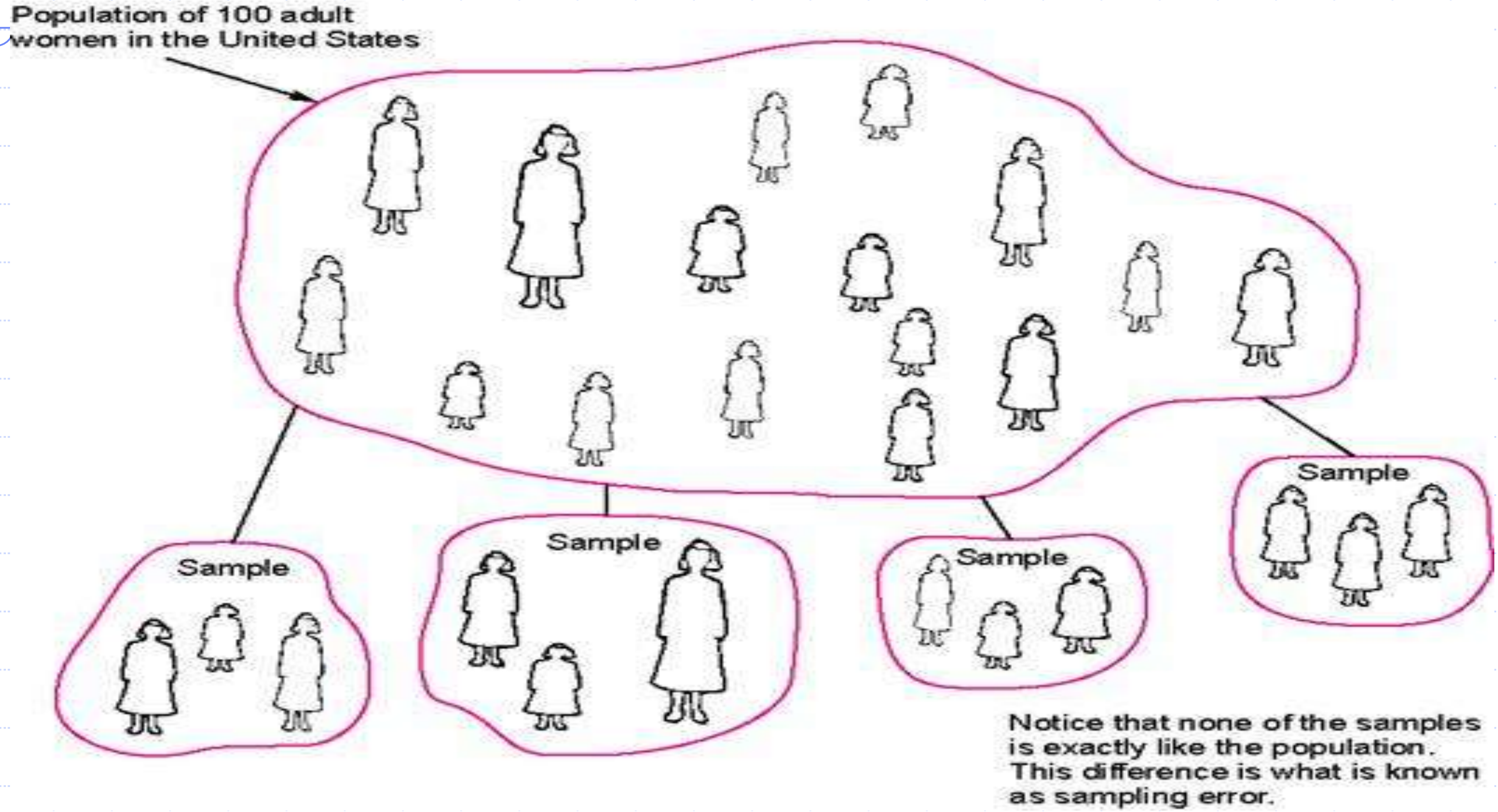
Two Samples from Two Distinct Populations



Sampling Error

- It is reasonable to assume that each sample will give you a fairly accurate picture of its population.
- However, samples are not likely to be identical to their parent populations.
- This difference between a sample and its population is known as Sampling Error. (see Figure 11.2)
- Furthermore, no two samples will be identical in all their characteristics.

Sampling Error

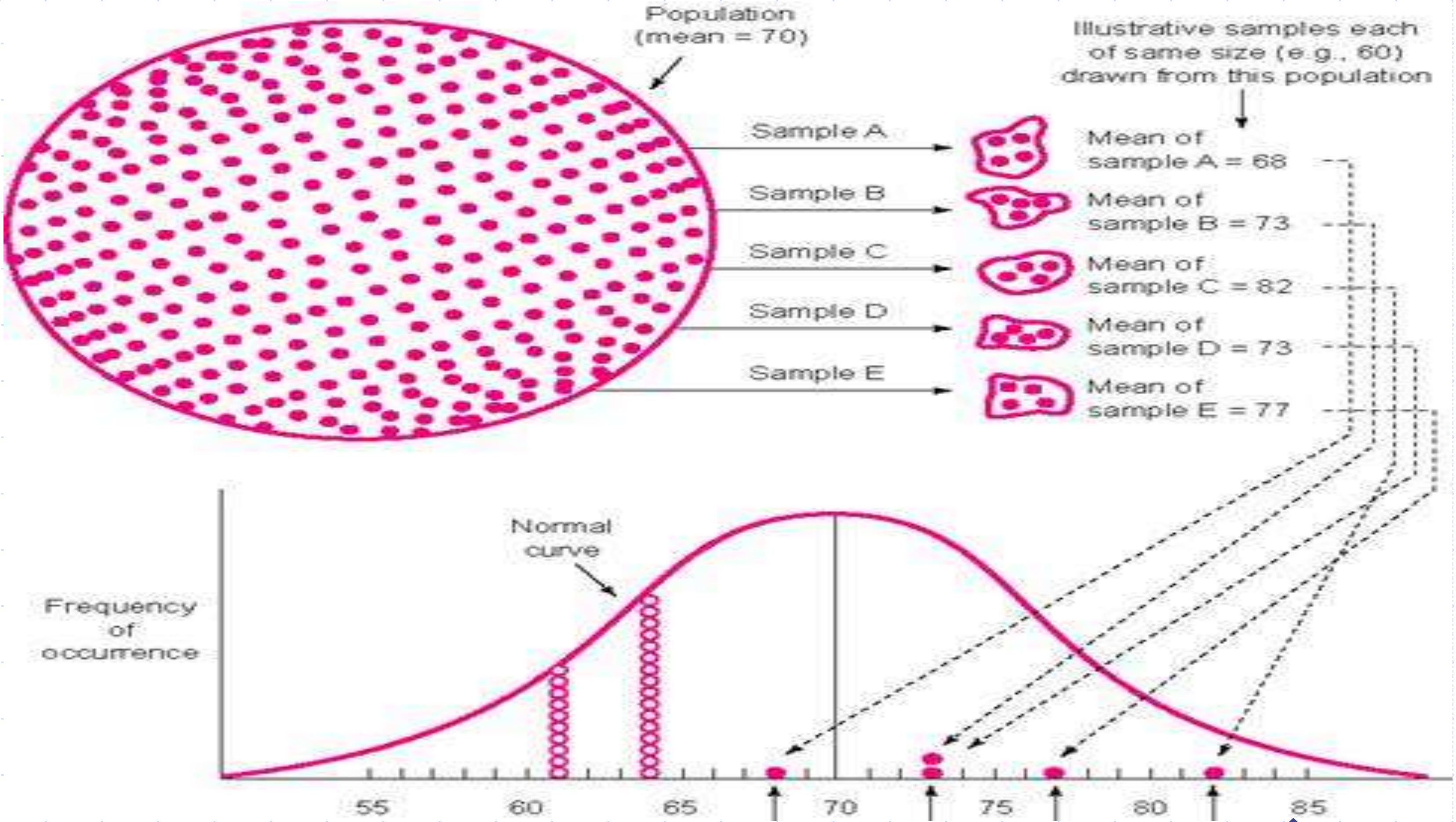


Distribution of Sample Means

- There are times where large collections of random samples do pattern themselves in ways that will allow researchers to predict accurately some characteristics of the population from which the sample was taken.
- A sampling distribution of means is a frequency distribution resulting from plotting the means of a very large number of samples from the same population

Refer to Figure 11.3

A Sampling Distribution of Means



Standard Error of the Mean

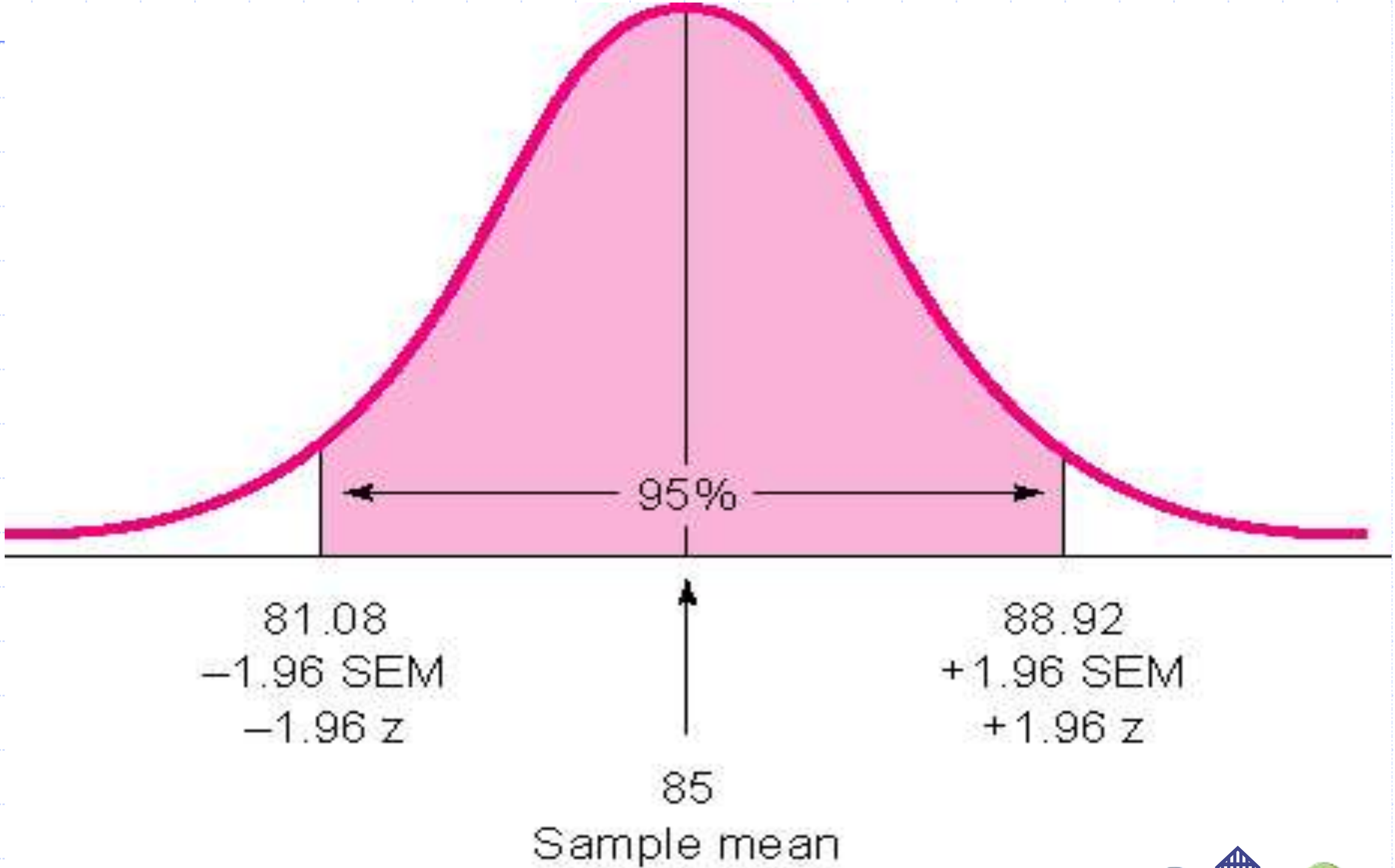
- The standard deviation of a sampling distribution of means is called the Standard Error of the Mean (SEM).
- If you can accurately estimate the mean and the standard deviation of the sampling distribution, you can determine whether it is likely or not that a particular sample mean could be obtained from the population.
- To estimate the SEM, divide the SD of the sample by the square root of the sample size minus one.

Refer to Figure 11.4

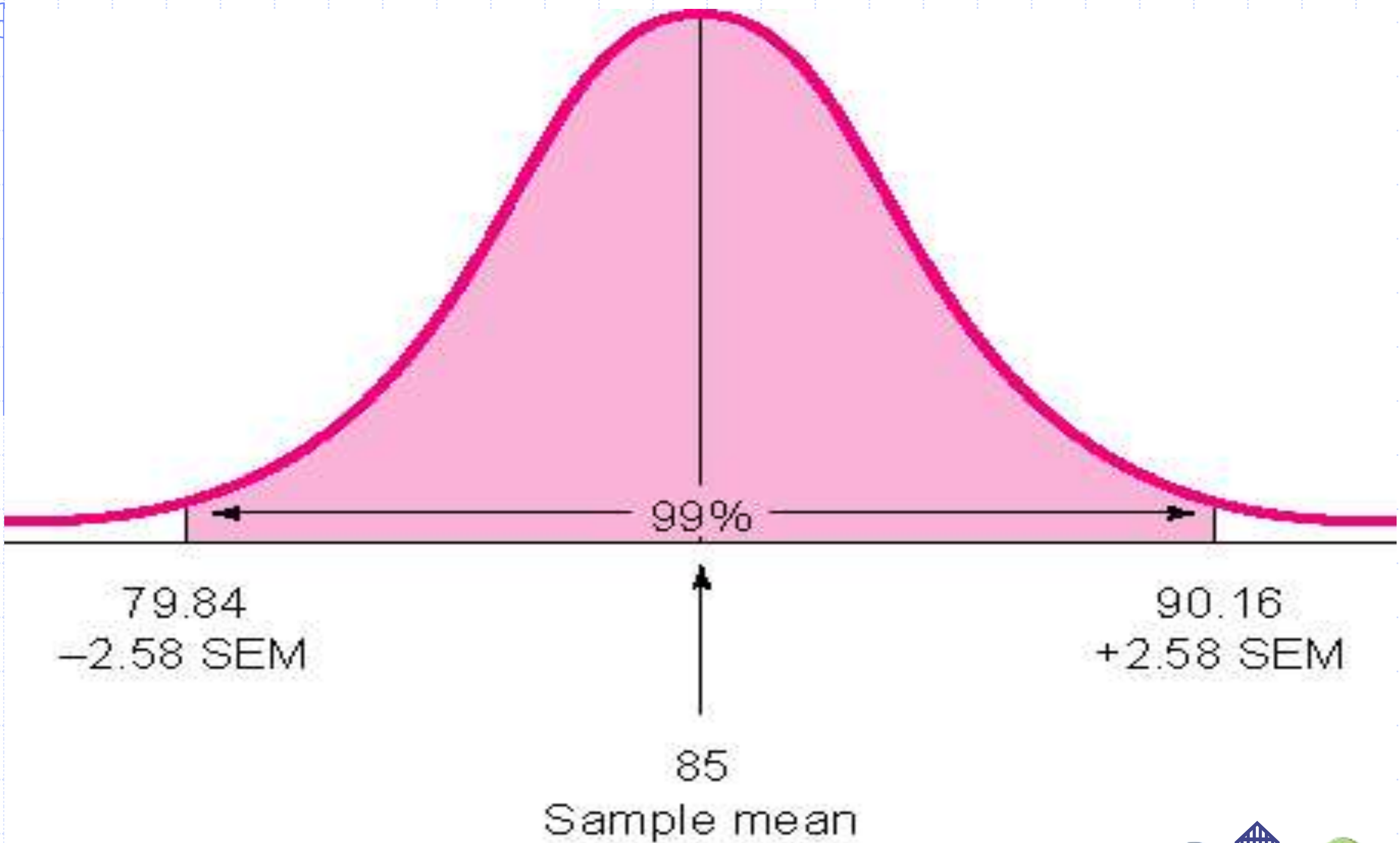
Confidence Intervals

- A Confidence Interval is a region extending both above and below a sample statistic within which a population parameter may be said to fall with a specified probability of being wrong.
- SEM's can be used to determine boundaries or limits, within which the population mean lies.
- If a confidence interval is 95%, there would be a 'probability' that 5 out of 100 (population mean) would fall outside the boundaries or limits.

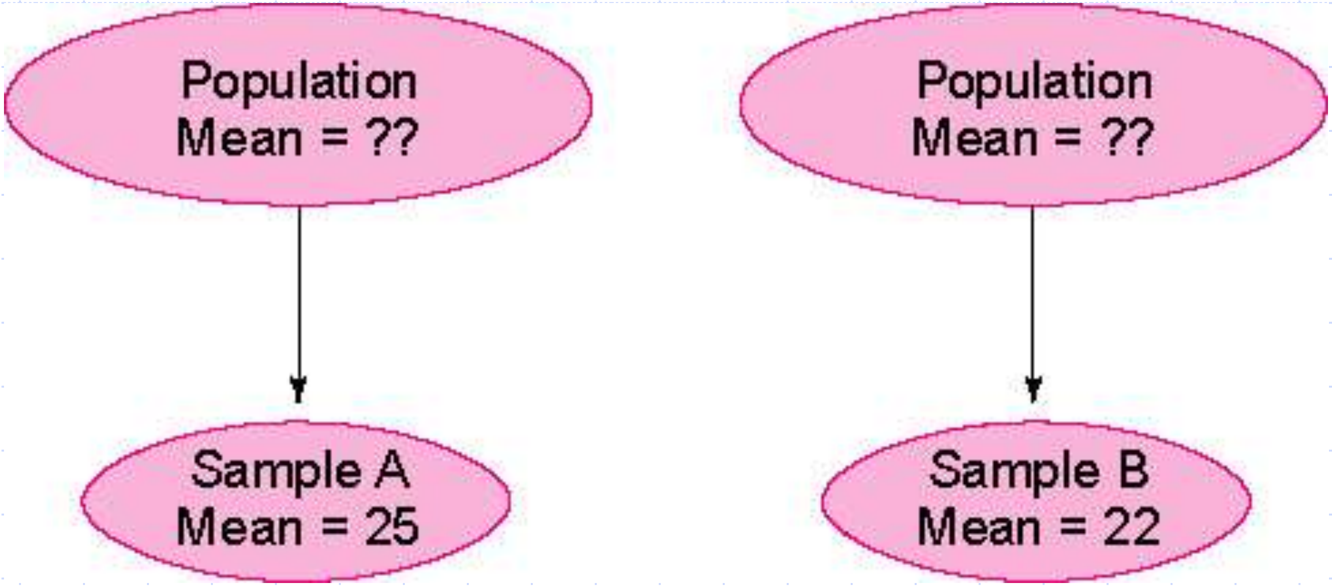
The 95 percent Confidence Interval



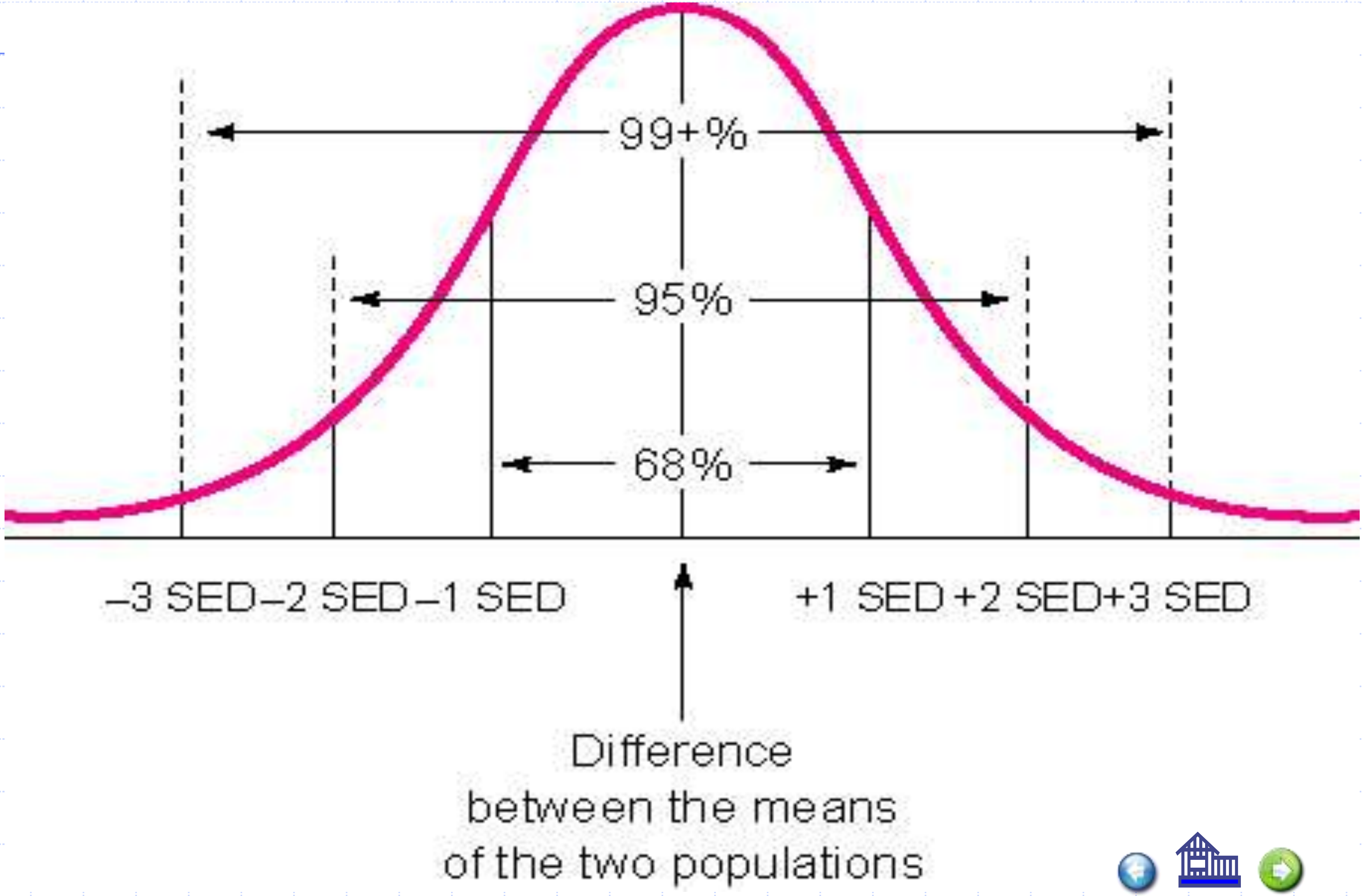
The 99 percent Confidence Interval



Does a Sample Difference Reflect a Population Difference?



Distribution of the Difference Between Sample Means



Confidence Intervals