THE NORMAL DISTRIBUTION, Z-SCORES, AND CONFIDENCE INTERVALS

FREQUENCY DISTRIBUTIONS

The histogram or distribution based on our real data

PROBABILITY DISTRIBUTIONS

Ideal/Theoretical

- Like frequency distribution *except* that its based on probability theory (not real data)
  - as if researcher (hypothetically) collected all possible scores

PROBABILITY THEORY

$$Probability\ of\ an\ event = \frac{number\ of\ times\ this\ event\ (object)\ appears}{total\ number\ of\ possible\ events\ (objects)}$$

PROBABILITY THEORY

Probability (*P*) varies from 0 to 1.0

- 0.0 = impossible
- 1.0 = certain

- Can be extremely unlikely (.01) to very probable (0.95)

DISTRIBUTIONS AND PROBABILITY

DIFFERENCES B/W PROBABILITY AND FREQUENCY DISTRIBUTIONS

Probability distribution is <u>theoretical</u> ideal

- If we flipped one coin an infinite number of times
- Things would even out to expected probability (.50 probability of heads)

- We can calculate the <u>empirical</u> probability (those based on trials)
  - Take the mean of the trials: heads == 1, tails == 0

ESTIMATES IN FREQUENCY AND PROBABILITY DISTRIBUTIONS

Probability we expect for a given event with (nearly) infinite numbers of any event

Frequency Distribution	Probability Distribution
$\bar{X}$	μ
SD or s	σ
SD^2	σ^2

NORMAL CURVE AS A PROBABILITY DISTRIBUTION

Normal Curve used to

- Demonstrate spread of scores (using Standard Deviation)
- Making statements of probability

CHARACTERISTICS OF THE NORMAL CURVE

Symmetrical (no skew)

Mesokurtic

Unimodal

Mean, median, mode all at center

Theoretical (probability distribution)

Probability Distribution
μ
σ
σ^2

AREA UNDER THE NORMAL CURVE

We know that 100% of all cases that exist fall under the curve

Normal curve has a mean of μ and a standard deviation of σ.

We can use the mean (μ) and standard deviation (σ) to break up normal curve

STANDARD DEVIATION AND AREA UNDER THE NORMAL CURVE

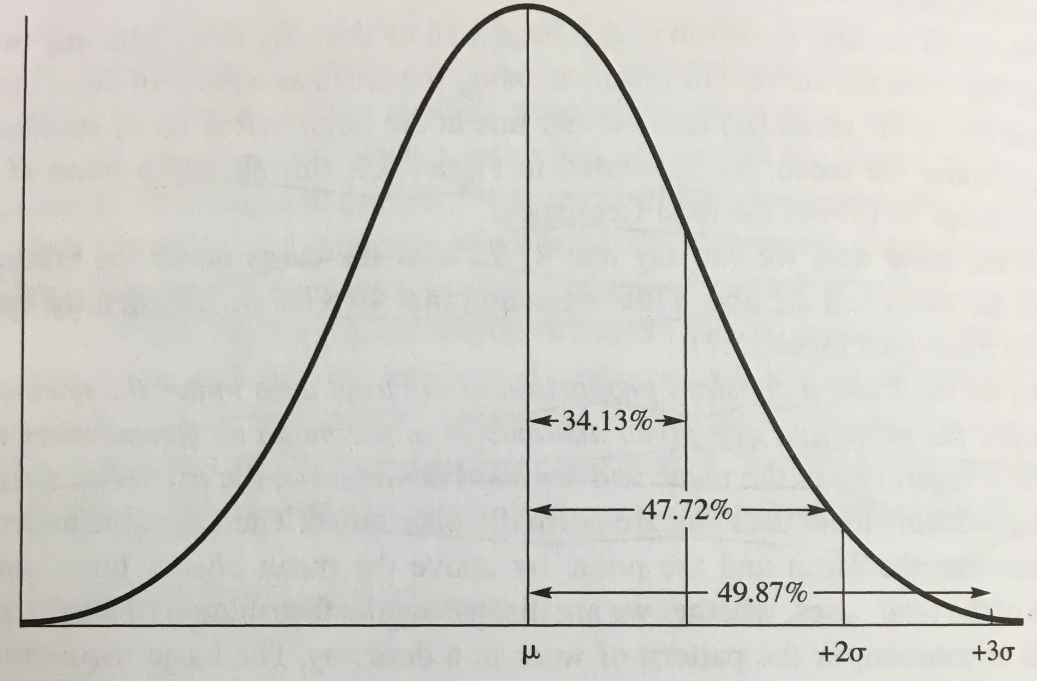

FIGURE 5.7 ***The Percent of Total Area under the Normal Curve between μ and the Points That are 1σ, 2σ, and 3σ from μ***

proportions? probabilities?

STANDARD DEVIATION AND AREA UNDER THE NORMAL CURVE

68.26% of cases fall between -1σ to +1σ

95.44% of cases fall between -2σ to +2σ

99.74% of cases fall between -3σ to +3σ

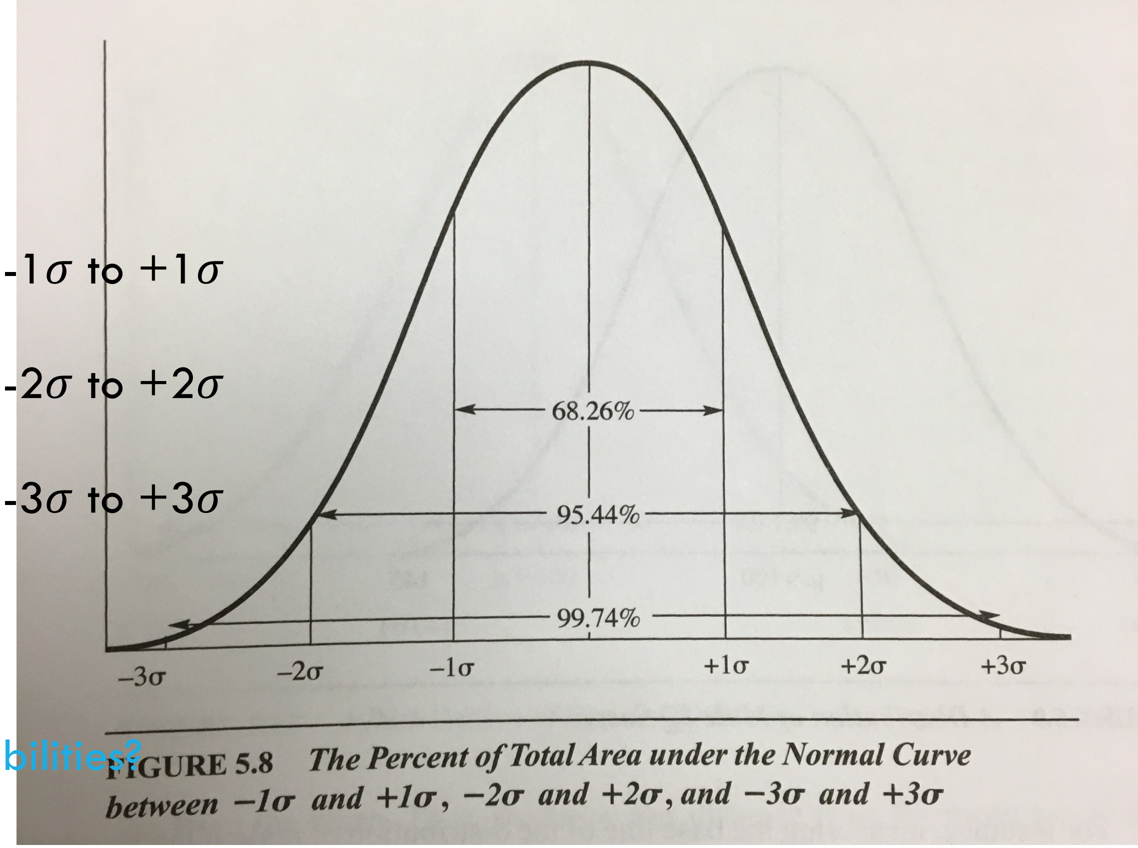

FIGURE 5.8 *The Percent of Total Area under the Normal Curve between -1σ and $+1\sigma$, -2σ and $+2\sigma$, and -3σ and $+3\sigma$*

proportions? probabilities?

STANDARD DEVIATION (BAND) AS A CONFIDENCE LEVEL

<u>68.26% confident</u> that the true population mean (μ) falls between -1σ to +1σ

<u>95.44% confident</u> that the true population mean (μ) falls between -2σ to +2σ

<u>99.74% confident</u> that the true population mean (μ) falls between -3σ to +3σ

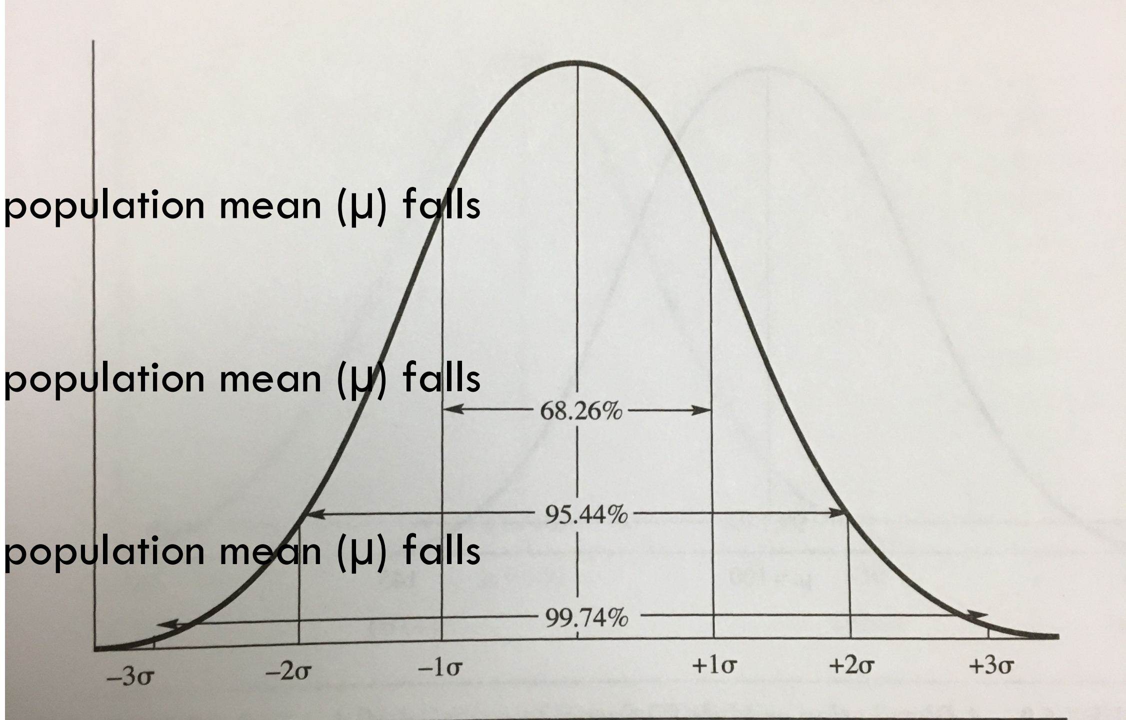

FIGURE 5.8 ***The Percent of Total Area under the Normal Curve between -1σ and $+1\sigma$, -2σ and $+2\sigma$, and -3σ and $+3\sigma$***

CONFIDENCE LEVELS AND THEIR CONFIDENCE INTERVALS

Confidence Levels

- Percentage of cases, around the mean, that are covered
- How confident that your CI range *includes* the population mean

Confidence Intervals

- Range, in <u>standard deviation units</u>, around the mean, that include the percentage of cases as defined by the Confidence Level

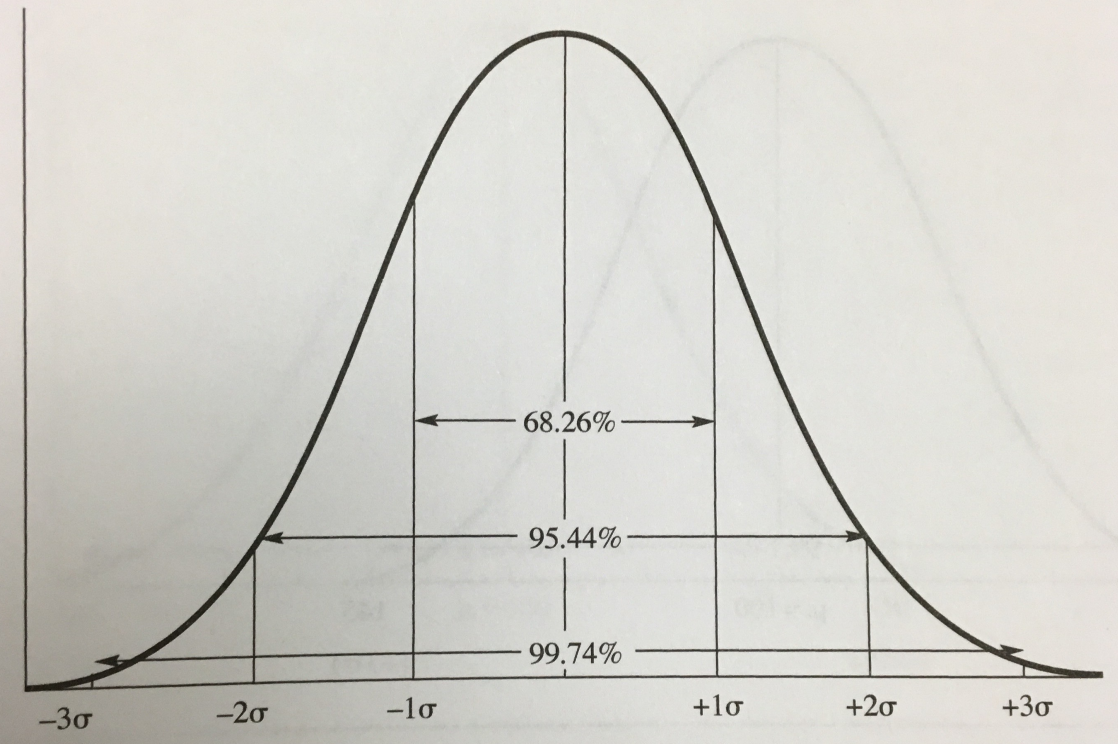

FIGURE 5.8 *The Percent of Total Area under the Normal Curve between -1σ and $+1\sigma$, -2σ and $+2\sigma$, and -3σ and $+3\sigma$*

CONFIDENCE LEVELS AND THEIR CONFIDENCE INTERVALS

68.26% Confidence Level

- Confidence Interval: -1σ to $+1\sigma$
  - CI Lower bound: $\mu - 1\sigma$
  - CI Upper bound: $\mu + 1\sigma$

95.44% Confidence Level

- Confidence Interval: $-\ 2\sigma$ to $+2\sigma$
  - CI Lower bound: $\mu - 2\sigma$
  - CI Upper bound: $\mu + 2\sigma$

99.74% Confidence Level

- Confidence Interval: $-\ 3\sigma$ to $+3\sigma$
  - CI Lower bound: $\mu - 3\sigma$
  - CI Upper bound: $\mu + 3\sigma$

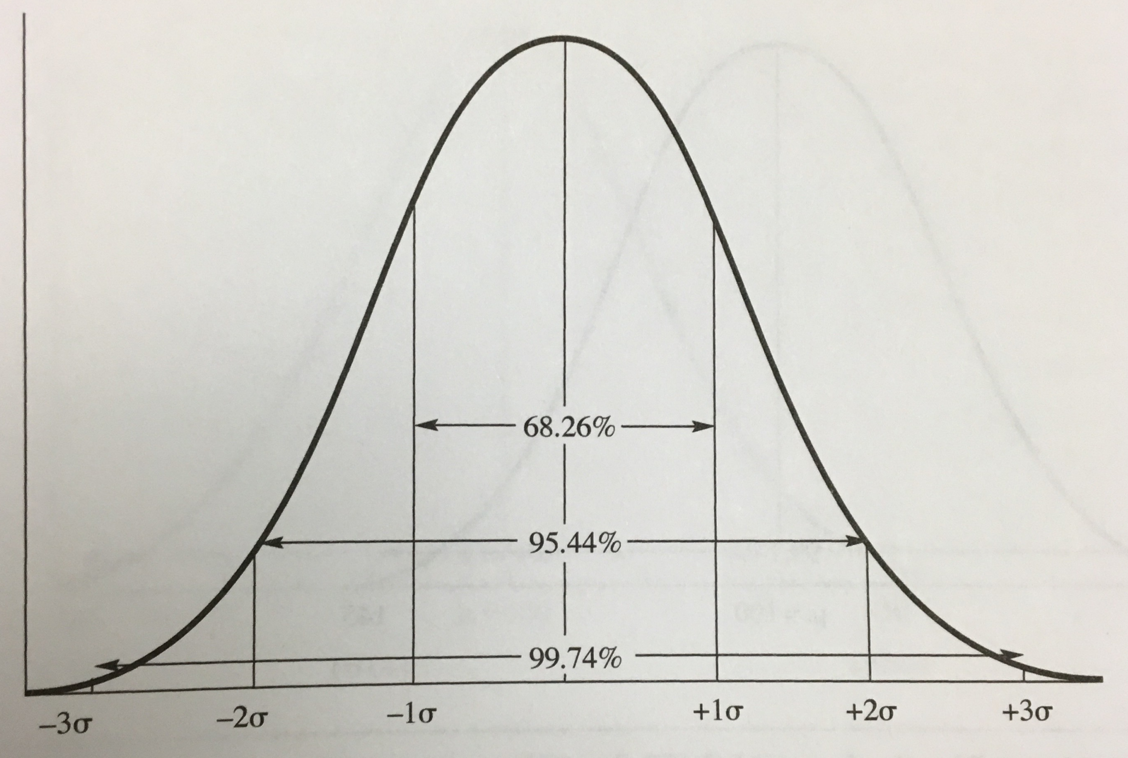

FIGURE 5.8 *The Percent of Total Area under the Normal Curve between -1σ and $+1\sigma$, -2σ and $+2\sigma$, and -3σ and $+3\sigma$*

CALCULATING CONFIDENCE INTERVALS FROM CONFIDENCE LEVELS

If you know the mean (μ) and standard deviation (σ), you can calculate actual values for

- the Confidence Interval (the Upper Bound and Lower Bound)
  - based on your Confidence Level

CONFIDENCE INTERVAL EXAMPLE: SEX DIFFERENCES IN IQ

Both men and women have mean IQ scores of 100, but men have more variability.

- Men have more low-IQ individuals and more high-IQ individuals
- Women have more individuals located toward the average

- Men $\sigma = 15$ points
- Women $\sigma = 10$ points

CONFIDENCE INTERVAL EXAMPLE: SEX DIFFERENCES IN IQ

68.26% of all male's IQ scores fall between which scores?

68.26% of all women's IQ scores fall between which scores?

95.44% of all male's IQ scores fall between which scores?

95.44% of all women's IQ scores fall between which scores?

99.74% of all male's IQ scores fall between which scores?

99.74% of all women's IQ scores fall between which scores?

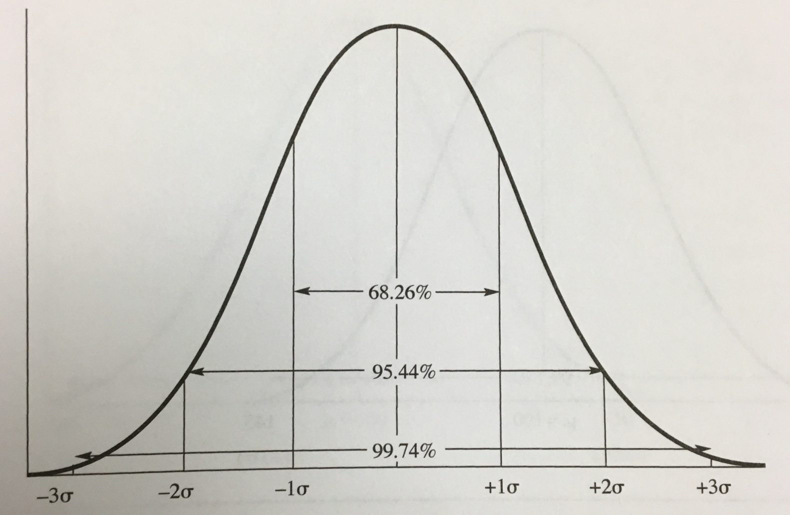

FIGURE 5.8 *The Percent of Total Area under the Normal Curve between −1σ and +1σ, −2σ and +2σ, and −3σ and +3σ*

CONFIDENCE INTERVAL EXAMPLE: SEX DIFFERENCES IN IQ

$\mu = 100$
$\sigma_{Male} = 15$
$\sigma_{Female} = 10$

68.26% of cases (68.26% Confidence Level)

- Lower bound: $\mu - 1\sigma$
- Upper bound: $\mu + 1\sigma$

95.44% of cases (95.44% Confidence Level)

- Lower bound: $\mu - 2\sigma$
- Upper bound: $\mu + 2\sigma$

99.74% of cases (99.74% Confidence Level)

- Lower bound: $\mu - 3\sigma$
- Upper bound: $\mu + 3\sigma$

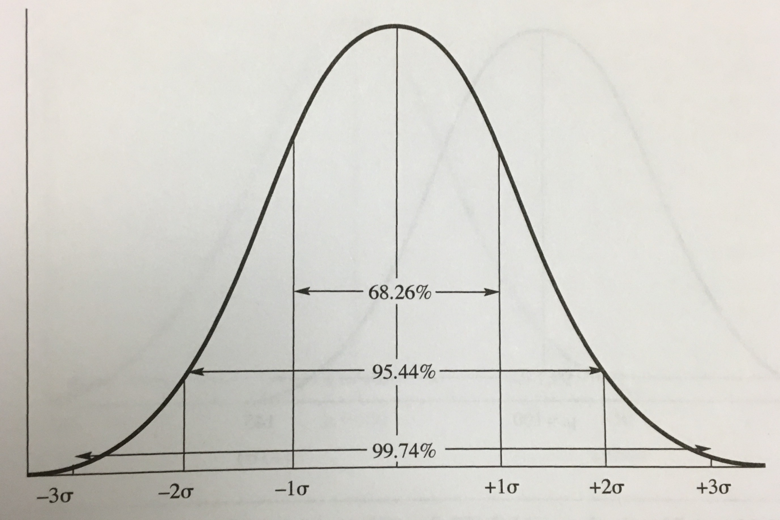

FIGURE 5.8 *The Percent of Total Area under the Normal Curve between -1σ and $+1\sigma$, -2σ and $+2\sigma$, and -3σ and $+3\sigma$*

CALCULATING CONFIDENCE INTERVALS FROM CONFIDENCE LEVELS

We can calculate confidence interval because we are using a standardized (standard deviation unit) score

Z-Score

CONFIDENCE LEVELS, CONFIDENCE INTERVALS, AND Z-SCORES

Z-score

- The number of Standard Deviation units (σ) a raw score (X) is from the mean (μ)

$$Z = \frac{X - \mu}{\sigma}$$

CONFIDENCE LEVELS, CONFIDENCE INTERVALS, AND Z-SCORES

$$Z = \frac{X - \mu}{\sigma}$$

68.26% Confidence Level ($Z = 1$)

- CI Lower bound: $\mu - 1\sigma$
- CI Upper bound: $\mu + 1\sigma$

95.44% Confidence Level ($Z = 2$)

- CI Lower bound: $\mu - 2\sigma$
- CI Upper bound: $\mu + 2\sigma$

99.74% Confidence Level ($Z = 3$)

- CI Lower bound: $\mu - 3\sigma$
- CI Upper bound: $\mu + 3\sigma$

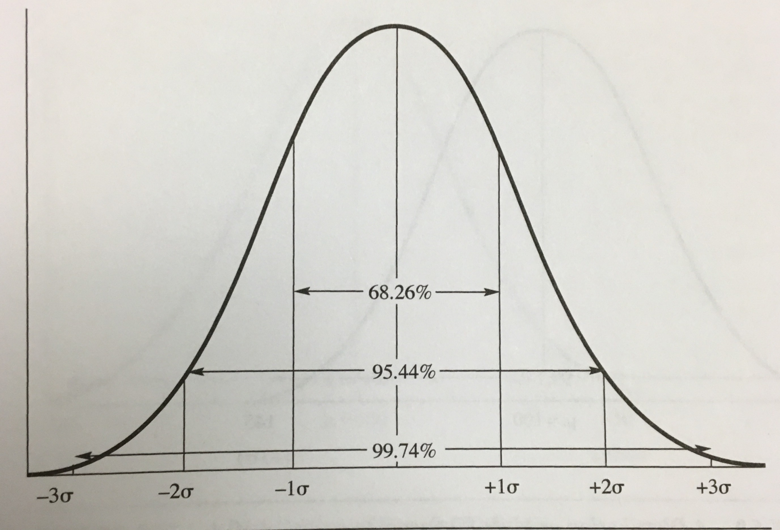

FIGURE 5.8 *The Percent of Total Area under the Normal Curve between -1σ and $+1\sigma$, -2σ and $+2\sigma$, and -3σ and $+3\sigma$*

CONFIDENCE LEVELS, CONFIDENCE INTERVALS, AND Z-SCORES

$$Z = \frac{X - \mu}{\sigma}$$

Calculate Confidence Intervals

- Use Z associated with Confidence Level

- $\mu \pm Z\sigma$

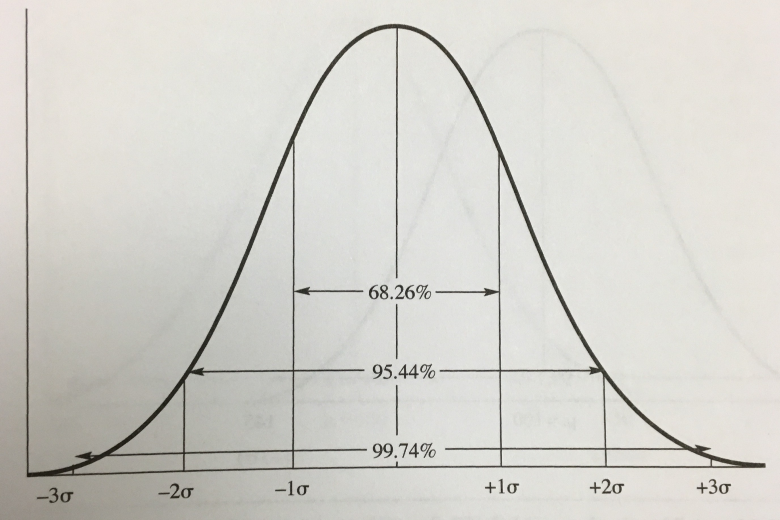

FIGURE 5.8 *The Percent of Total Area under the Normal Curve between -1σ and $+1\sigma$, -2σ and $+2\sigma$, and -3σ and $+3\sigma$*

CONFIDENCE INTERVAL EXAMPLE: SEX DIFFERENCES IN IQ

$\mu = 100$

$\sigma_{Male} = 15$

$\sigma_{Female} = 10$

68.26% of all male's IQ scores fall between which scores?
- Z?

68.26% of all women's IQ scores fall between which scores?
- Z?

95.44% of all male's IQ scores fall between which scores?
- Z?

95.44% of all women's IQ scores fall between which scores?
- Z?

99.74% of all male's IQ scores fall between which scores?
- Z?

99.74% of all women's IQ scores fall between which scores?
- Z?

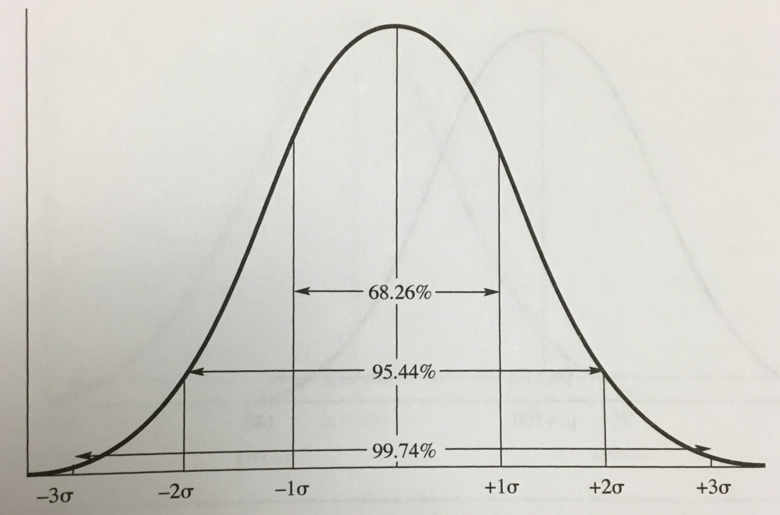

FIGURE 5.8 *The Percent of Total Area under the Normal Curve between -1σ and $+1\sigma$, -2σ and $+2\sigma$, and -3σ and $+3\sigma$*

PROBLEM: PROBABILITY DISTRIBUTIONS VS EMPIRICAL DISTRIBUTIONS

Often, we don't have data on the full population, so we won't have a probability distribution… so we don't know…

- μ
- σ

PROBLEM: PROBABILITY DISTRIBUTIONS VS EMPIRICAL DISTRIBUTIONS

Instead, we have data on a sample, so we know…

- $\bar{X}$ (not μ)
- SD (not σ)

PROBLEM: PROBABILITY DISTRIBUTIONS VS EMPIRICAL DISTRIBUTIONS

So we change our Z-Score formula to:

- $Z_i = \frac{X_i - \bar{X}}{SD}$ (instead of $Z = \frac{X - \mu}{\sigma}$)

Where Z_i is a Z score calculated for a specific raw score (X_i)

PROBLEM: PROBABILITY DISTRIBUTIONS VS EMPIRICAL DISTRIBUTIONS

We also change our CI formula to:

- $\bar{X} \pm Z\sigma_{\bar{X}}$ (instead of $\mu \pm Z\sigma$)

Where $\sigma_{\bar{X}}$ (*standard error of the mean of the sampling distribution*), or Standard Error (SE) is an estimate of the population standard deviation (σ), and is calculated as such:

$$\sigma_{\bar{X}} = \text{SE} = \frac{SD}{\sqrt{N}}$$

MORE ON CONFIDENCE INTERVALS (FOR SAMPLES)

However, 68.26% confident isn't that confident. We're only going to be right 68.26% of the time.

To have a better probability of finding the true population mean, we need a wider and less precise interval

- 95% confidence interval (only wrong 5% of the time)
- 99% confidence interval (only wrong 1% of the time)

MORE ON CONFIDENCE INTERVALS (FOR SAMPLES)

Confidence Interval for 95% Confidence Level is $\mu \pm 2\sigma$, not exactly 2σ

- exact Confidence Level is 95.44%

Confidence Interval for 99% Confidence Level is $\mu \pm 3\sigma$, not exactly 3σ

- exact Confidence Level is 99.74%

MORE ON CONFIDENCE INTERVALS (FOR SAMPLES)

CI for 95% Confidence Level

- Within $1.96\sigma_{\bar{X}}$ of mean
  - $\bar{X} \pm 1.96\sigma_{\bar{X}}$
- 95% confident that true population mean falls within this range

CI for 99% Confidence Level

- Within $2.58\sigma_{\bar{X}}$ of mean
  - $\bar{X} \pm 2.58\sigma_{\bar{X}}$
- 99% confident that true population mean falls within this range