Chapter 4
Variability

PowerPoint Lecture Slides

Essentials of Statistics for the Behavioral Sciences
Seventh Edition

by Frederick J Gravetter and Larry B. Wallnau

Chapter 4 Learning Outcomes

1. Understand purpose of measuring variability
2. Define range and interquartile range
3. Compute range and interquartile range
4. Understand standard deviation
5. Calculate SS, variance, standard deviation of population
6. Calculate SS, variance, standard deviation of sample

Concepts to review

- Summation notation (Chapter 1)
- Central tendency (Chapter 3)
  - Mean
  - Median

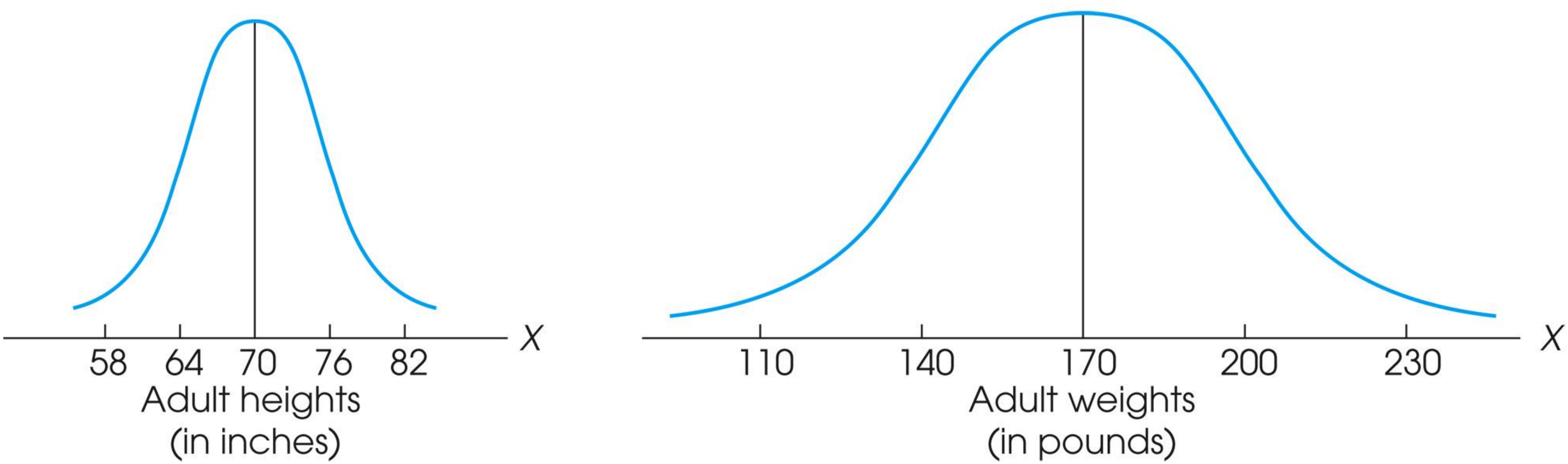
4.1 Overview

- **Variability** can be defined several ways
  - A quantitative measure of the differences between scores
  - Describes the degree to which the scores are spread out or clustered together
- Purposes of Measure of Variability
  - Describe the distribution
  - Measure how well an individual score represents the distribution

Three Measures of Variability

- The Range
- The Standard Deviation
- The Variance

Figure 4.1
Population Distributions: Height, Weight



4.2 The Range

- The distance covered by the scores in a distribution
  - From smallest value to highest value
- For continuous data, real limits are used

$$\text{range} = \text{URL for } X_{max} - LRL \textit{ for } X_{min}$$

- Based on two scores, not all the data
  - Considered a crude, unreliable measure of variability

4.3 Standard Deviation and Variance for a Population

- Most common and most important measure of variability
  - A measure of the standard, or average, distance from the mean
  - Describes whether the scores are clustered closely around the mean or are widely scattered
- Calculation differs for population and samples

Developing the Standard Deviation

- **Step One**: Determine the Deviation
- Deviation is distance from the mean

$$\text{Deviation score} = X - \mu$$

- **Step Two**: Calculate Mean of Deviations
  - Deviations sum to 0 because *M* is balance point of the distribution
  - The Mean Deviation will always equal 0; another method must be found

Developing the Standard Deviation (2)

- **Step Three**: Get rid of + and – in Deviations
  - Square each deviation score
  - Compute the Mean Squared Deviation, known as the Variance

Population variance equals the mean squared deviation. Variance is the average squared distance from the mean

- Variability is now measured in **squared units**

Developing the Standard Deviation (2)

- **Step Four:**
  - Goal: to compute a measure of the standard distance of the scores from the mean
  - Variance measures the average squared distance from the mean; not quite on goal
- Correct for having squared all the differences by taking the square root of the variance

$$Standard\ Deviation = \sqrt{Variance}$$

Figure 4.2
Calculation of the Variance

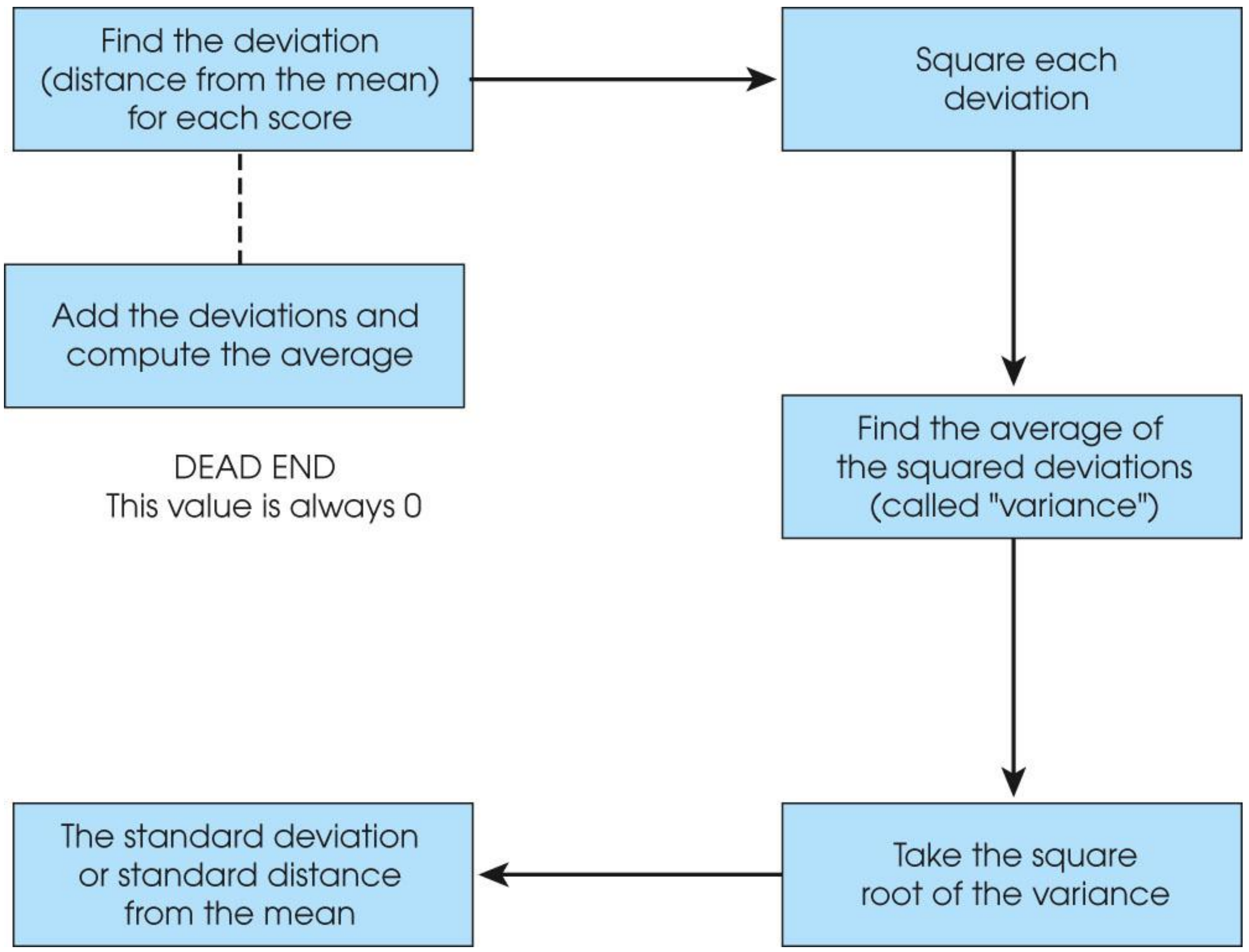
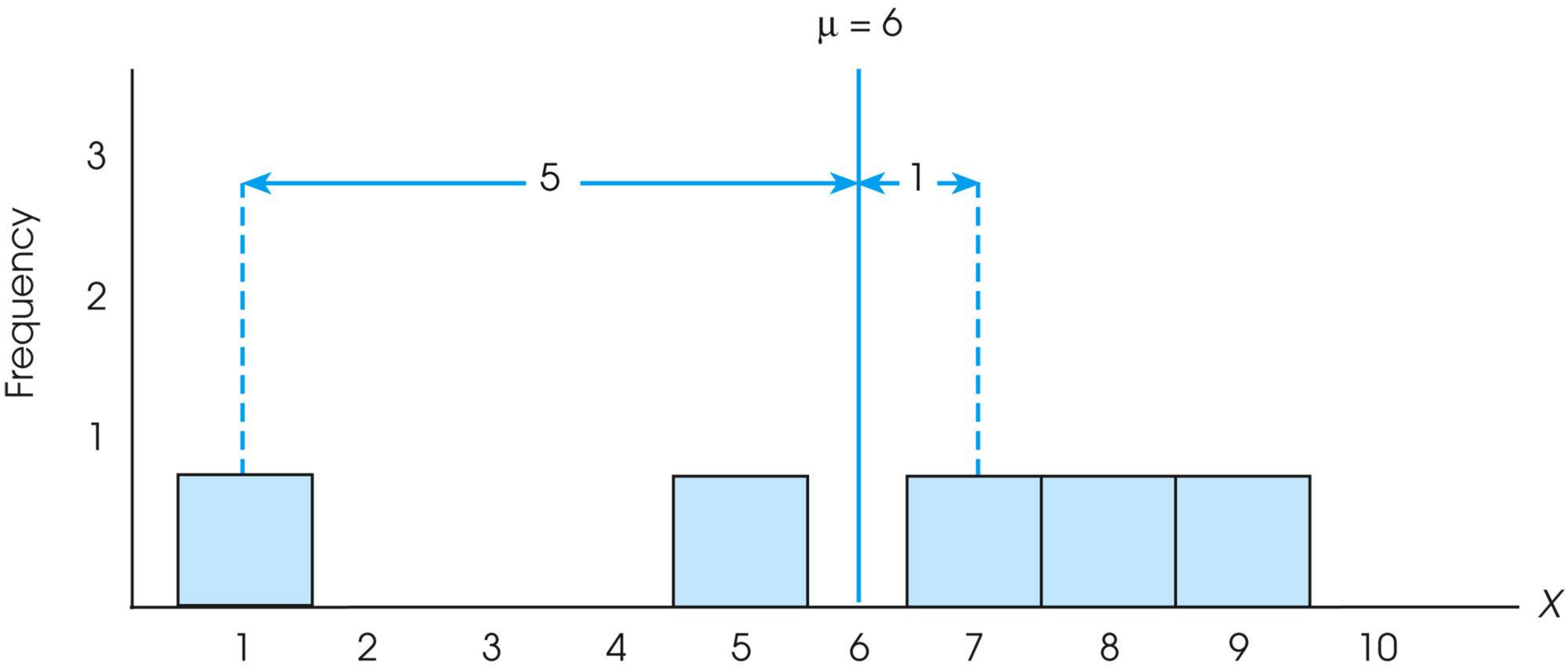


Figure 4.3

Frequency Distribution Histogram for Population



Formulas for Population Variance and Standard Deviation

- $$Variance = \frac{sum\ of\ squared\ deviations}{number\ of\ scores}$$

- *SS* (sum of squares) is the sum of the squared deviations of scores from the mean

- Two equations for computing *SS*

Two formulas for *SS*

Definitional Formula	Computational Formula
• Find each deviation score ($X-\mu$) • Square each deviation score, $(X-\mu)^2$ • Sum up the squared deviations	• Square each score and sum the squared scores • Find the sum of scores, square it, divide by N • Subtract the second part from the first
$SS = \sum (X-\mu)^2$	$SS = \sum X^2 - \frac{\left(\sum X\right)^2}{N}$

Population Variance: Formula and Notation

Formula

$$variance = \frac{SS}{N}$$

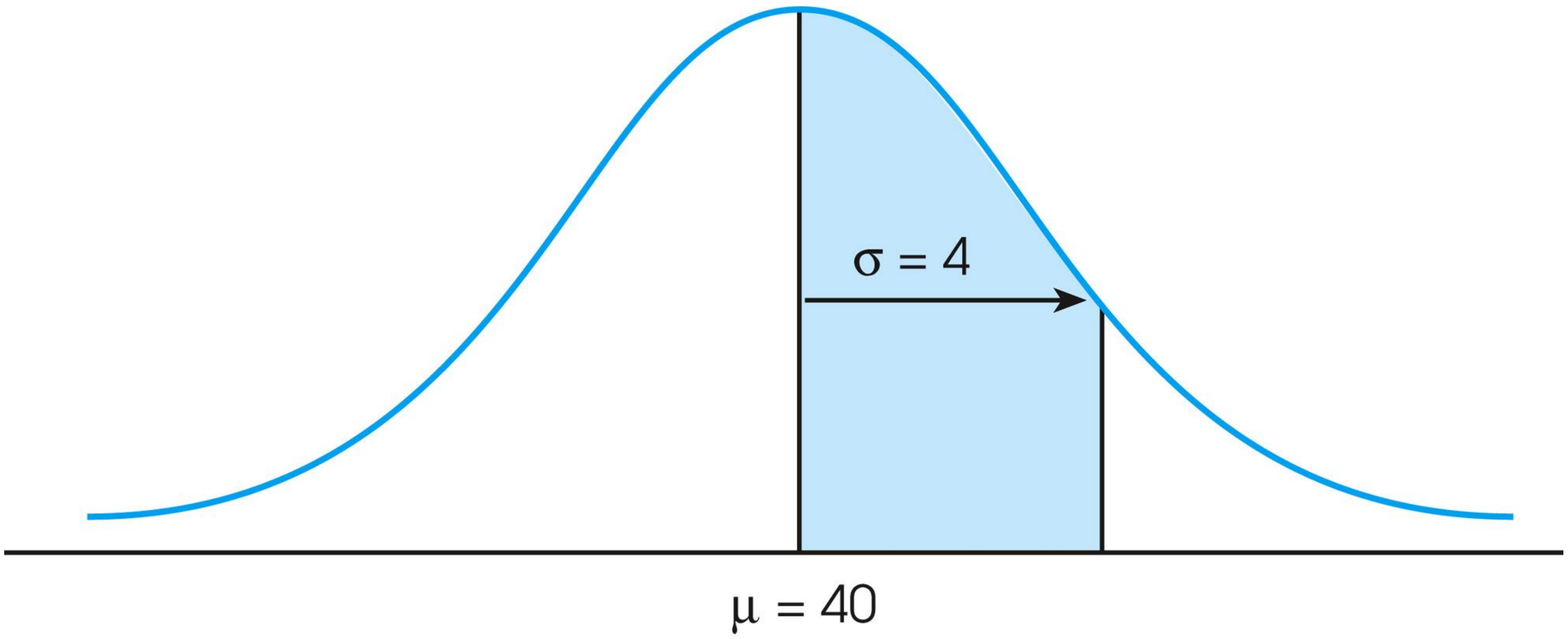
$$standard\,deviation = \sqrt{\frac{SS}{N}}$$

Notation

- Lowercase Greek letter sigma is used to denote the standard deviation of a population:
  σ
- Because the standard deviation is the square root of the variance, we write the variance of a population as σ^2

Figure 4.4

Graphic Representation of Mean and Standard Deviation



Learning Check

- Decide if each of the following statements is **True** or **False**.

T/F

- **The computational & definitional formulas for *SS* sometimes give different results.**

T/F

- **If all the scores in a data set are the same, the Standard Deviation is equal to 1.00.**

Learning Check - Answer

False

- **The computational formula is just an algebraic rearrangement of the definitional formula. Results are identical.**

False

- **When all the scores are the same, they are all equal to the mean. Their deviations = 0, as does their Standard Deviation.**

Learning Check

- The standard deviation measures …

A	Sum of squared deviation scores
B	Standard distance of a score from the mean
C	Average deviation of a score from the mean
D	Average squared distance of a score from the mean

Learning Check - Answer

- The standard deviation measures …

A
- Sum of squared deviation scores

B
- **Standard distance of a score from the mean**

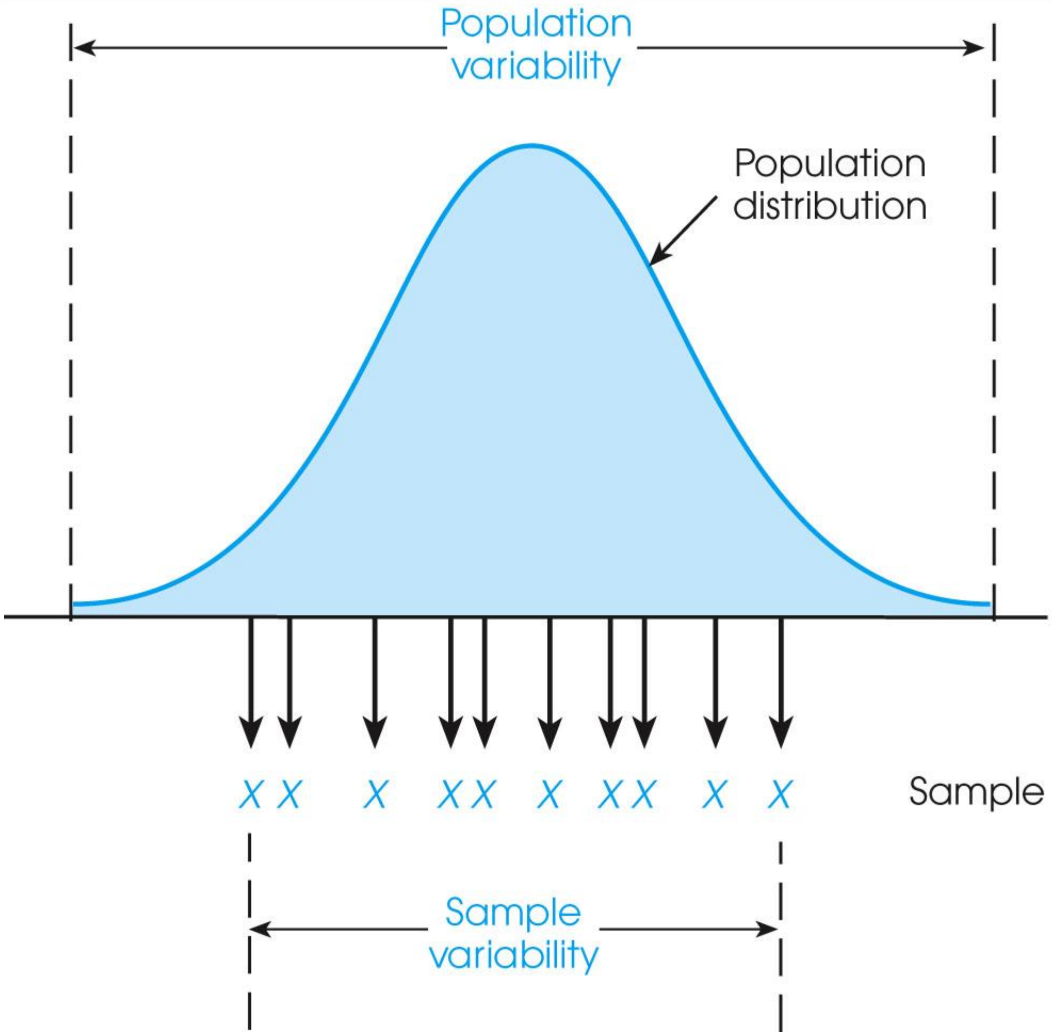
C
- Average deviation of a score from the mean

D
- Average squared distance of a score from the mean

4.4 Standard Deviation and Variance for a Sample

- Goal of inferential statistics:
  - Draw general conclusions about population
  - Based on limited information from a sample
- Samples differ from the population
  - Samples have less variability
  - Computing the Variance and Standard Deviation in the same way as for a population would give a *biased* estimate of the population values

Figure 4.5
Population of Adult Heights



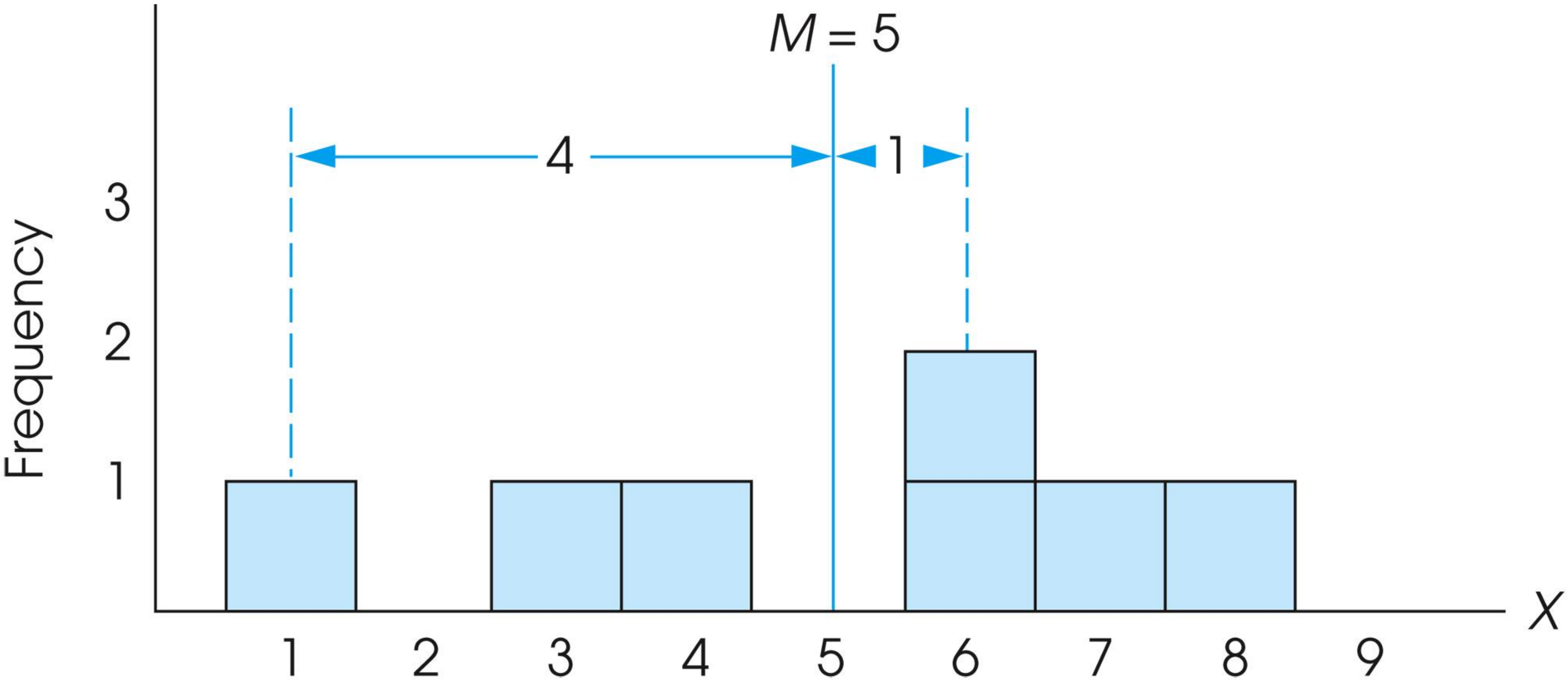
Variance and Standard Deviation for a Sample

- Sum of Squares (*SS*) is computed as before
- Formula for Variance has *n-1* rather than *N* in the denominator
- Notation uses *s* instead of σ

$$\text{variance of sample} = s^2 = \frac{SS}{n-1}$$

$$\text{standard deviation of sample} = s = \sqrt{\frac{SS}{n-1}}$$

Figure 4.6
Frequency Distribution of a Sample



Degrees of Freedom

- Population variance
  - Mean is known
  - Deviations are computed from a known mean
- Sample variance as estimate of population
  - Population mean is unknown
  - Using sample mean restricts variability
- Degrees of freedom
  - Number of scores in sample that are independent and free to vary
  - Degrees of freedom (*df*) = $n - 1$

Learning Check

- A sample of four scores has *SS* = 24. What is the variance?

A	• **The variance is 6**
B	• **The variance is 7**
C	• **The variance is 8**
D	• **The variance is 12**

Learning Check - Answer

- A sample of four scores has $SS = 24$. What is the variance?

A	The variance is 6
B	The variance is 7
C	**The variance is 8**
D	The variance is 12

Learning Check

- Decide if each of the following statements is **True** or **False**.

T/F

- **A sample systematically has less variability than a population**

T/F

- **The standard deviation is the distance from the Mean to the farthest point on the distribution curve**

Learning Check - Answer

True

- Extreme scores affect variability, but are less likely to be included in a sample

False

- The standard deviation extends from the mean approximately halfway to the most extreme score

4.5 More about Variance and Standard Deviation

- *Unbiased* estimate of a population parameter
  - Average value of statistic is equal to parameter
  - Average value uses all possible samples of a particular size *n*
- *Biased* estimate of a population parameter
  - Systematically overestimates or underestimates the population parameter

Figure 4.7
Sample of *n* = 20, *M* = 36, and *s* = 4

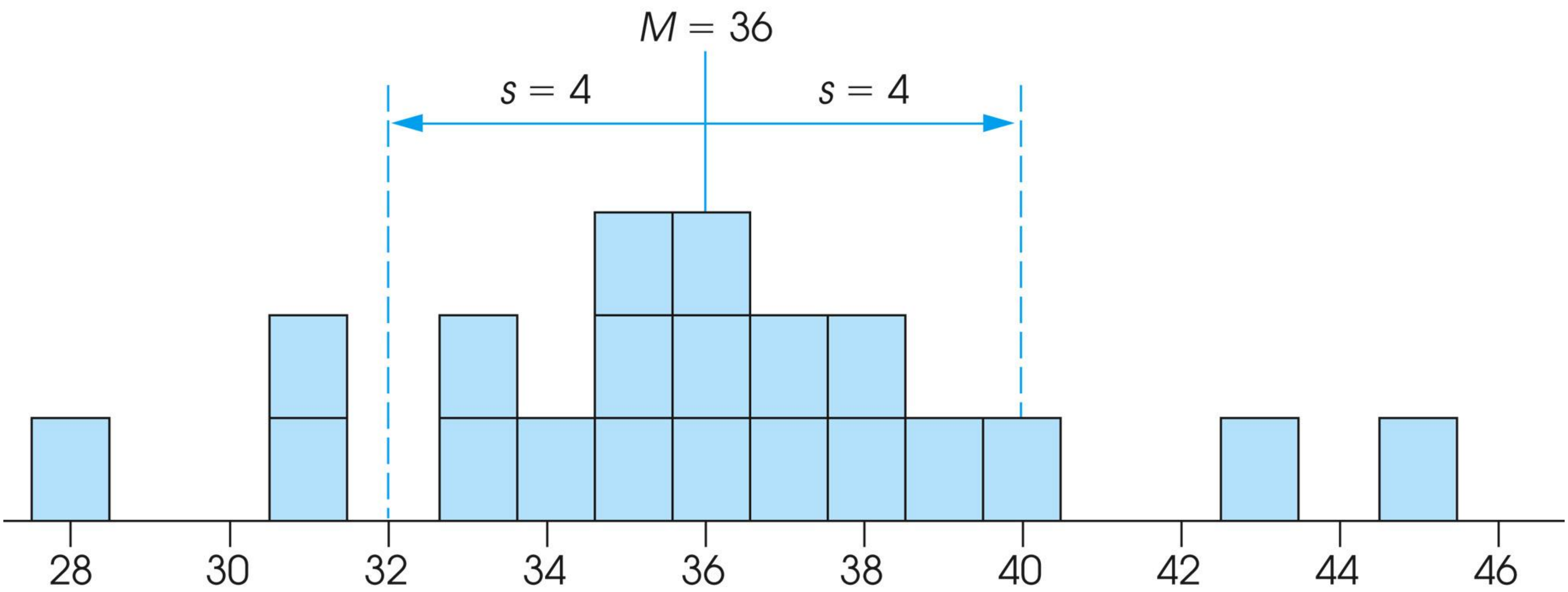


Table 4.1 Biased & Unbiased Estimates

			Sample Statistics		
Sample	1st Score	2nd Score	Mean	Biased Variance	Unbiased Variance
1	0	0	0.00	0.00	0.00
2	0	3	1.50	2.25	4.50
3	0	9	4.50	20.25	40.50
4	3	0	1.50	2.25	4.50
5	3	3	3.00	0.00	0.00
6	3	9	6.00	9.00	18.00
7	9	0	4.50	20.25	40.50
8	9	3	6.00	9.00	18.00
9	9	9	9.00	0.00	0.00
		Totals	**36.00**	**63.00**	**126.00**
					Actual $\boldsymbol{\sigma^2}$ **= 14**

This is an adaptation of Table 4.1

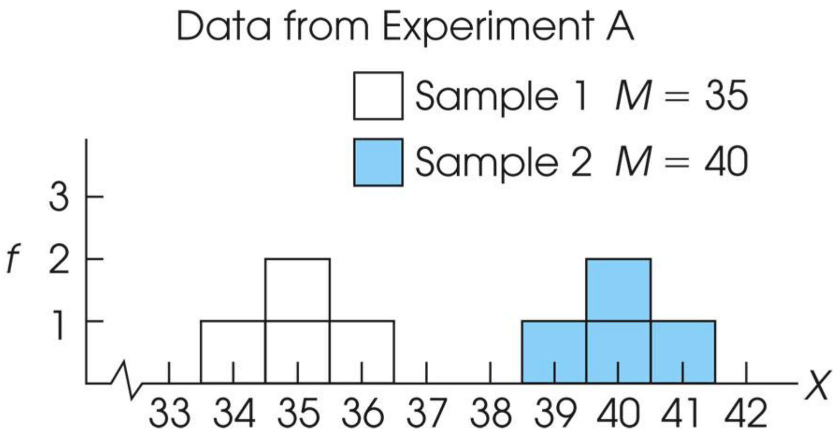
Transformations of Scale

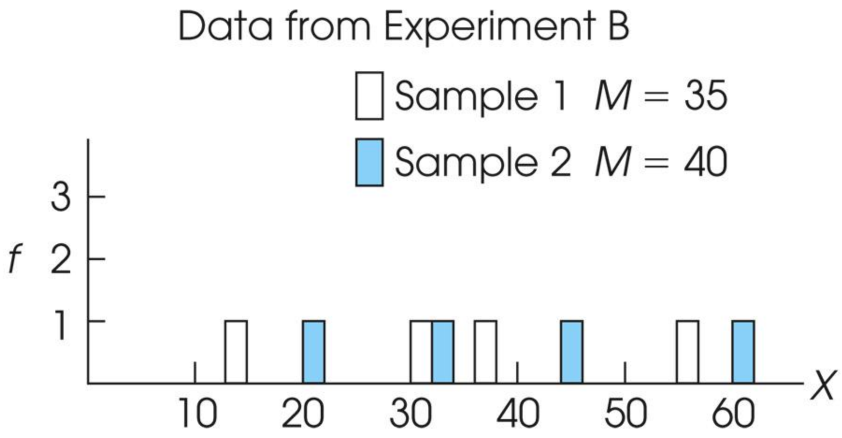
- Adding a constant to each score
  - The Mean is changed
  - The standard deviation is unchanged
- Multiplying each score by a constant
  - The Mean is changed
  - Standard Deviation is also changed
  - The Standard Deviation is multiplied by that constant

Variance and Inferential Statistics

- Goal of inferential statistics is to detect meaningful and significant patterns in research results
- Variability in the data influences how easy it is to see patterns
  - High variability obscures patterns that would be visible in low variability samples
  - Variability is sometimes called *error variance*

Figure 4.8
Experiments with high and low variability





Learning Check

A population has $\mu = 6$ and $\sigma = 2$. Each score is multiplied by 10. What is the shape of the resulting distribution?

A
- **$\mu = 60$ and $\sigma = 2$**

B
- **$\mu = 6$ and $\sigma = 20$**

C
- **$\mu = 60$ and $\sigma = 20$**

D
- **$\mu = 6$ and $\sigma = 5$**

Learning Check - Answer

A population has $\mu = 6$ and $\sigma = 2$. Each score is multiplied by 10. What is the shape of the resulting distribution?

A
- $\mu = 60$ and $\sigma = 2$

B
- $\mu = 6$ and $\sigma = 20$

C
- **$\mu = 60$ and $\sigma = 20$**

D
- $\mu = 6$ and $\sigma = 5$

Learning Check TF

- Decide if each of the following statements is **True** or **False**.

T/F

- **A biased statistic has been influenced by researcher error**

T/F

- **On average, an unbiased sample statistic has the same value as the population parameter**

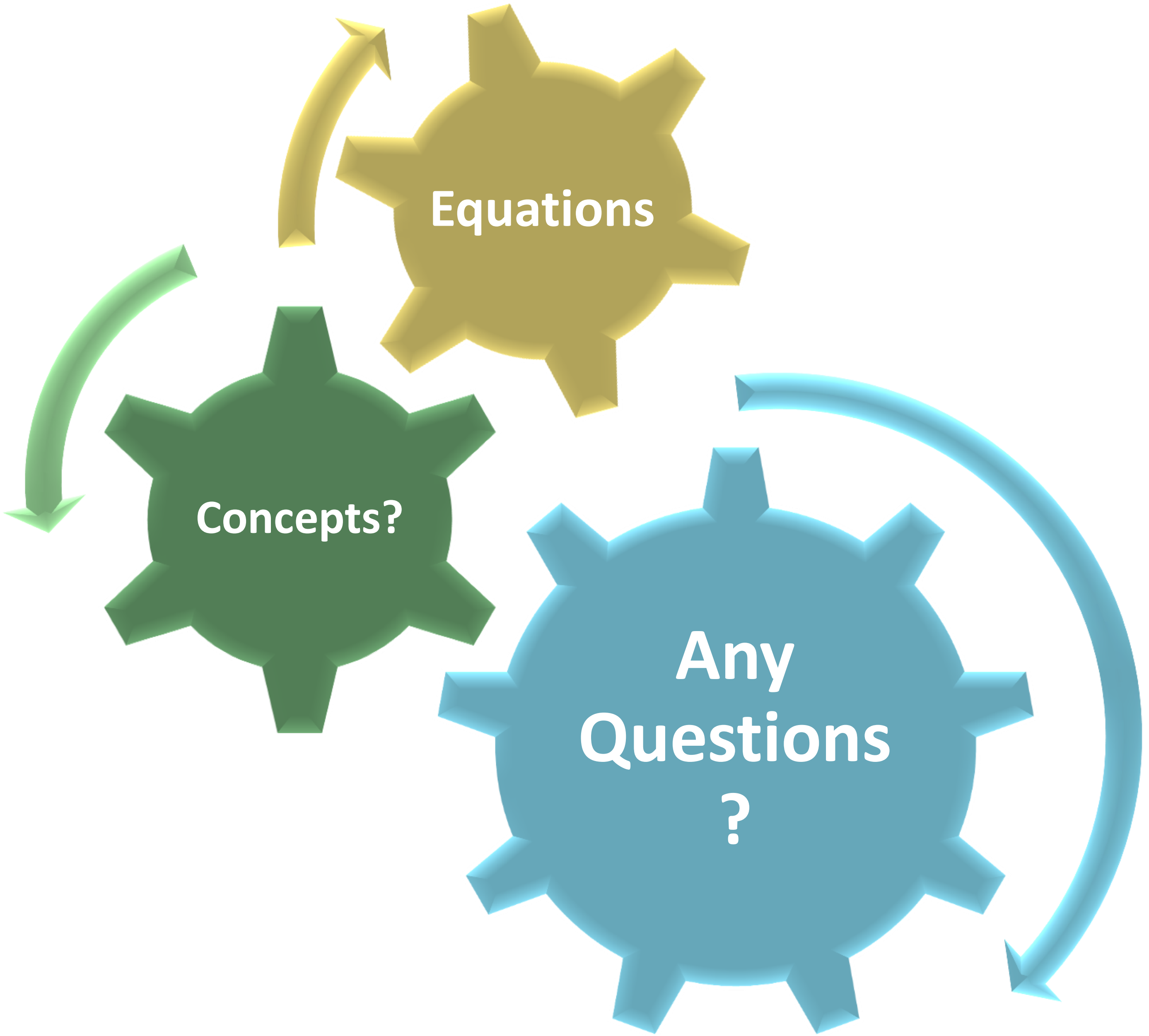
Learning Check - Answer

False

- Bias refers to the systematic effect of using sample data to estimate a population parameter

True

- Each sample's statistic differs from the population parameter, but the average of all samples will equal the parameter.


Equations
Concepts?
Any
Questions
?