Applied Statistics for Public Policy

Lecture #3

Cronbach's Alpha

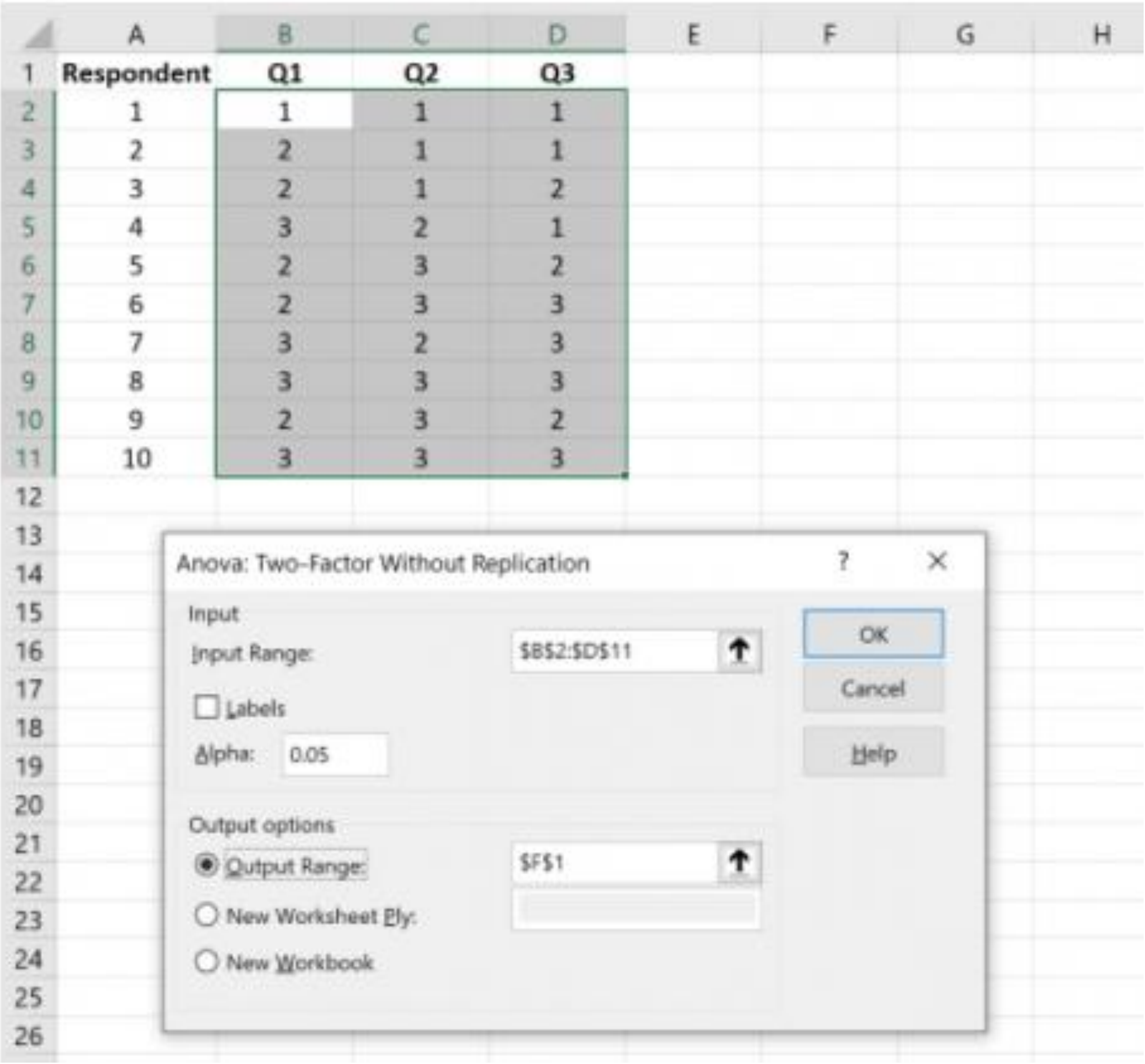
- Step #1: Enter the Data, e.g. Customer satisfaction survey on a questionnaire:

Respondent	Q1	Q2	Q3
1	1	1	1
2	2	1	1
3	2	1	2
4	3	2	1
5	2	3	2
6	2	3	3
7	3	2	3
8	3	3	3
9	2	3	2
10	3	3	3

Cronbach Alpha

- Step 2: **Perform a Two-Factor ANOVA Without Replication**
  - Click the **Data** tab along the top ribbon and then click the **Data Analysis** option under the **Analysis** group

Cont.



	A	B	C	D	E	F	G	H
1	Respondent	Q1	Q2	Q3				
2	1	1	1	1				
3	2	2	1	1				
4	3	2	1	2				
5	4	3	2	1				
6	5	2	3	2				
7	6	2	3	3				
8	7	3	2	3				
9	8	3	3	3				
10	9	2	3	2				
11	10	3	3	3				

Result of ANNOVA

Anova: Two-Factor Without Replication

SUMMARY	*Count*	*Sum*	*Average*	*Variance*
Row 1	3	3	1	0
Row 2	3	4	1.333333	0.333333
Row 3	3	5	1.666667	0.333333
Row 4	3	6	2	1
Row 5	3	7	2.333333	0.333333
Row 6	3	8	2.666667	0.333333
Row 7	3	8	2.666667	0.333333
Row 8	3	9	3	0
Row 9	3	7	2.333333	0.333333
Row 10	3	9	3	0
Column 1	10	23	2.3	0.455556
Column 2	10	22	2.2	0.844444
Column 3	10	21	2.1	0.766667

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	12.8	9	1.422222	4.413793	0.003591	2.456281
Columns	0.2	2	0.1	0.310345	0.737039	3.554557
Error	5.8	18	0.322222			
Total	18.8	29				

Cont.

- Step 3: **Calculate Cronbach's Alpha**

Anova: Two-Factor Without Replication

SUMMARY	*Count*	*Sum*	*Average*	*Variance*
Row 1	3	3	1	0
Row 2	3	4	1.333333	0.333333
Row 3	3	5	1.666667	0.333333
Row 4	3	6	2	1
Row 5	3	7	2.333333	0.333333
Row 6	3	8	2.666667	0.333333
Row 7	3	8	2.666667	0.333333
Row 8	3	9	3	0
Row 9	3	7	2.333333	0.333333
Row 10	3	9	3	0
Column 1	10	23	2.3	0.455556
Column 2	10	22	2.2	0.844444
Column 3	10	21	2.1	0.766667

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	12.8	9	1.422222	4.413793	0.003591	2.456281
Columns	0.2	2	0.1	0.310345	0.737039	3.554557
Error	5.8	18	0.322222			
Total	18.8	29				

Cronbach's Alpha	0.773438	=1-(I24/I22)

Cont.

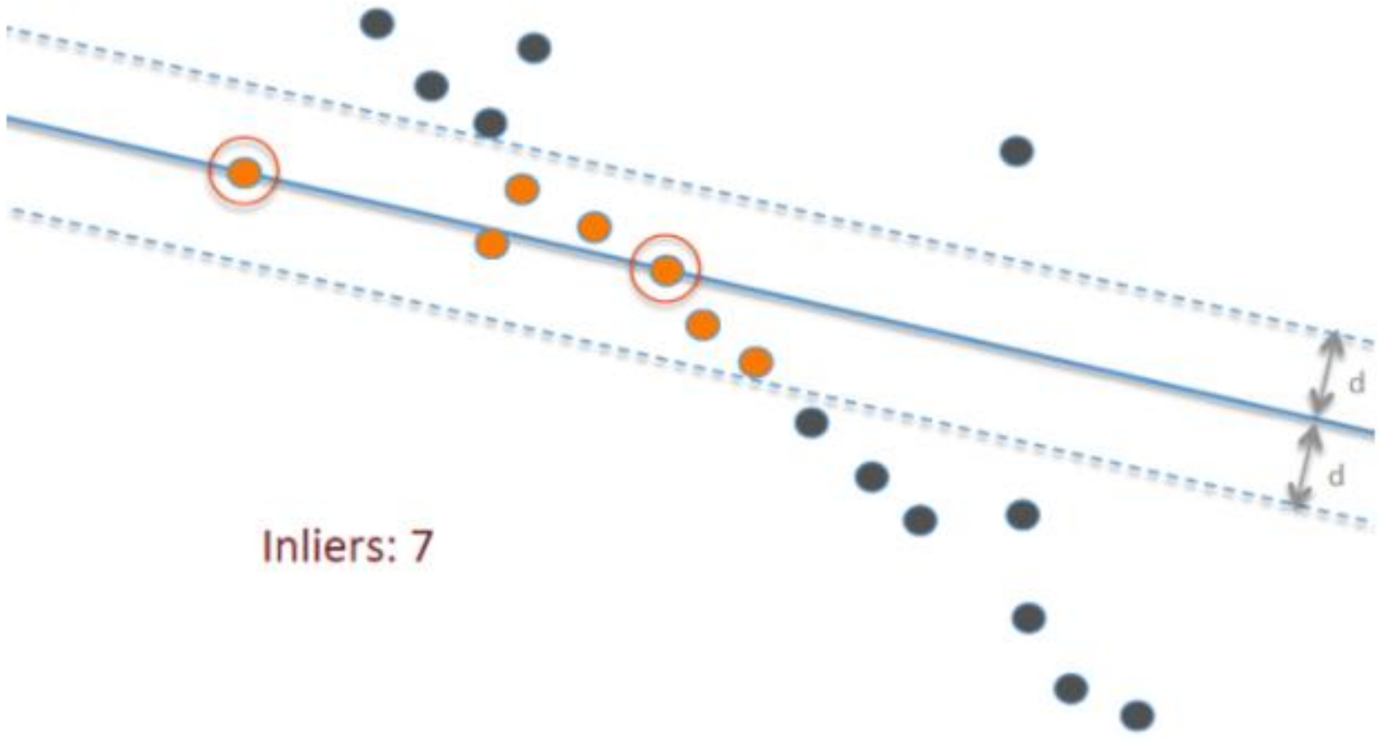
- Step 4: **Decision Rule**

Cronbach's Alpha	Internal consistency
$0.9 \leq \alpha$	Excellent
$0.8 \leq \alpha < 0.9$	Good
$0.7 \leq \alpha < 0.8$	Acceptable
$0.6 \leq \alpha < 0.7$	Questionable
$0.5 \leq \alpha < 0.6$	Poor
$\alpha < 0.5$	Unacceptable

Dealing with Errors in Measurement

- In reality, there is always some possibility that the number assigned does not reflect the true value for that case, i.e.:
  - Human Error e.g. 100 instead of 10
  - Mistakes in coding,
  - Subjective judgments,
  - Measuring instrument that lacks precision.
- How to overcome?
  - Bootstrapping
  - Detecting Outliers (keeping inliers)—see next slide
  - **Refer to practice session in STATA**

How many outliers to drop?

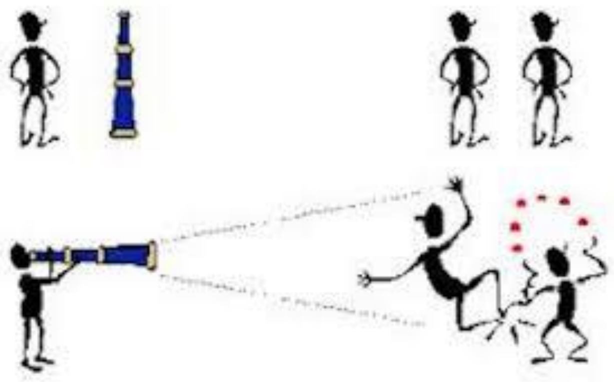


Types of Measures

- **Subjective Measure:** The subjective indicator requires some judgment to assign a value
- **Objective Measure:** Objective indicator seeks to minimize the discretion of judgment.
- **Example:**
- Assume that CDA wants to know the number of city services delivered to each Sector in Islamabad. Objective measures of city services would be ***acres of city parks***, ***number of tons of trash collected, number of police patrols***, and so on.
- Subjective measures of city services could be obtained by asking citizens whether the levels of various city services were **adequate**.

Types of Measures

- **Unobtrusive Indicator:**
- Is the third type of measure that is normally used to avoid the so-called ***Hawthorne effect*** *in the measurement.*
- In the Hawthorne studies, employees who were observed by a research team seemed to change their workplace behavior as a result of being observed.
- **Example:** Think about yourself? Presenting something in Infront of the mirror is much different than presenting the same thing in Infront of a live TV camera
- So what unobtrusive indicator does?



Example of Unobtrusive Measure

1. Indirect Measure

- **A study of radio station listening preferences**
- The researchers went to local auto dealers and garages and checked all cars that were being serviced to see what station the radio was currently tuned to.
- In a similar manner, if you want to know magazine preferences, you might rummage through the trash of your sample or even stage a door-to-door magazine recycling effort.
- **Caution!**
- In doing so, you may be violating their right to privacy, and you are certainly not using informed consent.

Types of Unobtrusive Measures

2. Content Analysis: *the analysis of text documents*

- Can be quantitative, qualitative or both

- **Sub-types:**

  1. **Thematic analysis of the text:** The identification of themes or major ideas in a document or set of documents.

  2. **Indexing text Document:** For instance, **Key Words in Context (KWIC)** analysis is a computer analysis of text data. A computer program scans the text and indexes all keywords.

Types of Unobtrusive Measures

3. Secondary Analysis of Data: Typically refers to the ***re-analysis*** of quantitative data rather than text.

- In our modern world, there is an unbelievable mass of data that is routinely collected by governments, businesses, schools, and other organizations.
- Much of this information is stored in **electronic databases** that can be accessed and analyzed.

Examples:

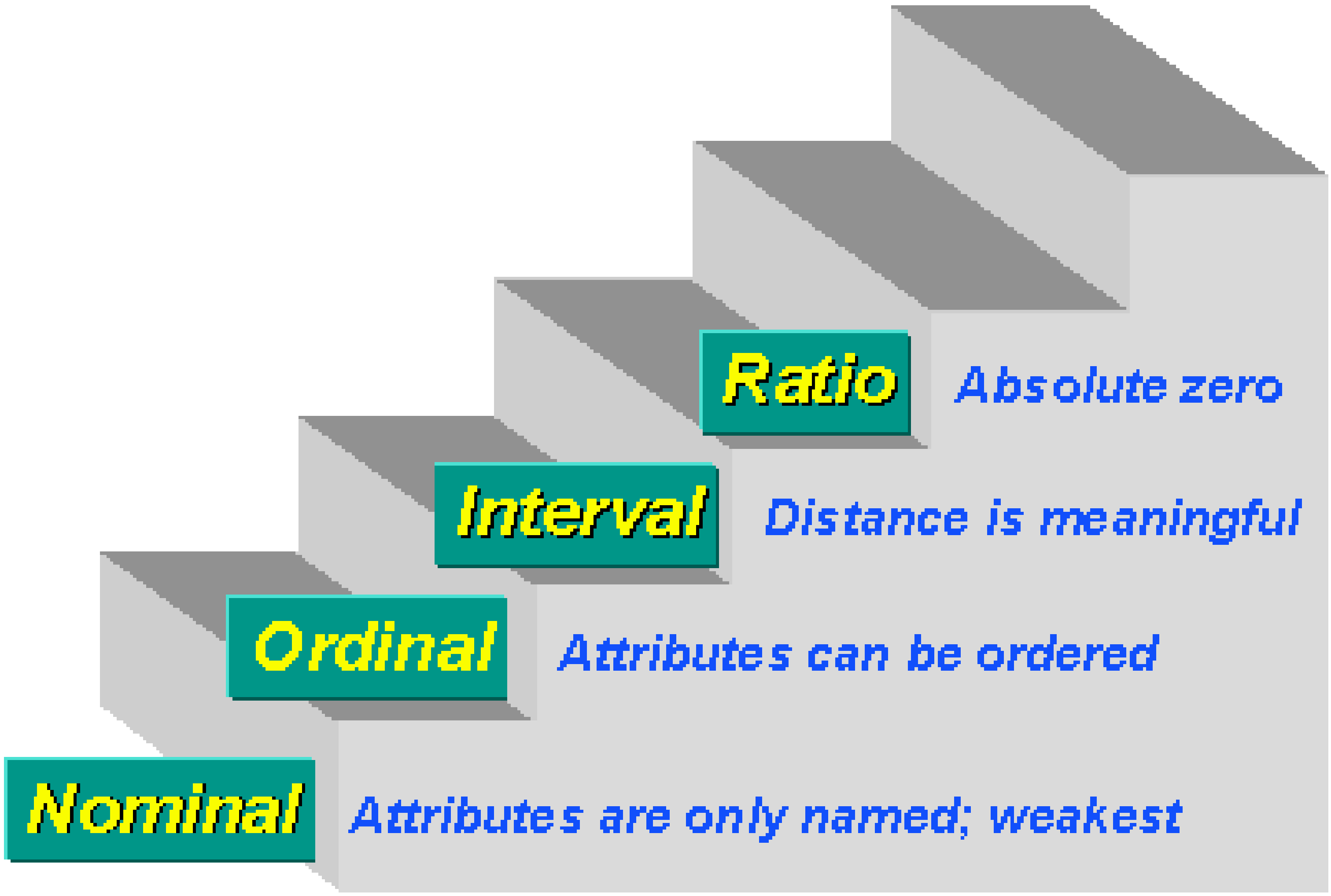
1. census bureau data
2. crime records
3. standardized testing data
4. economic data
5. consumer data

Levels of Measurement

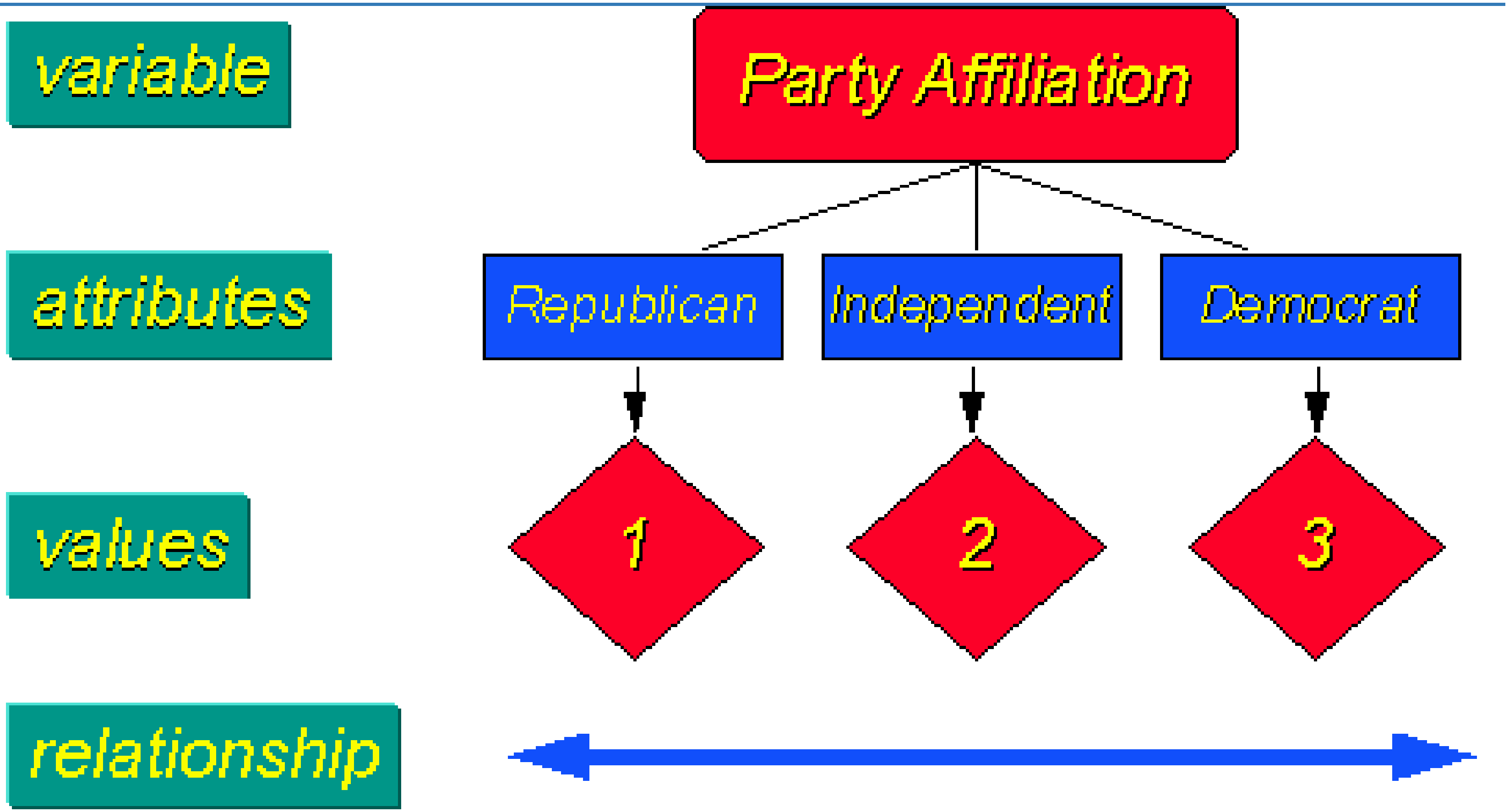
- **Interval:** based on a unit or interval that is accepted as a common standard. It is a Highly precise measure.
  - E.g. tons of garbage collected in a given town, number of arrests made by the police per week
- **Nominal:** Just name the attributes uniquely
  - E.g , Public Sector (=1), Private Sector (=2), etc.
- **Ordinal:** the attributes can be rank-ordered e.g. Scales of education less than grade 10(=0) , Grade ten (=1) etc.

- **Ratio:** You can construct a meaningful fraction (or ratio) with a ratio variable. Zero is used as a reference point for comparison
  - Age, money and weight are common ratio scale variables. For example, if you are 50 years old and your child is 25 years old, you can accurately claim you are twice their age.

Levels of Measurements



Nominal Measure



Ordinal Measure

Table 2.1	An Ordinal Measure of the Concept "Satisfaction"

1 = very satisfied
2 = satisfied
3 = neutral
4 = dissatisfied
5 = very dissatisfied

Table 2.2	An Example of an Ordinal Variable

Name	Satisfaction
Jones	2
R. Smith	3
Franklin	1
Barnes	2
A. Smith	3

Nominal Measure

Table 2.3	A Nominal Measure of Employee Gender

1 = female
0 = male

Table 2.4	An Example of a Nominal Variable

Name	Employee Gender
Jones	1
R. Smith	0
Franklin	1
Barnes	1
A. Smith	0

Table 2.5 **Some Variables: What Is the Level of Measurement?**

Variable	Level of Measurement
1. Number of children	
2. Opinion of the way the president is handling the economy (strongly approve, approve, neutral, disapprove, strongly disapprove)	
3. Age	
4. State of residence	
5. Mode of transportation to work	
6. Perceived income (very low, below average, average, above average, very high)	
7. Income in dollars	
8. Interest in statistics (low, medium, high)	
9. Sector of economy in which you would like to work (public, nonprofit, private)	
10. Hours of overtime per week	
11. Your comprehension of this book (great, adequate, forget it)	
12. Number of memberships in clubs or associations	
13. Dollars donated to nonprofit organizations	
14. Perceived success of animal rights association in advocacy (very high, high, moderate, low, very low)	
15. Years of experience as a supervisor	
16. Your evaluation of the level of "social capital" of your community (very low, low, moderate, high, very high)	

Table 2.6	Different Levels of Measurement for the Same Concept

Cost per Client in $ (interval)	Cost per Client (ordinal)	
57	2	where 1 = low (less than $50)
38	1	2 = moderate ($50 to $100)
79	2	3 = high ($100 or more)
105	3	
84	2	
159	3	
90	2	
128	3	
103	3	

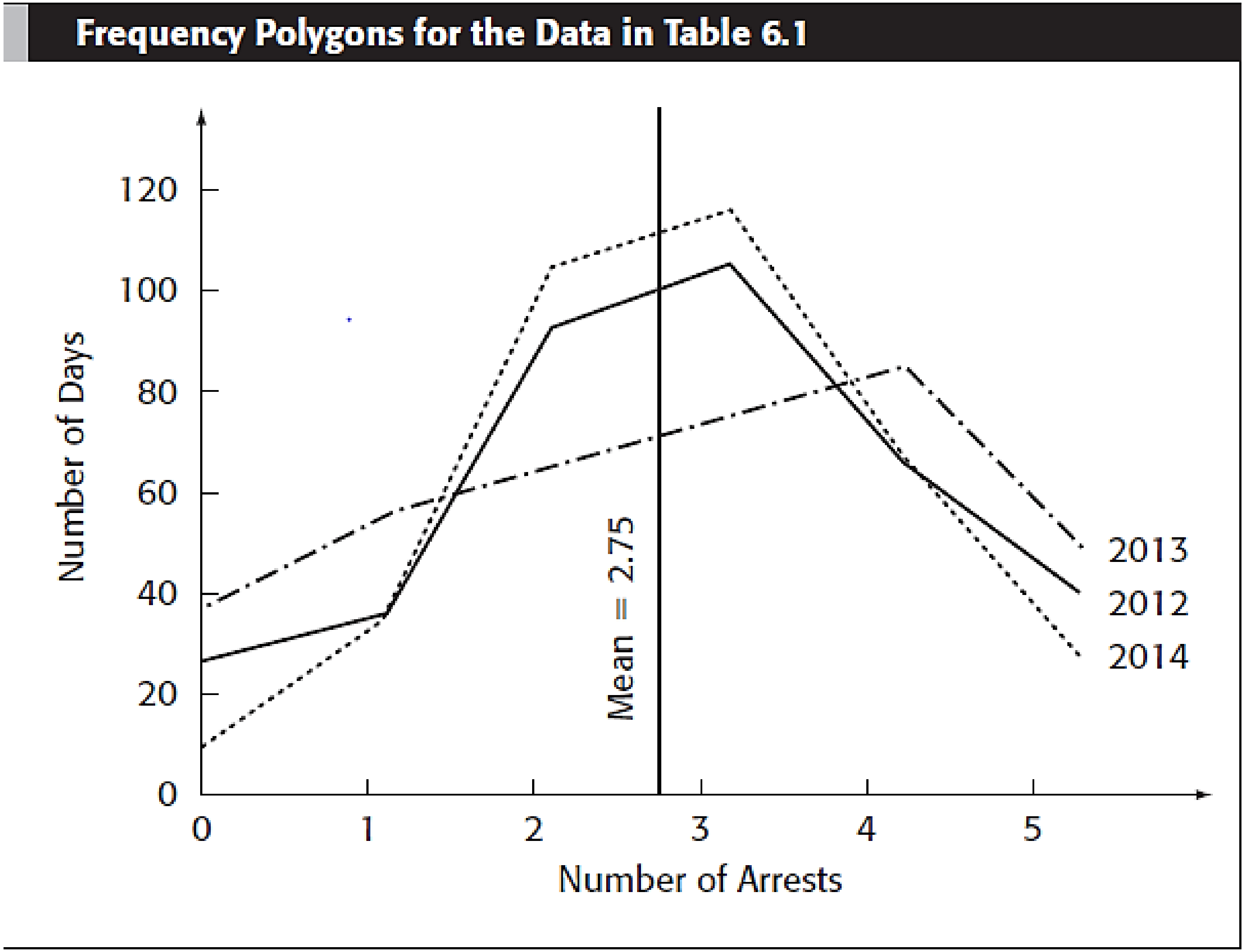
Measures of Dispersion

- Measures of dispersion indicate how closely the data cluster about the mean.
- For example, the data listed in Table 6.1 show the number of daily arrests in a city, for 2012, 2013, and 2014.
- The mean number of daily arrests for all 3 years is the same (2.75).
- How much the daily arrests cluster about the mean, however, varies.
- In 2013, the numbers cluster less about the mean than they do in 2012. In 2014, the arrests cluster closer to the mean than do either the 2013 or the 2012 arrests.
- This clustering is illustrated by the frequency polygons in Figure 6.1; the mean is depicted vertically to facilitate interpretation

Number of Daily Police Arrests in Wheezer, South Dakota

	Number of Days		
Number of Arrests	2012	2013	2014
0	24	36	10
1	36	54	36
2	95	65	109
3	104	74	118
4	66	84	66
5	40	52	26
	365	365	365

The Dispersion



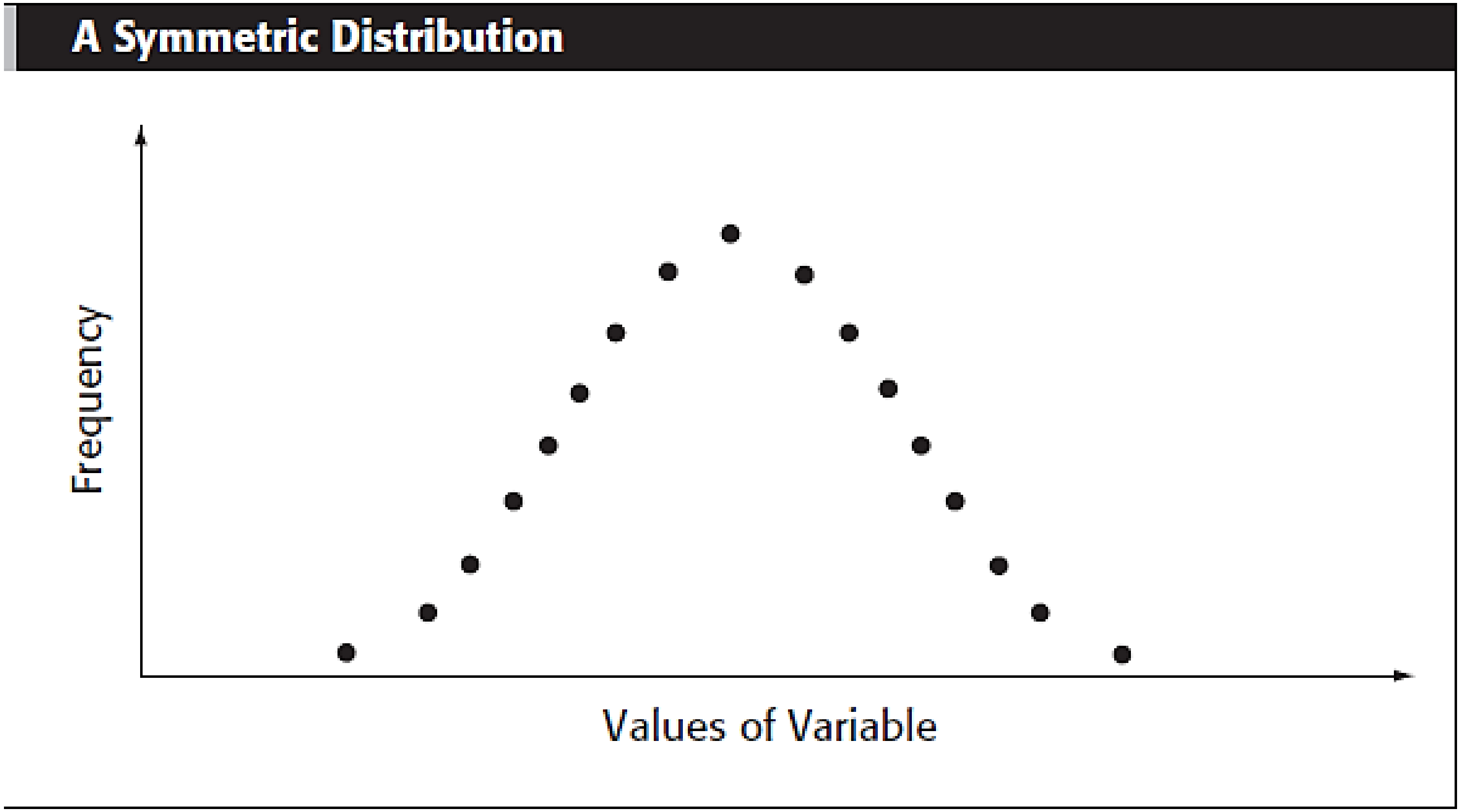
Some Quick Applications of the Standard Deviation

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}}$$

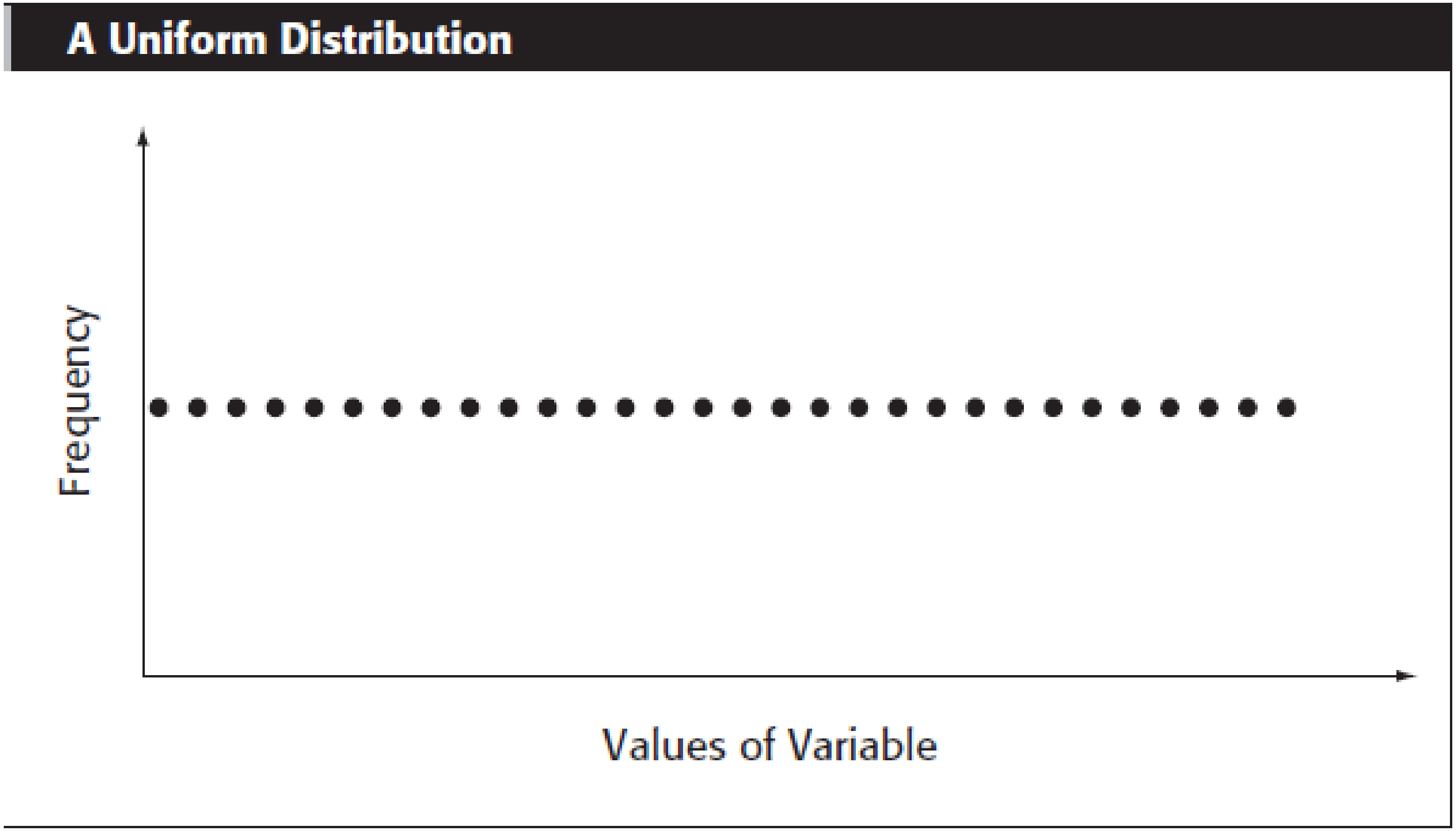
- It is so important, in fact, that the public or nonprofit manager should be cautious in drawing conclusions about the data if it is given only with the mean and not the standard deviation.

- In the absence of the standard deviation, the mean has limited value because it is difficult to see how much volatility or variability exists in the data.

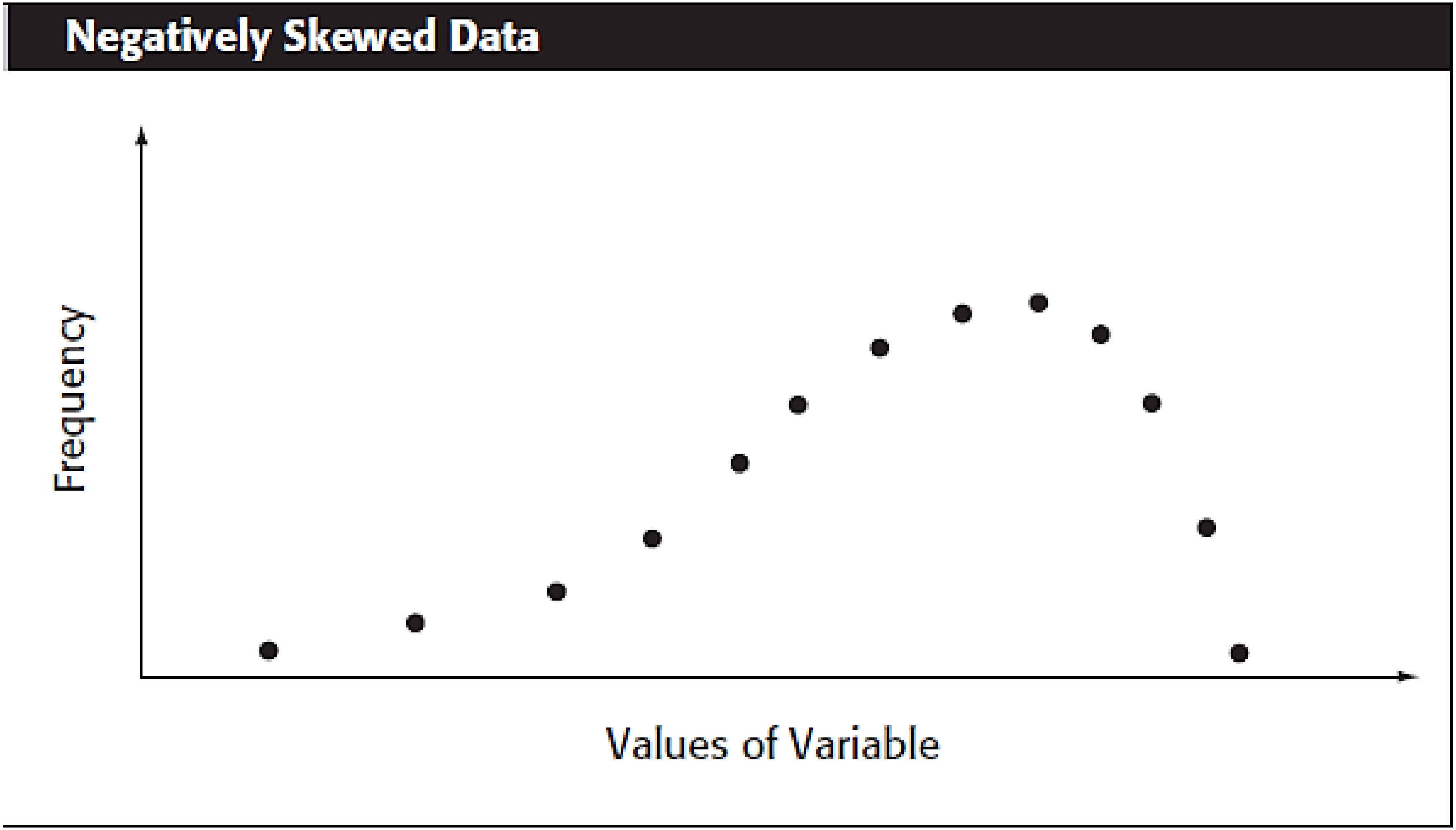
The shape of the distribution



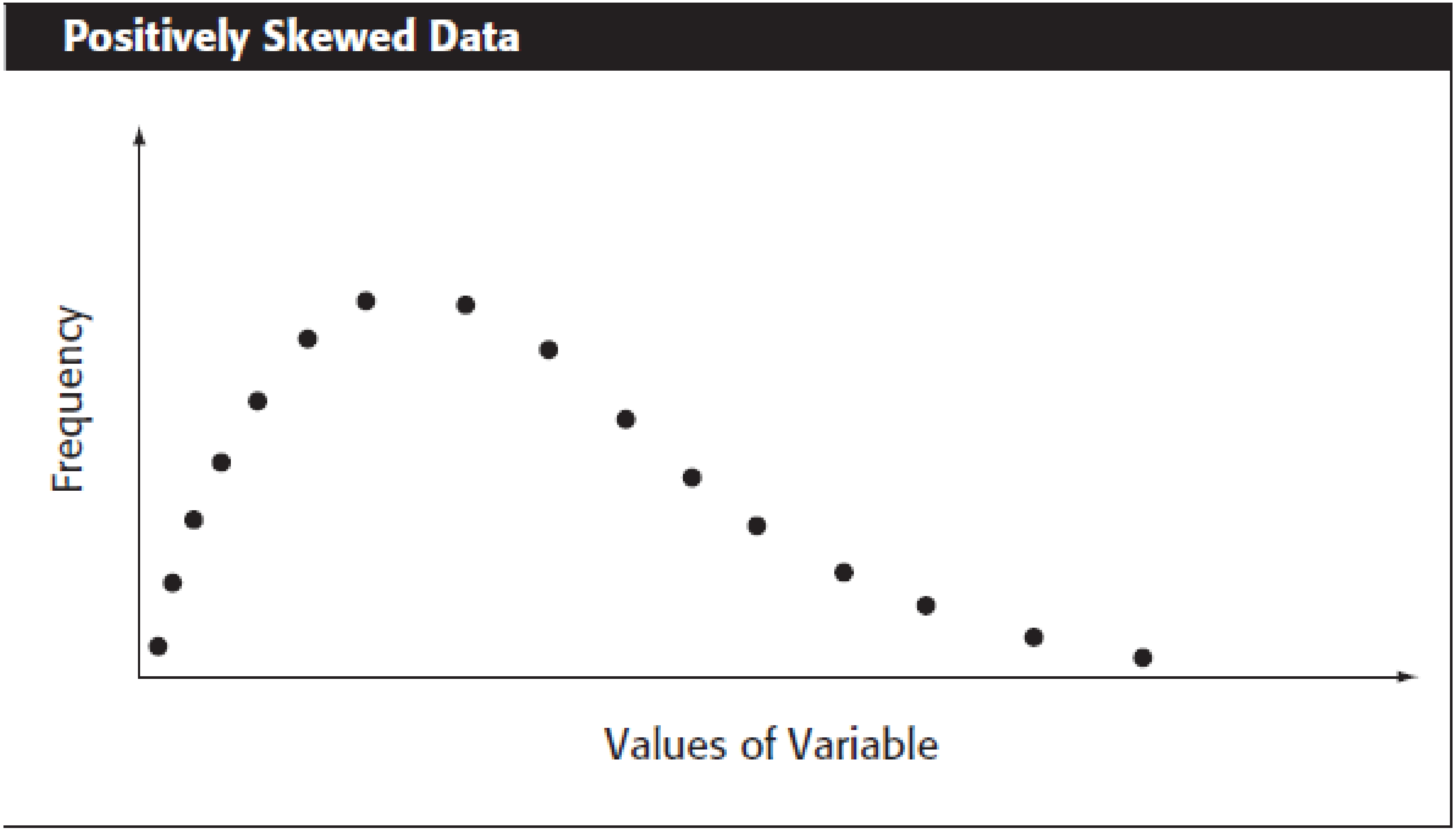
Uniform Distribution



Skewed Distribution



Skewed Distribution



Example

- Let’s say that a supervisor is informed that the mean score for a group of employees on a job skills test is 90%.
- The supervisor is very pleased with the test scores.
- To illustrate the value of the standard deviation when interpreting the mean, let’s look at a couple of different scenarios:

$$\mu = 90 \qquad \sigma = 2$$
$$\mu = 90 \qquad \sigma = 9$$

Practice Problem

- An officer of the Rescue Department wants to evaluate the efficiency of two different water pumps.
- Brand A pumps an average of 5,000 gallons per minute with a standard deviation of 1,000 gallons. Brand B pumps 5,200 gallons per minute with a standard deviation of 1,500 gallons.
- What can you say about the two types of pumps that would be valuable to the officer?

Variance

- Variance refers to the spread of a data set around its mean value.
- Variance is used in statistics to describe the spread between a data set from its mean value.
- It is calculated by finding the probability-weighted average of squared deviations from the expected value (mean).
- So, the larger the variance, the larger the distance between the numbers in the set and the mean.
- Conversely, a smaller variance means the numbers in the set are closer to the mean.

Variance

Population variance

$$\sigma^2 = \frac{\Sigma(x - \mu)^2}{N}$$

Sample Variance ?

Co-Variance

- Covariance provides insight into how two variables are related to one another.
- More precisely, covariance refers to the measure of how two random variables in a data set will change together.
- A positive covariance means that the two variables at hand are positively related, and they move in the same direction.
- A negative covariance means that the variables are inversely related, or that they move in opposite direction

$$COV(x,y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

Correlation

- Correlation also informs about the degree to which the variables tend to move together.
- Correlation standardizes the measure of interdependence between two variables and informs researchers as to how closely the two variables move together.

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

r = correlation coefficient

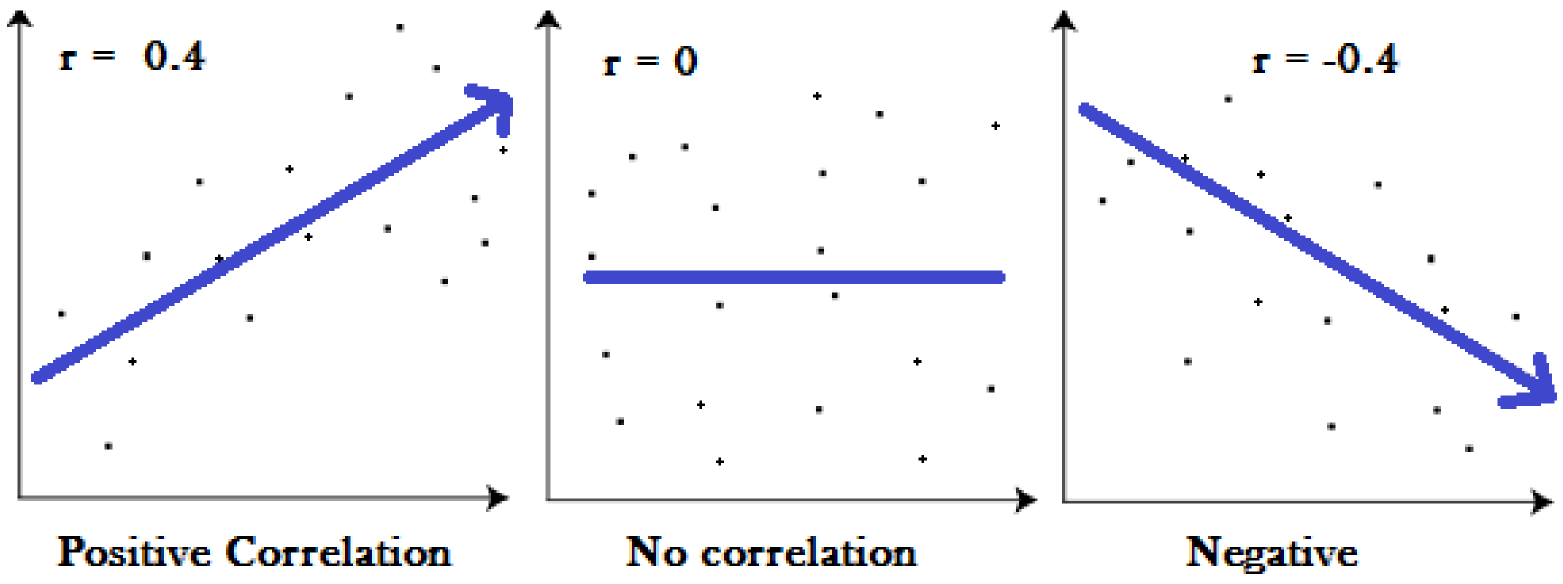
x_i = values of the x-variable in a sample

$\bar{x}$ = mean of the values of the x-variable

y_i = values of the y-variable in a sample

$\bar{y}$ = mean of the values of the y-variable

Coefficient of Correlation



Practice Example

Example question: Find the value of the correlation coefficient from the following table:

SUBJECT	AGE X	GLUCOSE LEVEL Y
1	43	99
2	21	65
3	25	79
4	42	75
5	57	87
6	59	81

Additional Details: Ratio vs Interval Scales Difference

Features	Interval scale	Ratio scale
Variable property	All variables measured in an interval scale can be added, subtracted, and multiplied. You cannot calculate a ratio between them.	Ratio scale has all the characteristics of an interval scale, in addition, to be able to calculate ratios. That is, you can leverage numbers on the scale against 0.
Absolute Point Zero	Zero-point in an interval scale is arbitrary. For example, the temperature can be below 0 degrees Celsius and into negative temperatures.	The ratio scale has an absolute zero or character of origin. Height and weight cannot be zero or below zero.
Calculation	Statistically, in an interval scale, the arithmetic mean is calculated.	Statistically, in a ratio scale, the geometric or harmonic mean is calculated.
Measurement	Interval scale can measure size and magnitude as multiple factors of a defined unit.	Ratio scale can measure size and magnitude as a factor of one defined unit in terms of another.
Example	A classic example of an interval scale is the temperature in Celsius. The difference in temperature between 50 degrees and 60 degrees is 10 degrees; this is the same difference between 70 degrees and 80 degrees.	Classic examples of a ratio scale are any variable that possesses an absolute zero characteristic, like age, weight, height, or sales figures.

Cont.

- contents

Cont.

- contents

Cont.

- contents

Cont.

- contents

Cont.

- contents

Cont.

- contents

Cont.

- contents

Cont.

- contents