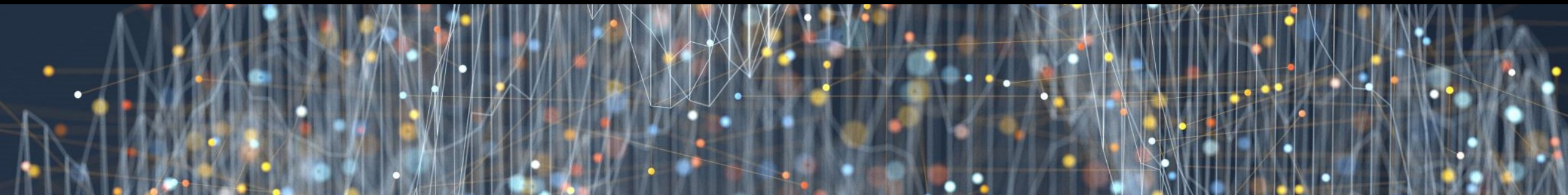
Math 114 Module 9: Data and Variation

Prof. Volk

Spring 2025



Schedule Updates

Test Review Due Tuesday Night

Test on Wednesday

Project 2 Past Due

Allowed Excel/Desmos/Calculator

Provided Formula Sheet

MATH 114
ASSORTED FORMULAS

$$Simple\ Interest = P * r * t$$

$$Principal + 200\%\ increase = triple$$

absolute change = new value – reference value

$$\text{rel diff} = \frac{\text{compare value} - \text{reference value}}{\text{reference value}} \times 100\%$$

COMPOUND INTEREST FORMULAS

$$\frac{log(1 + total\ return)}{log(1 + annual\ return)} = Years$$

	A	B
1	How much to save now	=PV(
2	How much will I have in the future	=FV(
3	How much to save/pay monthly	=PMT(
4	How many years/months will it take?	=NPER(

A = accumulated balance after *Y* years (FV)
P = starting principal or lump sum (PV)
APR = annual percentage rate (as a decimal)
n = number of compounding periods per year
Y = number of years
e = a special irrational number with a value
APY = annual percentage yield
nper = number of periods/months/years (n*Y)

$$A = \text{PMT} \times \frac{\left[\left(1 + \frac{\text{APR}}{n}\right)^{(nY)} - 1\right]}{\left(\frac{\text{APR}}{n}\right)}$$

$$\text{PMT} = \frac{A + \frac{\text{APR}}{n}}{\left[\left(1 + \frac{\text{APR}}{n}\right)^{(nY)} - 1\right]}$$

$$A = P \times (1 + \text{APR})^{Y}$$

$$\text{total return} = \frac{(A - P)}{P} \times 100\%$$

$$A = P\left(1 + \frac{\text{APR}}{n}\right)^{(nY)}$$

$$\text{annual return} = \left(\frac{A}{P}\right)^{(1/Y)} - 1$$

$$A = P \times e^{(\text{APR} \times Y)}$$

Suppose you are 25 years old and would like to retire at age 65. Furthermore, you would like to have a retirement fund from which you can draw an income of $398,421 per year-forever! How can you do it? Assume a constant APR of 6% compounded monthly. Hint: To draw $398,421 each year, the annual interest on the balance of the savings account must be $398,421.

- **A.** Deposit $2,871.52 per month.
- **B.** Deposit $2,391.25 per month.
- **C.** Deposit $3,334.36 per month.
- **D.** Deposit $4,191.71 per month.

You could take a 15-week, three-credit college course, which requires 9 hours per week of your time and costs $500 per credit-hour in tuition. Or during those hours you could have a job paying $15 per hour. What is the net cost of the class compared to working?

- A. $635
- B. $1635
- C. $2525
- D. $3525

Homework 9: 25 Questions

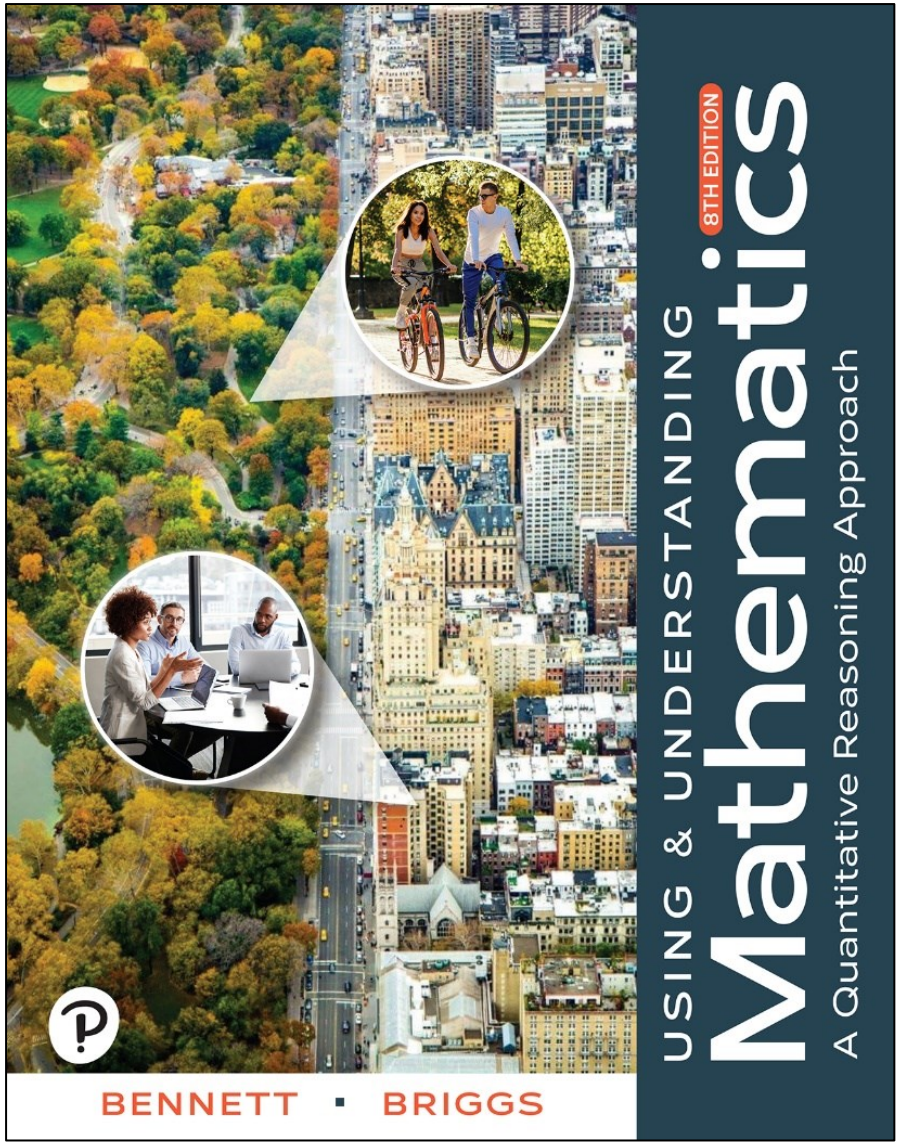
Due Sunday 11:59 pm

Questions 1 - 3	• Answer conceptual questions involving averages and shapes of distributions.
Questions 4 - 7	• Characterize the shape of a distribution by its number of modes, symmetry, and variation.
Questions 8 - 9	• Find and interpret the mean, median, and mode of a distribution.
Question 10	• Identify outliers and determine their effects on the mean and median.
Questions 11-16	• State whether the mean, median, or mode gives the best description of the averages.
Question 17	• Use histograms to classify the shapes of distributions.
Questions 18, 23-25	• Solve applications involving measures of center and variation.
Question 19	• Compute measures of center and variation.
Questions 20-21	• Answer conceptual questions involving measures of center and variation.
Question 22	• Calculate and interpret the five-number summary and boxplot for a distribution.

Quiz 9: 10 Questions

Due Monday
11:59 pm

Question 1	• Answer conceptual questions involving averages and shapes of distributions.
Questions 2 - 3	• Characterize the shape of a distribution by its number of modes, symmetry, and variation.
Question 4	• Find and interpret the mean, median, and mode of a distribution.
Question 5	• Identify outliers and determine their effects on the mean and median.
Question 6	• State whether the mean, median, or mode gives the best description of the averages.
Question 7	• Use histograms to classify the shapes of distributions.
Question 8	• Calculate and interpret the five-number summary and boxplot for a distribution.
Questions 9 & 10	• Solve applications involving measures of center and variation.

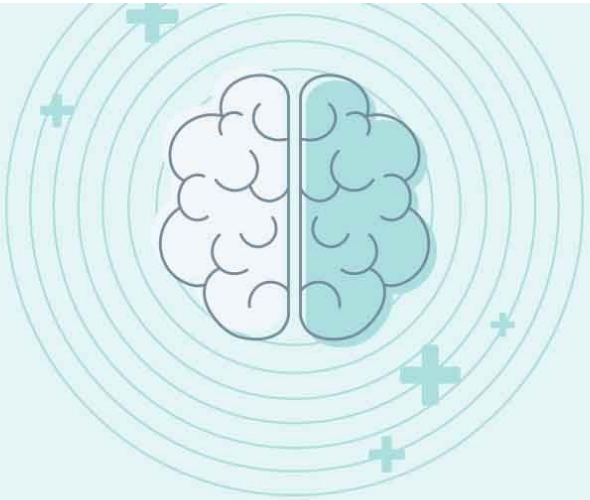


This Week's Reading

Chapter 6

(Sections A & B)

Data and Variation



Don’t copy the behavior and customs of this world, but let God transform you into a new person by ***changing the way you think***.

ROMANS 12:2

DISCUSSION QUESTION ONE

Definition

The **distribution** of a variable refers to the way its values are spread over all possible values. We can display a distribution with a table or graph.

Measures of Center in a Distribution

- The **mean** is what we most commonly call the average value. It is defined as follows:

$$\text{mean} = \frac{\text{sum of all values}}{\text{total number of values}}$$

- The **median** is the middle value in the sorted data set (or halfway between the two middle values if the number of values is even).
- The **mode** is the most common value (or group of values) in a distribution.

Notes

$$\text{Mean} = \frac{\text{Sum of all values}}{\text{Total number of all values}}$$

$$\text{Median} = \text{Middle value (when the data are arranged in order)}$$

$$\text{Mode} = \text{Most common value}$$

Mean Median Mode

The numbers of words defined on randomly selected pages from a dictionary are shown below. Find the mean, median, and mode of the listed numbers.

52 40 64 34 33 44 57 32 61 67

Mean → =Average( )

Median → =Median( )

Mode → =Mode( )

NO Mode

Example: Price Data (1 of 2)

Eight grocery stores sell the PR energy bar for the following prices:

$1.09 $1.29 $1.29 $1.35 $1.39 $1.49 $1.59 $1.79

Find the mean, median, and mode for these prices.

Solution

$$\text{mean} = \frac{\left(\begin{array}{l}\$1.09+\$1.29+\$1.29+\$1.35\\+\$1.39+\$1.49+\$1.59+\$1.79\end{array}\right)}{8} = \$1.41$$

The mean is $1.41

Example: Price Data (2 of 2)

Median: sort the data in ascending order:

$1.09 $1.29 $1.29 $1.35 $1.39 $1.49 $1.59 $1.79

3 values below | 2 middle values | 3 values above

Because there are eight prices (an even number), there are two values in the middle of the list: $1.35 and $1.39. The median lies halfway between these two values, which we calculate by adding them and dividing by 2:

$$\text{median} = \frac{\$1.35 + \$1.39}{2} = \$1.37$$

The mode is $1.29.

An **outlier** in a data set is a data value that is much higher or much lower than almost all other values. An outlier can change the mean of the data but generally does not affect the median or mode.

Consider the following data set of contract offers:

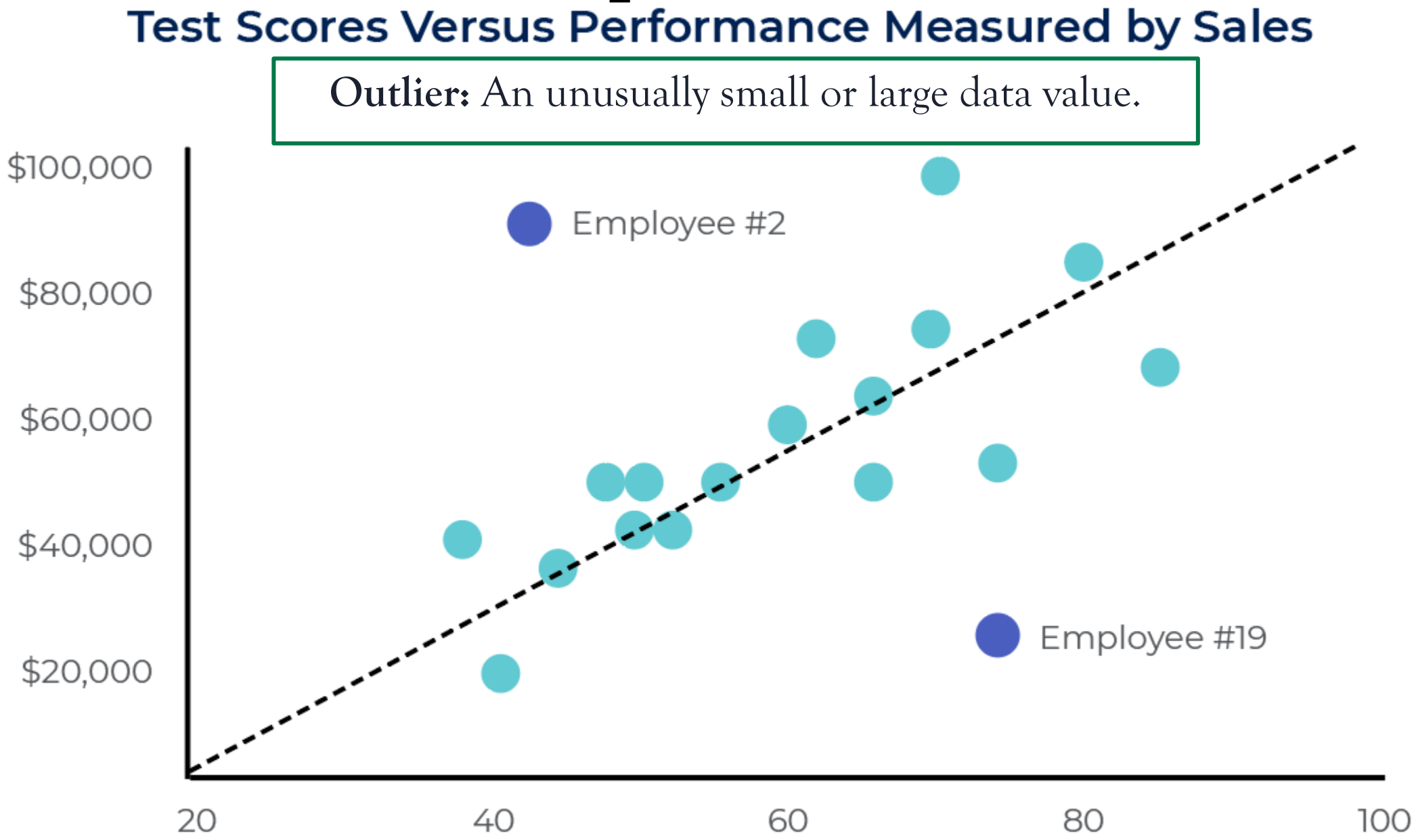
$0 $0 $0 $0 $10,000,000

The mean contract offer is

$$\text{mean} = \frac{\$0 + \$0 + \$0 + \$0 + \$10,000,000}{5} = \$2,000,000$$

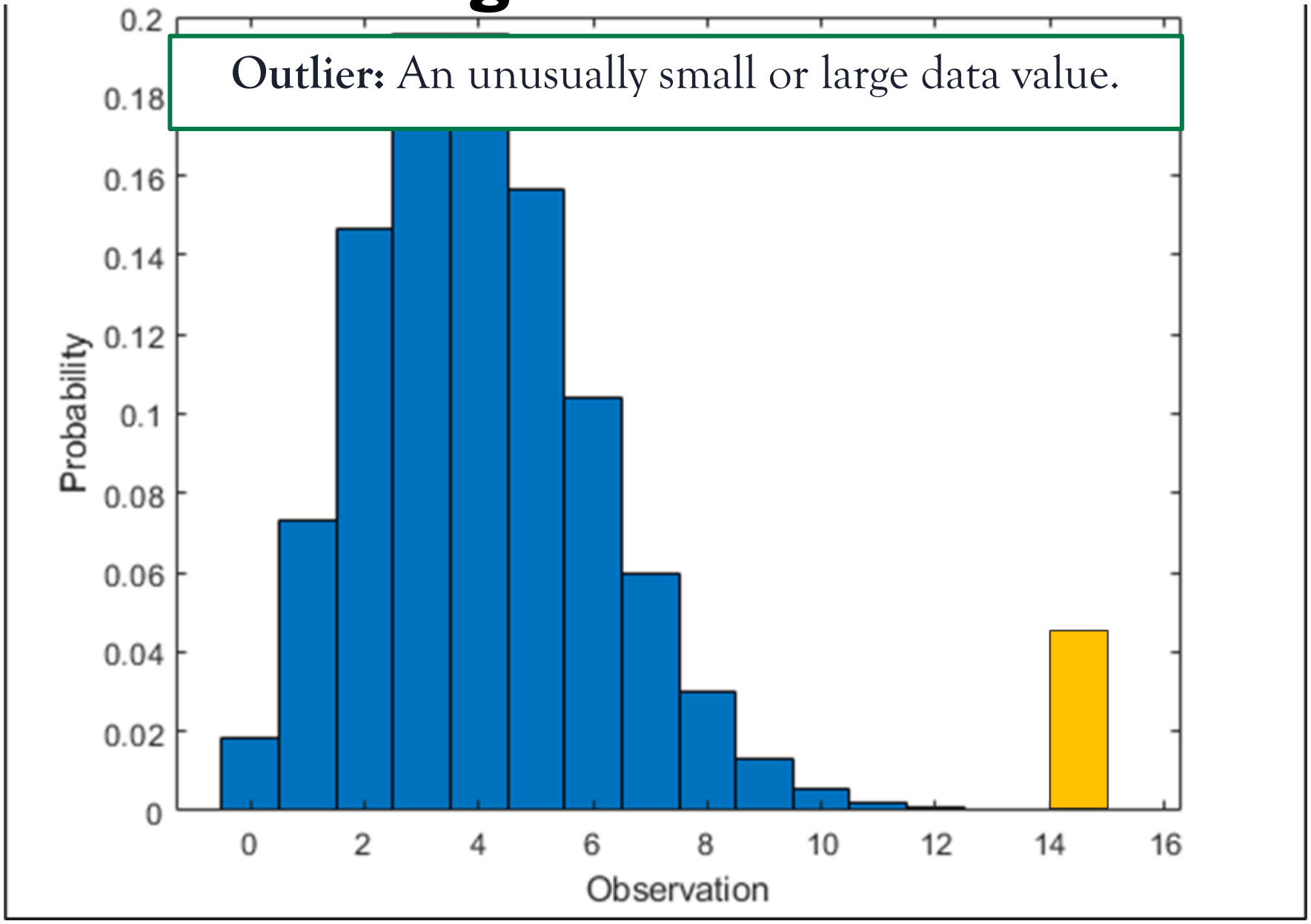
As displayed, outliers can pull the mean upward (or downward). The median and mode of the data are not affected.

Notes: Outlier Scatterplot

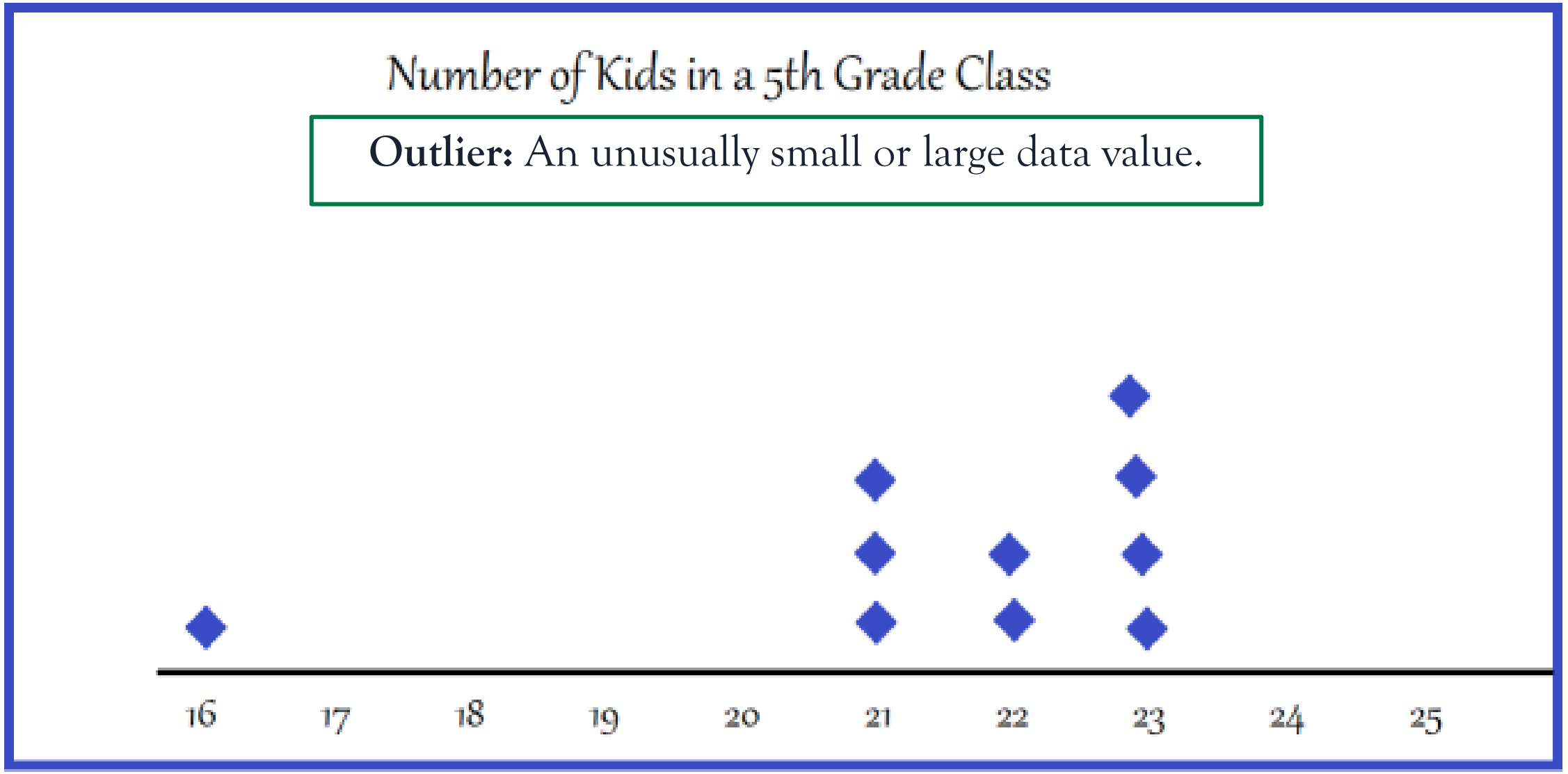


Notes: Outlier Histogram

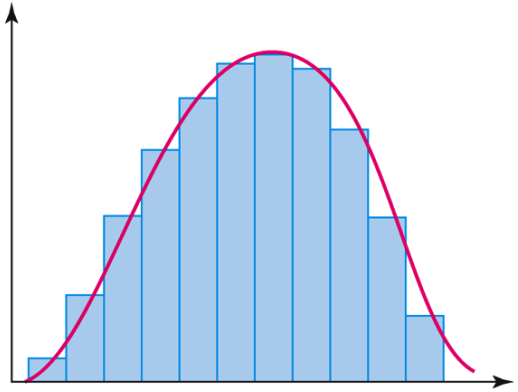
Outlier: An unusually small or large data value.

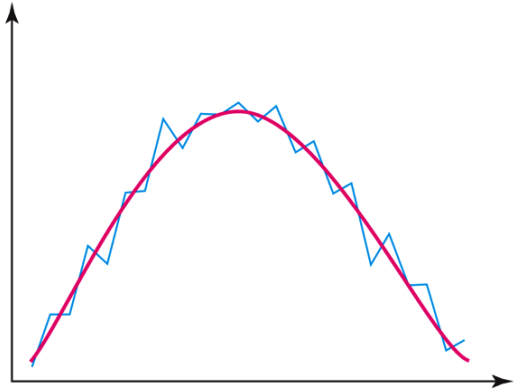


Notes: Outlier Dotplot

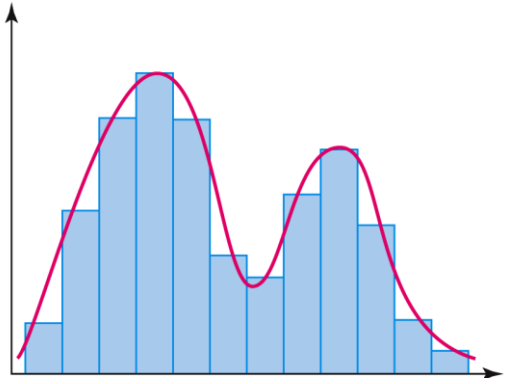


Shapes of Distributions





Two single-peaked (unimodal) distributions



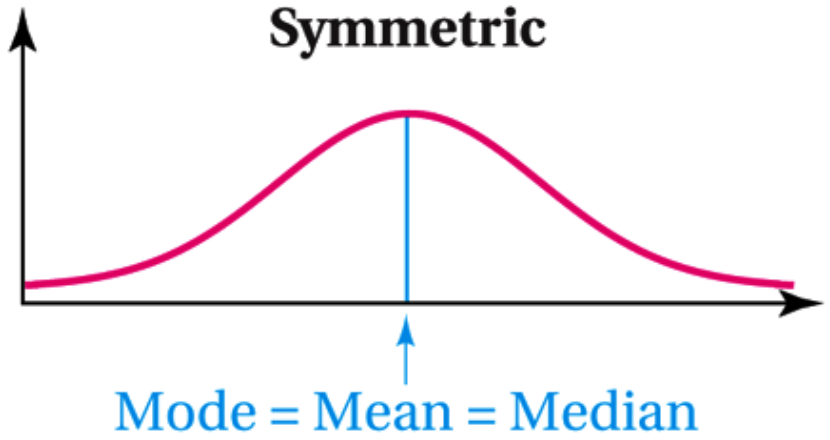
A double-peaked (bimodal) distribution

Discussion Question Two

- How have you utilized AI to improve your life as a student?
- How do you wish Professors/LU would use AI to improve?

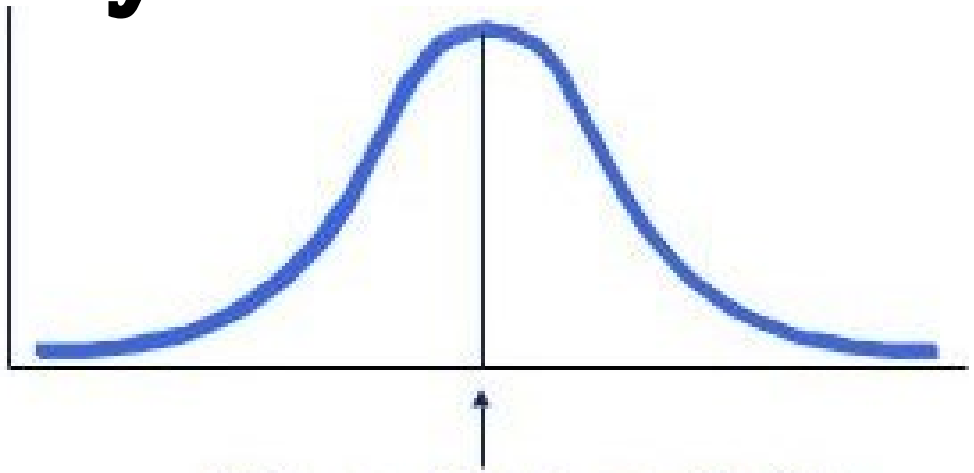
Symmetry

A distribution is **symmetric** if its left half is a mirror image of its right half.



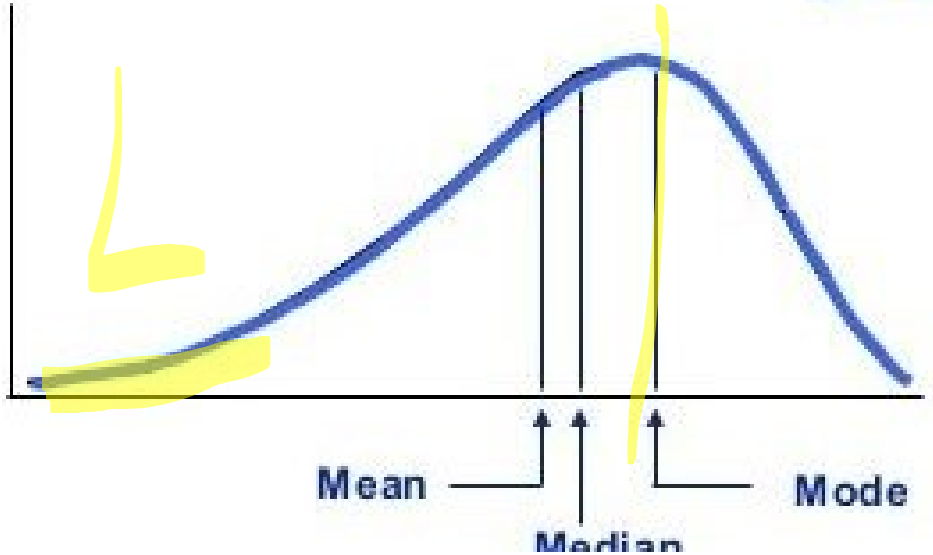
A distribution that is not symmetric must have values that tend to be more spread out on one side than the other. In this case we say the distribution is **skewed**.

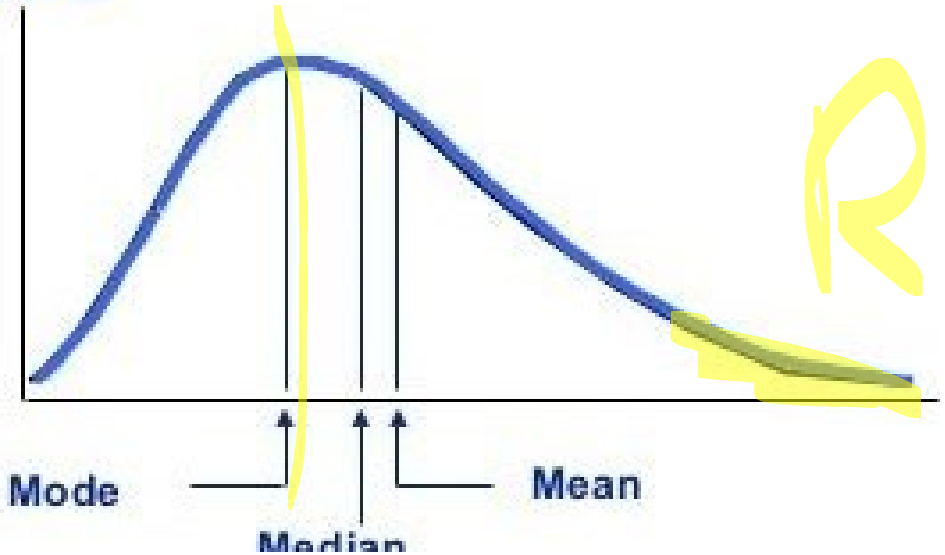
Notes: Symmetry and Skew



Mode = Mean = Median

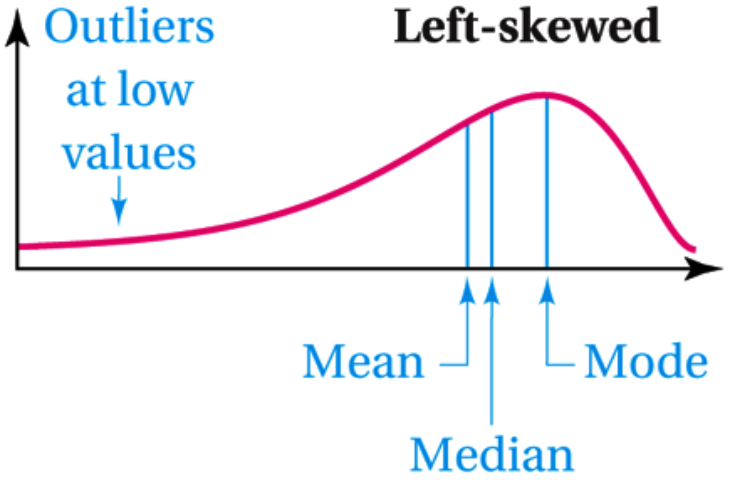
SYMMETRIC



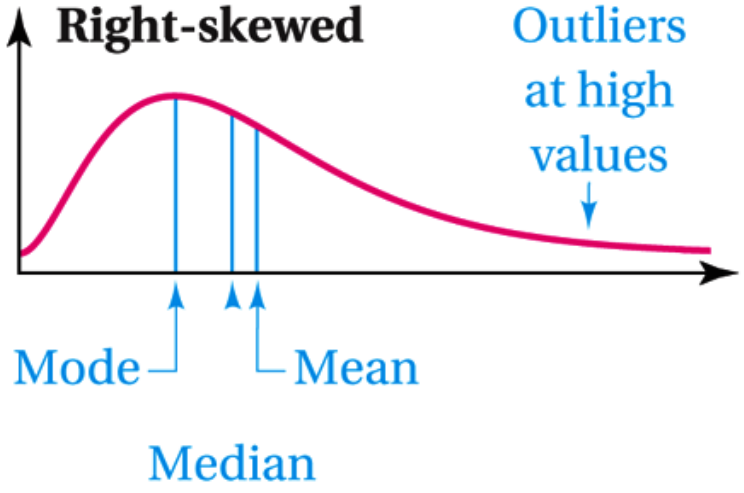


Skewness

A distribution is **left-skewed** if its values are more spread out on the left side.



A distribution is **right-skewed** if its values are more spread out on the right side.



Definition

Symmetry and Skewness

A single-peaked distribution is **symmetric** if its left half is a mirror image of its right half.

A single-peaked distribution is **left-skewed** if its values are more spread out on the left side of the mode.

A single-peaked distribution is **right-skewed** if its values are more spread out on the right side of the mode.

Discussion Question THREE

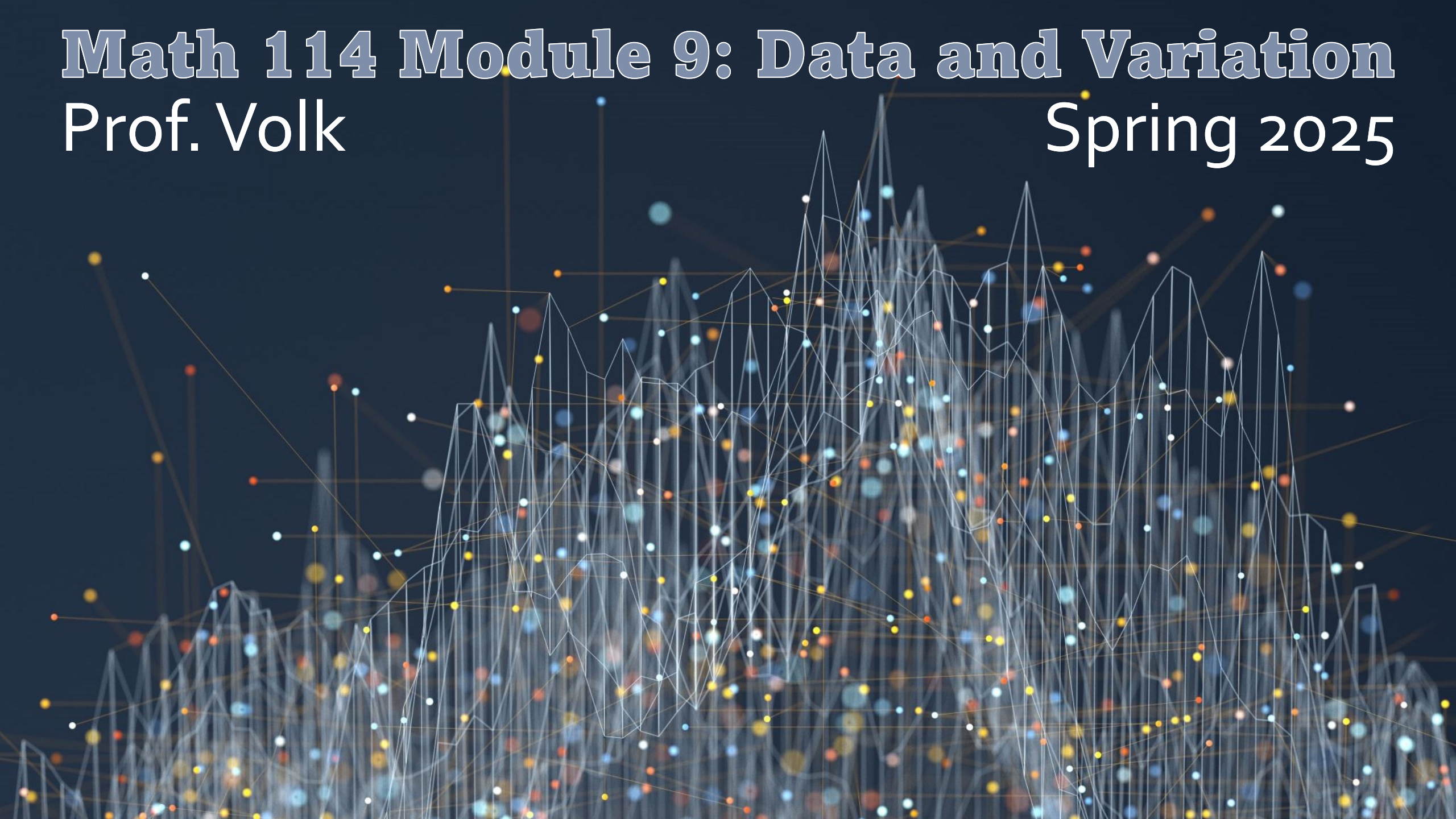
- Who do you see?

TOPHAT

Math 114 Module 9: Data and Variation

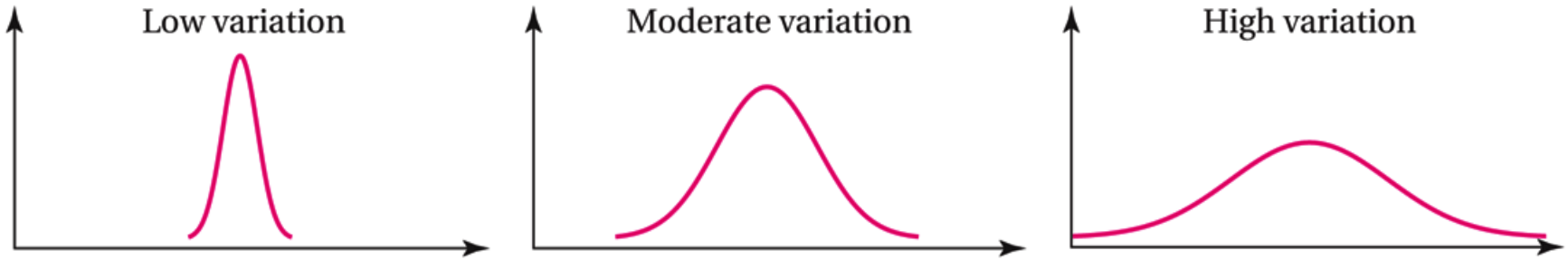
Prof. Volk

Spring 2025



Variation

Variation describes how widely data values are spread out about the center of a data set.



From left to right, these three distributions have increasing variation.

Why Variation Matters

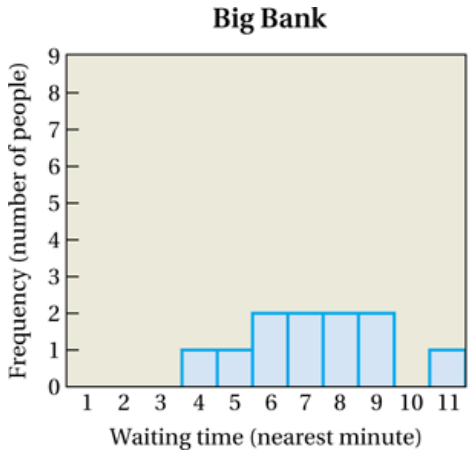
Consider the following waiting times for 11 customers at 2 banks.

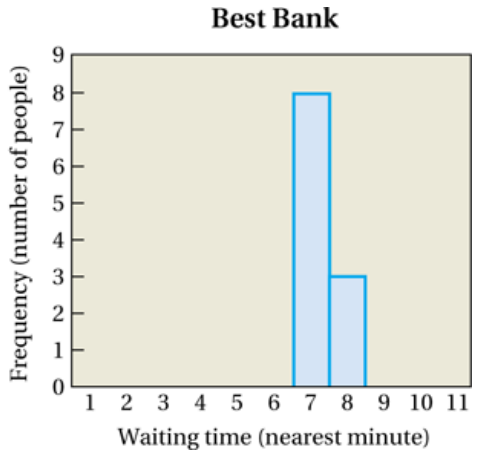
Big Bank (three lines):

4.1 5.2 5.6 6.2 6.7 7.2
7.7 7.7 8.5 9.3 11.0

Best Bank (one line):

6.6 6.7 6.7 6.9 7.1 7.2
7.3 7.4 7.7 7.8 7.8





Which bank is likely to have more unhappy customers?

→ Big Bank, due to more surprise long waits

Distribution Skewed?

Decide whether the following statement makes sense (or is clearly true) or does not make sense (or is clearly false). Explain your reasoning.

The distribution of grades was ~~left-skewed~~, but the mean, median, and mode were all the same.

Central Tendency

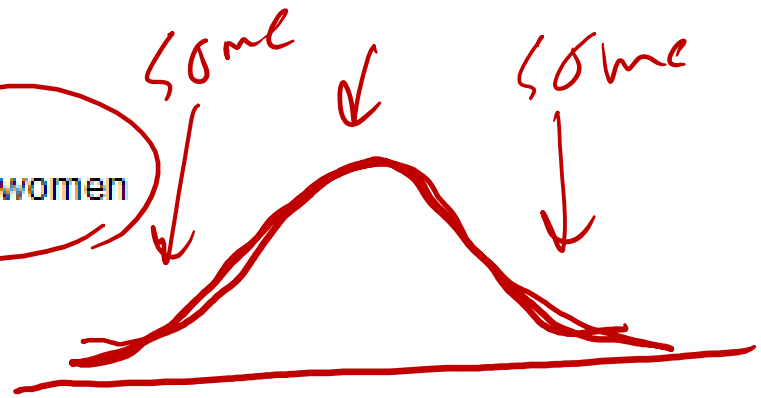
symmetric

Long Left Tail

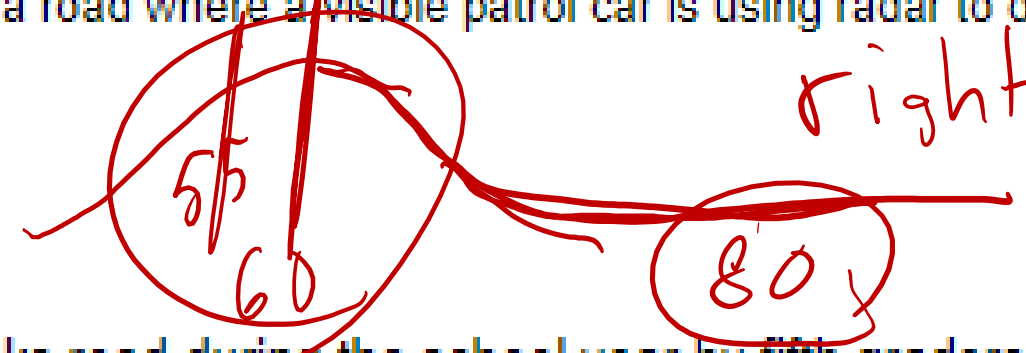
Distribution Examples

What do we mean when we say that a distribution is symmetric? Give simple examples of a symmetric distribution, a left-skewed distribution, and a right-skewed distribution.

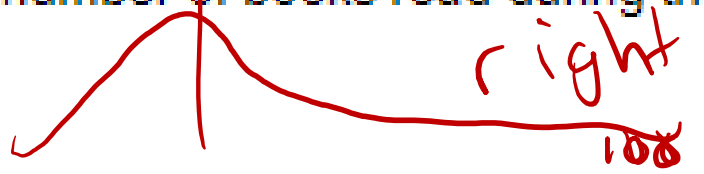
The heights of a sample of 100 women



The speed of cars on a road where a visible patrol car is using radar to detect speeders distribution.



The number of books read during the school year by fifth graders



Range

The **range** of a data set is the difference between its highest and lowest data values.

range = highest value (max) – lowest value (min)

Example: Misleading Range (1 of 2)

Consider the following two sets of quiz scores for nine students. Which set has the greater range? Would you also say that the scores in the set are more varied?

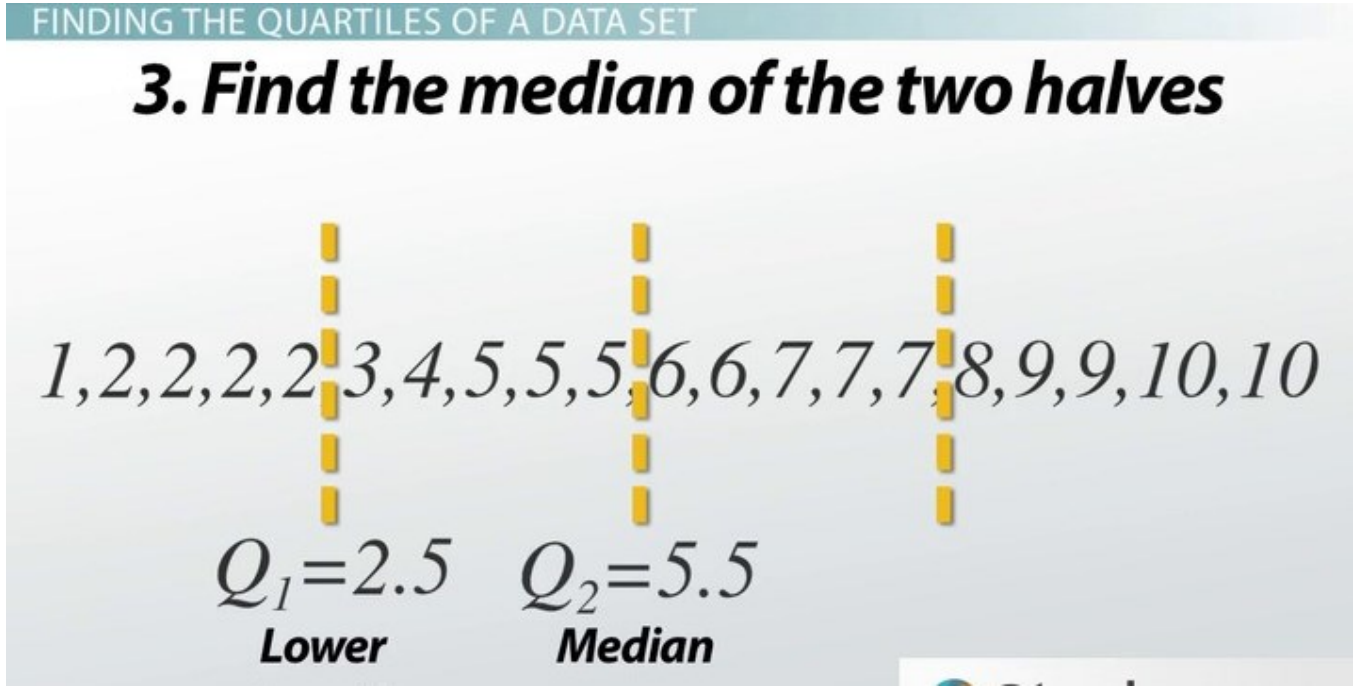
Quiz 1: 1 10 10 10 10 10 10 10 10

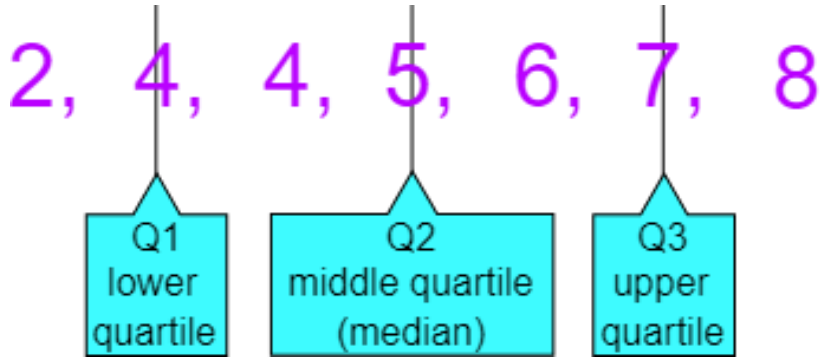
Quiz 2: 2 3 4 5 6 7 8 9 10

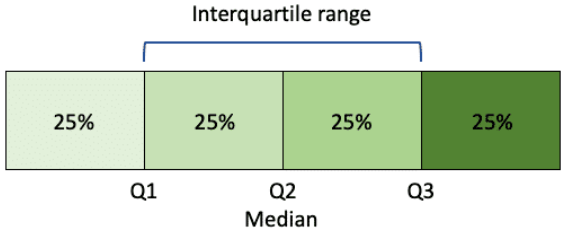
Quartiles

- The **lower quartile** (or first quartile) divides the lowest fourth of a data set from the upper three-fourths. It is the median of the data values in the *lower half* of a data set.
- The **middle quartile** (or second quartile) is the median of the entire data.
- The **upper quartile** (or third quartile) divides the lower three-fourths of a data set from the upper fourth. It is the median of the data values in the *upper half* of a data set.

Notes: Quartiles







The Five-Number Summary

- The **five-number summary** for a data set consists of the following five numbers:

  low value *lower quartile* *median* *upper quartile* *high value*

- A **boxplot** shows the five-number summary visually, with a rectangular box enclosing the lower and upper quartiles, a line marking the median, and whiskers extending to the low and high values.

The Five-Number Summary

Five-number summary of the waiting times at each bank:

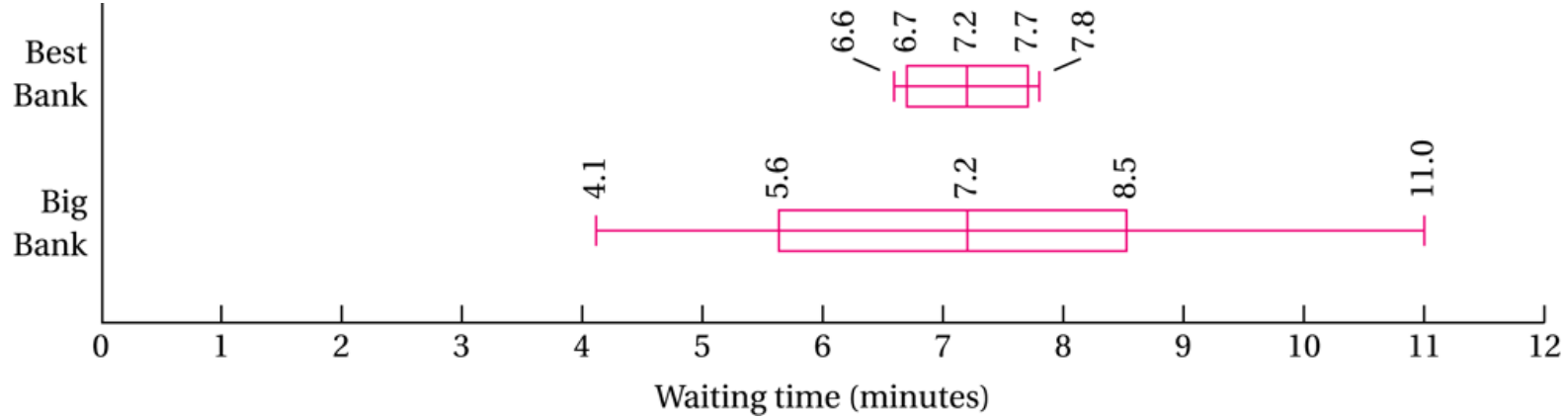
Big Bank

low value (min) = 4.1
lower quartile = 5.6
median = 7.2
upper quartile = 8.5
high value (max) = 11.0

Best Bank

low value (min) = 6.6
lower quartile = 6.7
median = 7.2
upper quartile = 7.7
high value (max) = 7.8

The corresponding boxplot:



The table gives the cost of living index (COLI) for six urban areas with a relatively high cost of living and six urban areas with a relatively low cost of living. The index is based on costs for housing, utilities, grocery items, transportation, health care, and miscellaneous goods and services; 100 represents the national average. Complete parts a through e.

High-Cost Urban Areas		Low-Cost Urban Areas	
A	245	U	82
B	201	V	82
C	198	W	78
D	158	X	77
E	152	Y	74
F	148	Z	72

a. Find the mean, median, and range for each of the two data sets.

The mean for the high-cost urban areas is ☐.
(Type an integer or decimal rounded to two decimal places as needed.)

The table gives the cost of living index (COLI) for six urban areas with a relatively high cost of living and six urban areas with a relatively low cost of living. The index is based on costs for housing, utilities, grocery items, transportation, health care, and miscellaneous goods and services; 100 represents the national average. Complete parts a through e.

High-Cost Urban Areas		Low-Cost Urban Areas
A	245	U
B	201	V
C	198	W
D	158	X
E	152	Y
F	148	Z

Open in StatCrunch

Copy to Clipboard

Open in Excel

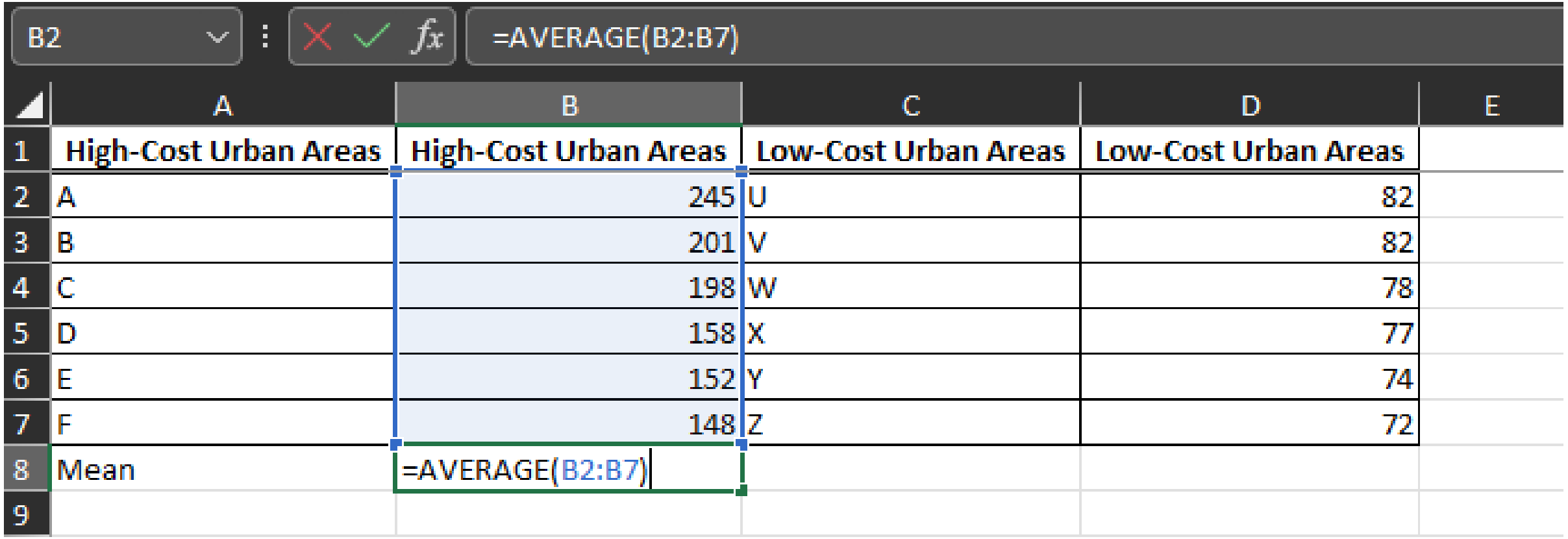
a. Find the mean, median, and range for each of the two data sets.

The mean for the high-cost urban areas is ☐.

(Type an integer or decimal rounded to two decimal places as needed.)

B24

	A	B	C	D	E
1	High-Cost Urban Areas	High-Cost Urban Areas	Low-Cost Urban Areas	Low-Cost Urban Areas	
2	A	245	U	82	
3	B	201	V	82	
4	C	198	W	78	
5	D	158	X	77	
6	E	152	Y	74	
7	F	148	Z	72	
8					
9					
10					
11					


B2
=AVERAGE(B2:B7)
A
B
C
D
E
1 High-Cost Urban Areas
High-Cost Urban Areas
Low-Cost Urban Areas
Low-Cost Urban Areas
2 A 245 U 82
3 B 201 V 82
4 C 198 W 78
5 D 158 X 77
6 E 152 Y 74
7 F 148 Z 72
8 Mean =AVERAGE(B2:B7)
9

	A	B	C	D	E
1	High-Cost Urban Areas	High-Cost Urban Areas	Low-Cost Urban Areas	Low-Cost Urban Areas	
2	A	245	U	82	
3	B	201	V	82	
4	C	198	W	78	
5	D	158	X	77	
6	E	152	Y	74	
7	F	148	Z	72	
8	Mean	183.6666667			
9	Median	=Median(			
10	Max	=MAX(			
11	Min	=Min(			
12	Range	=B10-B11			
13					

	A	B	C	D
1	High-Cost Urban Areas	High-Cost Urban Areas	Low-Cost Urban Areas	Low-Cost Urban Areas
2	A	245	U	82
3	B	201	V	82
4	C	198	W	78
5	D	158	X	77
6	E	152	Y	74
7	F	148	Z	72
8				
9	=QUARTILE(B2:B7,0)	148	Min	
10	=QUARTILE(B2:B7,1)	153.5	Q1	
11	=QUARTILE(B2:B7,2)	178	Median	
12	=QUARTILE(B2:B7,3)	200.25	Q3	
13	=QUARTILE(B2:B7,4)	245	Max	
14				

Standard Deviation

The **standard deviation** is the single number most commonly used to describe variation.

$$\text{standard deviation} = \sqrt{\frac{\text{sum of (deviations from the mean)}^2}{\text{total number of data values} - 1}}$$

Notes: Standard Deviation

Describe the process of calculating a standard deviation. Give a simple example of its calculation (such as calculating the standard deviation of the numbers 2, 3, 4, 4, and 6). What is the standard deviation if all of the sample values are the same?

Fill in the blanks to complete the process of calculating a standard deviation.

=stdev.S(rng)

Compute the mean of the data set. Then find the deviation from the mean for every data value by [ ] the data value.

Find the [ ] of all the deviations from the mean, and then [ ] them together.

Divide this sum by the [ ] n − 1

The standard deviation is the [ ] of this quotient.

The standard deviation of the numbers 2, 3, 4, 4, and 6 is approximately [ ].
(Round to three decimal places as needed.)

If all of the sample values are the same, then the standard deviation is [ ].

The Range Rule of Thumb

- The standard deviation is *approximately* related to the range of a distribution by the **range rule of thumb**:

$$\text{standard deviation} \approx \frac{\text{range}}{4}$$

- If we know the standard deviation for a data set, we estimate the low and high values as follows:

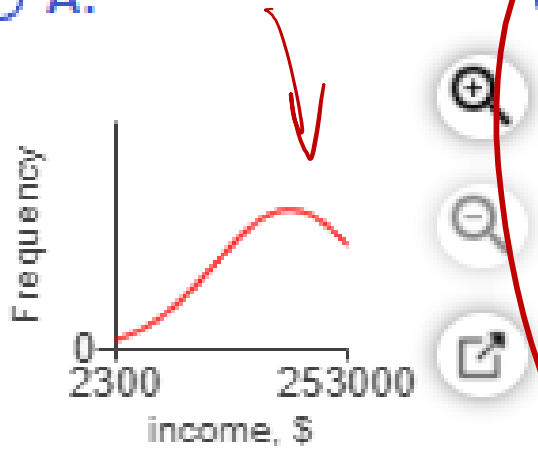
$$\text{low value} \approx \text{mean} - (2 \times \text{standard deviation})$$

$$\text{high value} \approx \text{mean} + (2 \times \text{standard deviation})$$

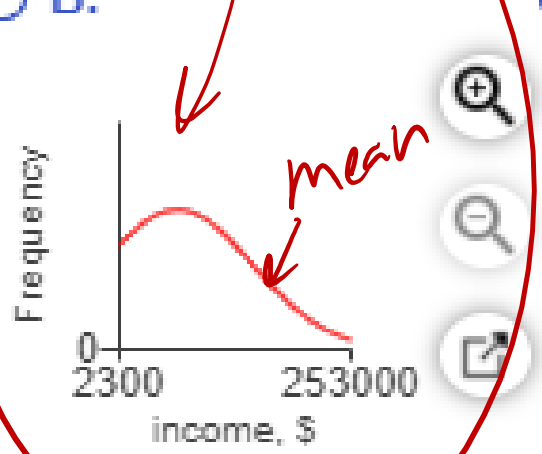
Suppose you study family income in a random sample of 300 families. Your results can be summarized as the mean family income was $47,000, the median family income was $32,000, the highest and lowest incomes were $253,000 and $2,300, respectively.

a. Draw a rough sketch of the income distribution, with clearly labeled axes. Choose the correct answer below.

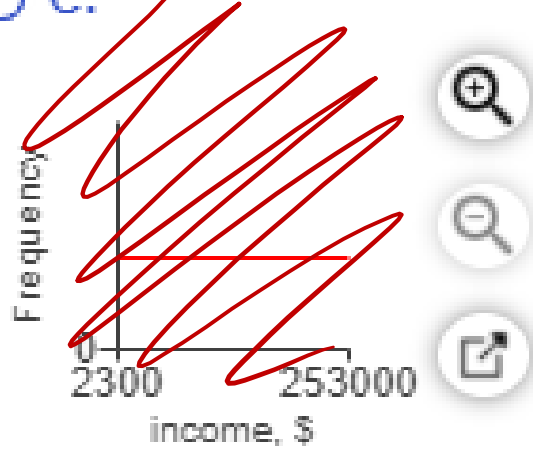
○ A.



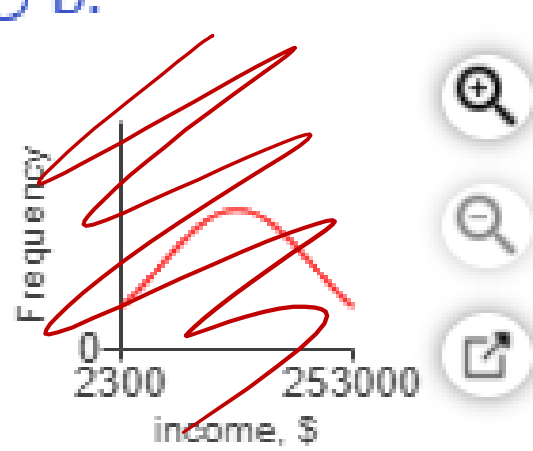
○ B.



○ C.



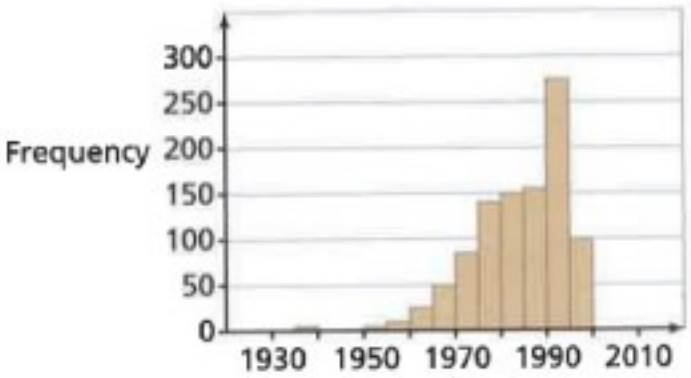
○ D.



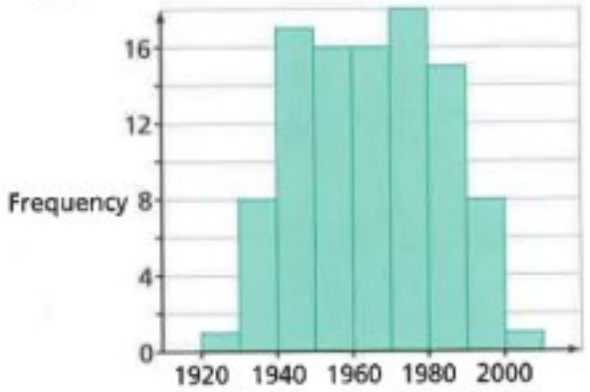
Match the following distribution curves to the random variables listed below. Also, describe the shape of the distribution.

- bowling scores
- the year shown on a penny
- the production year of the American Film Institute's top 100 films

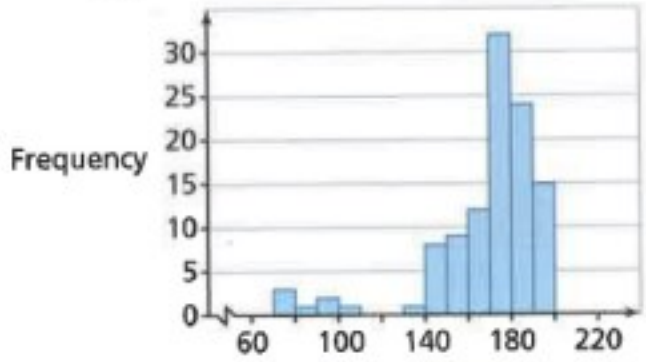
(a)



(b)



(c)



Using the following data:

13, 7, 5, 7, 9, 10, 5, 11, 8, 7, 9, 10, 10, 11, 14, 10, 6, 9, 7, 12, 9, 10, 6

Calculate a bin width that would form five uniform intervals

Discussion Questions:

- Did God stop creating after the first 7 days?

QUESTIONS?