

3.1 – Statistics Vocabulary and Measures of Central Tendency

- Statistics contains A LOT of vocabulary so below are some words that you NEED TO KNOW:

- **statistics** → the art and science of extracting meaning from data.
- **census** → acquiring data from an entire population
- **sampling** → polling a small portion of the population which is intended to show what the whole is like
 - **random sampling**: every possible sample has an EQUAL CHANCE of being selected (IDEAL)
 - **systematic sampling**: a random sample that is chosen through a specific order
 - **stratified sampling**: a random sample from a subgroup to obtain a particular type of a population
 - **multi-stage sampling**: a random sample from specific clusters of a population
 - **biased sampling**: some possible samples may have a preference for a specific outcome
 - **voluntary sampling**: possible samples choose whether or not they will participate
 - **convenience sampling**: most possible samples are selected because they're easiest to reach
- **variable** → the data collected on each subject and is broken up into two different types
 - **categorical (variable)**: a variable that is assigned a characteristic of each subject or group
 - **quantitative (variable)**: a variable that is assigned a numerical value of each subject or group
- **sample design** → the choice and specifics of how a researcher will collect data
 - **observation**: the researcher watches responses and tries to avoid influencing the subjects
 - **survey**: the researcher pose questions and record the subject's responses
 - **experiment**: the researcher intentionally imposes a treatment upon the subjects; contains a control group (placebo group) and a treatment group or double-blind group
 - **simulation**: the researcher uses a math model and probabilities to examine outcomes

Example 1: Random samples are ideal. Determine if each method produces a true random sample.

- asking every tenth person coming out of a health club how many times a week they exercise
not random ~ biased and convenient
- surveying people going into an Italian restaurant to find out people's favorite type of food
not random ~ biased
- the government sending a tax survey to everyone whose social security number ends in a particular digit
random
- surveying students in a honors chemistry classes to determine the average time students study each week
not random ~ biased, voluntary
- putting names of all seniors in a hat, then drawing names from the hat to select a sample of seniors
random ~ stratified
- selecting a student from 200 by picking the fifth student then every tenth student after that
not random ~ systematic
- selecting a resident of the United States to take survey by choosing a State, then a county, and a city
random ~ multi-stage

Example 2: Determine what type of sample design is being described.

- A teacher wants to make sure a pair of dice are fair (one side is not weighted more) by rolling them 1000 times.
Simulation
- A doctor is monitoring a patient after they have administered an injection to see if patient has a reaction.
observation
- A manufacturer asks random people in a town what products they would like to see sold in the area.
Survey
- A fitness instructor gives out a herbal supplement to her students to see if it enhances their performance.
Experiment

• A huge part of collecting data and interpreting it is finding the measures of central tendency:

– measures of central tendency → represents the center or middle of a set of data

• mean – the sum of numerical observation divided by the number of observations.

▪ a common notation for mean is \bar{x} (often called “x-bar”))

▪ it can also be written in this form $x_1 + x_2 + x_3 + \dots + x_n \rightarrow \frac{1}{n} \sum_{i=1}^n x_i$

• median – after arranging the numbers from least to greatest it represents the middle #

▪ if there are an odd number of numbers then take the middle number 3 7 (12) 13 16

▪ if there are an even number of numbers then take the average of two middle numbers

• mode – represents the number that occurs the most often

▪ if no number occurs more often than any other, then the data is amodal

▪ if there are two numbers that occur that same number of times, then the data is bimodal

3 7 (12 13) 16 20
↳ 12.5

Example 3: Find the measures of central tendency and answer its questions for each problem.

a.) Below are Mr. Dent’s Algebra 2 exam scores:

72	70	77	76	70	68	81	86	34	94
71	84	89	67	19	85	75	66	80	94

i.) mean = 73.9 median = 76.5 mode = 94

ii.) Mr. Dent’s students asked how they did on their exam. What measure of central tendency would best answer their question? Mean b/c it takes into account all the data.

c.) The table below lists the areas of some large shopping malls in the United States:

	Mall	Gross Leasable Area (ft ²)
1	Del Amo Fashion Center, Torrance, CA	3,000,000
2	South Coast Plaza/Crystal Court, Costa Mesa, CA	2,918,236
3	Mall of America, Bloomington, MN	2,472,500
4	Lakewood Center Mall, Lakewood, CA	2,390,000
5	Roosevelt Field Mall, Garden City, NY	2,300,000
6	Gurnee Mills, Gurnee, IL	2,200,000
7	The Galleria, Houston, TX	2,100,000
8	Randall Park Mall, North Randall, OH	2,097,416
9	Oakbrook Shopping Center, Oak Brook, IL	2,006,688
10	Sawgrass Mills, Sunrise, FL	2,000,000
10	The Woodlands Mall, The Woodlands, TX	2,000,000
10	Woodfield, Schaumburg, IL	2,000,000

i.) mean = 2,280,402.3 median = 2,150,000 mode = 2,000,000

ii.) You are a realtor who is trying to lease mall space in different areas of the country to a large retailer. Which measure would you talk about if the customer felt that the malls were too large for his store?
Mode b/c it shows the smallest square footage for leasing when compared to the mean and median.

b.) Use the stem-leaf plot below: Title

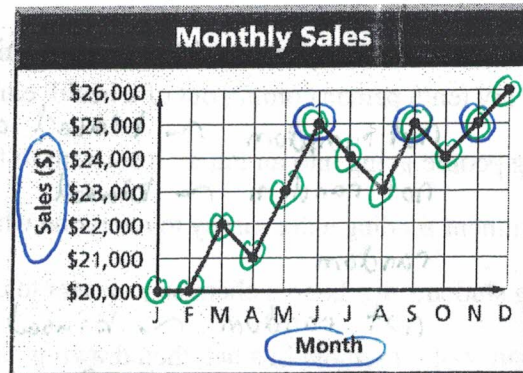
Stem	Leaf
4	4 5 6 7 7
5	3 5 6 <u>7</u> 8 9
6	7 7 8 9 9 9

Key 5|7 = 57

i.) mean = 57.4 median = 57 mode = 69

ii.) Why is having the data setup like this helpful?
Simple to read, it is in order, and you can see the mode and median quickly.

d.) Below is graph of a company’s monthly sales:



i.) mean = 23,200 median = 23,500 mode = 25,000

ii.) What are two types of situations that this graph could be used for?

a) Shows a company's annual sales (growth)
b) can be used to determine whether or not employees deserve a raise or bonus.