Section 1.8 – Measures of Central Tendency

Characteristics of center

Measures of center include mean and median, as tools for analyzing data. Not only determine the value of each measure of center, but also interpret those values.

Definition

A ***measure of center*** is a value at the center or middle of a data set

Mean

Definition

Definitions

The ***arithmetic mean*** of a variable is computed by adding all the values of the variable in the data set and dividing by the number of observations.

The ***population arithmetic mean***, μ (pronounced "mew"), is computed using all the individuals in a population. The population mean is a *parameter*.

The ***sample arithmetic mean***, $\bar{x}$ (pronounced "*x*-bar"), is computed using sample data. The sample mean is a *statistic*.

$$mean = \frac{\sum x}{N} \quad \begin{matrix} \leftarrow \text{ sum of all data values} \\ \leftarrow \text{ number of data values} \end{matrix}$$

$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{N} = \frac{\sum x}{N}$ is the mean of a set of ***sample values***.

$\mu = \frac{x_1 + x_2 + \cdots + x_n}{N} = \frac{\sum x}{N}$ is the mean of all values in a ***population***.

Notation

Σ denotes the ***sum*** of a set of values.

x is the *variable* usually used to represent the individual data values.

n represents the *number of data values* in a ***sample***. (***sample size***)

N represents the *number of data values* in a ***population***.

Example

Find the mean of these first five word counts from men: 27,531; 15,684; 5,638; 27,997; and 25,433

Solution

$$\bar{x} = \frac{\sum x}{n} = \frac{27{,}531 + 15{,}684 + 5{,}638 + 27{,}997 + 25{,}433}{5}$$

$$= \frac{102{,}283}{5}$$

$$= 20{,}456.6$$

The mean of the first five word counts is 20,456.6 words.

Example

The following data represent the travel times (in minutes) to work for all seven employees of a start-up web development company.

23, 36, 23, 18, 5, 26, 43

a) Compute the population mean of this data.

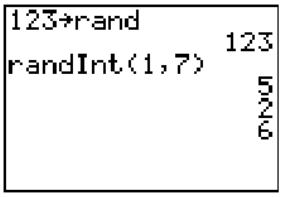
b) Then take a simple random sample of $n = 3$ employees. Compute the sample mean. Obtain a second simple random sample of $n = 3$ employees. Again compute the sample mean.

Solution

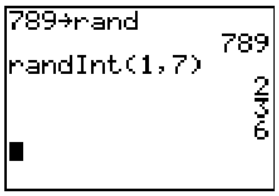
a) $\mu = \frac{\sum x}{N} = \frac{23 + 36 + 23 + 18 + 5 + 26 + 43}{7} = \frac{174}{7} = 24.9 \text{ minutes}$

b) Obtain a simple random sample of size $n = 3$ from the population of seven employees. Use this simple random sample to determine a sample mean. Find a second simple random sample and determine the sample mean.

1	2	3	4	5	6	7
23	36	23	18	5	26	43



$$\bar{x} = \frac{5 + 36 + 26}{3} = 22.3$$



$$\bar{x} = \frac{36 + 23 + 26}{3} = 28.3$$

Median

Definition

The ***median*** of a data set is the measure of center that is the ***middle value*** when the original data values are arranged in order of increasing (or decreasing) magnitude. The median is often denoted by $\tilde{x}$ (***x-tilde***)

Finding the Median

First ***sort*** the values (arrange them in order), the follow one of these

1. If the number of data values is odd, the median is the number located in the exact middle of the list.
2. If the number of data values is even, the median is found by computing the mean of the two middle numbers.

In order of ***even*** number of values: 5.40; 1.10; 0.42; 0.73; 0.48; 1.10

0.42 0.48 0.73 1.10 1.10 5.40

$$\text{Median} = \frac{0.73 + 1.10}{2} = 0.195$$

In order of ***odd*** number of values: 5.40; 1.10; 0.42; 0.73; 0.48; 1.10; 0.66

0.42 0.48 0.66 0.73 1.10 1.10 5.40

$$\text{Median} = 0.73$$

Example

Find the median for this sample of data values: 27,531; 15,684; 5,638; 27,997; and 25,433

Solution

First sort the data: 5,638 15,684 25,433 27,531 27,997
Median is 25,433

Example

Find the median for this sample of data values: 27,531, 15,684, 5,638, 27,997, 25,433 and 8,077

Solution

First sort the data: 5,638 8,077 15,684 25,433 27,531 27,997

$$\text{Median is} = \frac{15{,}684 + 25{,}433}{2} = 20{,}558.5$$

Example

The following data represent the travel times (in *minutes*) to work for all seven employees of a start-up web development company. 23, 36, 23, 18, 5, 26, 43

a) Determine the median of this data.

b) Suppose a new employee is hired who has a 130 minute commute. How does this impact the value of the mean and median?

Solution

a) ***Step 1:*** 5, 18, 23, 23, 26, 36, 43

Step 2: There are $n = 7$ observations.

Step 2: $\frac{n+1}{2} = \frac{7+1}{2} = 4$

Median is 23 5, 18, 23, **23**, 26, 36, 43

b) Mean before new hire: 24.9 minutes
Median before new hire: 23 minutes
Mean after new hire: 38 minutes
Median after new hire: 24.5 minutes

A numerical summary of data is said to be **resistant** if extreme values (very large or small) relative to the data do not affect its value substantially.

Relation Between the Mean, Median, and Distribution Shape

Distribution Shape	***Mean vs. Median***
Skewed left	Mean substantially smaller than median
Symmetric	Mean Roughly equal to median
Skewed right	Mean substantially larger than median

(a) Skewed Left
Mean < Median

(b) Symmetric
Mean = Median

(c) Skewed Right
Mean > Median

Example

The following data represent the birth weights (in pounds) of 50 randomly sampled babies.

5.8	7.4	9.2	7.0	8.5	7.6	7.9	7.8	7.9	7.7	9.0	6.7	8.2	7.0
8.7	7.2	6.1	7.2	7.1	7.2	7.9	5.9	7.0	7.8	7.2	7.6	7.4	7.1
7.3	6.4	7.4	8.2	9.1	7.3	9.4	6.8	7.0	8.1	8.0	7.2	7.0	8.7
7.3	6.9	6.9	6.4	7.8	7.5	7.1	7.5						

a) Find the mean and median.

b) Describe the shape of the distribution

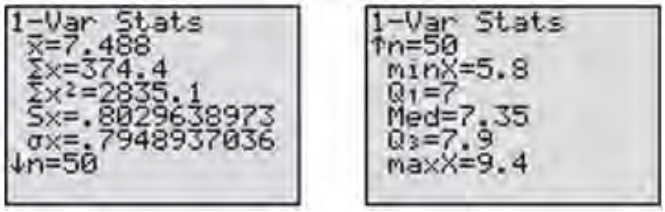
c) Which measure of central tendency better describes the average birth weight?

Solution

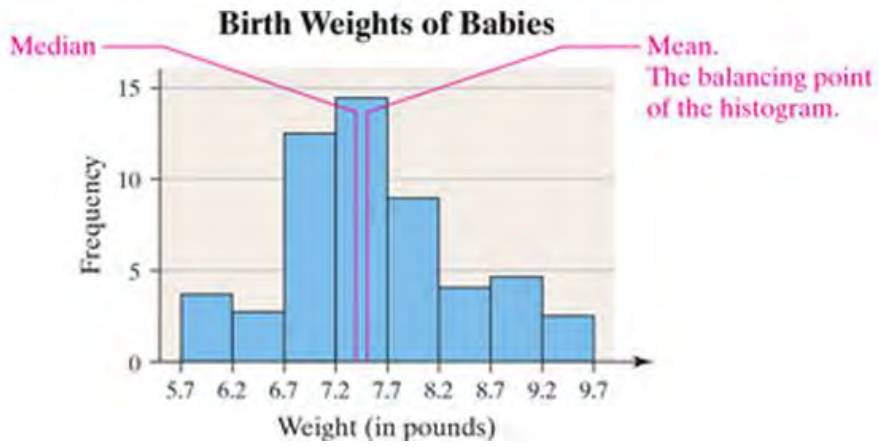
a) Using the calculator:

Mean: $\bar{x} = 7.488 \approx 7.49$

Median $M = 7.35$



b)



The distribution is bell shaped.

c) Because the mean and median are close in value, we use the mean as the measure of central tendency.

Example

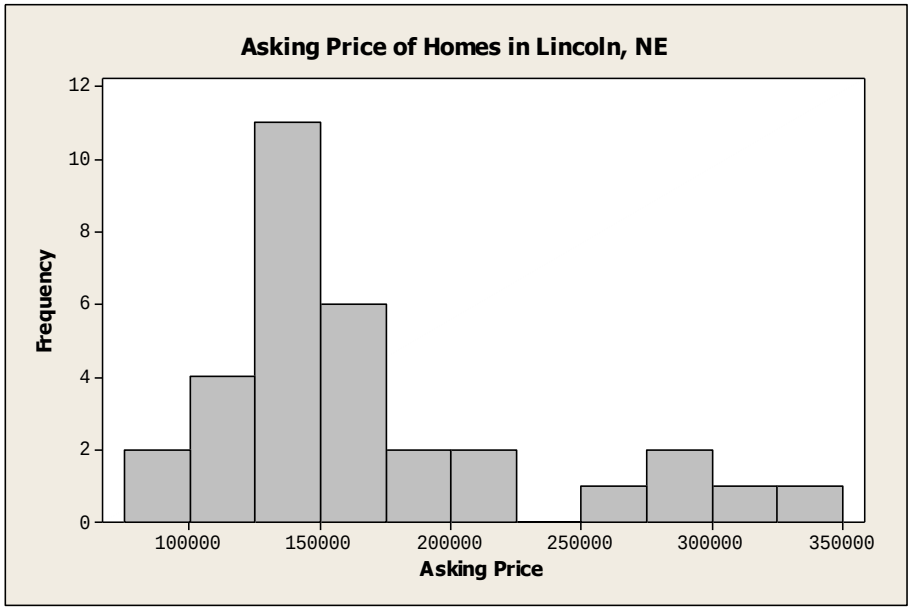
The following data represent the asking price of homes for sale in Lincoln, NE.

79,995	128,950	149,900	189,900
99,899	130,950	151,350	203,950
105,200	131,800	154,900	217,500
111,000	132,300	159,900	260,000
120,000	134,950	163,300	284,900
121,700	135,500	165,000	299,900
125,950	138,500	174,850	309,900
126,900	147,500	180,000	349,900

Find the mean and median. Use the mean and median to identify the shape of the distribution. Verify your result by drawing a histogram of the data.

Solution

The mean asking price is $168,320 and the median asking price is $148,700. Therefore, we would conjecture that the distribution is skewed right.



Mode

Definition

The ***mode*** of a data set is the value that occurs with the greatest frequency

A data set can have one, more than one, or no mode

Bimodal	two data values occur with the same greatest frequency
Multimodal	more than two data values occur with the same greatest frequency
No Mode	no data value is repeated

Example

a) Find the mode of: 5.40 1.10 0.42 0.73 0.48 1.10

Mode is 1.10

b) Find the mode of: 27 27 27 55 55 55 88 88 99

Mode is 27 & 55 (***bimodal***)

c) Find the mode of: 1 2 3 4 5 6 7

No Mode

Midrange

Definition

The midrange is the value midway between the maximum and minimum values in the original data set. It is found by adding the maximum data value to the minimum data value and then dividing the sum by 2:

$$Midpoint = Midrange = \frac{minimum\ data\ value + maximum\ data\ value}{2}$$

Example

Find the midrange of these values: 27,531; 15,684; 5,638; 27,997; and 25,433

Solution

$$Midrange = \frac{minimum\ data\ value + maximum\ data\ value}{2}$$

$$= \frac{5{,}638 + 27{,}997}{2}$$

$$= 16{,}817.5$$

Critical Thinking

- Think about whether the results are reasonable.
- Think about the method used to collect the sample data.

Example

For each of the following, identify a major reason why the mean and median are not meaningful statistics

a) Zip codes: 1260, 77573, 77574, 90210, 77550

b) Ranks of stress levels from different jobs: 2, 3, 1, 7, 9

Solution

a) The zip codes don't measure or count anything. The numbers are actually labels for geographic locations.

b) The ranks reflect an ordering, but they don't measure or count anything. The rank of 1 might come from a job that has a stress level substantially greater than the stress level from the job with a rank of 2, so the different numbers don't correspond to the magnitude of the stress levels.

Beyond the Basics of Measures of Center

Mean from a Frequency Distribution

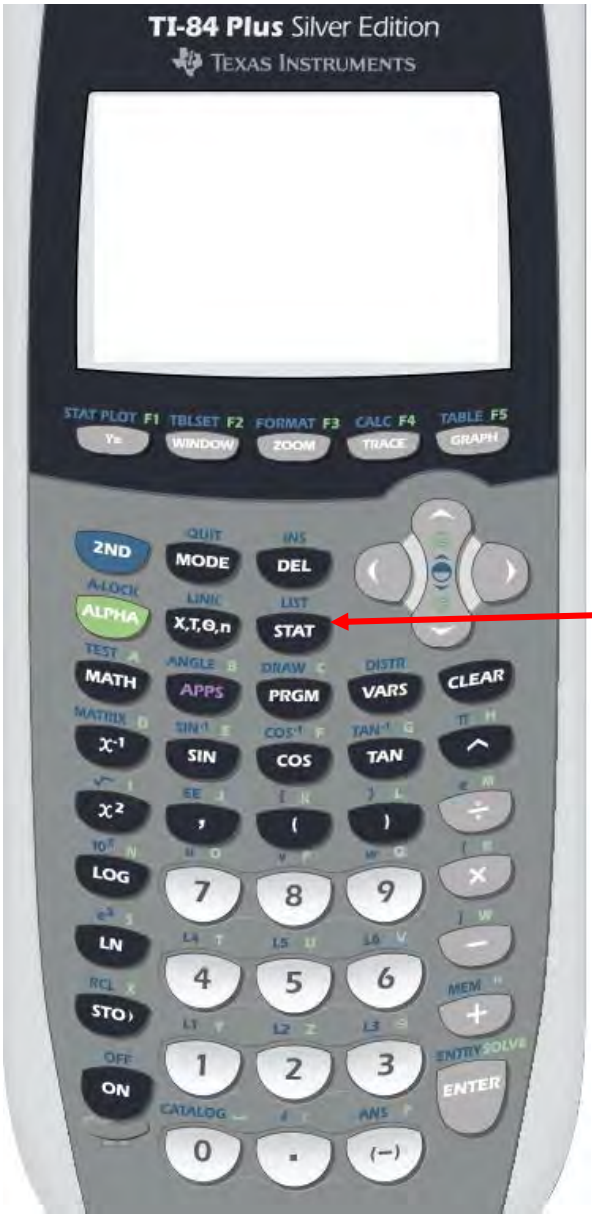
Assume that all sample values in each class are equal to the class midpoint.

multiply each frequency and class midpoin, then add the products

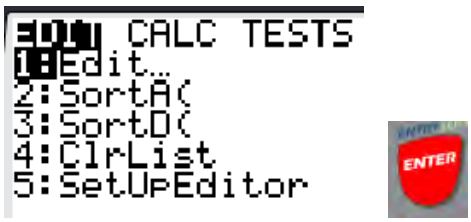
Mean from frequency distribution: $\bar{x} = \dfrac{\sum(f \cdot x)}{\sum f}$ ← *sum of frequencies*

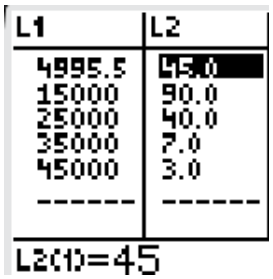
Example

Word Counts	Frequency f	Class Midpoint x	$f \cdot x$
0 – 9,9999	45	4,999.5	224,977.5
10,000 – 19,999	90	14,999.5	1,349,955
20,000 – 29,999	40	24,999.5	999,980
30,000 – 39,999	7	34,999.5	244,996.5
40,000 – 49,999	3	44,999.5	134,998.5
Totals:	$\Sigma f = 185$		$\sum f \cdot x = 2{,}954{,}907.5$
			$\bar{x} = \dfrac{\sum(f \cdot x)}{\sum f} = \dfrac{2{,}954{,}907}{185} = 15{,}972.5$



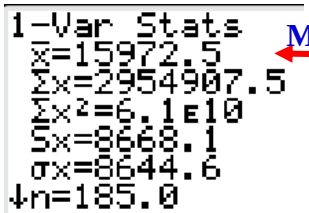
STAT





Click on 

→ **CALC** → **1**



Mean Value

Weighted Mean

When data values are assigned different weights, we can compute a ***weighted mean***.

$$\text{weighted mean:}\quad \bar{x} = \frac{\sum(w \cdot x)}{\sum w}$$

Example

In her first semester of college, a student of the author took five courses. Her final grades along with the number of credits for each course were: A (3 credits); A (4 credits); B (3 credits); C (3 credits) and F (1 credit). The grading system assigns quality points to letter grades as follows:

$A = 4$; $B = 3$; $C = 2$; $D = 1$; $F = 0$.

Compute her grade point average.

Solution

Weights = number of credits: $w = 3, 4, 3, 3, 1$.

Replace A, B, C, D, and F with the corresponding quality points: x = 4, 4, 3, 2, 0.

$$\bar{x} = \frac{\sum(w \cdot x)}{\sum w}$$

$$= \frac{(3\times 4)+(4\times 4)+(3\times 3)+(3\times 2)+(1\times 0)}{3+4+3+3+1}$$

$$= \frac{43}{14}$$

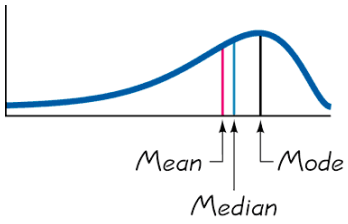
$$= 3.07$$

Skewed and Symmetric

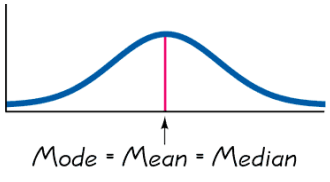
Definition

A distribution of data is ***skewed*** if it is not symmetric and extends more to one side than the other.
A distribution of data is ***symmetric*** if the left half of its histogram is roughly a mirror image of its right half.

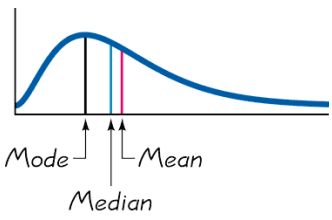
The mean and median cannot always be used to identify the shape of the distribution.



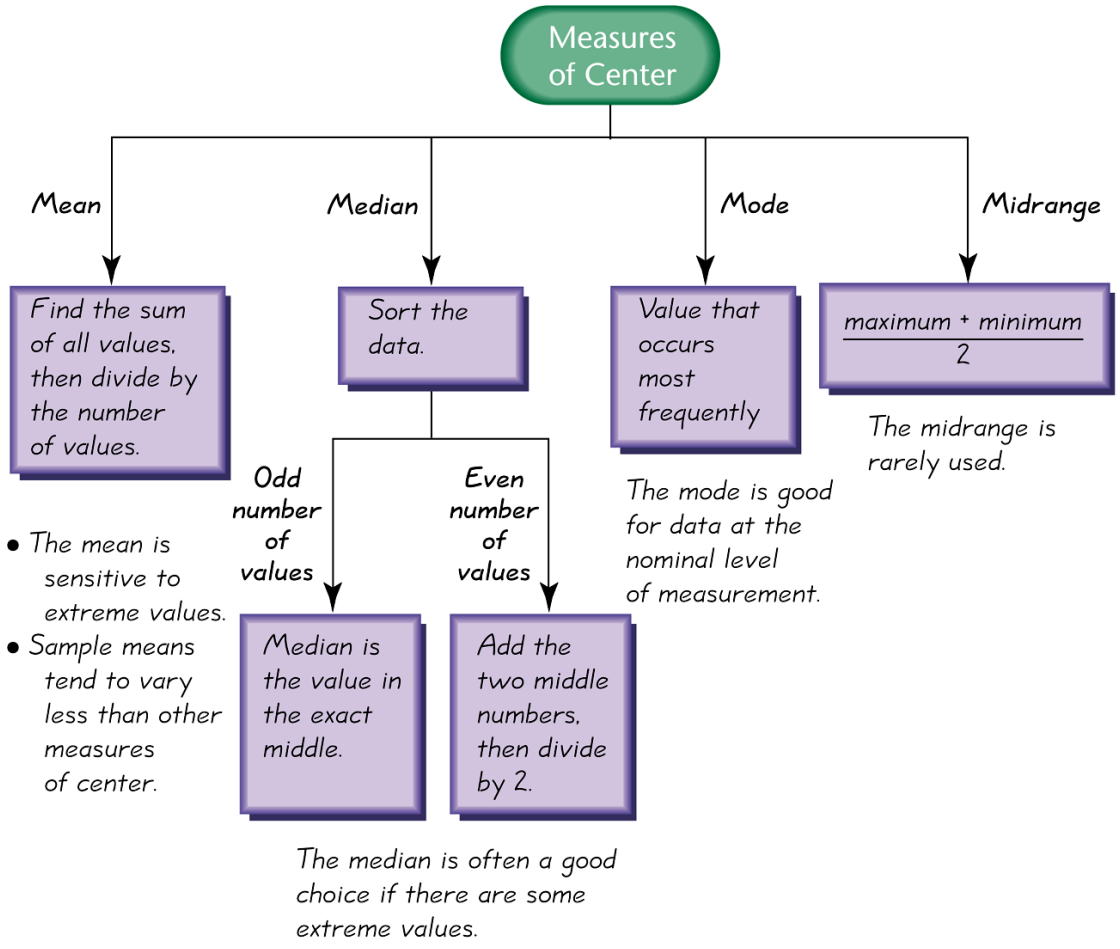
Skewed to the Left (Negatively)



Symmetric



Skewed to the Right (Positively)


Measures of Center
Mean
Median
Mode
Midrange
Find the sum of all values, then divide by the number of values.
Sort the data.
Value that occurs most frequently
maximum + minimum
2
The midrange is rarely used.
The mode is good for data at the nominal level of measurement.
Odd number of values
Even number of values
• The mean is sensitive to extreme values.
• Sample means tend to vary less than other measures of center.
Median is the value in the exact middle.
Add the two middle numbers, then divide by 2.
The median is often a good choice if there are some extreme values.

Exercises *Section* 1.8 – Measures of Central Tendency

1. In what sense are the mean, median, mode and midrange measures of "center"?

2. A headline in USA Today stated that "Average family income drops 2.3%." What is the role of the term average in statistics? Should another term be used in place of average?

3. In an editorial, the Poughkeepsie Journal printed this statement: "The median price – the price exactly in between the highest and lowest -- …" Does that statement correctly describes the median? Why or why not?

4. A simple random sample of pages from Merriam-Webster's Collegiate Dictionary, 11th edition, was obtained. Listed below are the numbers of words defined on those pages. Given that this dictionary has 1459 pages with defined words, estimate the total number of defined words in the dictionary.

51 63 36 43 34 62 73 39 53 79

Find the *a*) mean *b*) median *c*) mode *d*) midrange

e) Is that estimate likely to be an accurate estimate of the number of words in the English language?

5. The National Highway Traffic Administration conducted crash tests of child booster seats for cars. Listed below are results from those teats, with the measurements given in hic (standard head injury condition units.

774 249 1210 546 431 612

Find the *a*) mean *b*) median *c*) mode *d*) midrange

e) According to the safety requirement, the hic measurement should be less than 1000 hic. Do the results suggest that all of the child booster seats meet the specified requirement?

6. The insurance Institution for Highway Safety conducted tests with crashes of new cars traveling at 6 mi/h. The total cost of the damages was found for a simple random sample of the tested cars and listed below

$7448 $4911 $9051 $6374 $4277

Find the *a*) mean *b*) median *c*) mode *d*) midrange

e) Do the different measures of center differ very much?

7. Listed below are the durations (in hours) of a simple random sample of all flights (as of this writing) of NASA's Space Transport System (space shuttle).

73 95 235 192 165 262 191 376 259 235 381 331 221 244 0

Find the *a*) mean *b*) median *c*) mode *d*) midrange

e) How might that duration time be explained?

8. Listed below are the playing times (in seconds) of songs that were popular at the time of this writing.

448 242 231 246 246 293 280 227 213 262 239 213 258 255 257

Find the *a*) mean *b*) median *c*) mode *d*) midrange

e) Is there on time that is very different from the others?

9. Listed below are numbers of satellites in orbit from different countries.

158 17 15 17 7 3 5 1 8 3 4 2 4 1 2 3 1 1 1 1 1 1 1 1

Find the *a*) mean *b*) median *c*) mode *d*) midrange

e) Does on country have an exceptional number of satellites?

f) Can you guess which country has the most satellites?

10. Listed below are costs (in dollars) of roundtrip flights from JFK airport in NY City to San Francisco. (All flights involve one stop and a two-week stay.) The airlines are US Air, Continental, Delta, United, American, Alaska, and Northwest.

30 Days in Advance	244	260	264	264	278	318	280
1 Day in Advance	456	614	567	943	628	1088	536

a) Find the mean and median for each then compare the two sets of results.

b) Does it make much difference if the tickets are purchased 30 days in advance or 1 day in advance?

11. The trend of thinner Miss America winners has generated charges that the contest encourages unhealthy diet habits among young women. Listed below are body mass indexes (BMI) for Miss America winners from two different periods.

BMI (1920 – 1930)	20.4	21.9	22.1	22.3	20.3	18.8	18.9	19.4	18.4	19.1
BMI – (from recent winners)	19.5	20.3	19.6	20.2	17.8	17.9	19.1	18.8	17.6	16.8

Find the mean and median for each then compare the two sets of results.

12. Find the mean of the data summarized in the given frequency distribution.

a)

Tar (mg) in Nonfiltered Cigarettes	***Frequency***
10 – 13	1
14 – 17	0
18 – 21	15
22 – 25	7
26 – 29	2

b)

Pulse Rates of Females	***Frequency***
60 – 69	12
70 – 79	14
80 – 89	11
90 – 99	1
100 – 109	1
110 – 119	0
120 – 129	1

13. A student of the author earned grades of B, C, B, A, and D. Those courses has these corresponding numbers credit hours: 3, 3, 4, 4, and 1. The grading system assigns quality points to letter grades as follows: A = 4; B = 3; C = 2; D = 1; F =0. Compute the grade point average (GPA) and round the result with two decimal places. If the Dean's list requires a GPA 3.00 or greater, did this student make the Dean's list?

14. A student of the author earned grades of 92, 83, 77, 84, and 82 on her five regular tests. She earned grades of 88 on the final exam and 95 on her class projects. Her combined homework grade was 77. The five regular tests count for 60% of the final grade, the final exam counts for 10%, the project counts for 15%, and homework counts for 15%. What is her weighted mean grade? What letter grade did she earn? (A, B, C, D, or F)

15. You are taking a class in which your grade is determined from five sources: 50% from your test mean, 15% from your midterm, 20% from your final exam, 10% from your computer lab work, and 5% from your homework. Your scores are 86 (test mean), 96 (midterm), 82 (final exam), 98 (computer lab), and 100 (homework). What is the weighted mean of your scores? If the minimum average for an A is 90, did you get an A?

16. During a quality assurance check, the actual coffee contents (in ounces) of six jars of instant coffee were recorded as 6.03, 5.59, 6.40, 6.00, 5.99, and 6.02.

a) Find the mean and the median of the coffee content.

b) The third value was incorrectly measured and is actually 6.04. Find the mean and median of the coffee content again.

c) Which measure of central tendency, the mean or the median, was affected more by the data entry error?

17. The table below shows the U.S. exports (in billions of dollars) to 19 countries for a recent year.

U.S. Exports (in billions of dollars)		
Canada: 261.1	Mexico: 151.2	Germany: 54.5
Taiwan: 24.9	Netherlands: 39.7	China: 69.7
Australia: 22.2	Malaysia: 12.9	Switzerland: 22.0
Saudi Arabia: 12.5	United Kingdom: 53.6	Japan: 65.1
South Korea: 34.7	Singapore: 27.9	France: 28.8
Brazil: 32.3	Belgium: 28.9	Italy: 15.5
Thailand: 9.1		

a) Find the mean and the median.

b) Find the mean and median without the U.S. exports to Canada. Which measure of central tendency, the mean or the median, was affected more by the elimination of the Canadian exports?

c) The U.S. Exports to India were $17.7 billion. Find the mean and median with the Indian exports added to the original data set. Which measure of central tendency was affected more by adding the Indian exports?