Mathematics 13 : Elementary Statistics

Unit 1 : Section 7

Lesson 23 :

Important Characteristics to Describe

Important Characteristics to Describe

Descriptive statistics methods are used to summarize and describe important characteristics of data.

Primarily, there are three important characteristics statisticians focus on in order to provide an overall summary and useful description of results.

Important Characteristics to Describe

The three important characteristics to describe are

the distribution of the data,

the central tendency of the data, and

the dispersion of the data.

Important Characteristics to Describe

The distribution of the data describes the overall pattern or shape of the data values.

Graphical descriptive statistics methods such as a distribution table, Pareto chart, pie chart, histogram, and stem-and-leaf plot are well suited for displaying the distribution of the data.

Important Characteristics to Describe

When dealing with qualitative data, the construction of either a distribution table, Pareto chart, or pie chart is recommended to facilitate the description of the distribution of the data.

When dealing with quantitative data, the construction of either a distribution table, histogram, or stem-and-leaf plot is preferred for displaying the distribution of the data.

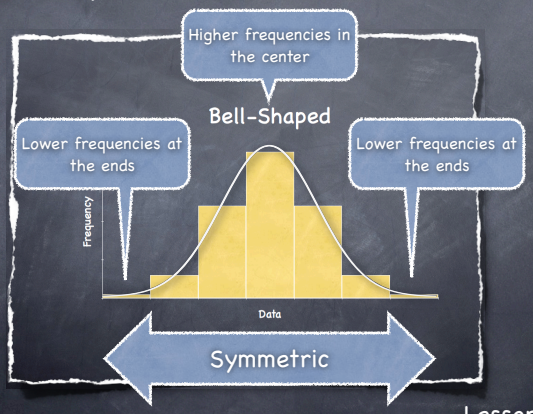
Important Characteristics to Describe

To describe the distribution of data that falls at the nominal level of measurement, each category is identified along with its corresponding frequency or proportion.

To describe the distribution of data that reaches the ordinal, interval, or ratio level of measurement, statistical terminology illustrating the pattern or shape revealed by the data values can be utilized.

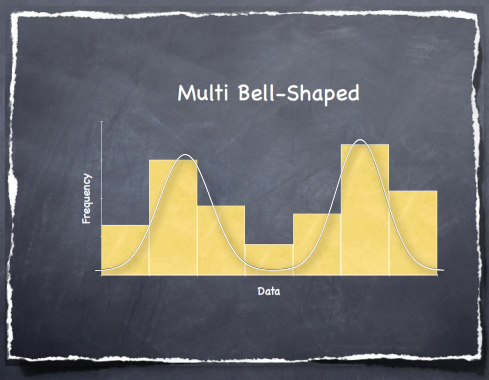
Important Characteristics to Describe

For instance,



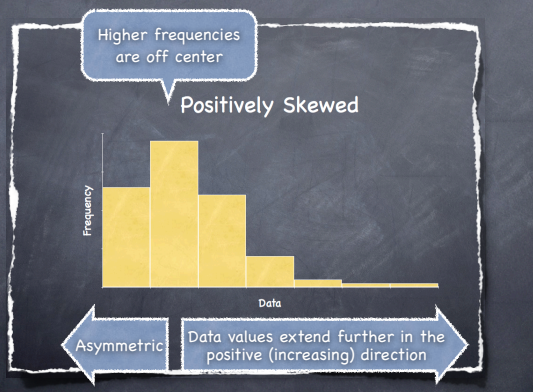
Important Characteristics to Describe

For instance,



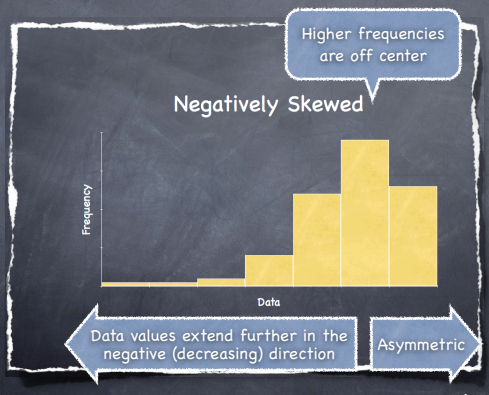
Important Characteristics to Describe

For instance,



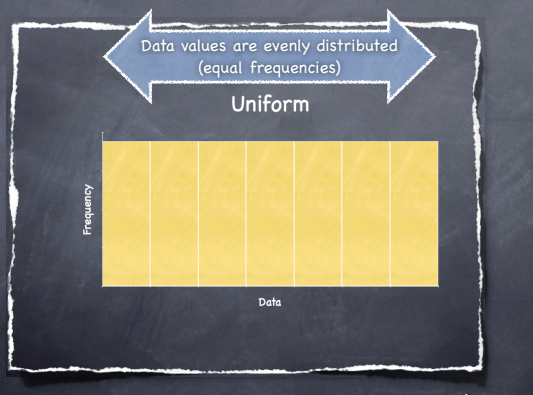
Important Characteristics to Describe

For instance,



Important Characteristics to Describe

For instance,



Important Characteristics to Describe

The central tendency of the data describes the overall average, center, or typical result of the data values.

Numerical descriptive statistics methods such as the mean, median, and mode are designed to measure the central tendency of the data.

Important Characteristics to Describe

The dispersion of the data describes the variation, spread, or diversity of the data values.

Numerical descriptive statistics methods such as the standard deviation, median absolute deviation, and variation ratio are designed to measure the dispersion of the data.

Important Characteristics to Describe

When dealing with qualitative data, determining the mode and the variation ratio is recommended to facilitate the description of the central tendency and the dispersion of the data.

This recommendation is necessary since the mean can not be calculated for qualitative data and the median can not be determined with qualitative data that falls at the nominal level of measurement.

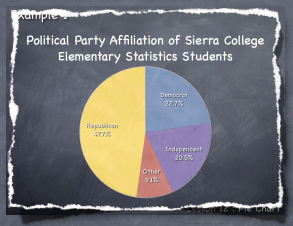
Important Characteristics to Describe

For instance,

use appropriate descriptive statistics methods to describe the important characteristics of the political party affiliation data (Q2) from our Sierra College Elementary Statistics Student Survey.

Important Characteristics to Describe

Distribution : Pie Chart



Central Tendency : Mode

Republican

Dispersion : Variation Ratio

$VR = 52.3\%$

Important Characteristics to Describe

Description:

More Sierra College Elementary Statistics Students identified themselves as republican, 47.7%, than any other political party. There were about the same percentage of students choosing democrat, 22.7%, as those who selected independent, 20.5%. The remaining 9.1% picked other. The variation ratio for the political party affiliation of Sierra College Elementary Statistics students was 52.3%.

Important Characteristics to Describe

When dealing with quantitative data, in most situations, calculating the mean and the standard deviation is preferred when describing the central tendency and the dispersion of the data.

The exception is when the mean (balancing point of the data) is distorted by the presence of outliers (extreme data values) in the data.

Important Characteristics to Describe

When dealing with quantitative data that includes outliers, calculating the median and the median absolute deviation is preferred when describing the central tendency and the dispersion of the data.

This preference is justified since both the median and the median absolute deviation are resistant to the presence of outliers.

Important Characteristics to Describe

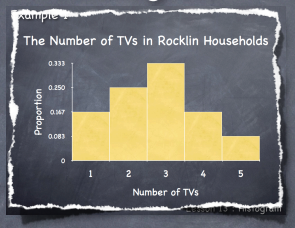
For instance,

use appropriate descriptive statistics methods to describe the important characteristics of the number of TVs in Rocklin households sample results.

Important Characteristics to Describe

Distribution : Histogram

Bell-Shaped



Central Tendency : Mean

$\bar{x} = 2.75$

Dispersion : Standard Deviation

$s_x \approx 1.22$

Important Characteristics to Describe

Description:

For this sample, the number of TVs in Rocklin households resembled a bell-shaped distribution with a mean of 2.75 TVs and a standard deviation of about 1.22 TVs.

Important Characteristics to Describe

For instance,

use appropriate descriptive statistics methods to describe the important characteristics of the age data (Q4) from our Sierra College Elementary Statistics Student Survey.

Important Characteristics to Describe

Distribution : Stem-and-Leaf Plot

Positively Skewed

Age of Sierra College Elementary Statistics Students

1	778888999
2	000000000001111122222223334
2	5699
3	4
3	8
4	03
4	
5	3

Central Tendency : Median

Med = 21

Dispersion : Median Absolute Deviation

MAD = 1

Important Characteristics to Describe

Description:

The age of Sierra College Elementary Statistics Students has a positively skewed shaped distribution with a median age of 21 years and a median absolute deviation of 1 year.

Important Characteristics to Describe

Distribution	Central Tendency	Dispersion
Quantitative Data with no Outliers		
Distribution Table, Histogram, or Stem-and-Leaf Plot	Mean	Standard Deviation
Quantitative Data with Outliers		
Distribution Table, Histogram, or Stem-and-Leaf Plot	Median	Median Absolute Deviation
Qualitative Data		
Distribution Table, Pareto Chart, or Pie Chart	Mode	Variation Ratio

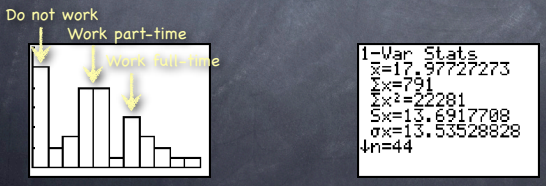
Example 1

Use appropriate descriptive statistics methods to describe the important characteristics of how many hours Sierra College Elementary Statistics students work in a typical week (Q9 of Student Survey).

Example 1

Distribution	Central Tendency	Dispersion
Quantitative Data with no Outliers		
Distribution Table, Histogram, or Stem-and-Leaf Plot	Mean	Standard Deviation

Multi Bell-Shaped $\bar{x} \approx 18.0$ $s_x \approx 13.7$



Example 1

Description:

The number of hours Sierra College Elementary Statistics students work in a typical week displays a multi bell-shaped distribution with a mean of 18.0 hours worked in a typical week and a standard deviation of 13.7 hours worked in a typical week.

Example 2

Use appropriate descriptive statistics methods to describe the important characteristics of the MPAA Ratings data in the Movie Database Sample.

Example 2

Distribution	Central Tendency	Dispersion
Qualitative Data		
Distribution Table, Pareto Chart, or Pie Chart	Mode	Variation Ratio

MPAA Rating	Frequency	Proportion
G	9	15.0%
PG	11	18.3%
PG-13	23	38.3%
R	17	28.3%

$$VR = 1 - \hat{p}_{\text{mode}} = 100\% - 38.3\% = 61.7\%$$

Mode = PG-13

Example 2

Description:

Most of the movies, 38.3%, in the Movie Database Sample had a PG-13 rating. The fewest movies, 15.0%, had a G rating. Whereas, R rated movies made up 28.3% of the sample, and 18.3% of the movies were rated PG. The variation ratio of the MPAA ratings data in the Movie Database Sample was 61.7%.

Example 3

A random sample of 20 sentences was selected from J. K. Rowling's book "Harry Potter and the Sorcerer's Stone". The length (in words) of each randomly selected sentence is given below.

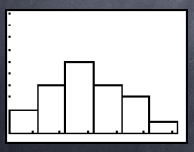
23, 12, 6, 25, 11, 20, 9, 19, 15, 8, 20, 3, 11, 30, 7, 11, 5, 14, 22, and 17

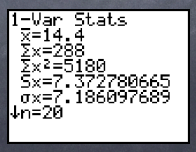
Use appropriate descriptive statistics methods to describe the important characteristics of this sample.

Example 3

Distribution	Central Tendency	Dispersion
Quantitative Data with no Outliers		
Distribution Table, Histogram, or Stem-and-Leaf Plot	Mean	Standard Deviation

Bell-Shaped $\quad \bar{x} \approx 14.4 \quad s_x \approx 7.4$





Example 3

Description:

The length of sentences sample from J. K. Rowling's book "Harry Potter and the Sorcerer's Stone" exhibits a bell-shaped distribution with a mean sentence length of 14.4 words and a standard deviation of 7.4 words.

Example 4

Use appropriate descriptive statistics methods to describe the important characteristics of the box office worldwide gross (in dollars) for the movies in the Movie Database Sample.

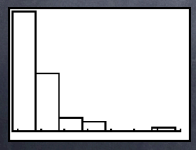
Example 4

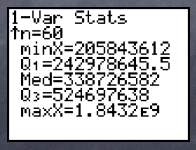
Distribution	Central Tendency	Dispersion
Quantitative Data with Outliers		
Distribution Table, Histogram, or Stem-and-Leaf Plot	Median	Median Absolute Deviation

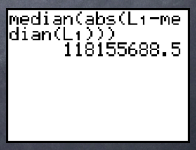
Med = \$338,726,582

Positively Skewed

MAD = \$118,155,688.5







Example 4

Description:

The box office worldwide gross for the movies in the Movie Database Sample has a positively skewed shaped distribution with a median of \$338,726,582 and a median absolute deviation of \$118,155,688.5.

Example 5

The randInt(minimum,maximum) random number generator function on the TI-84 calculator is programed to randomly generate an integer on the interval from a specified minimum value to a specified maximum value.

Example 5

For instance,

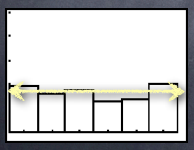
executing the randInt(1,6) command on the TI-84 calculator can simulate the random process of rolling a standard six sided die.

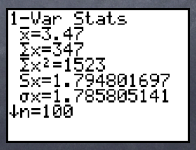
Example 5

Use appropriate descriptive statistics methods to describe the important characteristics of the results produced by using the randInt(1,6) random number generator function on the TI-84 calculator to simulate the random process of rolling a standard six sided die 100 times.

Example 5

Distribution	Central Tendency	Dispersion
Quantitative Data with no Outliers		
Distribution Table, Histogram, or Stem-and-Leaf Plot	Mean	Standard Deviation
Uniform	$\bar{x} \approx 3.5$	$s_x \approx 1.8$





Example 5

Description:

The results produced by using the randInt (1,6) random number generator function on the TI-84 calculator to simulate the random process of rolling a standard six sided die 100 times produced an approximate uniform shaped distribution with a mean of 3.5 and a standard deviation of 1.8.

Example 6

The heights (in inches) for a random sample of 30 men were measured. The following results were obtained.

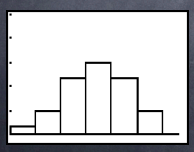
73.5, 69, 72, 74, 70.5, 71, 64.5,
69.5, 70.5, 75, 71.5, 66.5, 68,
76, 66.5, 70.5, 72, 67, 72.5,
69.5, 61.5, 73, 75.5, 69.5, 74.5,
66.5, 64, 67.5, 65, and 68

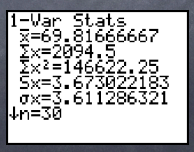
Example 6

Use appropriate descriptive statistics methods to describe the important characteristics of men's heights for this random sample.

Example 6

Distribution	Central Tendency	Dispersion
Quantitative Data with no Outliers		
Distribution Table, Histogram, or Stem-and-Leaf Plot	Mean	Standard Deviation
Bell-Shaped	$\bar{x} \approx 69.82$	$s_x \approx 3.67$





Example 6

Description:

For this random sample, men's heights followed a bell-shaped distribution with a mean height of 69.82 inches and a standard deviation of 3.67 inches.

Your solutions should be clear, complete, and sufficiently detailed in order to demonstrate your understanding and communicate your reasoning and method of solving the problem.

Exercise 1

The final grades received by students in Professor Brown's Experimental Psychology (PSYC 105) course last semester are given below.

B, B, C, B, C, A, A, B, D, C, B, F, B, A, C, D, C, and B

Use appropriate descriptive statistics methods to describe the important characteristics of this data.

Exercise 2

A nutritionist working for the United States Department of Agriculture (USDA) randomly selected three cartons of eggs from all of the available cartons of standard large eggs at a neighborhood grocery store. Each egg in the randomly selected cartons had their nutritional content analyzed. The data provide here are the amounts of milligrams of cholesterol in each of the sampled eggs.

186, 188, 179, 180, 192, 186, 183, 177, 184, 178, 191, 174,
189, 176, 190, 188, 196, 187, 184, 184, 192, 194, 198, 183,
183, 181, 187, 190, 186, 176, 183, 185, 191, 180, 184, 182

Use appropriate descriptive statistics methods to describe the important characteristics of this data.

Exercise 3

Use appropriate descriptive statistics methods to describe the important characteristics of the running time of movies in the Movie Database Sample.

Exercise 4

Use appropriate descriptive statistics methods to describe the important characteristics of how many hours Sierra College Elementary Statistics students study in a typical week (Q8 of Student Survey).

Exercise 5

University Hospital conducted a study in order to determine the blood type distribution of its patients. During each of the hospital's three laboratory work shifts (day, evening, and overnight), a random sample of fifteen patients who had blood tests performed during that shift was selected. The blood types for these randomly selected patients are given below.

Day Shift :	O, A, A, B, O, A, A, O, A, A, O, AB, O, O, A
Evening :	A, O, O, O, A, O, O, B, O, AB, O, O, A, A, B
Overnight :	O, O, O, O, A, B, A, A, O, A, A, A, B, O, O

Use appropriate descriptive statistics methods to describe the important characteristics of this data.

Exercise 6

A branch manager working for the Citywide Bank Corporation instructed a loan officer to record the FICO credit scores for a sample of its clients who applied for a home mortgage during the last month. The following data were collected.

782, 771, 835, 756, 804, 624, 765, 756, 590, 810, 692, 799, 795, 746,
772, 675, 735, 558, 758, 842, 691, 633, 736, 714, 663, 753, 755, 830,
685, 473, 814, 771, 704, 622, 729, 700, 758, 768, 750, 797, 708, 711,
742, 707, 679, 819, 790, 745, 623, 783, 723, 825, 684

Use appropriate descriptive statistics methods to describe the important characteristics of this sample.

Your solutions should be clear, complete, and sufficiently detailed in order to demonstrate your understanding and communicate your reasoning and method of solving the problem.

Exercise 1

The final grades received by students in Professor Brown's Experimental Psychology (PSYC 105) course last semester are given below.

B, B, C, B, C, A, A, B, D, C, B, F, B, A, C, D, C, and B

Use appropriate descriptive statistics methods to describe the important characteristics of this data.

Since the final grades received by students involves qualitative data, a distribution table, the mode, and the variation ratio would be appropriate descriptive statistics methods to describe the distribution, central tendency, and dispersion of this data.

Final Grade Distribution in Professor Brown's Experimental Psychology (PSYC 105) Course Last Semester

Final Grade	Proportion
A	16.7%
B	38.9%
C	27.8%
D	11.1%
F	5.6%

Mode → B ← Most Frequent

$Mode = \text{"B"}$ $\qquad VR = 1 - p_{\text{mode}} = 100\% - 38.9\% = 61.1\%$

Last semester, 38.9% of the students in Professor Brown's Experimental Psychology (PSYC 105) course received a final grade of "B". It was the most frequently occurring final grade in the course. This was followed by 27.8% of students who received a "C", 16.7% received an "A", 11.1% a "D", and 5.6% an "F". The variation ratio for the final grades received by students in this course was 61.1%.

Exercise 2

A nutritionist working for the United States Department of Agriculture (USDA) randomly selected three cartons of eggs from all of the available cartons of standard large eggs at a neighborhood grocery store. Each egg in the randomly selected cartons had their nutritional content analyzed. The data provide here are the amounts of milligrams of cholesterol in each of the sampled eggs.

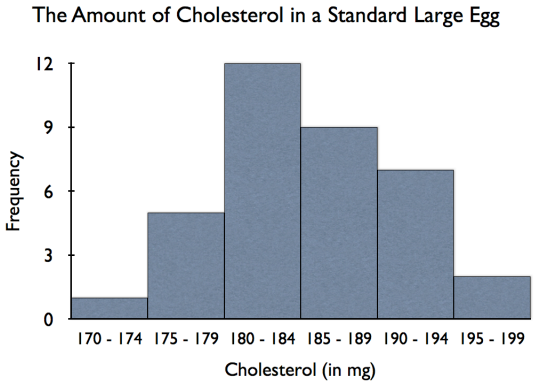
186, 188, 179, 180, 192, 186, 183, 177, 184, 178, 191, 174,
189, 176, 190, 188, 196, 187, 184, 184, 192, 194, 198, 183,
183, 181, 187, 190, 186, 176, 183, 185, 191, 180, 184, 182

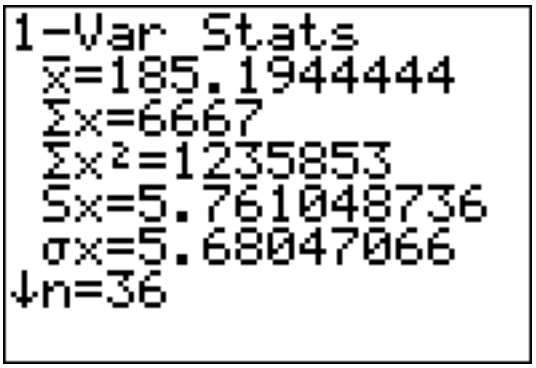
Use appropriate descriptive statistics methods to describe the important characteristics of this data.

$$185.2 - 3 \cdot 5.8 = 167.8 \qquad 185.2 + 3 \cdot 5.8 = 202.6$$

Since the amount of cholesterol in a standard large egg involves quantitative data with no outliers, a histogram, the mean, and the standard deviation would be appropriate descriptive statistics methods to describe the distribution, central tendency, and dispersion of this data.

Bell-Shaped





Based on the USDA sample, the amount of cholesterol in a standard large egg follows a bell-shaped distribution with a mean of 185.2 mg and a standard deviation of 5.8 mg.

Exercise 3

Use appropriate descriptive statistics methods to describe the important characteristics of the running time of movies in the Movie Database Sample.

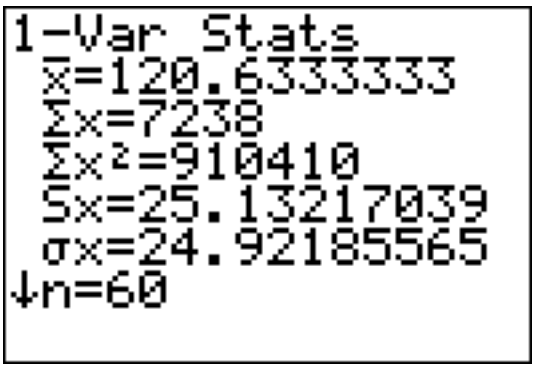
$$t = \frac{69 - 120.633}{25.132} \approx -2.05 \qquad t = \frac{194 - 120.633}{25.132} \approx 2.92$$

Since the running time of movies involves quantitative data with no outliers, a stem-and-leaf plot, the mean, and the standard deviation would be appropriate descriptive statistics methods to describe the distribution, central tendency, and dispersion of this data.

Bell-Shaped

Running Time (in minutes) of Movies in the Movie Database Sample

6	9												
7	8												
8	1	8											
9	1	1	2	4	5	8	8						
10	0	0	0	1	3	5	5	5	8				
11	0	0	1	3	5	5	7	8	8	9			
12	0	1	2	4	4	5	5	5	6	6	7	8	8
13	1	3	4	5	5	6	8						
14	1	2											
15	2	4											
16	1	2	9										
17	5	7											
18													
19	4												



The running time of movies in the Movie Database Sample exhibited a reasonably bell-shaped distribution with a mean of 120.6 minutes and a standard deviation of 25.1 minutes.

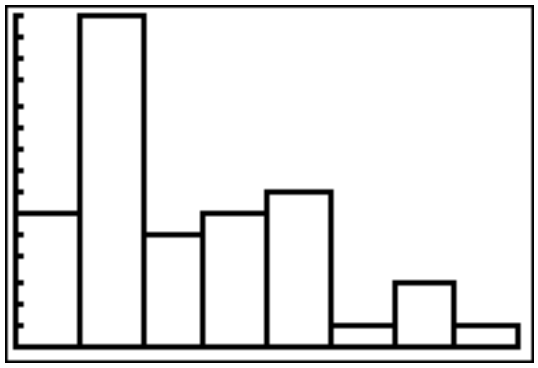
Exercise 4

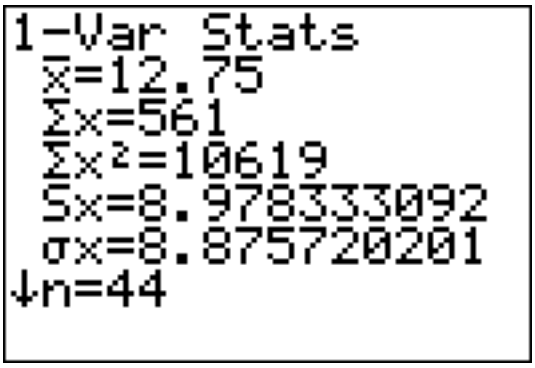
Use appropriate descriptive statistics methods to describe the important characteristics of how many hours Sierra College Elementary Statistics students study in a typical week (Q8 of Student Survey).

$$t = \frac{2 - 12.75}{8.978} \approx -1.20 \qquad t = \frac{35 - 12.75}{8.978} \approx 2.48$$

Since the hours studied in a typical week involves quantitative data with no outliers, a histogram, the mean, and the standard deviation would be appropriate descriptive statistics methods to describe the distribution, central tendency, and dispersion of this data.

Multi Bell-Shaped





The number of hours Sierra College Elementary Statistics students report studying in a typical week displayed a somewhat multi bell-shaped distribution with a mean of 12.75 hours per week and a standard deviation of 8.98 hours per week.

Exercise 5

University Hospital conducted a study in order to determine the blood type distribution of its patients. During each of the hospital's three laboratory work shifts (day, evening, and overnight), a random sample of fifteen patients who had blood tests performed during that shift was selected. The blood types for these randomly selected patients are given below.

Day Shift :	O, A, A, B, O, A, A, O, A, A, O, AB, O, O, A
Evening :	A, O, O, O, A, O, O, B, O, AB, O, O, A, A, B
Overnight :	O, O, O, O, A, B, A, A, O, A, A, A, B, O, O

Use appropriate descriptive statistics methods to describe the important characteristics of this data.

Since blood type (*O*, *A*, *B*, and *AB*) involves qualitative data, a distribution table, the mode, and the variation ratio would be appropriate descriptive statistics methods to describe the distribution, central tendency, and dispersion of this data.

Blood Type Distribution of Patients at University Hospital

Blood Type	Percentage
O	46.7%
A	37.8%
B	11.1%
AB	4.4%

Mode → *O* ← 46.7% ← Most Frequent

$$Mode = \text{Blood Type "O"}$$

$$VR = 1 - \hat{p}_{\text{mode}} = 100\% - 46.7\% = 53.3\%$$

Most of University Hospital's patients, 46.7%, have blood type "*O*". A little fewer, 37.8%, have blood type "*A*". A lot fewer, 11.1%, have type "*B*" blood, and very few, 4.4%, have blood type "*AB*". The variation ratio for the blood types of patients at University Hospital was 53.3%.

Exercise 6

A branch manager working for the Citywide Bank Corporation instructed a loan officer to record the FICO credit scores for a sample of its clients who applied for a home mortgage during the last month. The following data were collected.

782, 771, 835, 756, 804, 624, 765, 756, 590, 810, 692, 799, 795, 746,
772, 675, 735, 558, 758, 842, 691, 633, 736, 714, 663, 753, 755, 830,
685, 473, 814, 771, 704, 622, 729, 700, 758, 768, 750, 797, 708, 711,
742, 707, 679, 819, 790, 745, 623, 783, 723, 825, 684

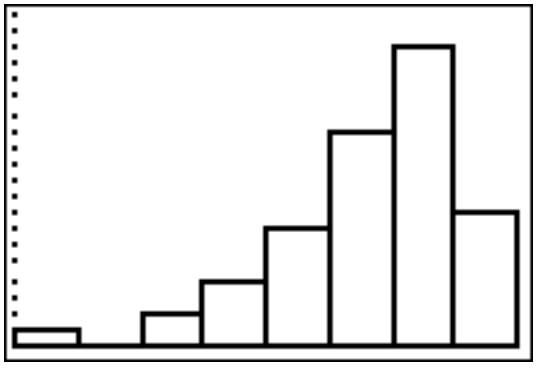
Use appropriate descriptive statistics methods to describe the important characteristics of this sample.

$$t = \frac{473 - 731.132}{73.953} \approx -3.49 \leq -3$$

Outlier = 473

Since FICO credit scores involves quantitative data with an outlier, a histogram, the median, and the median absolute deviation would be appropriate descriptive statistics methods to describe the distribution, central tendency, and dispersion of this data.

Negatively Skewed



Median = 746

MAD = 44

In this sample, the FICO credit scores of Citywide Bank Corporation clients who recently applied for a home mortgage exhibited a negatively skewed shaped distribution with a median of 746 and a median absolute deviation of 44.