MEASURES OF LOCATION AND VARIATION

The following set of (made-up) data will be used in the examples.

Data Set 1					Data Set 2					
6	19	18	15	10	4	5	6	2	5	
24	20	25	18	19	4	5	7	6	8	
12	20	18	20	14	4	4	5	5	5	5

1. **Mean**: To find the mean, add up all the data values and divide by the total number of data values.

Data Set 1: $(6 + 24 + 12 + 19 + \ldots. + 14)/15 = 17.2$ = the mean

2. **Median**: To find the median, order the data from smallest to largest (a stem and leaf plot is very helpful). Add 1 to the total number of values and divide by 2. This gives you the *position* of the median.

Data Set 1: 6 10 12 14 15 18 18 18 19 19 20 20 20 24 25
Position of the median: $(15 + 1)/2 = 8$. The median is the 8^{th} data value, which is 18.

Data Set 2: 2 4 4 4 4 5 5 5 5 5 5 5 6 6 7 8
Position of the median: $(16 + 1)/2 = 8.5$. The median is the number that is *halfway between* the 8^{th} and 9^{th} values. Add the 8^{th} and 9^{th} values and divide by 2 to get the median: $(5 + 5)/2 = 5$ = the median.

3. **Mode:** To find the mode, simply find the data value that occurs with the greatest frequency.

Data Set 2: The value 5 occurs most often. The mode is 5.
Data Set 1: The value 18 occurs 3 times but so does the value 20. We say "the data is *bimodal* with modes of 18 and 20". If there are more than 2 modes, perhaps the mode is not a good choice for describing central location. Consider using the mean or median instead if your data is *multimodal*.

4. **Pth Percentile:** To find the pth percentile, order the data from smallest to largest. Take *p* percent of the total number of data values. If this number is a fraction, round up to the nearest whole number. This is the *position* of the pth percentile. If *p* percent happens to be a whole number, average the value in that position with the value in the next position up to get the pth percentile.

Data Set 1: 6 10 12 14 15 18 18 18 19 19 20 20 20 24 25
Find the 30^{th} percentile.
Take 30% of 15: $.30(15) = 4.5$
Round up $= 5$
The 30^{th} percentile is the 5^{th} value = 15.

Data Set 1: 6 10 12 14 15 18 18 18 19 19 20 20 20 24 25
Find the 20^{th} percentile.
Take 20% of 15: .20(15) = 3
This is a whole number. Average the values in positions 3 and 4. The third value is 12 and the 4^{th} value is 14. (12 + 14)/2 = 13 = the 20^{th} percentile.

5. **Quartiles:** *The first quartile (Q1)* is the same as the 25^{th} percentile.
The third quartile (Q3) is the same as the 75^{th} percentile

6. **Range**: To find the range, subtract the smallest data value from the largest data value.
Data Set 1: 25 – 6 = 19 = the range.

7. **Interquartile Range:** To find the interquartile range, subtract Q1 from Q3.
Data Set 1: Q1 = 25^{th} percentile Q3 = 75^{th} percentile
.25(15) = 3.75 → 4^{th} value .75(15) = 11.25 → 12^{th} value
Q1 = 14 Q3 = 20

Q3 – Q1 = 20 – 14 = 6 = **IQR**

8. **Variance:** To find the variance, you must first calculate the mean. Determine whether your data is *population* data or *sample* data. This is important! The calculation formulas are slightly different. Subtract the mean from each data value and square the difference, Do this for every data value and add up the squared differences. Divide this result by the total number of data values (N) if this is population data. Divide by the total number of data values minus one (N-1) if this is sample data.

Data Set 2:

μ or $\bar{x}$ = 5

Data Value x_i	**Difference** $x_i - \mu$ **or** $x_i - \bar{x}$	**Squared Difference** $(x_i - \mu)^2$ **or** $(x_i - \bar{x})^2$
4	-1	1
5	0	0
6	1	1
5	0	0
5	0	0
4	-1	1
8	3	9
5	0	0
6	1	1
5	0	0
4	-1	1
5	0	0
2	-3	9
7	2	4
4	-1	1
5	0	0
		Total = 28

If the data is population data, divide 28 by 16 to get the population variance $\sigma^2 = 1.75$
If the data is sample data, divide 28 by (16 – 1) to get the sample variance $s^2 = 1.8667$

9. **Standard Deviation:** The standard deviation is the square root of the variance.

Population $\sigma = \sqrt{1.75} = 1.3229$

Sample $s = \sqrt{1.8667} = 1.3663$

10. **Coefficient of Variation:** The coefficient of variation tells us what percentage the standard deviation is of the mean, and is useful for comparing the variability of different populations. To find it, divide the standard deviation by the mean and multiply the result by 100.

$$(\sigma/\mu)(100) \quad \text{or} \quad (s/\bar{x})(100)$$

population: $(1.3229/5)(100) = 26.5 \rightarrow$ the standard deviation is 26.5% of the mean value
sample: $(1.3663/5)(100) = 27.3 \rightarrow$ the standard deviation is 27.3% of the mean value

11. **The z score:** The z score of a data value indicates how many standard deviations that value is away from the mean. To find it, subtract the mean from the data value, then divide the result by the standard deviation.

$$z = \frac{x_i - \mu}{\sigma} \qquad \textit{or} \qquad z = \frac{x_i - \bar{x}}{s}$$

$\frac{7-5}{1.3229} = 1.51 =$ z score for the data value 7 *or* $\frac{7-5}{1.3663} = 1.46$

Grouped Data

The following (made up) data will be used as an example.

Class	Frequency
10-19	5
20-29	8
30-39	10
40-49	6

1. To find the **mean** for grouped data:
   a) Find the class mark for each class. Class mark = (lower class limit + upper class limit)/2
   For example, the class mark for the first class is (10 + 19)/2 = 14.5.

   b) Multiply each class mark by its class frequency and add the results.

   c) Divide the sum by the total frequency. This is the mean.

Class	Frequency	Class Mark	(Class Mark)(Freq.)
10-19	5	14.5	72.5
20-29	8	24.5	196
30-39	10	34.5	345
40-49	6	44.5	267
Total	**29**		**880.5**
Mean = 880.5/29 = 30.4			

2. To find the **variance** for grouped data:
   a) Find the mean. For each class, subtract the mean from the class mark and square the result.

   b) Multiply each squared difference by its class frequency and add the results.

   c) Divide the sum by the total frequency if the data is a population, or by the total frequency minus one (N – 1) if the data is from a sample.

Class	Frequency	Class Mark – 30.4	(Class Mark – 30.4)^2	(Class Mark – 30.4)^2 (Freq.)
10-19	5	-15.9	252.81	1264.05
20-29	8	-5.9	34.81	278.48
30-39	10	4.1	16.81	168.1
40-49	6	14.1	198.81	1192.86
Total	**29**			**2903.49**
Variance = 2903.49/29 = 100.12 *or* 2903.49/28 = 103.70				
Standard Deviation = square root of the variance = 10.00 or 10.18				