MEASURES OF CENTRAL TENDENCY

values that describe the centre of a body of data

Mean:
measure of central tendency found by dividing the sum of all the data by the number of elements; calculated as:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \quad \text{or} \quad \bar{x} = \frac{\sum x}{n}$$

Example: Scott took 7 math tests in one marking period. What is the mean test score?
89, 73, 84, 91, 87, 77, 94

$$\bar{x} = \frac{594}{7}$$

$$\doteq 85$$

Outlier—an element of a data set that is very different; affects the mean more than the median or the mode

Deviation—difference between a data value and the mean; sum of the mean deviations is 0

Weighted Mean:
calculated by multiplying all the data values by their weights and dividing by the sum of the weights:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i} \text{ or } \frac{\sum xw}{w}$$

Example: A teacher weights student marks in her final calculation as follows:
Knowledge and Understanding, 25%; 78
Application, 15%; 75
Problem Solving, 20%; 80
Communication,10%; 85
and the Final Exam, 30%.
A student's marks in these categories are 78, 75, 80, and 85, respectively.
The final exam, has not, as yet, been written. (a) Calculate the student's term mark before the final exam.

$$\bar{x} = \frac{78(.25) + 75(0.15) + 80(0.2) + 85(0.1)}{0.25 + 0.15 + 0.2 + 0.1}$$

$$\doteq 78.9\%$$

(b) What mark must the student achieve on the final exam to earn a final grade of 82?

$$78.9(0.7) + E(0.3) = 82$$

$$E = \frac{82 - 78.9(0.7)}{0.3}$$

$$\doteq 89.2\%$$

Example: A sample of car owners was asked how old they were when they got their first car. The results were then reported in a frequency distribution. Calculate the mean.

Age	16–20	21–25	26–30	31–35	36–40
Frequency	10	18	12	8	2
Mean	18	23	28	33	38
Freq x Mean	180	414	336	264	76

$$\bar{x} = \frac{\text{Sum of (Freq x Mean)}}{\text{Sum of Freq}}$$

$$= \frac{\sum fm}{\sum f}$$

$$= \frac{180 + 414 + 336 + 264 + 76}{10 + 18 + 12 + 8 + 2}$$

$$\doteq 25.4$$

Median:
middle value in an ordered data set; the most appropriate measure of central tendency when outliers are present

Example: Monthly rents downtown and in the suburbs are collected from the classified section of a newspaper. Calculate the median rent in each district.

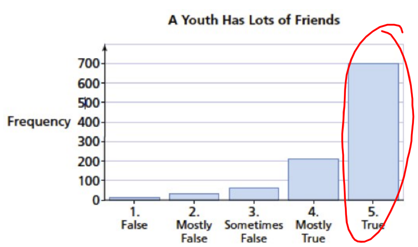
Downtown: \$850, \$750, \$1225, \$1000, \$800, \$1100, \$3200
Suburbs: \$750, \$550, \$900, \$585, \$220, \$625, \$500, \$800

Downtown → \$1000/month

Suburbs → $\frac{585+625}{2}$ = \$605/month

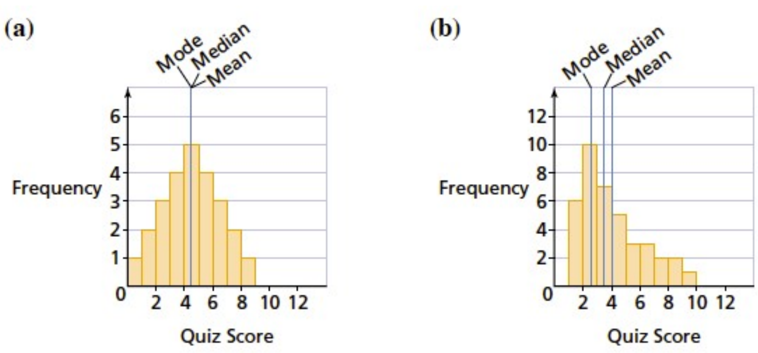
Mode:
most frequently occurring value; highest rectangle on a histogram;
Only measure of central tendency for qualitative data

Example: The graph represents the results when Ontario youths were asked if they have a lot of friends. Which measure of central tendency can be used to represent these data?



→ Mode

Compare the following data sets. What is the relationship between the shape of the distribution and the mean, median, and mode?



Skewed Right
"Mean is to the right of the median"

Symmetrical Distribution
Mean = Median = Mode

WHAT MEASURE SHOULD YOU USE?

While there can be no single rule governing which measure of central tendency you should use to describe a set of data, take the following into consideration:

- Outliers will affect the ___mean___ the most. If data contain outliers, use the ___median___ to avoid misrepresenting the data.

- If the data are strongly skewed, the ___median___ may best represent the central tendency of the data.

- If the data are roughly symmetric, the ___mean___ and the ___median___ will be close, so either is appropriate.

- If the data are not numeric (e.g., colour) or if the frequency of the data is more important than the value (e.g., shoe size), then the ___mode___ should be used.

Practice: #1-3ace, 5, 14-17