Measures of Central Tendency
§2.1

<u>Central Tendency</u> – a number that represents where data tends to center.

<u>Measures of Central Tendency</u>

1. <u>Mode</u> – the value that occurs most frequently in the data.
   *data may have no mode, 1 mode, 2 modes, etc.
   *mode is appropriate to determine the most common value is a set.

2. <u>Median</u> – the central (middle) value of an ordered data set.
   *if you know the median, there is an equal number of data values above and below it.

   Odd data, pick the middle number.
   Even data, add the two middle numbers and divide by 2.

3. Mean – the average of the data values.
   *add up the data values and divide by the number of entries.

<u>Example 1</u>

Find the mode, median, and mean of the following data.

2 7 9 3 5 7 8 2 1 6 8 9 7 3

Pg 84, 1-4, 9

Measures of Central Tendency
§2.1 (Day 2)

Notation for Mean

1. Sample Mean: $\bar{X} = \frac{\sum x}{n}$

$$\bar{X} = \frac{\text{the sum of the data values}}{\text{\# of data values}}$$

2. Population Mean: $\mu = \frac{\sum x}{N}$

$$\mu = \frac{\text{the sum of the data values}}{\text{\# of data values}}$$

*Formula is the same, but $\bar{X}$ and μ are used to distinguish between the mean for just a sample and the mean for an entire population.

<u>Resistant Measure</u> – a measure that is ***not influenced*** by extremely high or low data values.

Mean ?

Median?

Mode?

<u>Trimmed Mean</u> – a way of calculating the mean that makes it more resistant to extremely low or high data values.

order the data and then delete the top and bottom 5%
compute the mean of the remaining 90%

<u>Example 1</u>

14	20	20	20	20	23	25	30	30	30
35	35	35	40	40	42	50	50	80	80

Pg 85, 5-7

Measures of Variation
§2.2

Builder of measures of central tendency

Measures of Variations

1. Range – largest value – smallest value.

14	15	18	20	35

2. Standard Deviation – the average distance from the mean.

$$SD = s = \sqrt{\frac{\sum(x - \bar{X})^2}{n - 1}}$$

3. Variance – the average of the squares of the distance from the mean.

$$V = S^2 = \frac{\sum(x - \bar{X})^2}{n - 1}$$

Example 1

2	3	4	5	6	8	10	10

Example 2

5	8	11	11	13	13	17	20	22

Pg 98, 1a, 3ab, 4a, 5a

Measures of Variation
§2.2 (Day 2)

Coefficient of Variation
A disadvantage of the standard deviation as a comparative measure of variation is that it depends on the unit of measurement. This means it makes it difficult to compare measurements from different populations.

Therefore, we use coefficient of variation, which expresses the standard deviation as a percentage of the mean.

$$CV = \frac{s}{\bar{X}} \cdot 100 \quad \text{OR} \quad \frac{\sigma}{\mu} \cdot 100$$

Example 1
Find the standard deviation, variance, and the coefficient of variation.

2 6 8 9 4 3 7 1 5

Pg 98, 1b, 2ab, 3c, 4b, 5b-c, 6a-c

Measure of Variation
§2.2 (Day 3)

<u>Chebyshev's Theorem</u>
For any set of data (population or sample) and for any constant *k* greater than 1, the proportion of data that must be within *k* standard deviations on either side of the mean is at least:

$$1 - \frac{1}{k^2}$$

<u>Results of Chebyshevs Theorem</u>
For any set of data:
1. At least 75% of data fall between $\mu - 2\sigma$ and $\mu + 2\sigma$.
2. At least 88.9% of data fall between $\mu - 3\sigma$ and $\mu + 3\sigma$.
3. At least 93.8% of data fall between $\mu - 4\sigma$ and $\mu + 4\sigma$.

these are the minimum percentages, could be more

<u>Example 1</u>
Use Chebyshev's Law to find the 3 intervals.

5 5 6 6 6 7 7 8 9 10

<u>Example 2</u>
$\bar{X} = 525$ and $s(\sigma) = 30$. In what interval will at least 88.9% of the data fall?

Pg 100, 7, 8

Measure of Variation
§2.2 (Day 4)

Grouped Data
When data is grouped (such as a frequency table or histogram) we can estimate the mean and standard deviation.

Sample mean for Frequency Distribution

$$\bar{X} = \frac{\sum xf}{n}$$

x = midpoint of class
f = # of entries in that class
n = total # of entries

Sample Standard Deviation for Frequency Distribution

$$s = \sqrt{\frac{\sum (x - \bar{X})^2 f}{n - 1}}$$

Life Expectancy of Men	64-67	68-71	72-75
# of States	1	38	12

Pg 102, 12-15

Percentiles and Box-and-Whisker Plots
§2.3 (Day 1)

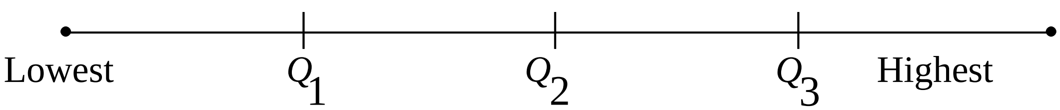
<u>Percentile</u> – the P^{th} percentile of a distribution of ordered data is a value such that P% of the data fall at or below it, and (100 – p)% fall at or above it.

Data must be ordered

Quartiles are percentiles that divide the data into **<u>fourths</u>**.

Q_1 (1st quartile) – 25th percentile
Q_2 (2nd quartile) – 50th percentile
Q_3 (3rd quartile) – 75th percentile

<u>3 quartiles divide the data into 4 parts.</u>



<u>Steps to compute quartiles</u>
1. Order data small to large
2. Find Q_2 (median)
3. Find Q_1 (median of the lower half of data…..lower quartile)
4. Find Q_3 (median of the upper half of data…..upper quartile)

<u>Interquartile Range</u> - Q_3 - Q_1

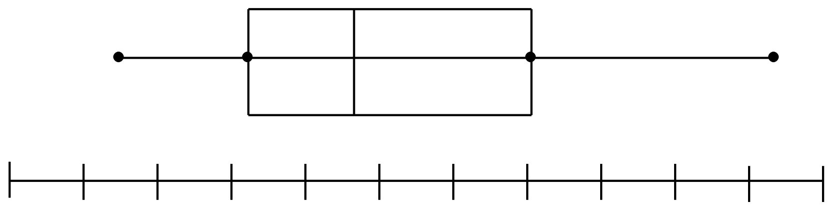
<u>Example 1</u>
Calories in Ice Cream Bars

342	379	319	353	295	234	182
310	439	111	201	190	151	131

Pg 112, 1-6

Percentiles and Box-and-Whisker Plots
§2.3 (Day 2)

Box-and-Whisker Plots



Example 1

Quiz Scores

14	18	18	19	21	22	24	24	24	25	26	27	27

Pg 112, 7-10
Appendix 2, Data 2 on Pg. A12