3.0 - Summation Notation

- In many statistical formulas, we are required to add several numbers (or products of numbers). So to show that type of addition in a general sense, **summation notation** is used.
  - For example consider the expression,

$$x_1 + x_2 + x_3 + x_4 + x_5 + x_6 =$$

  - Breaking down summation notation, or sigma notation –

$$\sum_{i=1}^{n} x_i$$

Expanding Summation Notation:

Ex 1). Expand each summation notation as an addition problem.

(1) $\sum_{i=1}^{8} (x_i + 2) =$

(2) $\sum_{i=3}^{10} x_i^2 f_i =$

(3) $\sum_{i=1}^{n} x_i y_i =$

An Alternate Method for Writing a Summation from *i = 1* to *n*.

Rewriting (or Condensing) an Addition Problem in Summation Notation:

Ex 2). Rewrite each sum in summation notation.

(1) $(x_1 - 5) + (x_2 - 5) + (x_3 - 5) + \ldots + (x_{12} - 5) =$

(2) $x_1^2 y_1 + x_2^2 y_2 + x_3^2 y_3 + \ldots + x_n^2 y_n =$

(3) $-3x_1 y_1 - 3x_2 y_2 - 3x_3 y_3 - \ldots - 3x_n y_n =$

Evaluating Summation Notation :

- Use the given data for examples and practice problems in this section.

$$x_1 = 1, x_2 = -3, x_3 = 7, x_4 = 4$$
$$y_1 = 2, y_2 = 0, y_3 = 5, y_4 = -2$$
$$f_1 = 5, f_2 = 3, f_3 = 1, f_4 = 6$$

Ex 3). Evaluate the following sums. Show a substitution step,

(1) $\sum x =$

(2) $\sum xf =$

(3) $\sum (x^2 - y) =$

(4) $(\sum x)(\sum f) =$

(5) $(\sum x)^2 =$

Important Properties: $\sum x^2 = (\sum x)^2$ $\qquad$ $2\sum x = \sum 2x$

$$\sum_{i=1}^{8} 4 = \qquad \sum_{i=1}^{n} k =$$

Homework: Summation Notation Worksheet

3.1 – Measures of Central Tendency

Three Sets of Measures necessary to describe data:

1. Measures of Central Tendency: _____

_____

2. Measures of Variation: _____

_____

3. Measures of Position: _____

_____

MEASURES OF CENTRAL TENDENCY:

A _____ is a characteristic or measure obtained by using the data values from a _____.

A _____ is a characteristic or measure obtained by using the data values from a specific _____.

General Rounding Rule – when computations are done in the calculation, rounding should not be done until the final answer is calculated. To avoid writing a long string of decimals, please give enough places (usually three or four) to obtain the same answer that a calculator would give after rounding on the last step. Round final answers to one more place than in the raw data.

MEAN: _____

SAMPLE: **POPULATION:** **GROUPED DATA:**

To find the mean for grouped data use the midpoints of the classes:

Ex 1):

Class Limits	Frequency	Midpoints
10 – 14	3	
15 – 19	1	
20 – 24	7	
25 – 29	5	
30 – 34	6	

MEDIAN: _____

MODE: _____

MiDRANGE: _____

Example pg. 118 "Applying the Concepts" #1 – 5:

Homework: pg. 118 #2 – 6 (e)

3.1 Day 2 – Weight Mean & Distribution Shapes

THE WEIGHTED MEAN:

_____

Ex 1) An instructor gives four 1-hour exams and one final exam, which counts as two 1-hour exams. Find a student's grade if she received 62, 83, 97, and 90 on the 1-hour exams and 82 on the final exam.

Ex 2). The costs of three models of helicopters are shown here. Find the weighted mean of the costs of the models.

Model	# sold	Cost
Sunscraper	9	$427,000
Skycoaster	6	$365,000
High-Flyer	12	$725,000

Review table of key ideas on page 116

DISTRIBUTION SHAPES:

SYMMETRIC/NORMAL BELL SHAPED CURVE:

NEGATIVELY/LEFT SKEWED:

POSITIVELY/RIGHT SKEWED:

Homework: pg. 119 #12, 18, 26, 30, 32

3.2 Day 1 – Measures of Variation

Answer the following questions while watching the video. *Against All Odds, Episode 6.*

1. It's important to know a distribution's measure of _____ and _____ in order to get a full picture of the data.

2. Standard Deviation measures how far the sales figures typically _____ from the _____.

3. Differences will almost always sum to _____.

4. What do you do to the sample variance to get the standard deviation?

<u>Ex 1).</u> Consider the following data sets:

Set 1	15	18	20	23	25	29	31
Set 2	1	3	4	23	24	43	63

- Find the mean and median of each data set.

- Comment on the data sets based on the measures of center.

RANGE: _____

VARIANCE: _____

STANDARD DEVIATION: _____

SHORTCUT FORMULAS:

Variance | Standard Deviation

Ex 1). Use the formula to find the standard deviation of Set 1 and Set 2. Describe what the standard deviation tells you about the data set.

Set 1: Set 2:

Ex 2). The number of yards gained in NFL playoff games by rookie quarterbacks is shown.

193	66	136	140	157
163	181	226	135	199

Find the range, variance, and standard deviation.

Homework: pg. 137 #1 – 6, 8, 10

3.2 Day 2 – Variance & Standard Deviation from Grouped Data

VARIANCE & STANDARD DEVIATION FOR GROUPED DATA: use midpoints

Ex 1). The costs per load (in cents) of 35 laundry detergents tested by a consumer organization are shown here. **Find the range, variance, and standard deviation.**

Class Limits	Frequency	Midpoint
13 – 19	2	
20 – 26	7	
27 – 33	12	
34 – 40	5	
41 – 47	6	
48 – 54	1	
55 – 61	0	
62 – 68	2	

COEFFICIENT OF VARIATION: _____

Ex 2). The mean of the number of sales of cars over a 3-month period is 87, and the standard deviation is 5. The mean of the commissions is $5225, and the standard deviation is $773. Compare the variations of the two.

RANGE RULE OF THUMB: _____

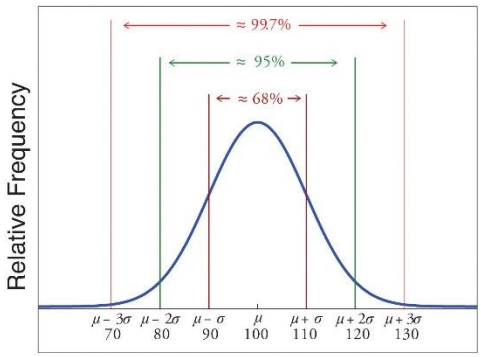
Homework: pg. 138 #20, 28 – 31

3.2 Day 3 – Chebyshev's Theorem

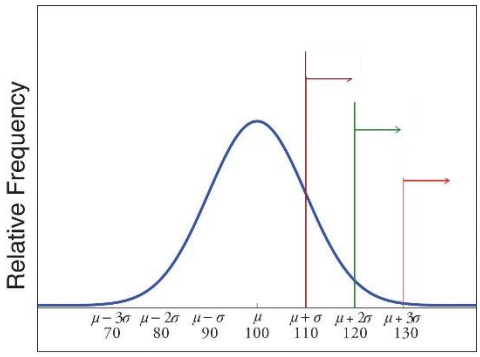
THE EMPIRICAL (NORMAL) RULE:

Ex 1). Suppose that the scores on a national achievement exam have a mean of 480 and a standard deviation of 90. If these scores are approximately normally distributed, then what values will define 1st, 2nd, and 3rd standard deviations from the mean?

Ex 2). Scores on IQ tests have a bell-shaped distribution with a $\mu = 100$ and standard deviation $\sigma = 10$. Discuss what the Empirical Rule implies concerning individuals with IQ scores of 110, 120, and 130.



(a) Whole Spectrum



(b) Higher End

CHEBYSHEV'S THEOREM: _____

_____

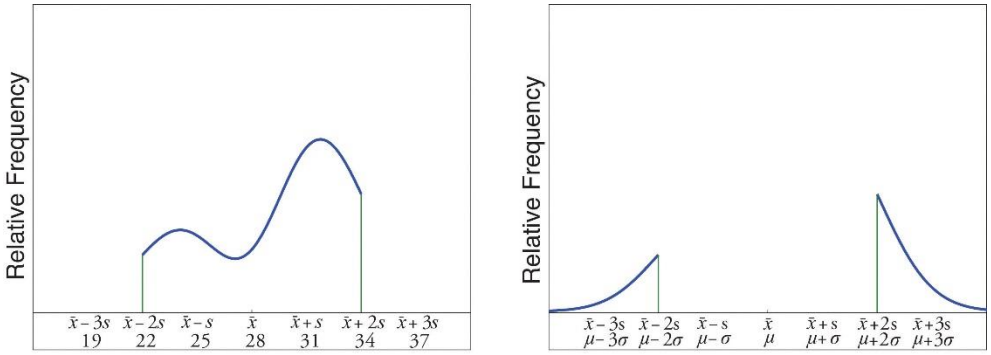
Ex 2). The mean height of a sample of men is 68.34 in and the standard deviation is 3.02 in.

(a) Find the interval that contains *at least 75%* of the heights of men.

(b) Find the interval that contains *at least 88.9%* of the heights of men.

(c) Find the interval that contains *at least 95%* of the heights of men.

Ex 3). A sample size of $n = 50$ has a mean of $\bar{x} = 28$ and standard deviation of $s = 3$. Without knowing anything else about the sample, what can be said about the number of observations that lie in the interval $[22,34]$? What can be said about the number of observations that lie outside of the interval?



(a) Within $\bar{x} \pm 2s$ (b) Outside $\bar{x} \pm 2$

Ex 4). Hockey pucks used in professional hockey games must weigh between 5.5 and 6 ounces. If the weight of the pucks manufactured by a particular process is bell-shaped and has a mean of 5.75 ounces and standard deviation of 0.125 ounces, what proportion of the pucks will be usable in professional games?

Ex 5). Speeds of vehicles on a section of highway have a bell-shaped distribution with mean of 60 mph and standard deviation of 2.5 mph. If the speed limit is 55 mph, what proportion of the cars are speeding?

Ex 6). Suppose that, as in the previous exercise, speeds of vehicles on a section of highway have a mean of 60 mph and standard deviation of 2.5 mph, but now the distribution is unknown.

a) If the speed limit is 55 mph, at least what proportion of vehicles must be speeding?

b) What can be said about the proportion of vehicles going 65 mph or faster?

Ex 7). Thirty-six students took an exam on which the average was 80 and the standard deviation was 6. A rumor says that five students had scores below 61. Can the rumor be true? Why or why not?

Homework: **Day 1 – pg. 140 #32 – 37 [due Thursday 10/6]**
Day 2 – pg. 140 #38 – 46 (no "g") [due Friday 10/7]
Day 3 – Chebyshev vs Empirical Rule Worksheet [due Monday 10/10]

3.3 – Measures of Position

Standard Score/z-score: _____

_____

Ex 1). A student scored 65 on an Algebra II test that had a mean of 50 and a standard deviation of 10; she scored 30 on a Chemistry test with a mean of 25 and a standard deviation of 5. Compare her relative positions on the two tests.

*Note: If the z-score is positive, the value is _____ the mean. If the z-score is negative, then the value is _____ the mean.

*When all data for a variable are transformed into z-scores, the resulting distribution will have a mean of ____ and a standard deviation of ____.

Percentiles: _____

Ex 2). A teacher gives a 20-point test to 10 students. The scores are shown here.

2, 3, 5, 6, 8, 10, 12, 15, 18, 20

a) Find the percentile rank of a score of 15.

b) Find the percentile rank of a score of 8.

c) Find the score corresponding to the 25th percentile.

d) Find the score corresponding to the 95th percentile.

Quartiles: _____

Ex 3). Find Q_1, Q_2, and Q_3 for the data set: 3, 7, 8, 9, 11, 12, 15, 16, 20, 23

Deciles: _____

Interquartile Range (IQR): _____

Outliers: _____

Ex 4). Check the following data for outliers: 5, 6, 12, 13, 15, 18, 22, 50

Reasons for Outliers:

1. May have been a measurement or observational _____
2. A recording error
3. The value may have been obtained from a subject that is not in the defined _____
4. The value may be a legitimate value that occurred by _____ (although the probability is extremely small)

*When a distribution is a _____ or bell-shaped, data values that are beyond ____ standard deviations of the mean can be considered suspected outliers.

Homework: pg. 153 #1, 2, 4 – 8, 10 – 16(e), 22, 30 (for #30 only do a & c)

3.4 ~ 5# SUMMARY & BOX PLOTS

In **Exploratory Data Analysis (EDA)** data can be organized using a _____. The measure of central tendency used is the _____. The measure of variation used is the _____. The data are represented graphically using a _____(box and whisker plot). The purpose: to examine data to find out what information can be discovered about the data such as the _____ and the _____.

The Five-Number Summary & Boxplots:

Ex 1). Construct a boxplot for the number of meteorites found in 10 states of the U.S.

30, 39, 47, 48, 78, 89, 138, 164, 215, 296

Information Obtained from a Boxplot:

1. If the median is near the _____, the distribution is approximately _____.
2. If the median falls to the _____ of the center, the distribution is _____ skewed.
3. If the median falls to the _____ of the center, the distribution is _____ skewed.
4. If the lines are about the _____ length, the distribution is approximately _____
5. If the right line is _____ than the left line, the distribution is _____ skewed.
6. If the right line is _____ than the left line, the distribution is _____ skewed.

Ex 2). A dietitian is interested in comparing the sodium content of real cheese with the sodium content of a cheese substitute. The data for two random samples are shown. Compare the distributions, using boxplots.

Real Cheese				**Cheese Substitute**			
310	420	45	40	270	180	250	290
220	240	180	90	130	260	340	310

Ex 3). The average number of hours per year that a driver is delayed by road congestion is listed here.

Orlando	***St. Louis***	***Denver***	***Austin***	***Washington***	***Dallas***	***Houston***	***Seattle***	***Atlanta***	***Los Angeles***
32	42	45	46	48	50	52	53	65	68

Check the data set for outliers.

Homework: pg. 166 #2 – 16(e)