Measures of Variation

Michael A. DeDonno Ph.D.

Florida Atlantic University

mdedonno@health.fau.edu

Agenda

- Explore measures of variation
  - Range
  - Variance
  - Standard deviation
- Compute the variance and standard deviation from raw data
- Explore the value of Standard Scores

Measures of Variation - Range

- An average is an attempt to summarize a set of data using just one number. However, an average taken by itself may not always be very meaningful.

  We need a statistical cross-reference that measures the spread of the data.
- **Range** – The difference between the largest and smallest values of a data distribution

Measures of Variation – Range (example)

- A large bakery regularly orders cartons of Maine blueberries.
- The average weight of the cartons is supposed to be 22 ounces. Random samples of cartons from two suppliers were weighed. The weights in ounces of the cartons were:
  - **Supplier I:** 17 22 22 22 27
  - **Supplier II:** 17 19 20 27 27
- Compute the range and the mean for each supplier

Measures of Variation – Range (example)

- **Supplier I:** 17 22 22 22 27
  - Range = 10
  - Mean = 22
- **Supplier II:** 17 19 20 27 27
  - Range = 10
  - Mean = 22.
- The samples have the same range and mean. How do they differ?
- Supplier I provides more cartons that have weights closer to the mean. Or, put another way, the weights of cartons from Supplier I are more clustered around the mean. The bakery might find Supplier I more satisfactory.

Measures of Variation
Variance and Standard Deviation

- Based on our review of blueberries in the cartons, it appears we need to measure the distribution, or spread of data around an expected value (either x or μ). *Variance (s^2)* and *standard deviation (s)* provide such measures.
- In statistics, the sample standard deviation and sample variance are used to describe the spread of data about the mean x .

Measures of Variation
Variance and Standard Deviation

- The defining formulas for variance (s^2)and standard deviation (s) emphasize the fact that the variance and standard deviation are based on the **<u>differences between each data value and the mean.</u>**

Defining Formulas (Sample Statistic)

$$\text{Sample variance} = s^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1} \qquad (1)$$

$$\text{Sample standard deviation} = s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}} \qquad (2)$$

where x is a member of the data set, $\bar{x}$ is the mean, and n is the number of data values. The sum is taken over all data values.

Measures of Variation
Variance and Standard Deviation

- You and your friends just measured the heights of your dogs with the following results (inches)
  - 27
  - 18
  - 6
  - 17
  - 12

Measures of Variation
Variance and Standard Deviation

- Several steps are involved in computing the variance and standard deviation. A table will be helpful.
- Since $n = 5$, take the sum of the entries in the first column and divide by 5 to find the mean x.

$$\overline{x} = \frac{\sum x}{n} = \frac{80}{5} = 16$$

Mean height

Total of all measurements

Result of how much each measurement deviates from the mean

Column 1	Column 2			Column 3		
x	$x - \overline{x}$			$(x - \overline{x})^2$		
27	27 - 16	=	11	$(11)^2$	=	121
18	18 - 16	=	2	$(2)^2$	=	4
6	6 - 16	=	-10	$(-10)^2$	=	100
17	17 - 16	=	1	$(1)^2$	=	1
12	12 - 16	=	-4	$(-4)^2$	=	16
$\sum x = 80$				$\sum(x - \overline{x})^2$	=	242

Measures of Variation
Variance and Standard Deviation

- Now one might think that we could just take the sum of the deviations from the mean and divide by *n* to obtain an average variance.
- Unfortunately, this is not possible as the sum of the deviations from the mean will always be zero.

Result of how much each measurement deviates from the mean

Column 1	Column 2			Column 3		
x	$x-\bar{x}$			$(x-\bar{x})^2$		
27	27 - 16	=	11	$(11)^2$	=	121
18	18 - 16	=	2	$(2)^2$	=	4
6	6 - 16	=	-10	$(-10)^2$	=	100
17	17 - 16	=	1	$(1)^2$	=	1
12	12 - 16	=	-4	$(-4)^2$	=	16
$\sum x = 80$				$\sum(x-\bar{x})^2$	=	242

The sum of the deviations from the mean will always be zero!

Measures of Variation
Variance and Standard Deviation

- Since the sum of the deviations from the mean will always be zero, we simply square each result and then take the sum of the squared values.
- In our example the sum of the squared values is 242

Square each deviation from the mean and then sum the values

Column 1	Column 2			Column 3		
x	$x - \bar{x}$			$(x - \bar{x})^2$		
27	27 - 16	=	11	$(11)^2$	=	121
18	18 - 16	=	2	$(2)^2$	=	4
6	6 - 16	=	-10	$(-10)^2$	=	100
17	17 - 16	=	1	$(1)^2$	=	1
12	12 - 16	=	-4	$(-4)^2$	=	16
$\sum x = 80$				$\sum(x - \bar{x})^2$	=	242

Measures of Variation
Variance and Standard Deviation

- Calculate the sample variance

$$s^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1} = \frac{242}{4} = \mathbf{60.5}$$

Sample Variance (s^2)

- Now obtain the sample standard deviation by taking the square root of the variance. *(remember in the last step we squared the deviations. Now by taking the square root of the result, we are returning the value back to original units)*

$$s = \sqrt{s^2} = \sqrt{60.5} = \mathbf{7.78}$$

Sample Standard Deviation (s)

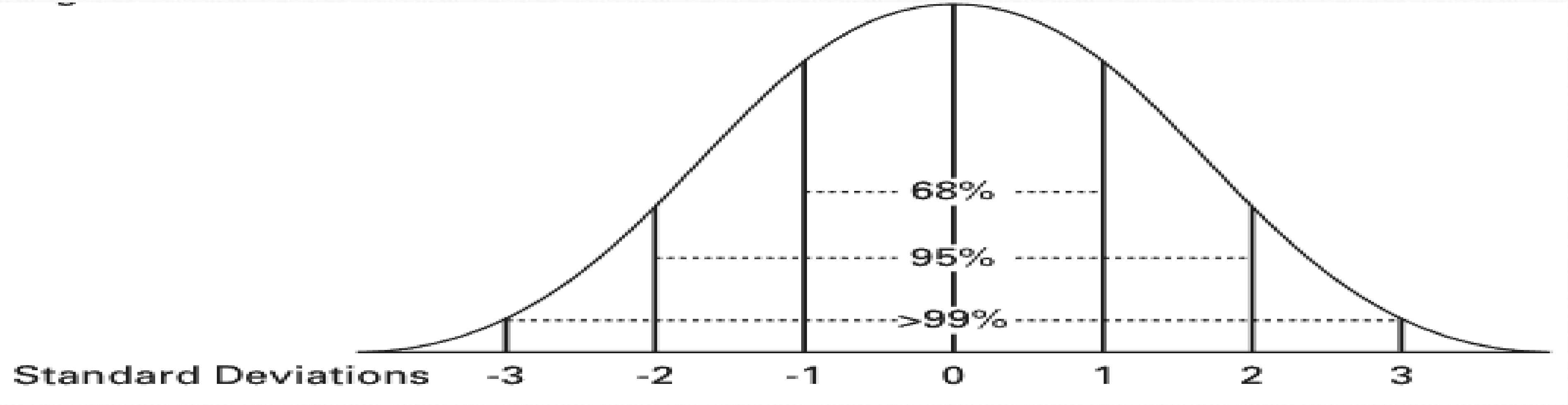
Standard Scores

- A challenge with Mean and standard deviation is that they can not effectively be used to compare factors beyond the sample.

- For example, It would be of little value to compare the mean and standard deviation of an English test with a Math test.

Standard Scores

- **Central Limit Theorem (CLT)** – The arithmetic mean of a sufficiently large number of iterates of independent random variables, each with a well-defined expected value and well-defined variance, will be approximately normally distributed.

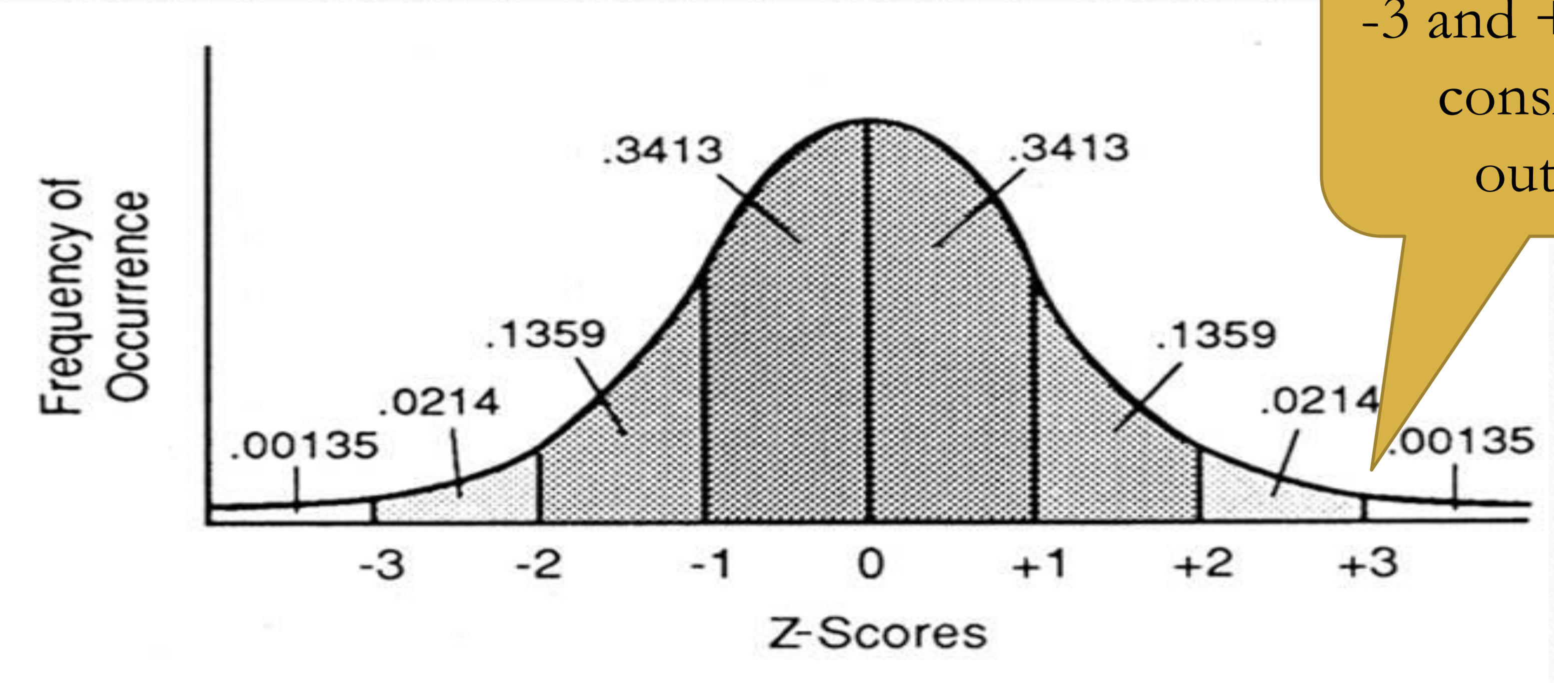


Normal Distribution *(aka, a bell curve)*

Standard Scores

- ***Z-score*** - A raw score converted into standard deviation (SD) units -- hence the term **standard score**.
  - Indicates how much a score deviates from the mean.
  - For example: Mean score = 80, Standard deviation = 5
    - Student one score = 85, z-score = 1 (1 standard deviation above the mean)
    - Student two score = 90, z-score = 2
    - Student three score = 92.5, z-score = 2.5 (2.5 standard deviations above the mean)
    - Student four score = 75, z-score = -1
    - Student five score = 70, z-score = -2 (2 standard deviations below the mean)
  - z scores are beneficial to educators because they allow comparisons to be made between different tests
    - Student performance in math and English. 70% on math test, 70% on English test does not tell us much. Converting scores to z scores will enable us to determine how well the student is doing overall in both classes.

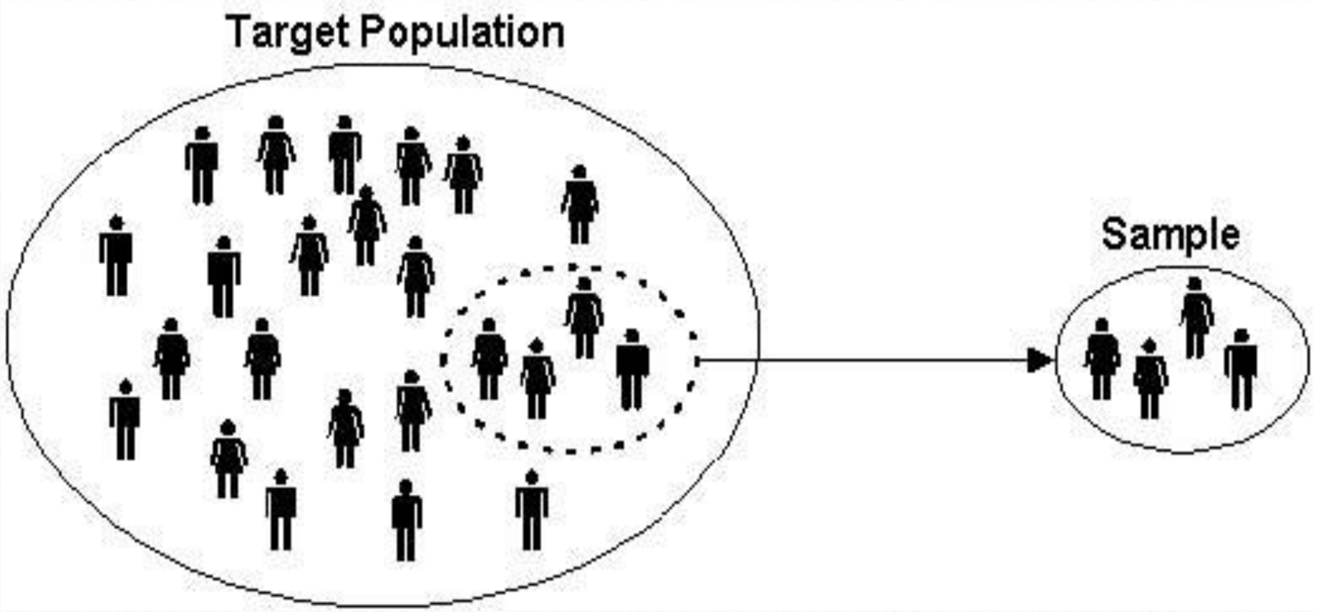
Standard scores as outliers



z-scores outside -3 and +3 can be considered outliers.

Additional Information

- In most applications of statistics, we work with a random sample of data rather than the entire population of *all* possible data values.



Additional Information

- Symbols used in formula's are different when we work with a Population or Sample

	Mean	Standard Deviation	Variance
Population	μ	σ	σ^2
Sample	$\bar{x}$	s	s^2

The End