# Chapter 3

Lecture 2

Section 3.3

# Measure of Variation

- *Standard Deviation:*

  The measure of variation of values about the mean. A type of average deviation of values from the mean. More practically, it is a measure of *statistical* dispersion. It measures how widely spread out the values in a data set are.

- *Sample Standard Deviation:*

$$s=\sqrt{\frac{\sum(x-\bar{x})^2}{n-1}},\ s=\sqrt{\frac{n\cdot\sum(x^2)-(\sum x)^2}{n(n-1)}}\quad \textit{Shortcut Formula}$$

- *Population Standard Deviation:*

$$\sigma=\sqrt{\frac{\sum(x-\mu)^2}{N}}\qquad \sigma = Lowercase\ "sigma"$$

## Properties of Standard Deviation

- Measure of variation of all the values from the mean.
- The standard deviation *"s"* is equal to zero only when all the values in a data set are the same number.
- The standard deviation *"s"* can increase considerably with the inclusion of one or more *outliers.*

(An element of the data that distinctly stands out from the rest of the data.)

- The units, such as feet, lbs., seconds, are the same as that of the data set.
- The larger the standard deviation, the more wide the data is spread out. The smaller the standard deviation, the closer the values in the data set are to the mean.

---

- *Variance:*

  Is measure of variation equal to the square of the standard deviation.

*Population Variance* $\rightarrow \sigma^2$

*Sample Variance* $\rightarrow s^2$

- *Range:*

  Maximum value – Minimum value

- *Round-Off Rule:*

  **Recall that when rounding, round to three decimal places.**

You should notice that to compute the standard deviation, you must first compute the variance.

Lets compute the standard deviation using the data from the previous lecture.

*1. The test scores of a class of seven students is given as:*

*77, 99, 56, 80, 70, 86, 66*

*The population mean was μ=76.286.*

| $x$ | $x-\mu$ | $(x-\mu)^2$ |
|---|---|---|
| 77 | 0.714 | 0.510 |
| 99 | 22.714 | 515.926 |
| 56 | -20.286 | 411.522 |
| 80 | 3.714 | 13.794 |
| 70 | -6.286 | 39.514 |
| 86 | 9.714 | 94.362 |
| 66 | -10.286 | 105.802 |

The sum of the last column is 1181.43, thus

$$\sigma^2 = \frac{(x-\mu)^2}{N} = \frac{1181.43}{7} = 168.776$$

$$\sigma = \sqrt{\frac{(x-\mu)^2}{N}} = \sqrt{\frac{1181.43}{7}} = \sqrt{168.776}$$

$$\sigma = 12.991$$

*2. The following is a sample of heights of men entering the Navy from the East Los Angeles area.*

*70 in., 72in., 76 in., 67in. The mean is* $\bar{x} = 71.25$

*Method 1:*

| x | x − $\bar{x}$ | $(x-\bar{x})^2$ |
|---|---|---|
| 70 | -1.25 | 1.563 |
| 72 | 0.75 | 0.563 |
| 76 | 4.75 | 22.563 |
| 67 | -4.25 | 18.063 |

$$s^2 = \frac{\sum (x-\bar{x})^2}{n-1} = \frac{42.752}{3} = 14.251$$

$$s = \sqrt{\frac{42.752}{3}} = \sqrt{14.251} = 3.775$$

*2. The following is a sample of heights of men entering the Navy from the East Los Angeles area.*

*70 in., 72in., 76 in., 67in. The mean is* $\bar{x} = 71.25$

*Method 2:(Sortcut Formula)*

| x | $x^2$ |
|---|---|
| 70 | 4900 |
| 72 | 5184 |
| 76 | 5776 |
| 67 | 4489 |

$$s^2 = \frac{n \cdot \sum (x^2) - (\sum x)^2}{n(n-1)}$$

$$s^2 = \frac{4(20349) - (81225)}{4(3)} = 14.25$$

$$\sum x = 285, \ \sum (x^2) = 20349$$

$$(\sum x)^2 = (285)^2 = 81225$$

$$s = \sqrt{14.25} = 3.775$$

*Compute the standard deviation for each of the two restaurants.*
*Then compare the data sets using the mean and the standard deviation.*

Students at ELAC usually will eat at the fast food restaurants that are conveniently located across the street in Atlantic Square. When investigating wait times of customers to receive their food, the following results (in seconds) were obtained:

El Pollo Loco: 190 229 74 377 300 481 428 255 328 270 109 266

Subway: 287 128 92 267 176 240 192 118 153 154 193 136

| El Pollo Loco | Subway |
|---|---|
| $\mu$ = 278.333sec → 4.639min | $\mu$ = 178sec → 2.967min |
| $\sigma$ = 119.791sec → 1.997min | $\sigma$ = 60.702sec → 1.012min |

## *Standard deviation of a frequency table:*

| Weight | ***f*** | ***x*** | ***f·x*** | ***x²*** | ***f·x²*** |
|---|---|---|---|---|---|
| 155 – 164 | 5 | 159.5 | 797.5 | 25440.25 | 127201.25 |
| 165 – 174 | 3 | 169.5 | 508.5 | 28730.25 | 86190.75 |
| 175 – 184 | 7 | 179.5 | 1256.5 | 32220.25 | 225541.75 |
| 185 – 194 | 7 | 189.5 | 1326.5 | 35910.25 | 251371.75 |
| 195 – 204 | 5 | 199.5 | 997.5 | 39800.25 | 199001.25 |

$\sum f = 27$ $\quad \sum(f \bullet x) = 4886.5$ $\quad \sum(f \bullet x^2) = 889306.75$

$$s = \sqrt{\frac{n\left[\sum(f \cdot x^2)\right] - \left[\sum(f \cdot x)\right]^2}{n(n-1)}} = \sqrt{\frac{27[8890306.75] - [4886.5]^2}{27(27-1)}} = 13.8 lbs$$

$$where \;\; n = \sum(f)$$

## *For interpreting a known value of the standard deviation:*

*If "s" is known, we will use it to estimate the minimum and maximum "usual" sample values.*

*Minimum "usual" value = mean – 2 × standard deviation*

*Maximum "usual" value = mean + 2 × standard deviation*

**Almost 100% of the data will fall within**

**3 standard deviation of the mean.**

# Computation of Standard Deviation

- A personal trainer has put six of his trainees on a new 1 month program. Their weight loss is as follows:

  11lbs, 9lbs, 15lbs, 10lbs, 5lbs, 11lbs.

  Compute the mean and the standard deviation.

$$\overline{x} = \frac{\sum x}{n} \qquad s = \sqrt{\frac{n \cdot \sum (x^2) - \sum (x)^2}{n(n-1)}}$$

***<u>Minitab Output:</u>***

**Descriptive Statistics: Weight Lost**

| Variable | N | Mean | Median | StDev |
|---|---|---|---|---|
| Weight L | 6 | 10.17 | 10.50 | 3.25 |

| Variable | Minimum | Maximum |
|---|---|---|
| Weight L | 5.00 | 15.00 |

*<u>Recall:</u>*

*Minimum "usual" value = mean – 2×standard deviation*

*Maximum "usual" value = mean + 2×standard deviation*

*Otherwise known as* $\overline{x} - 2s,\ \overline{x} + 2s$. *What this means is that, usually, we will have values between 2 standard deviations form the mean. It would be "unusual" to get values that are greater than 2 standard deviations from the mean or less than –2 standard deviations form the mean.*

*From our previous example,* $\overline{x} = 10.17$*lbs, s=3.25lbs.*

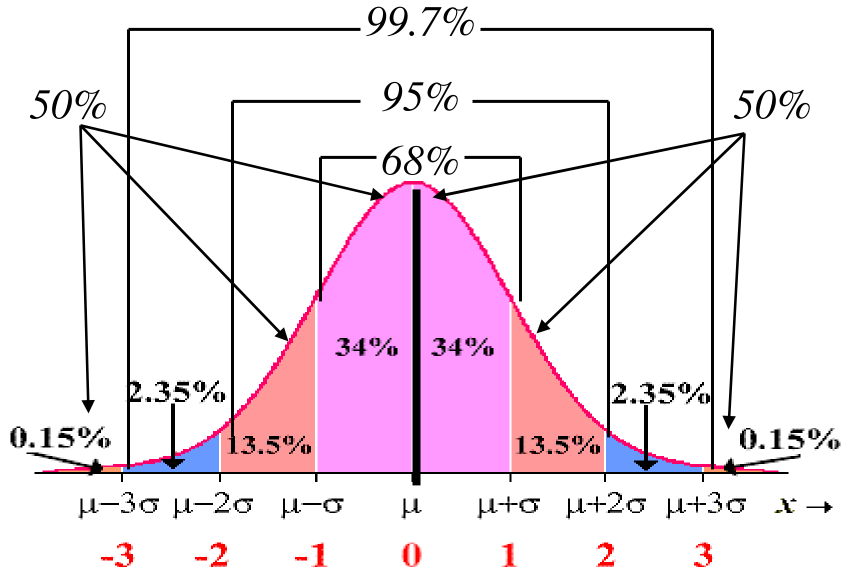
- *Our minimum "usual" value is 3.67lbs. and our maximum "usual" value is 16.67lbs. What does this tell us about the trainers new program?*
- *It is "unusual" for some one to lose less than 3.67lbs. or more than 16.67lbs. on this new program.*

# Empirical Rule

For data sets having a *normal, bell-shaped distribution*, the following properties apply:

- About 68% of the values are one standard deviation away from the mean. They fall in the interval $\bar{x} - s, \bar{x} + s$
- About 95% of the values are two standard deviation away from the mean. They fall in the interval $\bar{x} - 2s, \bar{x} + 2s$
- About 99.7% of the values are three standard deviation away from the mean. They fall in the interval $\bar{x} - 3s, \bar{x} + 3s$

# Normal Distribution



*It is known that the GPA of ELAC students is normally distributed with a mean of 3.0 with a standard deviation of 0.3. Use the empirical rule to come up with a conclusion about this information?*

Since the GPA of ELAC students is normally distributed, 68% of all ELAC students have a GPA between 3.3 and 2.7(*1s*). 95% of all ELAC students have a GPA between 3.6 and 2.4(*2s*). 99.7% of all ELAC students have a GPA between 3.9 and 2.1(*3s*).

*What can we say about the students with a GPA of 3.3 or less?*

Since GPA of ELAC Students are normally distributed and 3.3 is 1 positive standard deviations from the mean, this tells us that 84% of ELAC students have a GPA of 3.3 or less.

*Example:*

*The cost of homes in the greater Los Angeles area is found to be normally distributed with a mean of $525,000 with a standard deviation of $50,000? What can be said about the costs of homes in the greater Los Angeles area?*

*What can be said about homes that cost $625,000 or more in the greater Los Angeles area?*