4.4 Notes: Measures of Variation

Objectives:

4. Can you determine the range of a data set?
5. Can you determine the variance and standard deviation of a population and of a sample?
6. Can you use the Empirical Rule to interpret standard deviation?

Range: How Spread out **is the data?**

- The difference between the maximum and minimum data entries in the set.
- The data must be quantitative
- Range = Max. data entry – Min. data entry

Example 1: A corporation hired 10 graduates. The starting salaries for each graduate are shown. Find the range of the starting salaries.

Starting salaries (1000s of dollars)

41 38 39 45 47 41 44 41 37 42

max – min

47 – 32 = 15,000

Deviation: How far is an entry from the mean **?**

population mean μ

- The difference between the data entry, x, and the mean of the data set.
- Population data set:
  - Deviation of $x = x - \mu$

Sample mean $\bar{x}$

- Sample data set:
  - Deviation of $x = x - \bar{x}$

(Keep in Mind, we will use the Calculator to find Standard deviation, this just lets you know how)

Example 2: Use the starting salaries from example 1 to find the deviations for each entry.

Step 1: Find the sample mean $\bar{x}$.

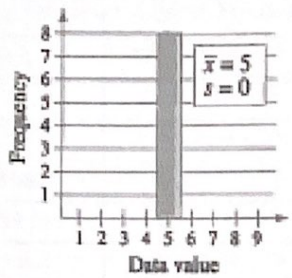
Step 2: Find the deviations: How far is each entry from the mean?

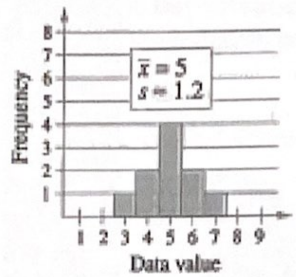
Salary ($1000s), x	Deviation: $x - \mu$	=
41	41 – 41.5 =	–0.5
38	38 – 41.5 =	–3.5
39	39 – 41.5 =	–2.5
45	45 – 41.5 =	3.5
47	47 – 41.5 =	5.5
41	41 – 41.5 =	-0.5
44	44 – 41.5 =	2.5
41	41 – 41.5 =	-0.5
37	37 – 41.5 =	-4.5
42	42 – 41.5 =	0.5

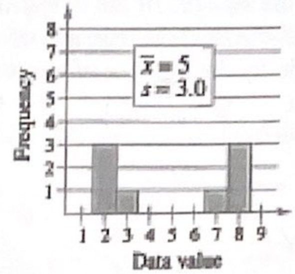
$\sum x =$ 415 $\qquad \sum(x - \bar{x}) =$ 0

Standard Deviation

- A measure of the typical amount an entry deviates from the ___mean___.
- The "average" of the deviations (kind of.)
- The more the entries are ___spread___ ___out___, the greater the standard deviation.
- Sample standard deviation uses the symbol ___s___.
- Population standard deviation uses the symbol ___σ___.







Variance

- Another measure of spread.
- It is the value of the standard deviation ___squared___.
- Not as useful to our purposes as standard deviation, because it is not in the same units as our data list (units ___squared___ instead.)
- Used at a higher level in more advanced statistics courses.
- Sample variance uses the symbol ___s^2___.
- Population variance uses the symbol ___σ^2___.

Formulas for Standard Deviation and Variance

	Standard Deviation	Variance
Population	$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\Sigma(x-\mu)^2}{N}}$ Step 1: Find the mean. Step 2: Find each deviation from the mean. Step 3: Square each deviation. Why? Step 4: Find the sum of those squares. Step 5: Divide by the number of entries. Step 6: Square root this value (the variance.)	$\sigma^2 = \frac{\Sigma(x-\mu)^2}{N}$ Step 1: Find the mean. Step 2: Find each deviation from the mean. Step 3: Square each deviation. Why? Step 4: Find the sum of those squares. Step 5: Divide by the number of entries.
Sample	$s = \sqrt{s^2} = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$ Step 1: Find the sample mean. Step 2: Find each deviation from the mean. Step 3: Square each deviation. Why? Step 4: Find the sum of those squares. Step 5: Divide by the sample size minus 1. Step 6: Square root this value (the variance.)	$s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$ Step 1: Find the sample mean. Step 2: Find each deviation from the mean. Step 3: Square each deviation. Why? Step 4: Find the sum of those squares. Step 5: Divide by the sample size minus 1.

(Keep in Mind, we will use the Calculator to find Standard deviation, this just lets you know how)

Example 3: A corporation hired 10 graduates. The starting salaries for each graduate are shown. Find the population variance and standard deviation of the starting salaries.

Starting salaries (1000s of dollars)

41 38 39 45 47 41 44 41 37 42

Step 1: Recall $\mu = 41.5$

Step 2: We already found the deviations are as shown.

Step 3: Find the squares of the deviations.

Step 4: Find the sum of the squares of the deviations.

Step 5: Divide by the number of entries.

Salary, x	Deviation: $x - \mu$	Squares: $(x - \mu)^2$
41	$41 - 41.5 = -0.5$	$(-0.5)^2 = .25$
38	$38 - 41.5 = -3.5$	$(-3.5)^2 = 12.25$
39	$39 - 41.5 = -2.5$	$(-2.5)^2 = 6.25$
45	$45 - 41.5 = 3.5$	$(3.5)^2 = 12.25$
47	$47 - 41.5 = 5.5$	$(5.5)^2 = 30.25$
41	$41 - 41.5 = -0.5$	$(-0.5)^2 = 0.25$
44	$44 - 41.5 = 2.5$	$(2.5)^2 = 6.25$
41	$41 - 41.5 = -0.5$	$(-0.5)^2 = 0.25$
37	$37 - 41.5 = -4.5$	$(-4.5)^2 = 20.25$
42	$42 - 41.5 = 0.5$	$(0.5)^2 = 0.25$
	$\Sigma(x - \mu) = 0$	$\Sigma(x - \mu)^2 = 88.5$

Step 5. $\sigma^2 = \dfrac{\Sigma(x-\mu)^2}{N} = \dfrac{88.5}{10} = 8.85$

$\sigma^2 \approx 8.9$

Standard deviation is the square root of the variance

$\sqrt{\sigma^2} = \sqrt{8.85} \quad \sigma = 2.97 \approx 3.0$

(Keep in Mind, we will use the Calculator to find Standard deviation, this just lets you know how)

Example 4: The starting salaries are for the Chicago branches of a corporation. The corporation has several other branches, and you plan to use the starting salaries of the Chicago branches to estimate the starting salaries for the larger population. Find the ***sample*** standard deviation of the starting salaries.

Starting salaries (1000s of dollars)

41 38 39 45 47 41 44 41 37 42

How is this problem different than Example 3? How can we adjust our results to find the sample standard deviation rather than the population standard deviation?

it's a sample, not a population, so use $n-1$ not N

$s^2 = \dfrac{88.5}{10-1} = \dfrac{88.5}{9} = 9.833$ variance

$\sqrt{s^2} = \sqrt{9.833} = 3.1358 \approx 3.14$ standard deviation

Using the TI-83 to find standard deviation and variance:

Step 1: Input your data in a list (STAT, EDIT…)

Step 2: Perform calculations (STAT, CALCULATION, 1-VAR STAT, ENTER)

Example 5: Sample office rental rates (in dollars per square foot per year) for Miami's central business district are shown in the table. Use a calculator to find the mean rental rate, the standard deviation, and the variance. *(Adapted from: Cushman & Wakefield Inc.)*

Office Rental Rates		
35.00	33.50	37.00
23.75	26.50	31.25
36.50	40.00	32.00
39.25	37.50	34.75
37.75	37.25	36.75
27.00	35.75	26.00
37.00	29.00	40.50
24.50	33.00	38.00

Note: What type of standard deviation and variance do we want?

$\bar{x} = 33.7 \quad \Sigma x = 809.5$

Sample standard deviation $S_x = 5.089$

$n = 24$

variance is $(S_x)^2 = 25.898$

population standard deviation $\sigma_x = 4.98$

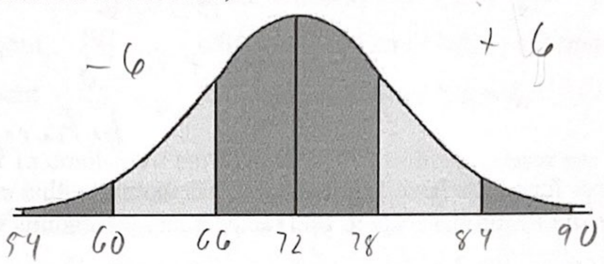
Working With Bell-Shaped Distributions:

- Often described as normal distributions.
- Symmetrical.
- Area under curve = 1.

Use the information above to predict the shaded area in the normal curves given to you in the 2.4 Activity.

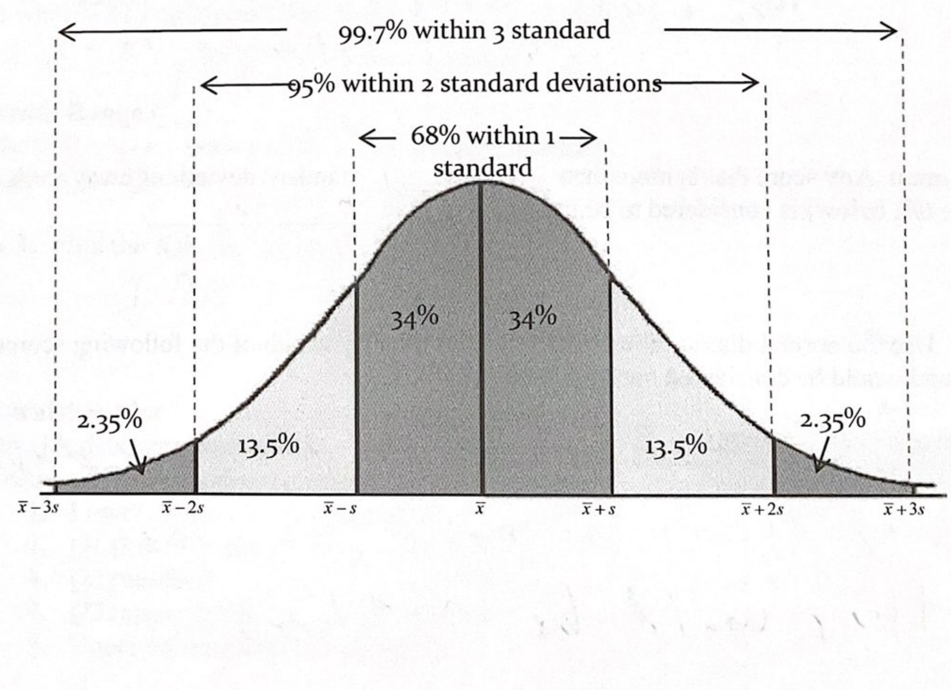
Creating a Diagram Using Mean and Standard Deviation:
For data with a (Symmetrical) bell-shaped distribution, the standard deviation can be used to separate the data into sections.

Example 2: A math test has a normal distribution, and the mean is 72% with a standard deviation of 6. Draw the distribution.



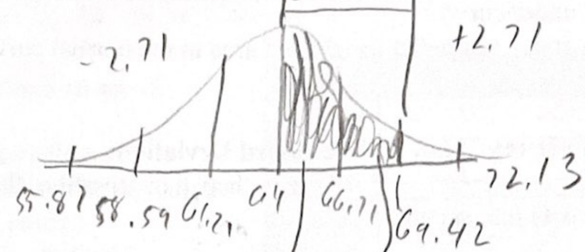
Empirical Rule (68 – 95 – 99.7 Rule)
For data with Symmetrical distributions, the standard deviation has the following characteristics:

- About 68% of the data lie within 1 standard deviation of the mean.
- About 95% of the data lie within 2 standard deviations of the mean.
- About 99.7% of the data lie within 3 standard deviations of the mean.



Example 2: In a survey conducted by the National Center for Health Statistics, the sample mean height of women in the United States (ages 20-29) was 64 inches, with a sample standard deviation of 2.71 inches. Estimate the percent of the women whose heights are between 64 inches and 69.42 inches.

Step 1: Draw a diagram.



Step 2: Use the Empirical Rule
(68 – 95 – 99.7 Rule)

34% + 13.5% = 47.5% of women are between 64 + 69.42 in tall

Example 3: IQ scores for adults have a bell-shaped distribution with a mean of 90 and a standard deviation of 12. Use the Empirical Rule to find the percentage of adults with scores between 78 and 102.

Step 1: Draw a diagram.

Step 2: Use the Empirical Rule.

34% + 34% = 68% of adults have IQ scores between 78 - 102

Unusual Scores: Any score that is more than ___2___ standard deviations away from the mean (above OR below) is considered to be unusual and is an ___outlier___.

Example 4: Use the normal distribution created in #3 to identify which of the following scores, if any, are unusual and would be considered outliers: 100, 60, 70, 117

60 + 117 lie outside 2 standard deviations away from the mean, therefore they would be outliers