Math 146 **3.2 - Measures of Variation**

Lets explore the similarities and difference of the following two data sets:

Data Set 1: 26 22 31 21 24 28 30

Data Set 2: 14 3 45 25 49 27 9 36

Example 1

a) Calculate the mean, median and mid-range

b) Calculate the range of each data set: range = max value – min value

c) Calculate the sum of the differences between each data value and the mean for Data Set 1

SAMPLE STANDARD DEVIATION

$$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} \quad \text{or} \quad s = \sqrt{\frac{n(\sum x^2)-(\sum x)^2}{n(n-1)}}$$

An **estimate** for the standard deviation is the *Range Rule of Thumb*

$$s \approx \frac{\text{range}}{4}$$

Unusual values are those values that lie 2 or more standard deviations from the mean.

Example 2 Estimate if any values in Data Set 2 might possibly be unusual.

The **variance** of a set of values is the square of the standard deviations:

sample variance = s^2

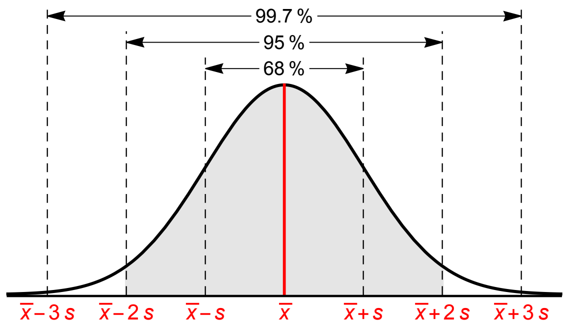
population variance = σ^2

(**Note**: the sample variance s^2 is an *unbiased estimator*, meaning it can be used to estimate the population variance, σ^2. More on this in chpater 6)

The Empirical Rule (or 68 – 95 – 99.7 rule) for *Normally* Distributed Data

One important property of the standard deviation is the *empirical rule* which states that for data that is approximately normal:

- about 68% of all values are within 1 standard deviation of the mean ($\overline{x} \pm 1\,s$)
- about 95% of all values are within 2 standard deviations of the mean ($\overline{x} \pm 2\,s$)
- about 99.7% of all values are within 3 standard deviations of the mean ($\overline{x} \pm 3\,s$)



(Note: the 95% two-standard deviation region is shaded.)

Example 3

Use the empirical rule on Data Set 2 for one and two standard deviations.

Example 4

Find the values that would be classified unusual for the following exam scores:

Exam Score	*f*
10–19	3
20–29	6
30–39	7
40 – 49	9
50–59	2

EXTRA:

- **Chebyshev's Theorem**
  - The proportion of any set of data lying within *K* standard deviations of the mean is always *at least* $1 - 1/K^2$, where $K > 1$.
- **Comparing Variation, e.g., Comparing Standard Deviations**
  - Only compare standard deviations that have the same units, and whose distribution means are similar, or
  - Use the **coefficient of variation** (**CV**) to compare when means are different: $\text{CV} = \frac{s}{n} \cdot 100\,\%$ (or $\frac{\sigma}{N} \cdot 100\,\%$)
- **Rounding** Round all standard deviations and variances to one decimal place further than the data values.