Topic-1
Describing Data
with Numerical Measures
Central Tendency (*Center*) and Dispersion (*Variability*)

- **Central tendency**: measures of the degree to which scores are clustered around the mean of a distribution

- **Dispersion**: measures the fluctuations (variability) around the characteristics of central tendency
Measures of Center

- A measure along the horizontal axis of the data distribution that locates the center of the distribution.
Arithmetic Mean or Average

- The **mean** of a set of measurements is the sum of the measurements divided by the total number of measurements.

\[
\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}
\]

where \( n \) = number of measurements

\[ \sum x_i \] = sum of all the measurements
Example

- The set: 2, 9, 1, 5, 6

\[
\bar{x} = \frac{\sum x_i}{n} = \frac{2 + 9 + 11 + 5 + 6}{5} = \frac{33}{5} = 6.6
\]

If we were able to enumerate the whole population, the population mean would be called \( \mu \) (the Greek letter “mu”).
The **median** of a set of measurements is the middle measurement when the measurements are ranked from smallest to largest.

The position of the median is \(0.5(n + 1)\) once the measurements have been ordered.
Example

- The set: 2, 4, 9, 8, 6, 5, 3  \( n = 7 \)
- Sort: 2, 3, 4, 5, 6, 8, 9
- Position: \( .5(n + 1) = .5(7 + 1) = 4^{th} \)
  
  Median = 4\(^{th}\) largest measurement

- The set: 2, 4, 9, 8, 6, 5  \( n = 6 \)
- Sort: 2, 4, 5, 6, 8, 9
- Position: \( .5(n + 1) = .5(6 + 1) = 3.5^{th} \)
  
  Median = \((5 + 6)/2 = 5.5\) — average of the 3\(^{rd}\) and 4\(^{th}\) measurements
Mode

- The **mode** is the measurement which occurs most frequently.
- The set: 2, 4, 9, 8, 8, 5, 3
  - The mode is **8**, which occurs twice
- The set: 2, 2, 9, 8, 8, 5, 3
  - There are two modes—**8** and **2** (bimodal)
- The set: 2, 4, 9, 8, 5, 3
  - There is **no mode** (each value is unique).
Example
The number of quarts of milk purchased by 25 households:

\[
\begin{array}{cccccccccccccccc}
0 & 0 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 3 & 3 \\
3 & 3 & 3 & 4 & 4 & 4 & 4 & 5 \\
\end{array}
\]

- **Mean?**
  \[
  \bar{x} = \frac{\sum x_i}{n} = \frac{55}{25} = 2.2
  \]

- **Median?**
  \[
  m = 2
  \]

- **Mode? (Highest peak)**
  \[
  \text{mode} = 2
  \]
Extreme Values

- The mean is more easily affected by extremely large or small values than the median.

- The median is often used as a measure of center when the distribution is skewed.
Extreme Values

Symmetric: Mean = Median

Skewed right: Mean > Median

Skewed left: Mean < Median
Measures of Variability

- A measure along the horizontal axis of the data distribution that describes the spread of the distribution from the center.

- **Range**
  - Difference between maximum and minimum values

- **Interquartile Range**
  - Difference between third and first quartile \( (Q_3 - Q_1) \)

- **Variance**
  - Average* of the squared deviations from the mean

- **Standard Deviation**
  - Square root of the variance

\[ \text{Definitions of population variance and sample variance differ slightly.} \]
The Range

• The range, $R$, of a set of $n$ measurements is the difference between the largest and smallest measurements.

• Example: A botanist records the number of petals on 5 flowers:

  5, 12, 6, 8, 14

• The range is $R = 14 - 5 = 9$.

Quick and easy, but only uses 2 of the 5 measurements.
Percentile

50\textsuperscript{th} Percentile  \equiv  \text{Median (Q}_2\text{)}
25\textsuperscript{th} Percentile  \equiv  \text{Lower Quartile (Q}_1\text{)}
75\textsuperscript{th} Percentile  \equiv  \text{Upper Quartile (Q}_3\text{)}

Interquartile Range:
\quad IQR=Q_3 - Q_1
The position of p-th percentile is \[ 0.\overline{p}(n + 1) \]

The position of \( Q_1 \) is \[ 0.25(n + 1) \]

The position of \( Q_3 \) is \[ 0.75(n + 1) \]

Once the measurements have been ordered. If the positions are not integers, find the quartiles by interpolation.
The prices ($) of 18 brands of walking shoes:

40  60  65  65  65  68  68  70  70
70  70  70  70  74  75  75  90  95

\[
\text{Position of } Q_1 = 0.25(18 + 1) = 4.75
\]

\[
\text{Position of } Q_3 = 0.75(18 + 1) = 14.25
\]

\[Q_1 \text{ is 3/4 of the way between the 4}^{th} \text{ and 5}^{th} \text{ ordered measurements, or } Q_1 = 65 + 0.75(65 - 65) = 65.\]
The prices ($) of 18 brands of walking shoes:

40  60  65  65  65  68  68  70  70
70  70  70  70  74  75  75  90  95

Position of $Q_1$ = 0.25(18 + 1) = 4.75
Position of $Q_3$ = 0.75(18 + 1) = 14.25

$Q_3$ is 1/4 of the way between the 14\textsuperscript{th} and 15\textsuperscript{th} ordered measurements, or

$Q_3 = 74 + .25(75 - 74) = 74.25$

and

$IQR = Q_3 - Q_1 = 74.25 - 65 = 9.25$
The 90-th percentile $P_{90}$

- The position of 90-th percentile is

$$0.9(18 + 1) = 17.1$$

The prices ($) of 18 brands of walking shoes:

$40 \ 60 \ 65 \ 65 \ 65 \ 68 \ 68 \ 70 \ 70 \ 70 \ 70 \ 74 \ 75 \ 75 \ 90 \ 95$

$P_{90} = 90 + .10 \times (95-90) = 90.5$
The Variance

- The variance is a measure of variability that uses all the measurements. It measures the average deviation of the measurements about their mean.
- Flower petals: 5, 12, 6, 8, 14

\[ \bar{x} = \frac{45}{5} = 9 \]
The Variance

- The variance of a population of \( N \) measurements is the average of the squared deviations of the measurements about their mean \( \mu \).

\[ \sigma^2 = \frac{\sum(x_i - \mu)^2}{N} \]

- The variance of a sample of \( n \) measurements is the sum of the squared deviations of the measurements about their mean, divided by \( (n - 1) \).

\[ s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} \]
The Standard Deviation

- In calculating the variance, we squared all of the deviations, and in doing so changed the scale of the measurements.
- To return this measure of variability to the original units of measure, we calculate the standard deviation, the positive square root of the variance.

\[
\text{Population standard deviation} : \sigma = \sqrt{\sigma^2} \\
\text{Sample standard deviation} : s = \sqrt{s^2}
\]
Two Ways to Calculate the Sample Variance

Use the Definition Formula:

\[ s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} \]

\[ s^2 = \frac{60}{4} = 15 \]

\[ s = \sqrt{s^2} = \sqrt{15} = 3.87 \]

<table>
<thead>
<tr>
<th>( x_i )</th>
<th>( x_i - \bar{x} )</th>
<th>( (x_i - \bar{x})^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>-4</td>
<td>16</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>-3</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>Sum</td>
<td>45</td>
<td>0</td>
</tr>
</tbody>
</table>
Two Ways to Calculate the Sample Variance

Use the calculation formula:

\[
s^2 = \frac{\sum x_i^2 - (\sum x_i)^2}{n}
\]

\[
s^2 = \frac{\sum x_i^2 - \left(\frac{\sum x_i}{n}\right)^2}{n - 1}
\]

<table>
<thead>
<tr>
<th>(x_i)</th>
<th>(x_i^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>12</td>
<td>144</td>
</tr>
<tr>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td>8</td>
<td>64</td>
</tr>
<tr>
<td>14</td>
<td>196</td>
</tr>
</tbody>
</table>

Sum: 45 465

\[
465 - \frac{45^2}{5} = \frac{5}{4} = 15
\]

\[
s = \sqrt{s^2} = \sqrt{15} = 3.87
\]
Using Measures of Center and Spread: The Empirical Rule

Given a distribution of measurements that is approximately mound-shaped:

- The interval $\mu \pm \sigma$ contains approximately 68% of the measurements.
- The interval $\mu \pm 2\sigma$ contains approximately 95% of the measurements.
- The interval $\mu \pm 3\sigma$ contains approximately 99.7% of the measurements.
The Empirical Rule: An Example

**EMPIRICAL RULE** For a symmetrical, bell-shaped frequency distribution, approximately 68 percent of the observations will lie within plus and minus one standard deviation of the mean; about 95 percent of the observations will lie within plus and minus two standard deviations of the mean; and practically all (99.7 percent) will lie within plus and minus three standard deviations of the mean.

**CHART 3-7** A Symmetrical, Bell-Shaped Curve Showing the Relationships between the Standard Deviation and the Observations
Approximating \( s \)

- From Chebyshev’s Theorem and the Empirical Rule, we know that
  \[ R \approx 4-6 \, s \]
- To approximate the standard deviation of a set of measurements, we can use:

\[
s \approx \frac{R}{4}
\]

or \( s \approx \frac{R}{6} \) for a large dataset.
**Approximating $s$**

The ages of 50 tenured faculty at a state university.

<table>
<thead>
<tr>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
</tr>
<tr>
<td>48</td>
</tr>
<tr>
<td><strong>70</strong></td>
</tr>
<tr>
<td>63</td>
</tr>
<tr>
<td>52</td>
</tr>
<tr>
<td>52</td>
</tr>
<tr>
<td>35</td>
</tr>
<tr>
<td>50</td>
</tr>
<tr>
<td>37</td>
</tr>
<tr>
<td>43</td>
</tr>
<tr>
<td>53</td>
</tr>
<tr>
<td>43</td>
</tr>
<tr>
<td>52</td>
</tr>
<tr>
<td>44</td>
</tr>
<tr>
<td>42</td>
</tr>
<tr>
<td>31</td>
</tr>
<tr>
<td>36</td>
</tr>
<tr>
<td>48</td>
</tr>
<tr>
<td>43</td>
</tr>
<tr>
<td><strong>26</strong></td>
</tr>
<tr>
<td>58</td>
</tr>
<tr>
<td>62</td>
</tr>
<tr>
<td>49</td>
</tr>
<tr>
<td>34</td>
</tr>
<tr>
<td>48</td>
</tr>
<tr>
<td>53</td>
</tr>
<tr>
<td>39</td>
</tr>
<tr>
<td>45</td>
</tr>
<tr>
<td>34</td>
</tr>
<tr>
<td>59</td>
</tr>
<tr>
<td>34</td>
</tr>
<tr>
<td>66</td>
</tr>
<tr>
<td>40</td>
</tr>
<tr>
<td>59</td>
</tr>
<tr>
<td>36</td>
</tr>
<tr>
<td>41</td>
</tr>
<tr>
<td>35</td>
</tr>
<tr>
<td>36</td>
</tr>
<tr>
<td>62</td>
</tr>
<tr>
<td>34</td>
</tr>
<tr>
<td>38</td>
</tr>
<tr>
<td>28</td>
</tr>
<tr>
<td>43</td>
</tr>
<tr>
<td>50</td>
</tr>
<tr>
<td>30</td>
</tr>
<tr>
<td>43</td>
</tr>
<tr>
<td>32</td>
</tr>
<tr>
<td>44</td>
</tr>
<tr>
<td>58</td>
</tr>
<tr>
<td>53</td>
</tr>
</tbody>
</table>

\[ R = 70 - 26 = 44 \]

\[ s \approx R/4 = 44/4 = 11 \]

Actual $s = 10.73$
Measures of Relative Standing

- Where does one particular measurement stand in relation to the other measurements in the data set?
- How many standard deviations away from the mean does the measurement lie? This is measured by the z-score.

\[
z = \frac{x - \bar{x}}{s}
\]

Suppose \( s = 2 \).

\[\bar{x} = 5 \quad x = 9\]

\( x = 9 \) lies \( z = 2 \) std dev from the mean.
z-Scores

- z-scores between $-2$ and $2$ are not unusual. z-scores should not be more than $3$ in absolute value. z-scores larger than $3$ in absolute value would indicate a possible outlier.
Constructing a Box Plot (Refer Lecture 2-1)

- Calculate $Q_1$, the median ($Q_2$), $Q_3$ and IQR.
- Draw a horizontal line to represent the scale of measurement.
- Draw a box using $Q_1$, the median, $Q_3$. 

![Diagram of a box plot with $Q_1$, m, and $Q_3$ marked on a horizontal line.]
Constructing a Box Plot cont.

- Isolate outliers by calculating
  - Lower fence: $Q_1 - 1.5 \times \text{IQR}$
  - Upper fence: $Q_3 + 1.5 \times \text{IQR}$

- Measurements beyond the upper or lower fence is are outliers and are marked (*).
Example

Amt of sodium in 8 brands of cheese:

260  290  300  320  330  340  340  520

Q₁ = 292.5  m = 325  Q₃ = 340
Example

IQR \quad = \quad 340 - 292.5 = 47.5

Lower fence \quad = \quad 292.5 - 1.5(47.5) = 221.25

Upper fence \quad = \quad 340 + 1.5(47.5) = 411.25

Outlier: \quad x = 520
Grouped and Ungrouped Data
Sample Mean

- **Ungrouped data:**
  \[
  \overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n}
  \]
  
  \(n=\text{number of observation}\)

- **Grouped data:**
  \[
  \overline{x} = \frac{\sum_{i=1}^{k} f_i x_i^*}{n} = \frac{1}{n} \left( \sum_{i=1}^{k} f_i x_i^* \right)
  \]
  
  \(n=\text{sum of the frequencies}\)
  
  \(k=\text{number class}\)
  
  \(f_i = \text{frequency of each class}\)
  
  \(x_i^* = \text{midpoint of each class}\)
Example: - ungrouped data

- Resistance of 5 coils:
  3.35, 3.37, 3.28, 3.34, 3.30 ohm.
- The average:

\[
\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{3.35 + 3.37 + 3.28 + 3.34 + 3.30}{5} = 3.33
\]
**Example: - grouped data**

Frequency Distributions of the life of 320 tires in 1000 km

<table>
<thead>
<tr>
<th>Boundaries</th>
<th>Midpoint $x_i^*$</th>
<th>Frequency, $f_i$</th>
<th>$f_i x_i^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>23.6-26.5</td>
<td>25.0</td>
<td>4</td>
<td>100</td>
</tr>
<tr>
<td>26.6-29.5</td>
<td>28.0</td>
<td>36</td>
<td>1008</td>
</tr>
<tr>
<td>29.6-32.5</td>
<td>31.0</td>
<td>51</td>
<td>1581</td>
</tr>
<tr>
<td>32.6-35.5</td>
<td>34.0</td>
<td>63</td>
<td>2142</td>
</tr>
<tr>
<td>35.6-38.5</td>
<td>37.0</td>
<td>58</td>
<td>2146</td>
</tr>
<tr>
<td>38.6-41.5</td>
<td>40.0</td>
<td>52</td>
<td>2080</td>
</tr>
<tr>
<td>41.6-44.5</td>
<td>43.0</td>
<td>34</td>
<td>1462</td>
</tr>
<tr>
<td>44.6-47.5</td>
<td>46.0</td>
<td>16</td>
<td>736</td>
</tr>
<tr>
<td>47.6-50.5</td>
<td>49.0</td>
<td>6</td>
<td>294</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>n = 320</strong></td>
<td></td>
<td><strong>11549</strong></td>
</tr>
</tbody>
</table>

\[
\bar{x} = \frac{\sum_{i=1}^{k} f_i x_i^*}{n} = \frac{11,549}{320} = 36.1
\]
Sample Variance

- Ungrouped data:

\[
s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - \left( \sum_{i=1}^{n} x_i \right)^2}{n}
\]

- Grouped data:

\[
s^2 = \frac{\sum_{i=1}^{k} (f_i x_i^*)^2 - \left( \sum_{i=1}^{k} f_i x_i^* \right)^2}{n} \div (n-1)
\]
Example- *ungrouped data*

- Sample: Moisture content (%) of kraft paper are:
  6.7, 6.0, 6.4, 6.4, 5.9, and 5.8.

\[
s = \sqrt{\frac{(231.26) - (37.2)^2}{(6-1)}} = 0.35
\]

- Sample standard deviation, \( s = 0.35 \) %
Calculating the Sample Standard Deviation - Grouped Data

- Standard deviation for a grouped sample:

Table: Car speeds in km/h

<table>
<thead>
<tr>
<th>Boundaries</th>
<th>( x_i^* )</th>
<th>( f_i )</th>
<th>( f_i x_i^* )</th>
<th>( f_i x_i^{*2} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>72.6-81.5</td>
<td>77.0</td>
<td>5</td>
<td>385</td>
<td>29645</td>
</tr>
<tr>
<td>81.6-90.5</td>
<td>86.0</td>
<td>19</td>
<td>1634</td>
<td>140524</td>
</tr>
<tr>
<td>90.6-99.5</td>
<td>95</td>
<td>31</td>
<td>2945</td>
<td>279775</td>
</tr>
<tr>
<td>99.6-108.5</td>
<td>104.0</td>
<td>27</td>
<td>2808</td>
<td>292032</td>
</tr>
<tr>
<td>108.6-117.5</td>
<td>113</td>
<td>14</td>
<td>1582</td>
<td>178766</td>
</tr>
<tr>
<td>Total</td>
<td>96</td>
<td>9354</td>
<td>920742</td>
<td></td>
</tr>
</tbody>
</table>

\[
s = \sqrt{\frac{\sum_{i=1}^{k} (f_i x_i^{*2}) - (\sum_{i=1}^{k} f_i x_i^*)^2}{n(n-1)}}
\]

\[
s = \sqrt{\frac{920742 - (9354)^2}{96(96-1)}} = 9.9
\]
For grouped data, class mode (or, modal class) is the class with the highest frequency.

\[
\text{Mode} = L + \frac{d_1}{d_1 + d_2} l
\]

Where:

\( L \) = Lower boundary of the modal class
\( d_1 \) = frequency of modal class – frequency of the pre modal class
\( d_2 \) = frequency of modal class – frequency of the post modal class
\( l \) = length of the modal class
Example of Grouped Data (Mode)

Based on the grouped data below, find the mode

<table>
<thead>
<tr>
<th>Class Limit</th>
<th>Boundaries</th>
<th>f</th>
<th>Cum Freq</th>
<th>Cum Rel Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>99 - 108</td>
<td>98.5 - 108.5</td>
<td>6</td>
<td>6</td>
<td>0.150</td>
</tr>
<tr>
<td>109 - 118</td>
<td>108.5 - 118.5</td>
<td>7</td>
<td>13</td>
<td>0.325</td>
</tr>
<tr>
<td>119 - 128</td>
<td>118.5 - 128.5</td>
<td>13</td>
<td>26</td>
<td>0.650</td>
</tr>
<tr>
<td>129 - 138</td>
<td>128.5 - 138.5</td>
<td>8</td>
<td>34</td>
<td>0.850</td>
</tr>
<tr>
<td>139 - 148</td>
<td>138.5 - 148.5</td>
<td>6</td>
<td>40</td>
<td>0.975</td>
</tr>
</tbody>
</table>

Modal Class = 119 – 128 (third)

\[
\begin{align*}
L &= 118.5 \\
I &= 10 \\
d_1 &= 13 - 7 = 6 \\
d_2 &= 13 - 8 = 5
\end{align*}
\]

Mode = \( L + \frac{d_1}{d_1 + d_2} \cdot I \)

\[
\begin{align*}
&= 118.5 + \frac{6}{10} \\
&= 123.95
\end{align*}
\]
Percentile (Ungrouped Data)

• The $p$-th percentile class from the frequency table:

$$k = \frac{n \times p}{100}$$

where $n =$ total frequency and $p$ is the $p$-th percentile

Find the $p$-th class in which the value of $k$ lies in the corresponding cumulative frequency group.
Percentile (Grouped Data)

\[ S_p = LBC_p + \frac{k - CFB}{f_p} \cdot l \]

\[ k = \frac{n \times p}{100} \]

- \( n \) = number of observations
- \( LBC_p \) = Lower Boundary of the percentile class
- \( CFB \) = Cumulative frequency of the class before \( C_p \)
- \( l \) = Class width
- \( f_p \) = Frequency of percentile class
Example- *Percentile for grouped data*

- Using the previous table of freq, find the median and $80^{th}$ percentile?

<table>
<thead>
<tr>
<th>Class Limit</th>
<th>Boundaries</th>
<th>$f$</th>
<th>Cum Freq</th>
<th>Cum Rel Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>99 - 108</td>
<td>98.5 - 108.5</td>
<td>6</td>
<td>6</td>
<td>0.150</td>
</tr>
<tr>
<td>109 - 118</td>
<td>108.5 - 118.5</td>
<td>7</td>
<td>13</td>
<td>0.325</td>
</tr>
<tr>
<td>119 - 128</td>
<td>118.5 - 128.5</td>
<td>13</td>
<td>26</td>
<td>0.650</td>
</tr>
<tr>
<td>129 - 138</td>
<td>128.5 - 138.5</td>
<td>8</td>
<td>34</td>
<td>0.850</td>
</tr>
<tr>
<td>139 - 148</td>
<td>138.5 - 148.5</td>
<td>6</td>
<td>40</td>
<td>0.975</td>
</tr>
<tr>
<td><strong>Median, ( S_{50} )</strong></td>
<td><strong>80th percentile, ( S_{80} )</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------</td>
<td>-------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( k = \frac{40\times 50}{100} = 20 )</td>
<td>( k = \frac{40\times 80}{100} = 32 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( C_{50} \ (119 - 128) ) \ (3rd class)</td>
<td>( C_{80} \ (129 - 138) ) \ (4th class)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( LBC_p = 118.5 )</td>
<td>( LBC_p = 128.5 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( CFB = 13 )</td>
<td>( CFB = 26 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( l = 10 )</td>
<td>( l = 10 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( f_p = 13 )</td>
<td>( f_p = 8 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( S_{50} = 118.5 + \left( \frac{20 - 13}{13} \right) \times 10 )</td>
<td>( S_{80} = 128.5 + \left( \frac{32 - 26}{8} \right) \times 10 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>= 136</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>= 128.885</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Coefficient of Variation (CV)

- When comparing between data sets with different units or widely different means, one should use the coefficient of variation for comparison instead of the standard deviation.
- The Coefficient of Variation can be written as

\[ CV = \frac{s}{\bar{x}} \]

- We express CV as a percentage by multiplying 100
- Less than 30% consider good