Sample Standard Deviation

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

Sample Variance

Almost the average squared deviation

\[ s^2 = \frac{\sum (y - \bar{y})^2}{n - 1} \]
Sample Variance: Golf Scores

\[ s^2 = \frac{(16 + 9 + 4 + 25 + 9 + 1)}{5} = \frac{64}{5} \]

\[ = 12.8 \text{ strokes}^2 \]

Sample Standard Deviation: Golf Scores

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

\[ s = \sqrt{12.8} = 3.58 \text{ strokes} \]

Sample Standard Deviation: Body Mass of Canidae

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

\[ s = \sqrt{64.36} = 8.02 \text{ kg} \]
Summary Measures

• Position
  – Sample quartiles
    • Five number summary
    • Sample inter-quartile range
    • Box and whiskers plot

Sample Quartiles

• Medians of the lower and upper halves of the data.
• Trying to split the data into fourths, quarters.

Sample Quartiles

Body Mass (kg) of Canidae

<table>
<thead>
<tr>
<th>0</th>
<th>1,3,3,4,4,4,4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0*</td>
<td>5,5,5,5,5,6,6,6,6,6,6,7,8,9,9</td>
</tr>
<tr>
<td>1</td>
<td>0,0,1,2,3</td>
</tr>
<tr>
<td>1*</td>
<td>2,3</td>
</tr>
<tr>
<td>2</td>
<td>2,3</td>
</tr>
<tr>
<td>2*</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>3*</td>
<td>6</td>
</tr>
</tbody>
</table>

Q₁ = (4+5)/2 = 4.5 kg
Q₃ = (10+11)/2 = 10.5 kg
Measure of Spread

- Inter-Quartile Range (IQR)
  - The distance between the quartiles.
  - $IQR = 10.5 - 4.5 = 6$ kilograms
  - The length of the interval that contains the central 50% of the data.

Five Number Summary

- Minimum: 1 kilogram
- $Q_1$: 4.5 kilograms
- Median: 6 kilograms
- $Q_3$: 10.5 kilograms
- Maximum: 36 kilograms

Box Plot

- Establish an axis with a scale.
- Draw a box that extends from $Q_1$ to $Q_3$.
- Draw a line from the $Q_1$ to the minimum and another line from the $Q_3$ to the maximum.
Outlier Box Plot

- Establishes boundaries on what are “usual” values based on the width of the box.
- Values outside the boundaries are flagged as potential outliers.
Standard Score
Look at the number of standard deviations a value is from the mean.

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

Comparing z-scores
- Body mass of *Canidae*  
  \[ \bar{y} = 9.3 \text{ kg} \]  
  \[ s = 8.02 \text{ kg} \]

- Body mass of *Felidae*  
  \[ \bar{y} = 24.2 \text{ kg} \]  
  \[ s = 42.51 \text{ kg} \]

Comparing z-scores
- Body mass of *Canis lupus*  
  \[ y = 36 \text{ kg} \]  
  \[ z = \frac{36.0 - 9.3}{8.02} \]  
  \[ z = 3.33 \]

- Body mass of *Panthera leo*  
  \[ y = 162 \text{ kg} \]  
  \[ z = \frac{162.0 - 24.2}{42.51} \]  
  \[ z = 3.24 \]