Box Plots

A Box Plot is a graphical display of a five-point summary.

- Example: Mammal speeds, 11,12,20,25,30,30,30,32,35,39,40,40,42,45,48,50,70.
  - $\min = 11$
  - $Q_1 = 30$
  - Median = 37
  - $Q_3 = 42$
  - $\max = 70$.
  - Range = 70 – 11 = 59
  - $IQR = 42 – 30 = 12$

Modified Box Plots

A Modified Box Plot also takes into account the outliers. An outlier is a value which is more than 1.5 times the IQR from the nearest quartile.

Example: Mammal speeds, 11,12,20,25,30,30,30,32,35,39,40,40,42,45,48,50,70.

- Example:
  - $(1.5)IQR = 18$
  - $11 < 30 – 18$, so 11 is an outlier.
  - $70 > 42 + 18$, so 70 is an outlier.
Modified Box Plots

Example: Mammal speeds, 11, 12, 20, 25, 30, 30, 32, 35, 39, 40, 40, 40, 42, 45, 48, 50, 70.

Example:
- \((1.5) \text{IQR} = 18\)
- \(11 < 30 - 18\), so 11 is an outlier.
- \(70 > 42 + 18\), so 70 is an outlier.

Box Plots (Modified)

- Box Plots and Modified Box Plots are useful when plotting a single quantitative variable and:
  - want to compare shape, center, and spread of two or more distributions.
  - The distribution has a large number of values
  - Individual values do not need to be identified.
  - (Modified) We want to identify outliers.

Section 2.3 Working with Summary Statistic

DISCUSSION: Which Summary Statistic?

- D22. Explain how to determine the total amount of property taxes for a city if you know the number of properties, the mean dollar value of all properties, and the tax rate. In what sense is knowing the mean equivalent to knowing the total?

- D23. When a measure of center for the income of a community's residents is given, that number is usually the median. Why do you think that is the case?

The effect of changing units

- **Recentering** a data set (adding the same number \(c\) to all the values in the set)
  - Shape or spread do not change.
  - It slides the entire distribution by the amount \(c\), adding \(c\) to the median and the mean.

- **Rescaling** a data set (multiplying all the values in the set by the same positive number \(d\))
  - Basic shape doesn't change.
  - It stretches or shrinks the distribution, multiplying the spread (IQR or standard deviation) by \(d\) and multiplying the center (median or mean) by \(d\).
The Effects of Changing Units

- D24. Suppose a U.S. dollar is worth 14.5 Mexican pesos.
  a. A set of prices, in U.S. dollars, has mean $20 and standard deviation $5. Find the mean and standard deviation of the prices expressed in pesos.
  b. Another set of prices, in Mexican pesos, has a median of 145.0 pesos and quartiles of 72.5 pesos and 29 pesos. Find the median and quartiles of the same prices expressed in U.S. dollars.

- D25. The median of the temperatures in Display 2.71 is 32°F. What is the median of the temperatures in Celsius? The standard deviation of the temperatures is 2.56°F. What is the standard deviation of the temperatures in Celsius?

The Influence of Outliers

A summary statistic is
- **resistant to outliers** if it does not change very much when an outlier is removed.
- **sensitive to outliers** if the summary statistic is greatly affected by the removal of outliers.

Example

<table>
<thead>
<tr>
<th>City</th>
<th>Country</th>
<th>Temperature (°F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addis Ababa</td>
<td>Ethiopia</td>
<td>32</td>
</tr>
<tr>
<td>Algiers</td>
<td>Algeria</td>
<td>32</td>
</tr>
<tr>
<td>Bangkok</td>
<td>Thailand</td>
<td>50</td>
</tr>
<tr>
<td>Madrid</td>
<td>Spain</td>
<td>14</td>
</tr>
<tr>
<td>Nairobi</td>
<td>Kenya</td>
<td>41</td>
</tr>
<tr>
<td>Brazil</td>
<td>Brazil</td>
<td>32</td>
</tr>
<tr>
<td>Warsaw</td>
<td>Poland</td>
<td>-22</td>
</tr>
</tbody>
</table>
Percentiles and CRF plots

- A value is at the $k$th percentile if $k\%$ of all values are less than or equal to it.
- Example: The 10th percentile is the value that separates the lowest 10\% of ordered values in a distribution from the rest.
- Percentiles are usually presented graphically by using a cumulative relative frequency plot, or cumulative percentage plot.

CRP for GMAT scores

<table>
<thead>
<tr>
<th>GMAT score</th>
<th>Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>760</td>
<td>99</td>
</tr>
<tr>
<td>750</td>
<td>98</td>
</tr>
<tr>
<td>700</td>
<td>90</td>
</tr>
<tr>
<td>650</td>
<td>80</td>
</tr>
<tr>
<td>600</td>
<td>66</td>
</tr>
<tr>
<td>550</td>
<td>51</td>
</tr>
<tr>
<td>500</td>
<td>36</td>
</tr>
<tr>
<td>450</td>
<td>23</td>
</tr>
<tr>
<td>400</td>
<td>13</td>
</tr>
<tr>
<td>350</td>
<td>7</td>
</tr>
<tr>
<td>300</td>
<td>3</td>
</tr>
<tr>
<td>250</td>
<td>1</td>
</tr>
</tbody>
</table>

Discussion

- D29. Display 2.78 shows the percentile scores that correspond to the life expectancies for the 43 countries of Europe.
  - Latvia has a mean life expectancy of 72 years. What is Latvia's approximate percentile? Ireland has a percentile of about 60. What is the mean life expectancy in Ireland?
  - How many countries have a mean life expectancy of 73 years?