Empirical Rule

We selected Q3.2.24 (p.142) as an example of using StatCrunch to calculate $\bar{x}$ and $s$, display data distribution shape, and calculate the actual percentage of a given interval for the sorted data.

**Q3.2.24**

**The Empirical Rule**  The following data represent the length of eruption for a random sample of eruptions at the Old Faithful geyser in Calistoga, California.

<table>
<thead>
<tr>
<th>108</th>
<th>108</th>
<th>99</th>
<th>105</th>
<th>103</th>
<th>103</th>
<th>94</th>
</tr>
</thead>
<tbody>
<tr>
<td>102</td>
<td>99</td>
<td>106</td>
<td>90</td>
<td>104</td>
<td>110</td>
<td>110</td>
</tr>
<tr>
<td>103</td>
<td>109</td>
<td>109</td>
<td>111</td>
<td>101</td>
<td>101</td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>102</td>
<td>105</td>
<td>110</td>
<td>106</td>
<td>104</td>
<td></td>
</tr>
<tr>
<td>104</td>
<td>100</td>
<td>103</td>
<td>102</td>
<td>120</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>113</td>
<td>116</td>
<td>95</td>
<td>105</td>
<td>103</td>
<td>101</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>101</td>
<td>107</td>
<td>110</td>
<td>92</td>
<td>108</td>
<td></td>
</tr>
</tbody>
</table>

*Source: Ladonna Hansen, Park Curator*

(a) Determine the sample standard deviation length of eruption. Express your answer rounded to the nearest whole number.

(b) Construct a histogram and comment on the appropriateness of using Empirical Rule to make any general statements about the length of eruptions.

(c) Using Empirical Rule to determine the percentage of eruptions that last between 192 and 116 seconds. Hint: $\bar{x} = 104$.

(d) Determine the actual percentage of eruptions that the last between 92 and 116 seconds inclusive.

(e) Use the Empirical Rule to determine the percentage of eruptions that last less than 98 seconds.

(f) Determine the actual percentage of eruptions that last less than 98 seconds.
(a) Determine the sample standard deviation length of eruption. Express your answer rounded to the nearest whole number.

Step 1: Download the data set.

Step 2: Click **Stat → Summary Stats → Columns**.

Step 3: 1) Click **Length of Eruption** under **Select column(s)**:

2) Choose **Std. dev.** under **Statistics**:

3) Click **Compute!**
The sample standard deviation is computed and shown below.

\[ s \approx 6 \]

(b) Construct a histogram and comment on the appropriateness of using Empirical Rule to make any general statements about the length of eruptions.

Use StatCrunch to construct a histogram starting at 90 with a class width of 5.

Step 1: Click **Graph → Histogram**.
Step 2: 1) Click **Length of Eruption** under **Select column(s)**:
   2) Choose **Frequency** under **Type**:
   3) Under **Bins**: enter 90 for **Start At** --> enter 5 for **Width**.
   4) Put a check \( \sqrt{ } \) for **Value above bar**.
   5) Under **Graph properties**: enter **Length of Eruption** for **X-axis label**, **Frequency** for **Y-axis label**, and **Histogram** for **Title**.
   6) Click **Compute!**
StatCrunch output of the histogram is shown below.

Although this is not a perfect bell-shape curve, it is close enough so the Empirical Rule can be applied to the data set.

(c) Using Empirical Rule to determine the percentage of eruptions that last between 92 and 116 seconds. Hint: $\bar{x} = 104$.
This part does not require the use of StatCrunch. The calculation is done by hand based on the Empirical Rule.

$\bar{x} = 104$ and $s = 6$

Lower bound: $92 = 104 - 2(6)$ which is 2 standard deviations below the mean ($\bar{x} - 2s$).
Upper bound: $116 = 104 + 2(6)$ which is 2 standard deviations above the mean ($\bar{x} + 2s$).

According to the empirical rule, 95% of the data fall within $\bar{x} \pm 2s$.

95% of the eruptions last between 92 and 116 seconds.

(d) Determine the actual percentage of eruptions that last between 92 and 116 seconds inclusive.

Part (c) is the theoretical percentage obtained from Empirical rule. Part (d) asked for the actual percentage of the given raw data that fall with $\bar{x} \pm 2s$. We will first sort the data in a ascending order then count how many data actually fall between 92 and 116 seconds inclusive.
Step 1: From StatCrunch, click Data → Sort.

Step 2: 1) Click Length of Eruption under Select column(s):
   2) Choose Ascending under Sort direction:
   3) Click Compute!

Step 3: Now browse Sort(Length of Eruption) column and count how many pieces of data are between 92 and 116 seconds. There are 41 out of 44 pieces of data. Thus the actual percentage is (41/44)x100 = 93.2%.
(e) Use the Empirical Rule to determine the percentage of eruptions that last less than 98 seconds.

This part does not require the use of StatCrunch. The calculation is done by hand based on the Empirical Rule. \( \bar{x} = 104 \) and \( s = 6 \)

\[ 98 = 104 - 6 \text{ which is one standard deviation below the mean } (\bar{x} - s). \]
According to the Empirical rule, 68% of the data fall within \( \bar{x} \pm s \). Thus 32% (100% - 68%) fall outside \( \bar{x} \pm s \). Due to the symmetrical shape of a bell-shape curve, 16% of the data fall below \( \bar{x} - s \) and 16% of the data fall above \( \bar{x} + s \).

According to the Empirical Rule, 16% of eruptions last less than 98 seconds.

(f) Determine the actual percentage of eruptions that last less than 98 seconds.

We just have to look at the sorted data and count the number of eruptions that last less than 98 seconds. There are 5 out of 44 eruptions had the length of eruption that last less than 98 seconds. Thus the actual percentage is \( (5/44) \times 100 \approx 11.4\% \).