Bootstrap Confidence Intervals

Start with a sample, x, in C1: 26.8, 31.0, 36.1, 29.4, 30.5, 26.6, 33.5, 29.4, 27.2, 30.6

Create 4000 bootstrap samples from this distribution, x*1, x*2, …, x*4000

- Begin by putting all 4000 samples (40000 observations) into C2
  
  Calculate → Random data → Sample from columns → Complete the dialogue → OK

- Add observation numbers 1,2,…..10, 1,2, ….10, …. in C3
  
  Calculate → Make pattered data → simple set of numbers → Complete the dialogue → OK

- Split C2 into 4000 bootstrap samples, x*1, x*2, …, x*4000. Each sample will be in a row, the column headings will be the observation numbers within each sample. the samples will be in rows in C4 to C13.
  
  Data → Unstack columns → Complete the dialogue → OK

- Calculate the 4000 bootstrap sample means. Put the sample means in C14.
  
  Calculate → Row statistics → Complete the dialogue → OK

- Calculate the 4000 bootstrap sample medians. Put the sample medians in C15.
  
  Calculate → Row statistics → Complete the dialogue → OK

- Calculate the 4000 bootstrap sample standard deviations. Put the sample stand deviations in C16. Calculate → Row statistics → Complete the dialogue → OK

Summary Statistics for the 3 bootstrap distributions

<table>
<thead>
<tr>
<th>Descriptive Statistics: means, medians, st.devs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>means</td>
</tr>
<tr>
<td>medians</td>
</tr>
<tr>
<td>st.devs</td>
</tr>
</tbody>
</table>
Finding the $p^{th}$ Sample Percentile.

- Calculate $\frac{p(n + 1)}{100}$

- Put the sample in order

- If $\frac{p(n + 1)}{100}$ is an integer, then the $p^{th}$ sample percentile is observation $\frac{p(n + 1)}{100}$

- If $\frac{p(n + 1)}{100}$ is not an integer, then round $\frac{p(n + 1)}{100}$ up and down to the nearest integers, and the $p^{th}$ sample percentile is the average of the observation $\frac{p(n + 1)}{100}$ rounded-up and observation $\frac{p(n + 1)}{100}$ rounded-down
Using Minitab to Order the $\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_B^*$ Values

Calc $\rightarrow$ Calculator $\rightarrow$ Complete the dialogue box $\rightarrow$ OK

The ordered sample means are in C17, the ordered sample medians are in C18 and the ordered sample standard deviations are in C19

95% Confidence Intervals

- For the 2.5$^{th}$ sample percentile observations #100 and #101
  \[ \frac{p(n + 1)}{100} = \frac{2.5(4001)}{100} = 100.025 \rightarrow \text{average} \]

- For the 97.5$^{th}$ sample percentile observations #3900 and #3901
  \[ \frac{p(n + 1)}{100} = \frac{97.5(4001)}{100} = 3900.975 \rightarrow \text{average} \]
95% Confidence Interval for the Mean

Use Minitab to Find Observations #100, #101, #3900, and #3901

Editor → check “enable commands” → the Minitab prompt, MTB >, will appear in the session window

```
MTB > let k1=c17(100)
MTB > let k2=c17(101)
MTB > let k3=c17(3900)
MTB > let k4=c17(3901)
MTB > prin k1 k2 k3 k4
```

Data Display

<table>
<thead>
<tr>
<th>K1</th>
<th>28.4400</th>
</tr>
</thead>
<tbody>
<tr>
<td>K2</td>
<td>28.4500</td>
</tr>
<tr>
<td>K3</td>
<td>31.9500</td>
</tr>
<tr>
<td>K4</td>
<td>31.9800</td>
</tr>
</tbody>
</table>

The 2.5th sample percentile is 28.445 and the 97.5th sample percentile is 31.965

A bootstrap 95% confidence interval for the mean is \( \left( \hat{\theta}^*_{.025}, \hat{\theta}^*_{.975} \right) = (28.445, 31.965) \)

Another 95% confidence interval for the mean is \( \left( 2\hat{\theta}^* - \hat{\theta}^*_{.975}, 2\hat{\theta}^* - \hat{\theta}^*_{.025} \right) = (28.253, 31.773) \)

95% Confidence Interval for the Median

```
MTB > let k1=c18(100)
MTB > let k2=c18(101)
MTB > let k3=c18(3900)
MTB > let k4=c18(3901)
MTB > print k1 k2 k3 k4
```

Data Display

<table>
<thead>
<tr>
<th>K1</th>
<th>27.2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>K2</td>
<td>27.2000</td>
</tr>
<tr>
<td>K3</td>
<td>32.0500</td>
</tr>
<tr>
<td>K4</td>
<td>32.0500</td>
</tr>
</tbody>
</table>

A bootstrap 95% confidence interval for the median is \( \left( \hat{\theta}^*_{.025}, \hat{\theta}^*_{.975} \right) = (27.200, 32.050) \)
95% Confidence Interval for the Standard Deviation

MTB > let k1=c19(100)
MTB > let k2=c19(101)
MTB > let k3=c19(3900)
MTB > let k4=c19(3901)
MTB > print k1 k2 k3 k4

Data Display

\begin{tabular}{|l|l|}
\hline
K1 & 1.56048 \\ 
K2 & 1.56493 \\ 
K3 & 3.93645 \\ 
K4 & 3.93667 \\
\hline
\end{tabular}

A bootstrap 95% confidence interval for the median is \( \left( \hat{\theta}_{0.025}, \hat{\theta}_{0.975} \right) = (1.562705, 3.93656) \)

Remember to uncheck “enable commands” in the editor menu