Suppose that you are an engineer at a plastic making company. You have recently been given two sets of data for which you need to find an interval for the center of the distribution. The data is the amount of aluminum contamination (in ppm) in plastic. Assume that the data was collected using simple random sampling techniques. Conduct a confidence interval for the mean.

Data Set: 30, 30, 60, 63, 70, 79, 87, 90, 101

Are the assumptions met for the t confidence interval?

T Confidence Intervals

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
<th>95.0 % CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>10</td>
<td>71.20</td>
<td>26.01</td>
<td>8.23</td>
<td>(52.59, 89.81)</td>
</tr>
</tbody>
</table>
What do you do when you don’t think that the data is Normal?
One option is the bootstrap method. This method uses the idea of repeated sampling to construct a sampling distribution of the sample mean.

1. You draw items from the original sample with replacement until you have a new sample. You compute its mean

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Sample 1</th>
<th>Sample 2</th>
<th>Sample 3</th>
<th>Sample 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>63</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>79</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>87</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>101</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>102</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. You then redraw another sample and compute its mean. You do these steps thousands of times until you have a list of sample means. You put these sample means in order.

3. You then try to find the middle 95% of the sample means. To do this, find the means at the 2.5% and 97.5% quartile.

Bootstrap Confidence Interval for the Mean Computing Instructions

To create a confidence interval for the mean using the bootstrap, use the instructions below.

If you are using Minitab version 12, the worksheet size is limited to 5000 cells. So, you need to use N=415.

If you are using Minitab version 13, you are not restricted in your worksheet size. Therefore, go ahead and use N=500.

1. Enter your data into column, C1.
2. Enter 0.10 a total of 10 times in the second data column.
4. In the “Values in” box, enter “C1”, enter “C2” in the “Probabilities in” box, and enter N in the number of rows box. Enter C3-C12 in the “Store in Columns” Box.
   • Each row from (C3 to C12) is a sample.
5. Label column C13, “Mean”. Find the mean of the samples that you just created by going to Row Statistics. Select mean and enter C3-C12 for input variables and store your results in C13.
   • This new column is now the sampling distribution of the sample mean.
6. Sort the means in increasing order. Go to Manip > Sort. A dialog box will appear. Enter C13 into all three boxes labeled, “Sort Column(s)”, “Store sort column(s) in:” and “Sort by Column”, and then click o.k.
7. Find the 2.5\textsuperscript{th} and 97.5\textsuperscript{th} percentile.
   a.) The 2.5\textsuperscript{th} percentile will be at the position \((0.025)(N +1) = \) ________.
   b.) The 97.5\textsuperscript{th} percentile will be at the position \((0.975)(N+1) = \) ________.
8. The values of these positions are the lower limit and upper limit of the 95% bootstrap interval for the mean.

   Bootstrap Interval : ______________________________

Every time that you run this procedure you will get a slightly different bootstrap interval.

Would it make sense to make a confidence interval about the median rather then the mean?

For homework, conduct a bootstrap interval around the median.

\textbf{Bootstrap Confidence Interval for the median Computing Instructions}

9. Label column C14, “Median”. Find the median of the samples that you just created by going to Row Statistics. Select median and enter C3-C12 for input variables and store your results in C14.
   • This new column is now the sampling distribution of the sample median.
9. Sort the means in increasing order. Go to Manip > Sort. A dialog box will appear. Enter C14 into all three boxes labeled, “Sort Column(s)”, “Store sort column(s) in:” and “Sort by Column”, and then click o.k.
10. Find the 2.5\textsuperscript{th} and 97.5\textsuperscript{th} percentile.
   a.) The 2.5\textsuperscript{th} percentile will be at the position \((0.025)(N +1) = \) ________.
   b.) The 97.5\textsuperscript{th} percentile will be at the position \((0.975)(N+1) = \) ________.
   Round to the nearest integer. (Ex. 10.5 rounds up to 11 and 96.2 rounds down to 92)
11. The values of these positions are the lower limit and upper limit of the bootstrap interval for the median.