Air France: Task 1: Find hours of Osaka museum

<table>
<thead>
<tr>
<th>Success?</th>
<th>Coded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td>Yes</td>
<td>1</td>
</tr>
</tbody>
</table>

\[
\hat{p} \pm z \left( \frac{1 - \hat{p}}{n} \right) \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \quad \text{Wald Confidence Interval}
\]

\[
\hat{p} \quad \text{Sample proportion}
\]

\[
\frac{z}{\sqrt{n}} \quad \text{Sample size}
\]

\[
\frac{z}{\sqrt{2}} \quad \text{Critical value of the normal distribution for the confidence level}
\]

MeasuringU percentile to Z score calculator: https://measuringu.com/zcalc.html

<table>
<thead>
<tr>
<th>Prop of Area</th>
<th>Z-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>95%</td>
<td>1.64</td>
</tr>
<tr>
<td>99%</td>
<td>2.58</td>
</tr>
</tbody>
</table>

What question can we answer with this data (i.e., population parameter we want to measure)?
- How long will it take for a user to perform the task?
- What kinds of problems in the UI are users likely to see?
- How successful will users be in performing the task?
- Which design is better, A or B?

Calculations:

\[ n: 15 \]

Sample proportion (successes/sample size): \[
\frac{10}{15} = 0.666
\]

\[ z: 1.96 \quad \text{(for 95%)} \]

\[
\hat{p} \left(\frac{1 - \hat{p}}{n}\right) = 0.014814
\]

\[
\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = 0.121715
\]

\[
\frac{z}{\sqrt{2}} = 0.23856
\]

Lower bound of the confidence interval:

\[
3.428
\]

Upper bound of the confidence interval:

\[
3.905
\]

The answer to the question is:
We are \( 95 \) \% confident that the actual population parameter mean value is between \( 0.428 \) and \( 0.905 \) (include units for these last 2 numbers.)

• FYI: Wald (90\%) \Rightarrow [0.4670, 0.8662]
Wald (99\%) \Rightarrow [0.3538, 0.9794]
Air France: Task 1: Find hours of Osaka museum

Adjusted Wald Confidence Interval

\[
\hat{p}_{adj} = \frac{z}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_{adj}}}}
\]

Adjusted Wald provides tighter coverage when the number of samples are less (≤150).

MeasuringU percentile to Z score calculator: https://measuringu.com/zcalcsp/

Calculations:

\[x: \quad 10\]
\[n_{adj}: \quad 19 \quad (18.84)\]

Adjusted sample proportion (successes/sample size): \[0.632\]

\[z: \quad 1.96 \quad (for \ 95\%)\]

\[\hat{p}_{adj}(1-\hat{p}_{adj}) \quad 0.0123\]

\[\frac{\hat{p}_{adj} - \hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_{adj}}}} \quad 0.11\]

\[\frac{\hat{p}_{adj} - \hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_{adj}}}} \quad 0.217\]

Lower bound of the confidence interval: \[0.4143\]

Upper bound of the confidence interval: \[0.849\]

Compare the intervals obtained from the Wald calculation and the adjusted Wald calculation:

Wald: \([0.428, 0.905]\)

Adjusted Wald: \([0.414, 0.849]\)

Compare intervals for at least 2 different confidence levels calculated from the adjusted Wald formula:

For 90%: \([0.454, 0.828]\)

For 95%: \([0.346, 0.884]\)

To achieve more confidence, we need to adjust the slack on the coverage interval.
Air France: Task 1: Find hours of Osaka museum

\[ \bar{x} \pm t \left( \frac{s}{\sqrt{n}} \right) \]

\[ \frac{5}{2} \text{ margins of error around the mean} \]

\[ \frac{s}{\sqrt{n}} \text{ Standard error (stdErr)} \]

\[ t \left( \frac{l + s}{\sqrt{n}} \right) \text{ stdErr} \text{ Margin of error (marginErr)} \]

\[ \bar{x} \text{ Sample mean } = 2.733 \]

\[ s \text{ Sample standard deviation } \text{http://www.usablestats.com/calcs/stdev} \]

\[ n \text{ Sample size} \]

\[ t \left( \frac{l + s}{\sqrt{n}} \right) \text{ Critical value from t-distribution for n-1 degrees of freedom and the specified confidence level} \]

\[ \alpha = 1 - \text{confidence level} \text{ http://www.usablestats.com/calcs/tiny degrees of freedom } = n - 1 \]

\[ \begin{array}{|c|c|c|}
\hline
\text{Proportion of Area} & \text{Degrees of Freedom} & t\text{-Score} \\
\hline
0.05 & 11 & 2.201 \\
\hline
\end{array} \]

Example:

assume confidence level of 95\%, \( n = 12 \)

Excel 2013:

= T.INV.2T(0.05, 11)
returns 2.2

Older Excel versions:

= TINV(0.05, 11)
returns 2.2

What question can we answer with this data (i.e., population parameter we want to measure)?

- How long will it take for a user to perform the task?
- What kinds of problems in the UI are users likely to see?
- How successful will users be in performing the task?
- How easy will users find this task to do?

Calculations:

\[ n: 15 \]

\[ s: 1.486 \]

\[ t: 2.1448 \]

\[ \text{stdErr: } 0.3836 \]

\[ \text{marginErr: } 0.8231 \]

Lower bound of the confidence interval: \( 1.8101 \)

Upper bound of the confidence interval: \( 3.5565 \)

The answer to the question is:

We are 95% confident that the actual population parameter mean value is between \( 1.81 \) and \( 3.56 \) (Include units for these last 2 numbers.)

for 90\% \( \rightarrow [2.057, 3.409] \)

for 99\% \( \rightarrow [1.590, 3.875] \)
Air France: Task 1: Find hours of Osaka museum

<table>
<thead>
<tr>
<th>Time taken (min)</th>
<th>( \ln \text{ of raw data} )</th>
<th>( \overline{x}_n ) Mean of natural log (ln) values of raw data</th>
<th>( n ) Sample size</th>
<th>( s_n ) Standard deviation of ln values</th>
<th>( \frac{s_n}{\sqrt{n}} ) Standard error of ln values (sErr)</th>
<th>( t_{(\alpha/2)} ) Critical value from t-distribution for n-1 degrees of freedom and the specified confidence level</th>
<th>( t_{(\alpha/2)} \cdot \text{sErr} ) Margin of error (marginErr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.69</td>
<td>1.50988</td>
<td>15</td>
<td>0.42996</td>
<td>0.111</td>
<td>[1.271, 1.748]</td>
<td>0.238</td>
</tr>
<tr>
<td>7.5</td>
<td>2.01</td>
<td>1.74</td>
<td>15</td>
<td>0.42996</td>
<td>0.111</td>
<td>[1.271, 1.748]</td>
<td>0.238</td>
</tr>
<tr>
<td>5.75</td>
<td>1.74</td>
<td>1.74</td>
<td>15</td>
<td>0.42996</td>
<td>0.111</td>
<td>[1.271, 1.748]</td>
<td>0.238</td>
</tr>
<tr>
<td>4.33</td>
<td>1.465</td>
<td>1.74</td>
<td>15</td>
<td>0.42996</td>
<td>0.111</td>
<td>[1.271, 1.748]</td>
<td>0.238</td>
</tr>
<tr>
<td>7.66</td>
<td>2.036</td>
<td>1.74</td>
<td>15</td>
<td>0.42996</td>
<td>0.111</td>
<td>[1.271, 1.748]</td>
<td>0.238</td>
</tr>
<tr>
<td>3.5</td>
<td>1.252</td>
<td>1.74</td>
<td>15</td>
<td>0.42996</td>
<td>0.111</td>
<td>[1.271, 1.748]</td>
<td>0.238</td>
</tr>
<tr>
<td>4.6</td>
<td>1.526</td>
<td>1.74</td>
<td>15</td>
<td>0.42996</td>
<td>0.111</td>
<td>[1.271, 1.748]</td>
<td>0.238</td>
</tr>
<tr>
<td>5.33</td>
<td>1.673</td>
<td>1.74</td>
<td>15</td>
<td>0.42996</td>
<td>0.111</td>
<td>[1.271, 1.748]</td>
<td>0.238</td>
</tr>
<tr>
<td>6.75</td>
<td>1.609</td>
<td>1.74</td>
<td>15</td>
<td>0.42996</td>
<td>0.111</td>
<td>[1.271, 1.748]</td>
<td>0.238</td>
</tr>
<tr>
<td>4.8</td>
<td>1.568</td>
<td>1.74</td>
<td>15</td>
<td>0.42996</td>
<td>0.111</td>
<td>[1.271, 1.748]</td>
<td>0.238</td>
</tr>
<tr>
<td>5.5</td>
<td>1.704</td>
<td>1.74</td>
<td>15</td>
<td>0.42996</td>
<td>0.111</td>
<td>[1.271, 1.748]</td>
<td>0.238</td>
</tr>
<tr>
<td>4.66</td>
<td>1.539</td>
<td>1.74</td>
<td>15</td>
<td>0.42996</td>
<td>0.111</td>
<td>[1.271, 1.748]</td>
<td>0.238</td>
</tr>
<tr>
<td>1.75</td>
<td>0.539</td>
<td>1.74</td>
<td>15</td>
<td>0.42996</td>
<td>0.111</td>
<td>[1.271, 1.748]</td>
<td>0.238</td>
</tr>
<tr>
<td>5.66</td>
<td>1.733</td>
<td>1.74</td>
<td>15</td>
<td>0.42996</td>
<td>0.111</td>
<td>[1.271, 1.748]</td>
<td>0.238</td>
</tr>
<tr>
<td>3.4</td>
<td>1.223</td>
<td>1.74</td>
<td>15</td>
<td>0.42996</td>
<td>0.111</td>
<td>[1.271, 1.748]</td>
<td>0.238</td>
</tr>
</tbody>
</table>

What question can we answer with this data (i.e., population parameter we want to measure)?

- How long will it take for a user to perform the task?
- What kinds of problems in the UI are users likely to see?
- How successful will users be in performing the task?
- How easy will users find this task to do?

Calculations:

\( n = 15 \)

\( t = 2.1448 \)

\( \text{sErr} = 0.111 \)

\( \text{marginErr} = 0.238 \)

Natural log of the lower bound of the confidence interval: 1.271

Natural log of the upper bound of the confidence interval: 1.748

Lower bound of the confidence interval: 3.567

Upper bound of the confidence interval: 5.743

The answer to the question is:
We are 95% confident that the actual population parameter mean value is between 3.567 and 5.743 (min).

FYI: For 90% → [3.722, 5.504] min
For 99% → [3.252, 6.299] min