Bonus

Confidence Intervals

Jean-Yves Le Boudec
May 15, 2015
Here are the results of 10 independent simulation runs (throughput in Mb/s).

13.3  
12.9  
13.1  
13.5  
12.8  
12.9  
13.2  
189.6  
12.9  
13.0  

Give a 95% confidence interval.

A. [ 12.9 ; 189.6]  
B. [ 12.9 ; 13.5]  
C. [ 13.0 ; 13.1]  
D. [ 0 ; 200]  
E. [ 12.9 ; 63.55]  
F. [-2.11 ; 63.55]  
G. I don’t know
Method 1: CI for median
\[ [x_{(2)} ; x_{(9)}] = [12.9 ; 13.5] \]

Method 2: CI for mean \( m \pm 1.96 \frac{s}{\sqrt{n}} \)
mean \( m = 30.72 \)
Standard deviation \( s = 52.96 \) with \( \frac{1}{n} \) formula
Standard deviation \( s = 55.83 \) with \( \frac{1}{n-1} \) formula
CI = \([-2.11 ; 63.55]\) or \([-3.88 ; 65.32]\)
Method 2 may be hard to justify with \( n = 10 \)

Answer B is correct.
We have tested a system for errors and found 0 error in 36 runs. Give a confidence interval for the probability of failure

A. [0% ; 0%]
B. [0%; 1.74%]
C. [0% ; 9.74%]
D. [0% ; 33.74%]
E. [0 ; 100%]
F. I don’t know
CI is \([0 \; p_0(n)]\) with \(p_0(n) = 1 - \left(\frac{1-\gamma}{2}\right)^{\frac{1}{n}} \approx \frac{3.689}{n}\)

Exact formula gives \(p_0(n) = 9.74\%\)

Approximation gives \(p_0(n) \approx \frac{3.689}{36} = 10.2\%\)

Answer C is correct.
We do \( n = 100 \) experiments and compute a 95%-confidence interval \([L, U]\) for the median. We find \( L = 7.4 \) and \( U = 8.3 \). Which statement is correct.

A. The probability that the median is in \( L ; U \) is at least 95%
B. The probability that the median is in \([7.4 ; 8.3]\) is at least 95%
C. Both
D. None
E. I don’t know
Solution

A is correct. In other words, we assume that the median is well defined, but hidden to us. The probability that one experiment gives a 95%-confidence interval $[L; U]$ that contains the median is at least 95%.

B is not correct. In classic statistics, there is no probability assigned to events expressed in terms of the parameter of the distribution.

Answer A.
A. ... a 95% confidence interval to be narrower than a 99% confidence interval
B. ... a 95% confidence interval to be wider than a 99% confidence interval
C. It depends on the data
D. It depends on the type of confidence interval
E. I don’t know
Solution

Answer A

If we want to be more certain, we have to enlarge the confidence interval. Check this with the confidence interval for the median ($n = 37$):

\[
\begin{align*}
95\%: & \quad [x_{(13)}; x_{(25)}] \\
99\%: & \quad [x_{(11)}; x_{(27)}]
\end{align*}
\]

for the mean:

\[
\begin{align*}
95\%: & \quad m \pm 1.96 \frac{s}{\sqrt{n}} \\
99\%: & \quad m \pm 2.58 \frac{s}{\sqrt{n}}
\end{align*}
\]
A data set $x_i > 0$ is such that $y_i = 1/x_i$ looks normal. A 95% confidence interval for the mean of $y_i$ is $[L; U]$.

Is it true that a confidence interval for the mean of $x_i$ is $\left[\frac{1}{U}; \frac{1}{L}\right]$?

A. Yes
B. No
C. I don’t know

33% 33% 33%
Solution

Answer B

\( Y_i = 1/X_i \) is a non linear transformation, the mean is not preserved:

\[
m_X \neq \frac{1}{m_Y}
\]

in general.

\( 1/m_Y \) (i.e. the inverse of the mean of the inverses) is called the harmonic mean.

Note that with probability 95%, \( L \leq m_Y \leq U \)

\[ \left[ \frac{1}{U} ; \frac{1}{L} \right] \] is a confidence interval for the harmonic mean of \( x_i \)
A data set \( x_i > 0 \) is such that \( y_i = 1/x_i \) looks normal. A 95% confidence interval for the median of \( y_i \) is \([L ; U]\).

Is it true that a confidence interval for the median of \( x_i \) is \([1/U ; 1/L]\)?

A. Yes
B. No
C. I don’t know

33%  33%  33%
Solution

Answer A

\( Y_i = 1/X_i \) is a monotonic decreasing transformation, the median is preserved: 
\[
\text{median}_X = \frac{1}{\text{median}_Y}
\]

Further:
\[
P(L \leq \text{median}_Y \leq U) \geq 0.95
\]
\[
(L \leq \text{median}_X \leq U) \iff \left( \frac{1}{U} \leq \frac{1}{\text{median}_Y} \leq \frac{1}{L} \right)
\]
\[
\iff \left( \frac{1}{U} \leq \text{median}_X \leq \frac{1}{L} \right)
\]
Thus \( P \left( \frac{1}{U} \leq \text{median}_X \leq \frac{1}{L} \right) \geq 0.95 \)