Chance or Tolerance Probability of an Confidence Interval

Chao Cheng

June 15, 2023

1 Introduction

When a Confidence Interval(CI) is constructed, usually its upper and lower limits (for a two-sided interval) is random variables. And the probability that this interval will cover (or more precisely, the lower limit will be smaller than and at the same time the upper limit will be greater than) the value in interest, e.g population mean, is the confidence level, usually denoted by $1-\alpha$.

Since both upper and lower limits are random variables, the length (or some times, half of the interval length) is also a random variable. And in some projects, we want to make some statement about the property of this length. Usually we want to quantify the probability that this (half) width of CI is smaller than a specified value. In different literature, this is called the **chance** of CI lies in the width, or the **tolerance probability** of the CI at given width.

In this notes, we will be using CI for normal mean as an example.

2 Construct $(1-\alpha)$ CI

For a random sample

$$x_1, \cdots, x_n \overset{i.i.d}{\sim} N\left(\mu, \sigma^2\right),$$

where μ and σ^2 are both unkonwn, we konw that

$$\frac{\bar{x}-\mu}{\sqrt{S^2/n}} \sim t_{n-1},$$

where $S^2 = \frac{1}{n-1}\sum_{i=1}^{2}\left(x_i-\bar{x}\right)^2$ is sample variance. Then the $1-\alpha$ CI for μ is just

$$\left[\bar{x} - t_{1-\alpha/2,n-1}\cdot\sqrt{\frac{S^2}{n}}, \quad \bar{x} + t_{1-\alpha/2,n-1}\cdot\sqrt{\frac{S^2}{n}}\right].$$

And the **half** length d of this interval is just

$$d \triangleq t_{1-\alpha/2,n-1}\cdot\sqrt{\frac{S^2}{n}}.$$

3 Computation of the chance at given d_0

3.1 When population variance σ^2 is known

Note that for a normal sample,

$$\frac{(n-1)\,S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Therefore

$$\begin{aligned} P\left(d \leq d_0\right) =& P\left(t_{1-\alpha/2,n-1} \cdot \sqrt{\frac{S^2}{n}} \leq d_0\right) \\ =& P\left(\frac{(n-1)\,S^2}{\sigma^2} \leq \frac{n\,(n-1)\,d_0^2}{t^2_{1-\alpha/2,n-1}\sigma^2}\right) \\ =& P\left(X^2_{n-1} \leq \frac{n\,(n-1)\,d_0^2}{t^2_{1-\alpha/2,n-1}\sigma^2}\right). \end{aligned}$$

Hence the chance of this CI falls in the half width limit of d_0 is the probability of a χ^2 random variable X^2_{n-1} with $n-1$ degree of freedom being smaller than

$$\frac{n\,(n-1)\,d_0^2}{t^2_{1-\alpha/2,n-1}\sigma^2}.$$

Note: Actually if population variance σ^2 is known, then we know that

$$\bar{x} \sim N\left(\mu, \sigma^2/n\right),$$

for normal sample. This even holds asymptotically for non-normal sample by CLT. So the $1-\alpha$ CI can be constructed as

$$\left[\bar{x} - z_{1-\alpha/2} \cdot \sigma, \quad \bar{x} + z_{1-\alpha/2} \cdot \sigma\right].$$

And the half width is just $z_{1-\alpha/2} \cdot \sigma$, a scalar, not a random variable and no need for the concept of chance.

3.2 When population variance σ^2 is from previous estimation

If σ^2 is unknown, but previously we have m samples from the same normal distribution and the sample variance is S_0^2. Then again

$$\frac{(m-1)\,S_0^2}{\sigma^2} \sim \chi^2_{m-1},$$

and

$$\frac{S^2}{S_0^2} = \frac{\frac{(n-1)S^2}{\sigma^2}/\left(n-1\right)}{\frac{(m-1)S_0^2}{\sigma^2}/\left(m-1\right)} \sim F_{n-1,m-1}$$

Therefore

$$\begin{aligned}
P\left(d \leq d_0\right) =& P\left(t_{1-\alpha/2,n-1} \cdot \sqrt{\frac{S^2}{n}} \leq d_0\right) \\
=& P\left(\frac{S^2}{S_0^2} \leq \frac{nd_0^2}{t_{1-\alpha/2,n-1}^2 S_0^2}\right) \\
=& P\left(F_{n-1,m-1} \leq \frac{nd_0^2}{t_{1-\alpha/2,n-1}^2 S_0^2}\right).
\end{aligned}$$

Hence the chance of this CI falls in the half width limit of d_0 is the probability of a $F_{n-1,m-1}$ random variable being smaller than

$$\frac{n\left(n-1\right)d_0^2}{t_{1-\alpha/2,n-1}^2 \sigma^2}$$

4 Additional notes and resources

- In Pharmocokinect(PK) analysis, the samples $z_1, \cdots, z_n$ for PK parameters are often assumed to follow log-normal distribution, which means $x_i = \log\left(z_i\right)$ follows normal distribution and the analysis is often down on this log-scale, then transformed back to original scale. Note that for $\mathrm{E}X = \mu$ and $\mathrm{Var}X = \sigma^2$, we have

  $$\mathrm{E}Z = e^{\mu+\frac{\sigma^2}{2}}, \quad \mathrm{Var}Z = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right).$$

  And the Coefficient of Variation(CV) at original scale satisfies

  $$CV = \frac{\sqrt{\mathrm{Var}Z}}{\mathrm{E}Z} = \sqrt{e^{\sigma^2} - 1}.$$

  Note that in PK, CV is often offerd instead of σ^2.

- This analysis is available in PASS as *Confidence Intervals for One Mean with Tolerance Probability.*

- This analysis is available in SAS, for example *The POWER Procedure example: Confidence Interval Precisio.*