STAT22000 Autumn 2013 Lecture 18

Yibi Huang

November 8, 2013

6.1 Confidence Intervals

Example: Lifetime of Light Bulbs

A certain brand of light bulbs claims that **mean lifetime** of its bulbs is 1200 hours with a SD σ of 100 hours.

As a statistician you are **skeptical about the mean** lifetime (which can be overstated), but ready to believe the SD is correctly quoted.

To estimate the mean lifetime μ, you may conduct the following experiment:

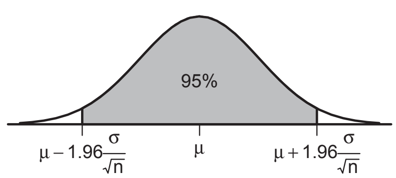
- taking a simple random sample of 100 light bulbs and burn them out, and then
- finding the **average lifetime in the sample** ($\overline{X}$)

Recall that CLT says, for large n,

$$\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = N\left(\mu, \frac{100}{\sqrt{100}}\right) = N(\mu, 10).$$

Often, in an experiment like the lifetime of light bulbs, the actual mean μ is **unknown**.

Recall that CLT says, for large n, $\overline{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$. For a normal curve, 95% of its area is within 1.96 SDs from the center. That means, **for 95% of the time,** $\overline{X}$ **will be within** $1.96\frac{\sigma}{\sqrt{n}}$ **from** μ.



Alternatively, we can also say, **for 95% of the time,** μ **will be within** $1.96\frac{\sigma}{\sqrt{n}}$ **from** $\overline{X}$.

Hence, we call the interval

$$\overline{X} \pm 1.96\frac{\sigma}{\sqrt{n}} = \left(\overline{X} - 1.96\frac{\sigma}{\sqrt{n}}, \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

a **95% confidence interval for** μ.

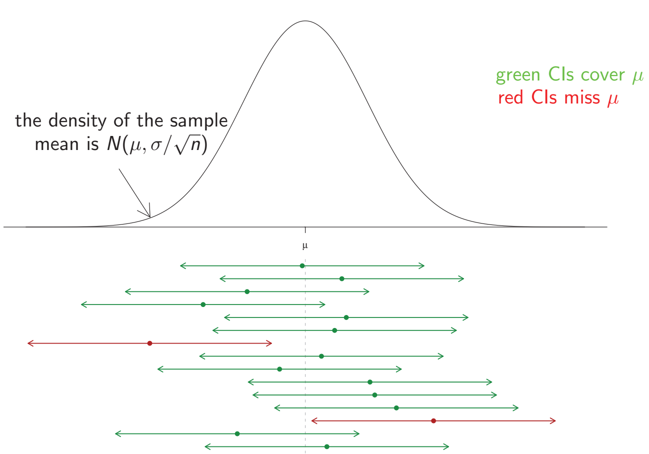
Procedures to find a 95% Confidence Interval for μ (σ Known)

1. Take a simple random sample (or i.i.d. sample) of size n and find the sample mean $\overline{X}$.
2. If n is large, the 95% confidence interval for μ is given by

$$\overline{X} \pm 1.96\frac{\sigma}{\sqrt{n}}$$

Interpretation of confidence intervals: If we repeat the following procedure above multiple times, 95% of the intervals thus constructed will cover the true (unknown) population mean.

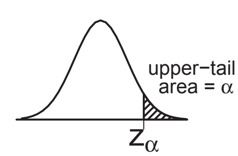
The value "95%" is called the **confidence level** of the interval.



Notation z_α

Let z_α be the value that the area to the right of z_α under the standard normal curve is α. I.e.,

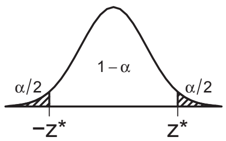
$P(Z > z_\alpha) = \alpha$ or



Confidence Intervals at Other Confidence Levels

For a given confidence level $(1-\alpha)$, we want to find a z^* such that

$$P(-z^* < Z < z^*) = 1-\alpha \quad \text{or}$$



Clearly, such a z^* is simply $z_{\alpha/2}$.

In general, a confidence intervals at confidence level $(1-\alpha)$ is

$$\overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

- 90% C.I.: $\alpha = 0.1$, $z_{\alpha/2} = z_{0.05} = 1.645 \Rightarrow \overline{X} \pm 1.645\frac{\sigma}{\sqrt{n}}$
- 95% C.I.: $\alpha = 0.05$, $z_{\alpha/2} = z_{0.025} = 1.96 \Rightarrow \overline{X} \pm 1.96\frac{\sigma}{\sqrt{n}}$
- 99% C.I.: $\alpha = 0.01$, $z_{\alpha/2} = z_{0.005} = 2.58 \Rightarrow \overline{X} \pm 2.58\frac{\sigma}{\sqrt{n}}$

Example: Utility Company Survey

A utility company serves 50,000 households. As a part of a survey of customer attitudes, they take a SRS of 400 of these households. The average number of TV sets in the sample households turns out to be 1.86, and the SD is known to be 0.90. Find a 95%-confidence interval for the average number of TV sets in all 50,000 households.

Solution. 95% confidence interval is

$$\begin{aligned}&\overline{X} \pm 1.96\frac{\sigma}{\sqrt{n}}\\ &= 1.86 \pm 1.96\frac{0.9}{\sqrt{400}} \approx 1.86 \pm 0.09 = (1.77, 1.95).\end{aligned}$$

Back to the Light Bulb Example

Suppose the average lifetime of 100 randomly selected light bulbs is found to be $\overline{X} = 1150$ hours. Recall the SD is $\sigma = 100$ hours. So a 95% confidence interval for the population mean lifetime μ is

$$\begin{aligned}\overline{X} \pm 1.96\frac{\sigma}{\sqrt{n}} &= 1150 \times 1.96\frac{100}{\sqrt{100}}\\ &= 1150 \pm 19.6 = (1130.4, 1169.6) \text{ hours.}\end{aligned}$$

True or false, and explain:

- The interval $(1130.4, 1169.6)$ contains the sample mean with probability 0.95.
  False. The confidence interval $\overline{X} \pm 1.96\frac{\sigma}{\sqrt{n}}$ definitely (100%) contains the sample mean $\overline{X}$, not just with probability 95%.

Back to the Light Bulb Example (2)

- About 95% of the light bulbs have lifetime between 1130.4 hours and 1169.6 hours.
  False. The confidence interval is for covering the population mean μ, not for covering the entire population. If 95% of the light bulbs have lifetime in the short range 1130.4 - 1169.6 hours, the SD of the lifetimes won't be as large as 100 hours.
- This interval $(1130.4, 1169.6)$ has probability of 0.95 of enclosing the true mean lifetime μ of all light bulbs.
  False. The population mean μ is a fixed number, not random. It is either in the interval 1130.4, 1169.6), or not in the interval. There is no uncertainty involved.

Remark: So what is the thing that is true for 95% of the time?
Ans. It is how the interval might have turned out. About 95% of the intervals constructed in this way (taking a SRS and then calculating $\overline{X} \pm 1.96\sigma/\sqrt{n}$) turn out to cover the population mean μ.

Factors Affecting Length of Confidence Intervals

The half-width $m = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ of the confidence interval, is called the **margin of error**.

The length of the confidence interval decreases if we

1. decrease the confidence level $1-\alpha$
2. increase the sample size n
3. reduce the standard deviation σ

Sample Size Calculation

Before conducting a study, we may decide a confidence level $(1-\alpha)$ and an **upper bound** m **for the margin of error**. In that case we need a sample of size n at least:

$$\left(\frac{z_{\alpha/2}\,\sigma}{m}\right)^2$$