Confidence Intervals Notes

Lt Col Ken Horton Professor Bradley Warner

18 July, 2020

Objectives

Objectives

1) Define the term confidence interval.
2) Using asymptotic methods, obtain and interpret a confidence interval for an unknown parameter, based on a random sample.
3) Describe the relationships between confidence intervals, confidence level, and sample size.

confidence interval

A point estimate provides a single plausible value for a parameter. However, a point estimate is rarely perfect; usually there is some error in the estimate. In addition to supplying a point estimate of a parameter, a next logical step would be to provide a plausible **range of values** for the parameter.

Capturing the population parameter

A plausible range of values for the population parameter is called a **confidence interval**. Using only a point estimate is like fishing in a murky lake with a spear, and using a confidence interval is like fishing with a net. We can throw a spear where we saw a fish, but we will probably miss. On the other hand, if we toss a net in that area, we have a good chance of catching the fish.

If we report a point estimate, we probably will not hit the exact population parameter. On the other hand, if we report a range of plausible values – a confidence interval – we have a good shot at capturing the parameter.

> **Exercise**: If we want to be very certain we capture the population parameter, should we use a wider interval or a smaller interval?[1]

Constructing a confidence interval

A point estimate is our best guess for the value of the parameter, so it makes sense to build the confidence interval around that value. The standard error, which is a measure of the uncertainty associated with the point estimate, provides a guide for how large we should make the confidence interval.

Generally, what you should know about building confidence intervals is laid out in the following steps:

[1] If we want to be more certain we will capture the fish, we might use a wider net. Likewise, we use a wider confidence interval if we want to be more certain that we capture the parameter.

1. Identify the parameter you would like to estimate (for example, μ).
2. Identify a good estimate for that parameter (sample mean, $\bar{X}$).
3. Determine the distribution of your estimate or a function of your estimate. This tells us where our estimate *should* be if we knew the value of our parameter. (According to the central limit theorem, $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathsf{Norm}(0,1)$ and $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim \mathsf{t}(n-1)$).
4. Use this distribution to obtain a range of feasible values (confidence interval) for the parameter. (For μ, we can solve for μ to find a reasonable range).

> Constructing a 95% confidence interval for the mean When the sampling distribution of a point estimate can reasonably be modeled as normal, the point estimate we observe will be within 1.96 standard errors of the true value of interest about 95% of the time. Thus, a **95% confidence interval** for such a point estimate can be constructed:

$$\hat{\theta} \pm \ 1.96 \times SE_{\hat{\theta}}$$

We can be **95% confident** this interval captures the true value.

> **Exercise**:
> Compute the area between -1.96 and 1.96 for a normal distribution with mean 0 and standard deviation 1.

```r
pnorm(1.96)-pnorm(-1.96)
```

```
## [1] 0.9500042
```

In mathematical terms, the derivation of this confidence is as follows:

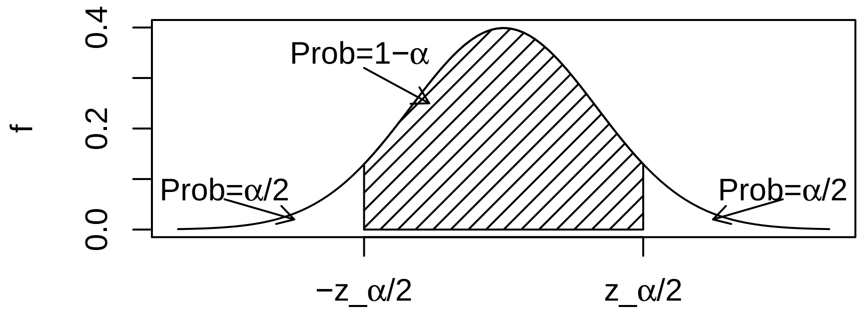
Let $X_1, X_2, ..., X_n$ be an iid sequence of random variables, each with mean μ and standard deviation σ. The central limit theorem tells us that

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \overset{approx}{\sim} \mathsf{Norm}(0,1)$$

Let $0 \le \alpha \le 1$ with the confidence level being $1-\alpha$, yes α is the same as the significance level in hypothesis testing.. Then,

$$\mathrm{P}\left(-z_{\alpha/2} \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le z_{\alpha/2}\right) = 1-\alpha$$

where $z_{\alpha/2}$ is such that $\mathrm{P}(Z \ge z_{\alpha/2}) = \alpha/2$, where $Z \sim \mathsf{Norm}(0,1)$. A picture would help:



So, I know that $(1-\alpha)*100\%$ of the time, $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ will be between $-z_{\alpha/2}$ and $z_{\alpha/2}$.

By rearranging the expression above and solving for μ, we get:

$$\mathrm{P}\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

Be careful with the interpretation of this expression. As a reminder $\bar{X}$ is the random variable here. The population mean, μ, is NOT a variable. It is an unknown parameter. Thus, the above expression is NOT a probabilistic statement about μ, but rather about $\bar{X}$.

Nonetheless, the above expression gives us a nice interval for "reasonable" values of μ given a particular sample.

A $(1-\alpha)*100\%$ *confidence interval for the mean* is given by:

$$\mu \in \left(\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$