Statistical Inference

A parameter is a numerical summary calculated from a population data set. A statistic is a numerical summary calculated from a sample data set.

Confidence interval is a range of values within which one would expect a parameter to fall. The calculation is based on the point estimator (which should be an unbiased estimator) and the margin of error (sometimes called the precision). In general, confidence intervals are constructed as follows:

\[ \text{point estimator} \pm \text{margin of error} \]

Confidence Interval for One Population Mean
(Population Standard Deviation Unknown)

We use the t-distribution with the degrees of freedom, \( df = n - 1 \). The confidence interval has the following form:

\[ \bar{x} \pm t \frac{s}{\sqrt{n}} \]

Where \( s \) is the sample standard deviation, \( n \) is the sample size, \( t \) is the t-value determined by the degrees of freedom and the preset confidence level. Here we need to know that the sample was obtained from a normally distributed population.

Example 1

Suppose a sample of size 23 (degrees of freedom \( df = 22 \)) produced a mean of 122.45 and a standard deviation of 8.55. Assuming that the sample was drawn from a normally distributed population, we calculate the 95% confidence interval as follows:

\[ \bar{x} \pm t \frac{s}{\sqrt{n}} = 122.45 \pm 2.074 \frac{8.55}{\sqrt{23}} = 122.45 \pm 3.698 \]

Therefore, the confidence limits are 118.752 and 126.148. In other words, one can be 95% that the population mean falls between 118.752 and 126.148. More correctly, if one took many samples of the same size, and constructed the corresponding confidence intervals, then 95% of these confidence intervals would contain the true population mean, and 5% of them would not.

The t-value of 2.074 was located in the t-distribution table; note that you have to match the correct degrees of freedom (22) with the correct confidence level (0.95).
Example 2

Same as above, with the sample size of 374. Now, the degrees of freedom is 373. Note that for such a high value, the table gives the z-values (values from the standard normal distribution). Therefore, the confidence interval can be calculated as:

$$\bar{x} \pm t \frac{s}{\sqrt{n}} = 122.45 \pm 1.96 \frac{8.55}{\sqrt{374}} = 122.45 \pm 0.867$$

Therefore, the limits are 121.583 and 123.317. Note the huge reduction in the margin of error. This range gives us a possible range of values within which we expect the true population mean to fall.