Chemometrics

MA4605

Week 2. Lecture 4. Confidence Intervals

September 13, 2011

Confidence Intervals for a population mean using the Normal distribution

Using the properties of the Normal distribution, we know that 95% of the sample means, will lie within the range of the population mean $\mu \pm$ 1.96 standard errors of the mean i.e.

$$\mu \pm 1.96 SE(\overline{x}) = [\mu - 1.96 SE(\overline{x}), \mu + 1.96 SE(\overline{x})]$$

Or equivalently

$$\mu - 1.96 SE(\overline{x}) < \overline{x} < \mu + 1.96 SE(\overline{x})$$

Since in practice $\overline{x}$ (sample mean) is known and μ (population mean) is unknown, the equations above can be rearranged to obtain a range that contains the values of the true parameter μ with 95% confidence.

$$\overline{x} - 1.96SE(\overline{x}) < \mu < \overline{x} + 1.96SE(\overline{x})$$

The 95% CI of the mean= [$\overline{x} \pm$ 1.96 SE($\overline{x}$)]

This range is called a 95% confidence interval for μ .
It is a range of values which contains the true population mean with a probability of 0.95 or 95%. We can expect that a 95% confident interval will not include the true mean 5% of the time.
From the **Central Limit Theorem** we know that the standard error is

$$SE(\overline{x}) = \frac{\sigma}{\sqrt{n}}$$

When σ is unknown (most likely) use s, the sample standard deviation

$$SE(\overline{x}) = \frac{s}{\sqrt{n}}$$

- When σ is unknown (most likely) and the sample size n is large, replace σ by s (sample standard deviation).

- When σ is unknown (most likely) and the sample size n is large, replace σ by s (sample standard deviation).
- What happens when the sample size n is small?

- When σ is unknown (most likely) and the sample size n is large, replace σ by s (sample standard deviation).
- What happens when the sample size n is small?
- Use the t distribution.

The Student's t distribution

The formula for confidence intervals may be inaccurate when the sample size n is small and the population standard deviation σ is replaced by the sample standard deviation s.

- t distribution is symmetric like the Normal distribution but it is flatter and the tails of the distribution are more spread out.
- As n gets large (> 30), the t distribution is almost identical to the normal distribution.
- The t distribution is characterised by its number of degrees of freedom (df). The number of degrees of freedom for a single sample of size n is always $n - 1$.

The 95% CI of the mean=$[\overline{x} \pm\ t_{\frac{\alpha}{2};n-1}$ SE($\overline{x}$)]
where $t_{\frac{\alpha}{2};n-1}$ is a the value producing an area of $\frac{\alpha}{2}$ in the upper tail of a t distribution with $n-1$ degrees of freedom.

Example

Calculate the 95% CI of the mean of a population based on the results obtained from a sample of size n=15 in which

- the sample mean $\overline{x}$=3
- the sample standard deviation $s = 0.23$

The confidence interval can be obtained as

$$
\begin{aligned}
&[\overline{x} \pm \; t_{(\frac{\alpha}{2};n-1)} \cdot SE(\overline{x})] \\
&= [\overline{x} \pm \; t_{(\frac{\alpha}{2};n-1)} \cdot \frac{\sigma}{\sqrt{n}}] \\
&= [\overline{x} \pm \; t_{(\frac{0.05}{2};14)} \cdot \frac{s}{\sqrt{n}}] \\
&= [3 \pm \; t_{(0.025;14)} \cdot \frac{0.23}{\sqrt{15}}] \\
&= [3 \pm \; t_{(0.025;14)} \cdot 0.0594]
\end{aligned}
$$

How to read quantiles from the Student *t* distribution.

Use the *R* function **qt** to calculate the quantile for which the probability that a *t* distributed random value with 14 degrees of freedom is 0.025.

$> qt(0.975, 14)$

[1] 2.144787

So the confidence interval is:

$$
\begin{aligned}
&= [3 \pm \ t_{(0.025;14)} \cdot 0.0594] \\
&= [3 \pm \ 2.144787 \cdot 0.0594] \\
&= [3 \pm \ 0.13] \\
&= [2.87, 3.13]
\end{aligned}
$$