Confidence intervals

Paula Moraga

Confidence interval

A $(c \times 100)\%$ confidence interval for a population parameter based on sample observations is an interval (A, B) such that we are $(c \times 100)\%$ confident that the population parameter is within the interval.

- When we use a different sample, we obtain a different confidence interval
- If we take many samples from the population and calculate 95% CIs for each sample, then 95% of these CIs will contain the true population parameter.
- Confidence level $c = 1 - \alpha$. Significance level $\alpha = 1 - c$
- 95% CI. Confidence level $c = 0.95$. Significance level $\alpha = 0.05$
- 90% CI. Confidence level $c = 0.90$. Significance level $\alpha = 0.10$
- 99% CI. Confidence level $c = 0.99$. Significance level $\alpha = 0.01$

Construct a confidence interval

A $(c \times 100)\%$ confidence interval for a population parameter is

$$\text{sample statistic} \pm \text{critical value} \times \text{SE}$$

- **Sample statistic** used to estimate the population parameter. Sample mean ($\bar{X}$) to estimate the population mean (μ) Sample proportion ($\hat{P}$) to estimate the population proportion (p)
- **Standard error SE** is the standard deviation of the sampling distribution of the sample statistic
- **Critical value** (z^*, $t^*(n-1)$).
  Measures the number of SE to be added and subtracted from the sample statistic to achieve the desired confidence level

Confidence interval for a proportion

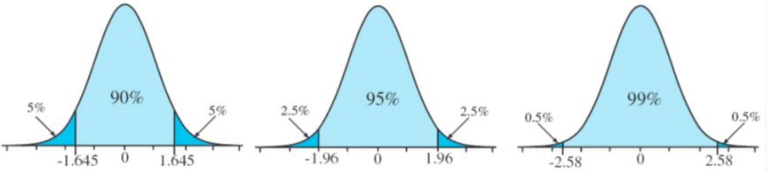
$(1-\alpha) \times 100\%$ confidence interval for the population proportion p

$$\hat{P} \pm z^* \times SE = \hat{P} \pm z^* \times \sqrt{\hat{P}(1-\hat{P})/n}$$

Assumptions:

1. Sample observations are independent (random sample)
2. In the sample, there are at least 10 successes and 10 failures

z^*: value such that $c = 1 - \alpha$ of the probability in the N(0,1) falls between $-z^*$ and z^*.
Value such that $\alpha/2$ of the probability in the $N(0,1)$ is below z^*



Example. Confidence interval for a proportion

In a random sample of 40 students, 24 said they bought a copy of the textbook. Estimate the proportion of all students who bought the book and give a 95% confidence interval.

Solution

$$\hat{p} \pm z^* \times SE = \hat{p} \pm z^* \times \sqrt{\hat{p}(1-\hat{p})/n}$$

Assumptions:

- Sample observations independent
- At least 10 successes and 10 failures

Example. Confidence interval for a proportion

A random sample of 826 people living in UK was surveyed to better understand their political preferences. 70% of the responses supported the political party A. Calculate a 95% confidence interval for the proportion of people that support the political party A.

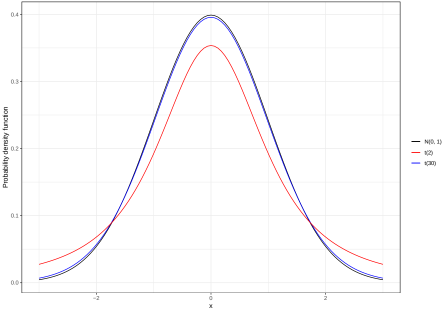
Solution

$$\hat{p} \pm z^* \times SE = \hat{p} \pm z^* \times \sqrt{\hat{p}(1-\hat{p})/n}$$

Assumptions:

- Observations are independent (they are from a random sample)
- In the sample at least 10 support and at least 10 do not support

t distribution



- Symmetric and bell-shaped like the normal distribution.
- Mean 0, Heavier tails than the normal distribution. Observations are more likely to fall beyond two standard deviations from the mean than under the normal distribution.

The t distribution has a parameter called degrees of freedom (df). The df describes the form of the t-distribution. When the df increases, the t-distribution approaches the standard normal N(0, 1)

Confidence interval for a mean

$(1-\alpha) \times 100\%$ confidence interval for the population mean μ

$$\bar{X} \pm t^*_{n-1} \times SE = \bar{X} \pm t^*_{n-1} \times \frac{S}{\sqrt{n}}$$

t^*_{n-1} value such that $\alpha/2$ of the probability in $t(n-1)$ is below t^*_{n-1}

Assumptions:

1. Sample observations are independent.
2. Sample observations from a normally distributed population.

When the population standard deviation is known:

$$\bar{X} \pm z^* \times SE = \bar{X} \pm z^* \times \frac{\sigma}{\sqrt{n}}$$

z^* value such that $\alpha/2$ of the probability in the N(0,1) is below z^*

Example. Confidence interval for a mean

In a class survey, students are asked how many hours they sleep per night. In a sample of 52 students, the mean was 5.77 hours with a standard deviation of 1.572 hours. Construct a 95% confidence interval for the mean number of hours slept per night in the population from which this sample was drawn.

Example. Confidence interval for a mean

Elevated mercury concentrations are an important problem for both dolphins and other animals who occasionally eat them. Calculate a 95% CI for the average mercury content in dolphins using a sample of 19 dolphins in Japan. In the sample, $n = 19$, $\bar{x} = 4.4$, $s = 2.3$, minimum $= 1.7$ and maximum $= 9.2$ μg/wet gram (micrograms of mercury per wet gram of muscle).