## Confidence intervals

Confidence intervals are a useful way to capture variation around point estimators. A confidence interval gives the range of values that we would expect our point estimator to be in, for a certain percentage of the time.

Consider a 95% confidence interval for $\mu$. This means that if we calculated a large number of confidence intervals, we would expect 95% of them to contain $\mu$, the population mean.

This is not the same as saying that our confidence interval has a 95% chance of containing the population mean. Our confidence interval either contains the population mean or it does not contain it. There is no probability associated with that (other than 0 or 1). We just do not know unless we know the population mean!

To calculate a confidence interval for the mean of a Normal distribution with **known** variance, use the formula:

$$\text{CI} = \bar{x} \pm z\sqrt{\frac{\sigma^2}{n}}$$

### Example

Bill is a long-jumper. He completes 10 jumps with the following results (measured in metres):

6·21, 6·33, 6·02, 6·11, 6·13,
6·40, 6·51, 6·29, 6·16, 6·44.

Assume the distances are a random sample from a Normal distribution with mean $\mu$ and standard deviation 0·1 metres. Calculate a 95% CI for $\mu$.

$\bar{x} = 6{\cdot}26$

$\text{CI} = 6{\cdot}26 \pm 1{\cdot}96\sqrt{\frac{0{\cdot}1^2}{10}} = [6{\cdot}198, 6{\cdot}322]$

If the variance is **unknown** and the sample is **small**, use the unbiased estimator for population variance, $s^2$, and the Student's $t$-distribution with $n-1$ degrees of freedom.

$$\text{CI} = \bar{x} \pm t\sqrt{\frac{s^2}{n}}$$

### Example

Ian is a discus thrower. He throws the discus 10 times with the following results (measured in metres):

25·3, 23·8, 24·7, 24·9, 23·7,
25·6, 24·6, 24·0, 25·3, 24·1

You may assume that the distances (in metres) are Normally distributed with mean $\mu$ and variance $\sigma^2$. Calculate a 95% CI for $\mu$.

$\bar{x} = 24{\cdot}6 \qquad s^2 = 0{\cdot}46$ (See KO 5.1)

$\text{CI} = 24{\cdot}6 \pm 2{\cdot}262\sqrt{\frac{0{\cdot}46}{10}} = [24{\cdot}115, 25{\cdot}085]$

2·262 is from the Student's $t$-distribution with 9 degrees of freedom.

If the variance is **unknown** and the sample is **large**, use the unbiased estimator for population variance, $s^2$. (See KO 5.1 for formula for $s^2$).

$$\text{CI} = \bar{x} \pm z\sqrt{\frac{s^2}{n}}$$

To calculate a confidence interval for the difference between two means from Normal distributions with **known** variance, use the formula:

$$\text{CI} = \bar{x} - \bar{y} \pm z\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$$

The confidence interval for a proportion or probability when $n$ is large is given by:

$$\text{CI} = p \pm z\sqrt{\frac{p(1-p)}{n}}$$

### Example

In order to estimate the proportion, $p$, of a certain population who are bilingual, a random sample of 1200 members of the population are questioned. It is found that 498 of them are bilingual. Calculate a 90% CI for the proportion of the population who are bilingual.

$\hat{p} = \frac{498}{1200} = 0{\cdot}415$

$\text{CI} = 0{\cdot}415 \pm 1{\cdot}645\sqrt{\frac{0{\cdot}415 \times 0{\cdot}585}{1200}} = [0{\cdot}392, 0{\cdot}438]$