

## 17. Confidence intervals (CIs)

In [Chapter 15](#), we saw how it is possible to calculate the probability of sampling values from a specific interval of the normal distribution (e.g., the probability of sampling a value within 1 standard deviation of the mean). In this chapter, we will see how to apply this knowledge to calculating intervals that express confidence in the mean value of a population.

Remember that that we almost never really know the true mean value of a *population*,  $\mu$ . Our best estimate of  $\mu$  is the mean that we have calculated from a *sample*,  $\bar{x}$  (see [Chapter 4] for a review of the difference between populations and samples). But how good of an estimate is  $\bar{x}$  of  $\mu$ , really? Since we cannot know  $\mu$ , one way of answering this question is to find an interval that expresses a degree of confidence about the value of  $\mu$ . The idea is to calculate 2 numbers that we can say with some degree confidence that  $\mu$  is between (i.e., a lower confidence interval and an upper confidence interval). The wider this interval is, the more confident that we can be that the true mean  $\mu$  is somewhere within it. The narrower the interval is, the less confident we can be that our confidence intervals (CIs) contain  $\mu$ .

Confidence intervals are notoriously easy to misunderstand. We will explain this verbally first, focusing on the general ideas rather than the technical details. Then we will present the calculations before coming back to their interpretation again. The idea follows a similar logic to the standard error from [Chapter 12](#).

Suppose that we want to know the mean body mass of all domestic cats (Figure 16.1). We cannot weigh every living cat in the world, but maybe we can find enough to get a sample of 20. From these 20 cats, we want to find some interval of weights (e.g., 3.9-4.3 kg) within which the *true* mean weight of the population is contained. The only way to be 100% certain that our proposed interval *definitely* contains the true mean would be to make the interval absurdly large. Instead, we might more sensibly ask what the interval would need to be to contain the mean with 95% confidence. What does “with 95% confidence” actually mean? It means when we do the calculation to get the interval, the true mean should be somewhere within the interval 95% of the time that a sample is collected.

In other words, if we were to go back out and collect another sample of 20 cats, and then another, and another (and so forth), calculating 95% CIs each time, then in 95% of our samples the true mean will be within our CIs (meaning that 5% of the time it will be outside the CIs). Note that this is slightly different than saying that there is

## 17. Confidence intervals (CIs)



Figure 17.1.: Two domestic cats sitting side by side with much different body masses.

a 95% probability that the true mean is between our CIs.<sup>1</sup> Instead, the idea is that if we were to repeatedly resample from a population and calculate CIs each time, then 95% of the time the true mean would be within our CIs (Sokal and Rohlf, 1995). If this idea does not make sense at first, that is okay. The calculation is actually relatively straightforward, and we will come back to the statistical concept again afterwards to interpret it. First we will look at CIs assuming a normal distribution, then the special case of a binomial distribution.

### 17.1. Normal distribution CIs

Remember from the Central Limit Theorem in Chapter 15 that as our sample size  $N$  increases, the distribution of our sample mean  $\bar{x}$  will start looking more and more like a normal distribution. Also from Chapter 15, we know that we can calculate the

---

<sup>1</sup>The reason that these two ideas are different has to do with the way that probability is defined in the frequentist approach to statistics (see Chapter 14). With this approach, there is no way to get the probability of the true mean being within an interval, strictly speaking. Other approaches to probability, such as Bayesian probability, do allow you to build intervals in which the true mean is contained with some probability. These are called “credible intervals” rather than “confidence intervals” (e.g., Ellison, 2004). The downside to credible intervals (or not, depending on your philosophy of statistics) is that Bayesian probability is at least partly subjective; i.e., based in some way on the subjective opinion of the individual researcher.

probability associated with any interval of values in a normal distribution. For example, we saw that about 68.2% of the probability density of a normal distribution is contained within a standard deviation of the mean. We can use this knowledge from [Chapter 15](#) to set confidence intervals for any percentage of values around the sample mean ( $\bar{x}$ ) using a standard error (SE) and z-score (z). Confidence intervals include 2 numbers. The **lower confidence interval** (LCI) is below the mean, and the **upper confidence interval** (UCI) is above the mean. Here is how they are calculated,

$$LCI = \bar{x} - (z \times SE),$$

$$UCI = \bar{x} + (z \times SE).$$

Note that the equations are the same, except that for the LCI, we are subtracting  $z \times SE$ , and for the UCI we are adding it. The specific value of z determines the confidence interval that we are calculating. For example, about 95% of the probability density of a standard normal distribution lies between  $z = -1.96$  and  $z = 1.96$  (Figure 16.2). Hence, if we use  $z = 1.96$  to calculate LCI and UCI, we would be getting 95% confidence intervals around our mean.

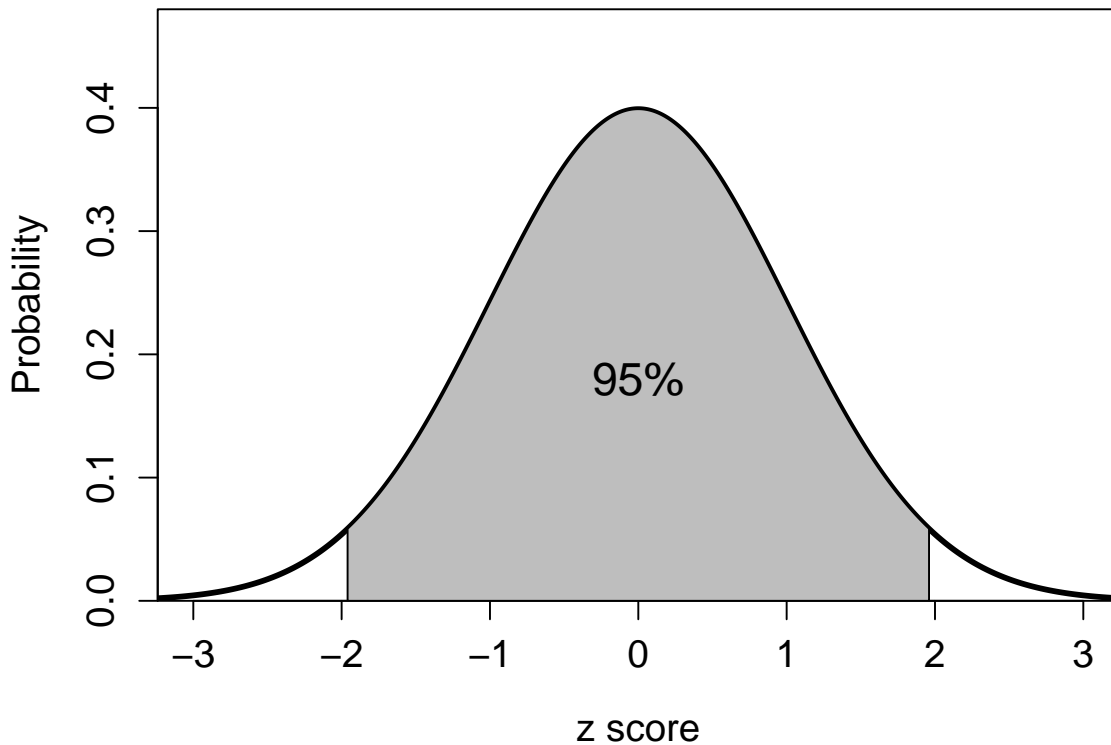


Figure 17.2.: A standard normal probability distribution showing 95 per cent of probability density surrounding the mean.

## 17. Confidence intervals (CIs)

An [interactive application](#) helps visualise the relationship between probability intervals and z-scores more generally (make sure to set ‘Tailed’ to ‘Two-tailed’ using the pulldown menu).

Now suppose that we want to calculate 95% CIs around the sample mean of our  $N = 20$  domestic cats from earlier. We find that the mean body mass of cats in our sample is  $\bar{x} = 4.1$  kg, and that the standard deviation is  $s = 0.6$  kg (suppose that we are willing to assume, for now, that  $s = \sigma$ ; that is, we know the true standard deviation of the population). Remember from [Chapter 12](#) that the sample standard error can be calculated as  $s/\sqrt{N}$ . Our lower 95% confidence interval is therefore,

$$LCI_{95\%} = 4.1 - \left(1.96 \times \frac{0.6}{20}\right) = 4.041$$

Our upper 95% confidence interval is,

$$UCI_{95\%} = 4.1 + \left(1.96 \times \frac{0.6}{20}\right) = 4.159$$

Our 95% CIs are therefore  $LCI = 4.041$  and  $UCI = 4.159$ . We can now come back to the statistical concept of what this actually means. If we were to go out and repeatedly collect new samples of 20 cats, and do the above calculations each time, then 95% of the time our true mean cat body mass would be somewhere between the LCI and UCI.

Ninety-five per cent confidence intervals are the most commonly used in biological and environmental sciences. In other words, we accept that about 5% of the time (1 in 20 times), our confidence intervals will not contain the true mean that we are trying to estimate. Suppose, however, that we wanted to be a bit more cautious. We could calculate 99% CIs; that is, CIs that contain the true mean in 99% of samples. To do this, we just need to find the z-score that corresponds with 99% of the probability density of the standard normal distribution. This value is about  $z = 2.58$ , which we could find with the [interactive application](#), a [z table](#), some maths, or a quick online search<sup>2</sup>. Consequently, the upper 99% confidence interval for our example of cat body masses would be,

$$LCI_{99\%} = 4.1 - \left(2.58 \times \frac{0.6}{20}\right) = 4.023$$

Our upper 99% confidence interval is,

$$UCI_{99\%} = 4.1 + \left(2.58 \times \frac{0.6}{20}\right) = 4.177$$

---

<sup>2</sup>While it is always important to be careful when searching, typing “z score 99 percent confidence interval” will almost always get the intended result.

Notice that the confidence intervals became wider around the sample mean. The 99% CI is now 4.023-4.177, while the 95% CI was 4.041-4.159. This is because if we want to be more confident about our interval containing the true mean, we need to make a bigger interval.

We could make CIs using any percentage that we want, but in practice it is very rare to see anything other than 90% ( $z = 1.65$ ), 95% ( $z = 1.96$ ), or 99% ( $z = 2.58$ ). It is useful to see what these different intervals actually look like when calculated from actual data, so this [interactive application](#) illustrates CIs on a histogram with red dotted lines next to the LCI and UCI equations.

[Click here](#) for an interactive application demonstrating confidence intervals.

Unfortunately, the CI calculations from this section are a bit of an idealised situation. We assumed that the sample means are normally distributed around the population mean. While we know that this *should* be the case as our sample size increases, it is not quite true when our sample is small. In practice, what this means is that our z-scores are usually not going to be the best values to use when calculating CIs, although they are often good enough when a sample size is large<sup>3</sup>. We will see what to do about this in [Chapter 18](#), but first we turn to the special case of how to calculate CIs from binomial proportions.

## 17.2. Binomial distribution CIs

For a binomial distribution, our data are counts of successes and failures (see [Chapter 14](#)). For example, we might flip a coin 40 times and observe 22 heads and 18 tails. Suppose that we do not know in advance the coin is fair, so we cannot be sure that the probability of it landing on heads is  $p = 0.5$ . From our collected data, our estimated probability of landing on heads is,  $\hat{p} = 22/40 = 0.55$ .<sup>4</sup> But how would we calculate the CIs around this estimate? In this case, the formula is similar to ones for LCI and UCI from the normal distribution shown earlier. We just need to note that the variance of  $p$  for a binomial distribution is  $\sigma^2 = p(1 - p)$  ([Box et al., 1978](#); [Sokal and Rohlf, 1995](#)).<sup>5</sup> This means that the standard deviation of  $p$  is  $\sigma = \sqrt{p(1 - p)}$ , and  $p$  has a standard error,

---

<sup>3</sup>What defines a ‘small’ or a ‘large’ sample is a bit arbitrary. A popular suggestion (e.g., [Sokal and Rohlf, 1995](#), page 145) is that any  $N < 30$  is too small to use z-scores, but any cut-off  $N$  is going to be somewhat arbitrary. Technically, the z score is not **completely** accurate until  $N \rightarrow \infty$ , but for all intents and purposes, it is usually only trivially inaccurate for sample sizes in the hundreds. Fortunately, you do not need to worry about any of this when calculating CIs from continuous data in Jamovi because Jamovi applies a correction for you, which we will look at in [Chapter 18](#).

<sup>4</sup>The hat over the  $P$ , ( $\hat{P}$ ) is just being used here to indicate the *estimate* of  $P(\text{heads})$ , rather than the *true*  $P(\text{heads})$ .

<sup>5</sup>Note, the variance of total *successes* is simply  $np(1 - p)$ ; i.e., just multiply the variance of  $p$  by  $n$ .

## 17. Confidence intervals (CIs)

$$SE(p) = \sqrt{\frac{p(1-p)}{N}}.$$

We can use this standard error in the same equation from earlier for calculating confidence intervals. For example, if we wanted to calculate the lower 95% CI for  $\hat{p} = 0.55$ ,

$$LCI_{95\%} = 0.55 - 1.96\sqrt{\frac{0.55(1-0.55)}{40}} = 0.396$$

Similarly, to calculate the upper 95% CI,

$$UCI_{95\%} = 0.55 + 1.96\sqrt{\frac{0.55(1-0.55)}{40}} = 0.704.$$

Our conclusion is that, based on our sample, 95% of the time we flip a coin 40 times, the true mean  $p$  will be somewhere between 0.396 and 0.704. These are quite wide CIs, which suggests that our flip of  $\hat{p} = 0.55$  would not be particularly remarkable even if the coin was fair ( $p = 0.5$ ).<sup>6</sup>

We can do another example, this time with our example of the probability of testing positive for Covid-19 at  $\hat{p} = 0.025$ . Suppose this value of  $\hat{p}$  was calculated from a survey of 400 people ( $N = 400$ ). We might want to be especially cautious about estimating CIs around such an important probability, so perhaps we prefer to use 99% CIs instead of 95% CIs. In this case, we use  $z = 2.58$  as with the normal distribution example from earlier. But we apply this  $z$  score using the binomial standard error to get the LCI,

$$LCI_{99\%} = 0.025 - 2.58\sqrt{\frac{0.025(1-0.025)}{400}} = 0.00486$$

Similarly, we get the UCI,

$$UCI_{99\%} = 0.025 + 2.58\sqrt{\frac{0.025(1-0.025)}{400}} = 0.0451.$$

---

<sup>6</sup>You might ask, why are we doing all of this for the binomial distribution? The central limit theorem is supposed to work for the mean of any distribution, so should that not include the distribution of  $p$  too? Can we not just indicate success (heads) with a 1 and failures (tails) with a 0, then estimate the standard error of 22 values of 1 and 18 values of 0? Well, yes! That actually does work and gives an estimate of 0.079663, which is very close to the  $\sqrt{\hat{p}(1-\hat{p})/N} = 0.078661$ . The problem arises when the sample size is low, or when  $p$  is close to 0 or 1, and we are trying to map the  $z$  score to probability density. For this reason, it is best to stick with  $\sqrt{\hat{p}(1-\hat{p})/N}$ . There are other methods that attempt to give even better estimates (the one we are using is called the Wald method), but we will not consider these here.

Notice that the LCI and UCI differ here by about an order of magnitude (i.e., the UCI is about 10 times higher than the LCI).

In summary, this chapter has focused on what confidence intervals are and how to calculate them. [Chapter 18](#) will turn to the t-interval, what it is and why it is used.