# Inf2 – Foundations of Data Science 2021
# Topic: Confidence intervals

David C. Sterratt

17th May 2022

**Recommended reading:**

- *Computational and Inferential Thinking*, Chapter 13
- *Modern Mathematical Statistics with Applications*, Sections 8.1–8.3 and 8.5

## 1 Video: Principle of confidence intervals

**Illustration: confidence intervals for the mean** From the Central Limit Theorem we know that for large samples, the distribution of the mean is normal, and that the estimated standard error of the mean should be close to the standard error in the mean. We can then ask "if we looked at an interval around our estimate for the mean, how often would the true value be contained in that interval"?

Figure 1 gives an illustrated answer to this question. Each blue horizontal line corresponds to one sample of size $n$ from a population, and shows a range of estimates for the population mean based on that sample – in other words a **confidence interval**. We can see that the true value of the mean (black vertical line) is contained in most of the intervals, but not all of them.

**Size of confidence interval** We have chosen the length of the intervals to ensure that, if we carried on estimating the mean and the interval, about 95% of intervals would contain the true mean. To determine this length, we use the $z$ critical values of the normal distribution (Figure 2). We define the $z$ **critical value** $z_\alpha$ as the value of $z$ in a normal distribution which has the area $\alpha$ under the curve to its right. If we want the intervals to contain the true mean 95% of the time, we need to make sure that the mean is within the central 95% of the distribution. This implies that we need 2.5% of the area under the curve to the right of the upper bound, so we look up $z_{0.025}$ in a statistical table or a function in a stats package and find that $z_{0.025} = 1.96$ – we will show how to do this later. The $z$ critical value of 1.96

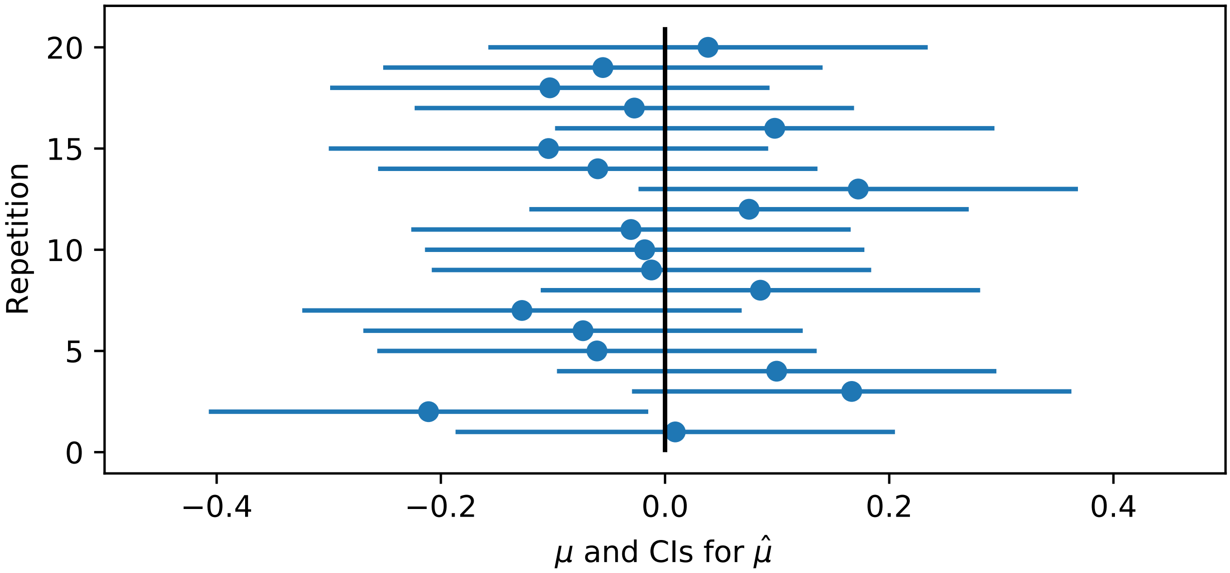

Figure 1: Principle of confidence intervals. We repeat a simulation using a sample size of $n = 100$ to estimate the sample mean of a normal distribution with mean 0 and standard deviation 1. The black vertical line indicates the true mean, the blue dots indicate the sample means, and the blue horizontal lines indicate the 95% confidence intervals obtained in each of the 20 repetitions. It can be seen that 19 of the confidence intervals do contain the population mean, but one of them does not.

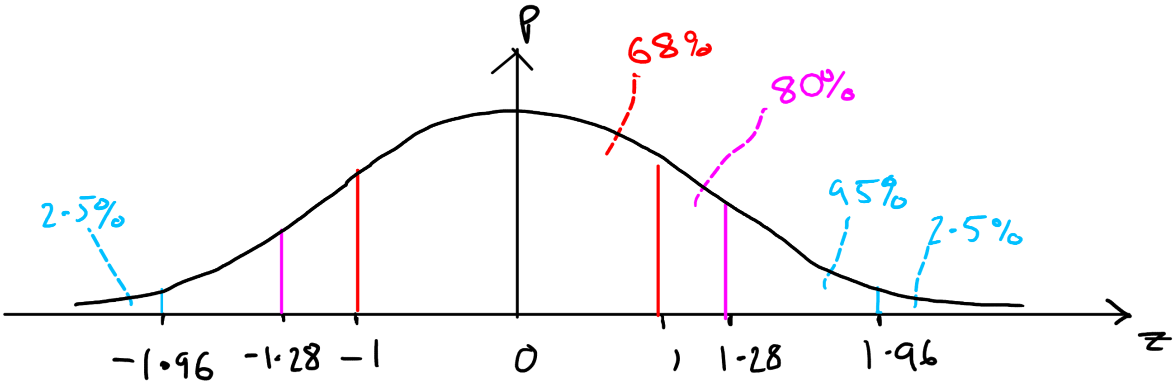

Figure 2: Confidence intervals of a normal distribution. The intervals containing various amounts of probability mass under a normal distribution are shown. The 95% confidence interval (blue) is $[-1.96, 1.96]$ and has 2.5% of the probability mass in each tail. The 80% confidence interval is $[-1.28, 1.28]$. The amount of probability mass contained in one standard deviation is 68%. In general for a confidence interval of $100(1 - \alpha)\%$, the upper and lower boundaries are determined by the $z$ critical value $z_{\alpha/2}$. E.g. with the 95% confidence interval $\alpha = 0.05$ and there is 2.5% of the area of the curve above the upper boundary of the confidence interval.

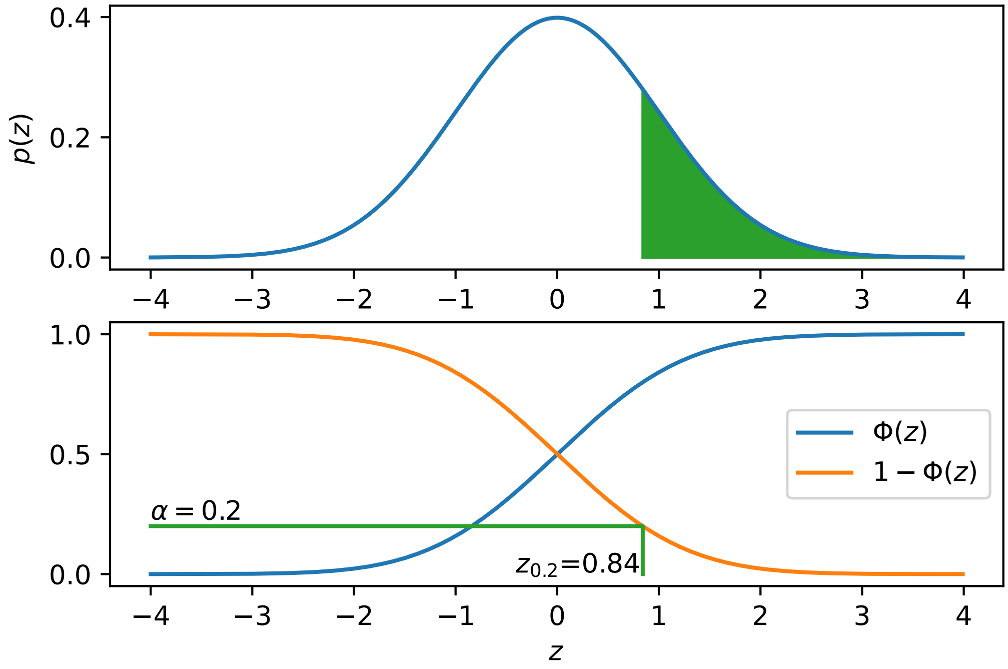

Figure 3: Concept of the $z$ critical value. Top: $z_\alpha$ is the value of $z$ such in a normal distribution such that the area under the curve to the right of the $z_\alpha$ (green) is equal to $\alpha$. i.e. $\alpha = \int_{z_\alpha}^{\infty} p(z)dz$. Bottom: the blue curve shows the cumulative distribution function $\Phi(z) = \int_{-\infty}^{z} p(z)dz$. The orange curve shows the "survival function" $\mathrm{sf}(z) = 1 - \Phi(z)$. The survival function of $z$ is exactly the area to the right of $z$ under the pdf. Therefore we want to look up the inverse survival function to determine $z_\alpha$ from $\alpha$, as indicated by the green lines.

tells us that the length of the lines on the side of each estimate of the mean should all be 1.96 times the standard error of the mean (SEM).

We may want to be more or less certain of whether the mean is contained in a confidence interval. In this case we can look up the $z$ critical value for our chosen level of confidence. We can also decide to express the confidence interval in terms of the multiples of the SEM. For example confidence intervals of plus or minus one SEM correspond to a 68% confidence interval.

**Reminder** It is worth remembering that these simulations are artificial in the sense that we can repeat many samples. In real life we only get one sample, which does or does not contain the true value – but we don't know.

**Looking up a $z$ critical value (added after 2021/22 lectures)** To look up a $z$ critical value, you can use the python `scipy` package. For example to find $z_{0.2}$ you would use:

```
from scipy.stats import norm
alpha = 0.2
print(norm.isf(alpha))
```

The function name `isf` stands for **inverse survival function**. As illustrated in Figure 3, it's the inverse of one minus the cumulative distribution function (cdf).

# 2 Video: Definition of confidence intervals

**Definition of confidence intervals** We define a **confidence interval** as an interval $(\hat{\vartheta} - a\hat{\sigma}_{\hat{\vartheta}}, \hat{\vartheta} + b\hat{\sigma}_{\hat{\vartheta}})$ that has a specified chance $1 - \alpha$ of containing the parameter, and where the positive numbers $a$ and $b$ defining the lower and upper bounds of the interval depend on $\alpha$. The smaller $\alpha$ is, the larger the values of $a$ and $b$ can be for the statement to hold. A common value for $\alpha$ is 0.05 (i.e. 5%), which gives a 95% confidence interval. However, we could set $\alpha = 0.2$, which would give a narrower 80% confidence interval. Often $a$ and $b$ are equal, but we have given them distinct symbols for full generality.

We can express the definition in terms of a probability statement as follows:

$$P\left(\hat{\vartheta} - a\hat{\sigma}_{\hat{\vartheta}} < \vartheta < \hat{\vartheta} + b\hat{\sigma}_{\hat{\vartheta}}\right) = 1 - \alpha \tag{1}$$

In this probability statement, the upper and lower bounds of the interval are random variables, since they are based on the estimators and the estimated standard error, which are themselves random variables derived from the sample.

**Expression in terms of random variable in fixed interval** We can rearrange the definition of the confidence interval in terms of a standardised variable $(\hat{\vartheta} - \vartheta)/\hat{\sigma}_{\hat{\vartheta}}$:

$$P\left(-b < \frac{\hat{\vartheta} - \vartheta}{\hat{\sigma}_{\hat{\vartheta}}} < a\right) = 1 - \alpha \tag{2}$$

Because this standardised variable is derived from the sample, it fits our definition of a statistic. Furthermore, it is composed of *two* statistics, the estimator $\hat{\vartheta}$ and the estimated standard error $\hat{\sigma}_{\hat{\vartheta}}$.

# 3 Video: Method of estimating confidence interval of the mean of a large sample

**Methods of estimating confidence intervals** There are two main methods of estimating confidence intervals:

1. Under some assumptions about the distribution of the data $X_i$ and the number of samples $n$ we can derive the distribution of $(\hat{\vartheta} - \vartheta)/\hat{\sigma}_{\hat{\vartheta}}$, which will then tell us the values of $-b$ and $a$ at the $100\alpha/2$th centile and the $100(1 - \alpha/2)$th centiles.

2. More generally we can use a type of statistical simulation called a bootstrap estimator to derive the confidence interval.

We'll demonstrate the first approach by continuing with our simplified example of a normal distribution with known parameters. In the following section we'll then cover the bootstrap estimator.

**Example: confidence interval for the mean of a normal distribution with known variance** In the example of sampling from a normal distribution introduced in the last video, we *know* the population variance $\sigma$, and by definition, the standard estimate of the mean is $\hat{\sigma}_{\hat{\vartheta}} = \sigma/\sqrt{n}$. Because the population variance $\sigma$ is known, it's not a random variable, and therefore the SEM $\hat{\sigma}_{\hat{\vartheta}}$ isn't a random variable either. The standardised variable in Equation 2 is therefore

$$\frac{\hat{\vartheta} - \vartheta}{\hat{\sigma}_{\hat{\vartheta}}} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \tag{3}$$

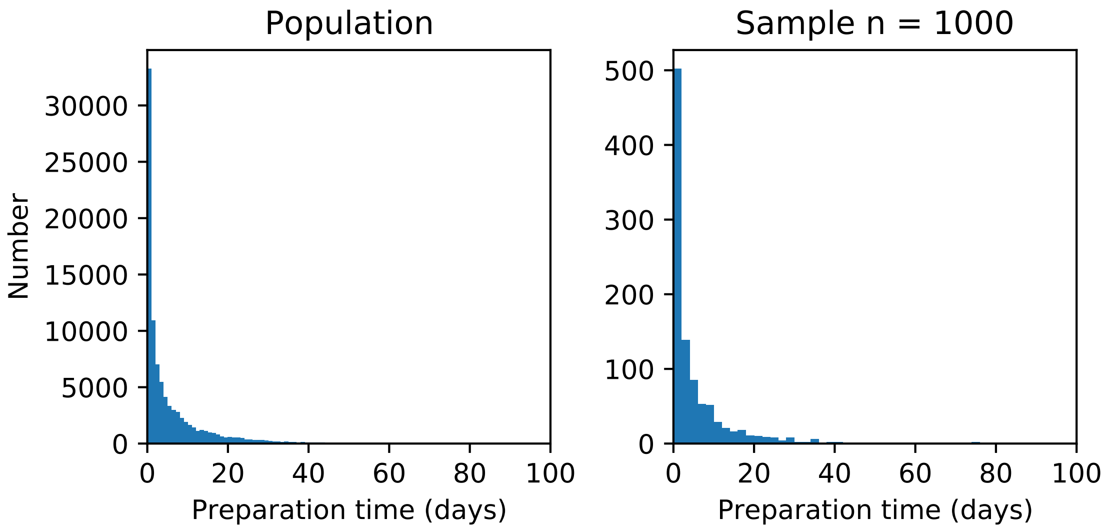

Figure 4: Distribution of time from making a reservation to the reservation time ("preparation time" ) in restaurants using the "air" booking system in Japan in the period January 2016–April 2017.

and only contains one random variable, $\overline{X}$. This makes it quite easy to deal with, since we know that this distribution is a standard normal distribution, so we can define a 95% confidence interval by setting $a$ and $b$ to be values at which the cumulative distribution function (cdf) is equal to $2.5\%(= \alpha/2)$ and $97.5\%(= 1 - \alpha/2)$. In this case the values $a = b = 1.96$ satisfy these conditions. Generally we set $a$ and $b$ symmetrically, so that there is equal weight in the "tails" of the distribution (Figure 2).

**Confidence intervals for the mean of a large sample** The central limit theorem states that the distribution of the sample mean of a "large" sample from any distribution should be normal. How large the sample needs to be depends on the distribution, but Figure 1 demonstrates that sample means of $n = 100$ samples from an exponential distribution already appear to fairly normally distributed with the SEM as predicted to be the standard deviation of the exponential distribution divided by $\sqrt{n}$. This means that we can use the procedure above to find confidence intervals.

**Confidence intervals for the mean of an empirical distribution** Up until now, we have considering estimating parameters from theoretical probability distributions, such as the normal distribution or the exponential distribution. We'll now consider how we can estimate the parameters of an **empirical distribution**, i.e. real-world data, from a sample of that distribution.

As an example, we will take the population of times between making a reservation and the time of the reservation itself in Japanese restaurants using the "air" booking system. The full population contains 92378 times (Figure 4, left) and we've created a random sample of 1000 of these times (Figure 4, right). In real life, if we had the full set of data, there would not be any point in creating this random sample of times, but we do so here to demonstrate how well we can estimate confidence intervals. From now on imagine that the sample of 1000 times is all that we have available to us. It's important to notice that the distribution of the sample resembles the population distribution, even though it is rougher.

Table 1 shows the summary statistics for the population and the sample. We can see that the estimates for the mean, standard deviation and centiles from the sample are all similar to the true population values. From the table we can see that the population mean is $\mu = 8.30$ days and the sample mean is $\overline{x} = 8.06$ days. The sample mean would be different if we'd happened to have taken a random different

Table 1: Summary statistics of population and sample of preparation times, generated by the pandas `describe` function.

|  | Population | Sample |
|---|---:|---:|
| count | 92378.00 | 1000.00 |
| mean | 8.30 | 8.06 |
| std | 25.65 | 27.72 |
| min | 0.00 | 0.00 |
| 25% | 0.21 | 0.17 |
| 50% | 2.08 | 1.96 |
| 75% | 7.88 | 6.92 |
| max | 393.12 | 364.96 |

sample.

From the summary statistics from the sample, we have $\overline{x} = 8.06$ days and the standard deviation $s = 27.72$ days. Our estimator for the mean is $\hat{\vartheta} = \overline{x} = 8.06$ days. Our estimator for the standard error in the mean is $\hat{\sigma}_{\hat{\vartheta}} = s/\sqrt{n} = 27.72/\sqrt{1000} = 0.88$ days. The 95% confidence interval for the mean in days is therefore $(\hat{\vartheta} - 1.96\hat{\sigma}_{\hat{\vartheta}}, \hat{\vartheta} + 1.96\hat{\sigma}_{\hat{\vartheta}}) = (6.34, 9.78)$.

**Reporting confidence intervals** When reading scientific papers, there are various ways of reporting confidence intervals:

- M=8.06, CI=6.34–9.78. Here "M" stands for mean and "CI" stands for confidence interval.
- 8.06 ± 1.72 (95% confidence interval)
- 8.06 ± 0.88 (± 1 SEM). This is a 68% confidence interval, though the confidence interval isn't specified in terms of area under the curve.
- 8.06 ± 1.76 (± 2 SEM).

# 4 Video: Bootstrap estimation of confidence intervals

**Principle of a bootstrap estimator** We want to estimate the standard error in an estimator. In the topic on sampling, we have already seen what happens when we sample a mean repeatedly from a theoretical distribution, and that this can give us a measure of the standard error of the estimator.

The name "bootstrap estimator" arises because it appears to do something physically impossible, such as "pulling ourselves up by our own bootstraps". (Equivalently we could pull ourselves up by our pigtail, Figure 5).

In a bootstrap estimator we treat the sample that we have available as a population, and resample from it to give the sampling distribution of the estimator. From the sampling distribution we can compute the standard error of the estimator. It feels as though we shouldn't be able to treat the sample as a population, but it works because if we have a large enough sample, the distribution of the sample will resemble the population itself.

Figure 5: Bootstrapping: Baron Münchhausen pulls himself and his horse out of a swamp by his pigtail. Public domain image from Wikipedia's article on bootstrapping.

**Bootstrap procedure for finding a confidence interval for the mean** We will start with a large sample $n$ from the data, which has a mean $\overline{x}$. By large, we mean large enough that the sample resembles the population distribution. Of course, this is not possible to know exactly, so the larger the better. We decide to take $B$ bootstrap samples. Common numbers are 1000 or 5000, or 10000. More samples are generally better, but bootstrapping can be computationally expensive, and fewer samples can also give reasonable results.

Here is the procedure:

- For $j$ in $1, \ldots, B$
  - Take sample $x^*$ of size $n$ from the sample *with replacement*
  - Compute the sample mean of the new sample $\overline{x_j^*}$
- To compute the bootstrap confidence interval, we find the centiles of the distribution at $100\alpha/2$ and $100(1 - \alpha/2)$. We can do this by arranging the sample means $\overline{x_j^*}$ in order from lowest to highest, and pick $\overline{x_j^*}$ at $k = \alpha(B+1)/2$ to be the lower end of the CI and pick $\overline{x_j^*}$ at $k = B - \alpha(B+1)/2$ to be the upper end of the CI.
- We can also compute the bootstrap estimator of the variance of the mean:

$$s^2_{\text{boot}} = \frac{\sum_{j=1}^{B}(\overline{x_j^*} - \overline{x})^2}{B - 1}$$

The advantages of the bootstrap procedure are that we can use it for any estimator, e.g. the median, and that we do not need to make any assumptions about the distribution of the estimator.

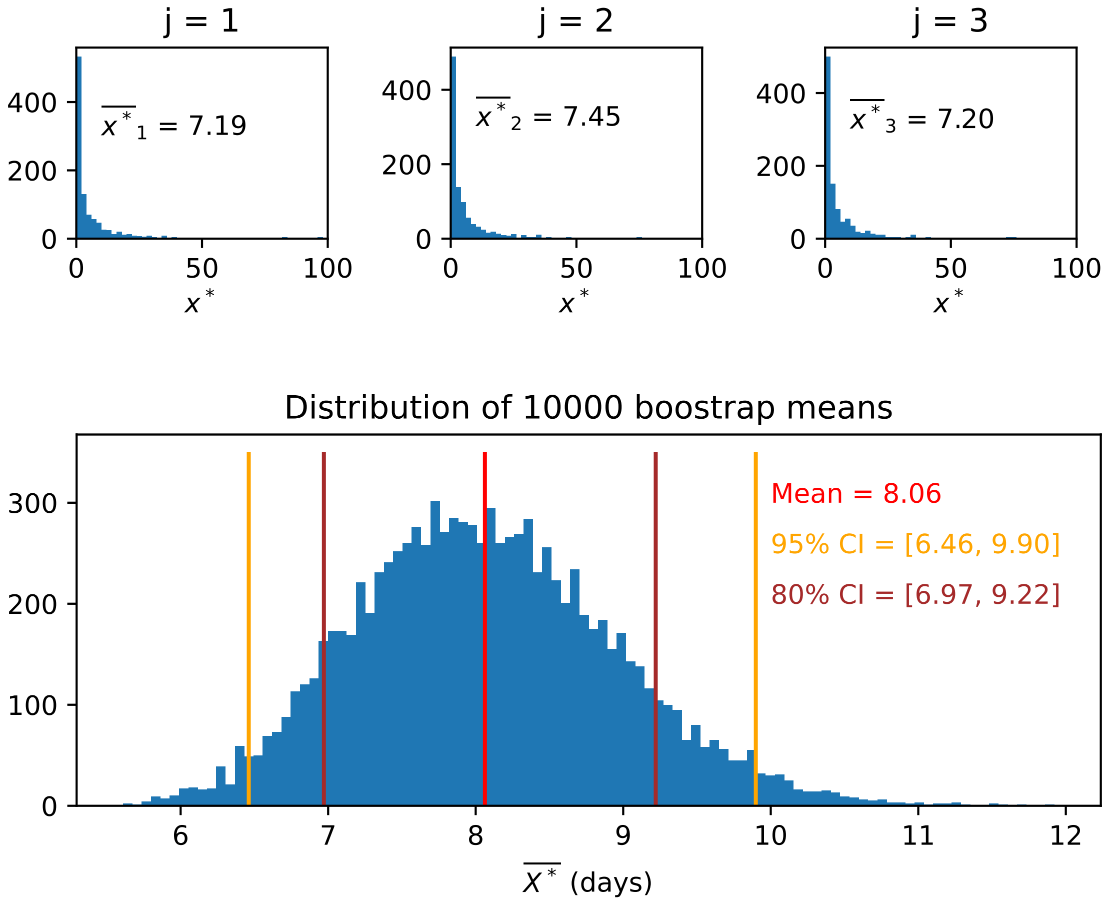

Figure 6: Demonstration of bootstrap mean applied to restaurant reservation time data (Figure 4). The top row shows the distributions obtained from the first 3 of 10000 bootstrap samples. Although the distributions are similar to each other, they are not exactly the same, and the sample mean of each is different. The bottom figure is the distribution of all 10000 of these bootstrap sample means. The mean of the original sample is shown, as is the 95% and 80% confidence intervals.

**Example of bootstrap estimator applied to mean** We'll now apply the bootstrap estimator to give us a confidence interval for the mean (Figure 6). For each of our 10,000 bootstrap samples, we'll resample 1000 samples *with replacement* from our sample of 1000. Each of these samples will be a distribution (top row of Figure 6), from which we can compute the 10,000 bootstrap means. Then we'll plot the distribution of the bootstrap means (bottom row of Figure 6) and find the 95% and 80% confidence intervals. In this case we can see both the 95% and 80% confidence intervals contain the *population* mean (8.30, Table 1). However, if we replicate the experiment with a different initial random sample of 1000, in around 5% of cases we should expect that the 95% confidence interval does not contain the mean.

We'll leave this as a lab exercise for you to implement, though you will find that you get different answers for the confidence intervals, depending on the state of the random number generator.

**Comparison of bootstrap confidence intervals with normal approximation** The 95% confidence interval obtained via the bootstrap procedure is (6.46, 9.90) days, which is very similar to the confidence interval obtained by the normal approximation, (6.34, 9.78) days. The bootstrap interval is slightly shifted to the right, suggesting that the normal approximation is quite accurate at a sample size of $n = 1000$.

**General formulation of bootstrap estimator** A great advantage of the bootstrap is that we can easily apply it to statistics other than the mean. Here is the general procedure for estimating the confidence interval of a generic estimator $\hat{\vartheta}$:

- For $j$ in $1 \dots B$
  - Take sample $x^*$ of size $n$ from the sample *with replacement*
  - Compute the sample statistic of the new sample $\hat{\vartheta}^*_j$
- Then compute the bootstrap estimator of the variance of the statistic:

$$s^2_{\text{boot}} = \frac{\sum_{j=1}^{B} (\hat{\vartheta}^*_j - \hat{\vartheta})^2}{B - 1}$$

- To compute the bootstrap confidence interval, we find the centiles of the distribution at $100\alpha/2$ and $100(1 - \alpha/2)$.

This procedure works well for measures of centrality such as the median, and for the variance. It doesn't work so well for statistics of extremes of the distribution, such as the maximum or minimum.

# 5 Video: Interpretation of confidence intervals

**Interpretation of confidence intervals** Although we have only computed confidence intervals in a simple artificial example, we are already at a stage where we can consider how to interpret confidence intervals. From Equation 1 we can see that confidence intervals are a random interval – whenever we take a new sample, we will end up with a new interval, as illustrated in Figure 1. The interpretation (according to the frequentist interpretation of statistics) is that if we performed a long run of experiments (i.e. repeatedly took samples) the parameter (the mean in this case) would be in around 95% of the confidence intervals.

**How big should a confidence interval be?** Should we choose the 95% confidence interval or the 80% confidence interval? The answer to this question depends on the problem. For example, suppose we have a machine that makes tens of thousands of ball bearings for aircraft jet engines every day. Each ball bearing needs to have a diameter of $2 \pm 0.0001$mm for the engine to work safely. We measure the diameter of a sample of the ball bearings every day. Because this is a safety-critical application, we need to have high confidence (say 99.999%) that the ball bearings are in the range $2 \pm 0.0001$mm. This might require a large sample size, but it's worthwhile because the consequences of getting it wrong could be catastrophic.

On the other hand, suppose we are estimating the number of red squirrels in a population so that we know how much red-squirrel friendly food to put out for them over winter. We might want to leave out a bit more than we expect they need, we're happy to accept a 10% chance that the true number of squirrels might be greater than the upper end of a confidence interval, so we compute the 80% confidence interval, and put out enough food for the number of squirrels at the upper end of the interval. There's a 10% chance that we might not be providing for enough squirrels, but it's not as catastrophic as in the aircraft situation (depending on how much you value red squirrels compared to humans).

**Upper and lower confidence bounds** In this case, we're not worried about our estimate being too low, so we only need to compute the upper confidence bound – we would quote a mean number of squirrels and an upper limit.

**Aside: Capture-recapture** Suppose we want to estimate the number of squirrels $N$ in a population. We can do this with a clever method called capture-recapture:

1. Capture $n$ of the squirrels, tag them so that they can be identified if caught again, then release them.

2. Wait for the squirrels to move around.

3. Recapture $K$ of the squirrels and record the number $k$ of these recaptured squirrels that have tags.

4. The estimator of the number of squirrels in the population is

$$\hat{N} = \frac{nK}{k}$$

This should work if the capturing and recapturing processes are random. If this is the case, the expected proportion of tagged squirrels in the whole population $n/N$ is equal to the proportion in the recaptured sample $k/K$, hence the estimator. It's possible to derive theoretical confidence intervals for this estimator.

# 6 Video: Confidence intervals on the mean for small samples

**Small samples** We'll now consider the distribution of the mean based on a sample of a "small" number (usually $n < 40$) of data that appears to be distributed normally and whose variance we are not given – we can only estimate it from the data.

For example, suppose we want to estimate the mean weight of a population of female squirrels from a sample of $n = 32$ squirrels (Wauters and Dhondt, 1989). The sample mean is $\overline{x} = 341.0$g and the estimated standard error in the mean is $\hat{\sigma}_{\overline{X}} = 3.9$g.

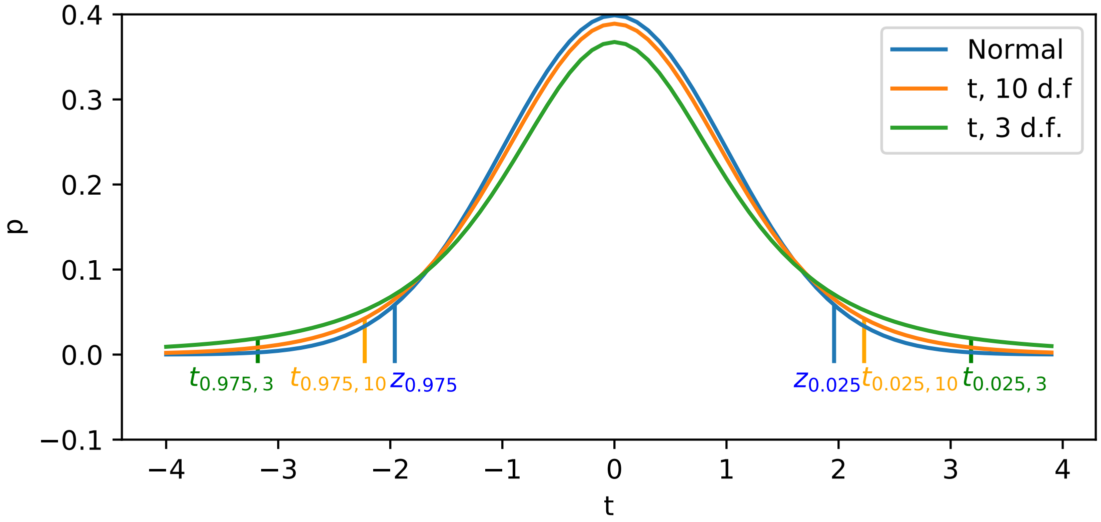

Figure 7: The $t$-distribution for 3 degrees of freedom and 10 degrees of freedom, with normal distribution for comparison. 2.5% $t$ critical values and $z$ critical values are shown.

We can imagine that if we had taken a different sample of $n = 32$ squirrels, we would have found both a different sample mean and a different estimate for the standard error in the mean. Thus, the standardised statistic $(\overline{X} - \mu)/\hat{\sigma}_{\overline{X}}$ itself contains *two* statistics, $\overline{X}$ and $\hat{\sigma}_{\overline{X}}$, which are, in general, random variables derived from the sample. As we are estimating the standard deviation, rather than knowing it, the normal approximation to the distribution of the mean begins to break down.

**The *t*-distribution** We could use the bootstrap estimator to estimate confidence intervals. However, in this special case, there is another option. There's a theorem that states that when $X$ is a random sample of size $n$ from a normal distribution with mean $\mu$, the random variable

$$T = \frac{\overline{X} - \mu}{\hat{\sigma}_{\overline{X}}}$$

is distributed as a ***t*-distribution** with $n - 1$ degrees of freedom, where the $t$-distribution with $\nu$ degrees of freedom has the probability density function depicted in Figure 7, which is given by the equation:

$$p_\nu(t) = \frac{1}{\sqrt{\pi\nu}} \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)} \frac{1}{(1 + t^2/\nu)^{(\nu+1)/2}}$$

where $\Gamma(x)$ is a gamma function. We will not prove this theorem here; in *Modern Mathematical Statistics with Applications* Section 6.4 there is the sketch of a proof.

The $t$-distribution is very similar in shape to the normal distribution: it is bell-shaped, symmetrical, and centred on 0. However, for small numbers of degrees of freedom, has longer tails. This means that the tails contain more of the weight of the distribution. We define the ***t* critical value** $t_{\alpha,\nu}$ as the value of $t$ in a $t$-distribution with $\nu$ degrees of freedom which has the area $\alpha$ under the curve to its right.

For small degrees of freedom, the $t$ critical values are considerably bigger than the $z$ critical values of the normal distribution (Figure 7). As the number of degrees of freedom increases, the $t$-distribution approaches a normal distribution. The distribution with 40 degrees of freedom (not shown in the figure) looks very similar to a normal distribution.

Table 2: Abbreviated table of critical values for $t$ and $z$ distributions.

| $\alpha$ | 0.100 | 0.050 | 0.025 | 0.010 | 0.005 | 0.001 |
|---|---|---|---|---|---|---|
| $\nu$ | | | | | | |
| 1 | 3.078 | 6.314 | 12.706 | 31.821 | 63.657 | 318.309 |
| 2 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | 22.327 |
| 3 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 10.215 |
| 4 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 7.173 |
| 5 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 5.893 |
| 6 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 | 5.208 |
| 7 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 | 4.785 |
| 8 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 | 4.501 |
| 9 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 | 4.297 |
| 10 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | 4.144 |
| 20 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 | 3.552 |
| 30 | 1.310 | 1.697 | 2.042 | 2.457 | 2.750 | 3.385 |
| 40 | 1.303 | 1.684 | 2.021 | 2.423 | 2.704 | 3.307 |
| $\infty$ | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | 3.090 |

**Looking up a *t* critical value (added after 2021/22 lectures)** To look up a $t$ critical value, you can use the python `scipy` package. For example to find $t_{0.025,10}$ you would use:

```
from scipy.stats import t
alpha = 0.025
nu = 10
t_cv = t(nu).isf(alpha)
print(t_cv)
```

You can also look up $t$ critical values and $z$ critical values in statistical tables, such as the ones in the appendices of *Modern Mathematical Statistics with Applications*. Table 2 shows an abbreviated example of such a table. Each row contains $t$ critical values for degree various levels of $\alpha$. The final row, with infinite number of degrees of freedom, is the $z$ critical values for these values of $\alpha$. The full tables include values for more degrees of freedom.

**Using the *t*-distribution to derive a confidence interval** The $100(1-\alpha)$ percent confidence interval around a mean $\overline{x}$ of a sample of $n$ values with estimated SEM $\hat{\sigma}_{\overline{X}}$ derived using a $t$-distribution is:

$$(\overline{x} - t_{\alpha/2,n-1}\hat{\sigma}_{\overline{X}}\ ,\ \overline{x} + t_{\alpha/2,n-1}\hat{\sigma}_{\overline{X}}) \tag{4}$$

Note that we have used the $t$ critical value $t_{\alpha/2,\nu}$. Here the number of degrees of freedom is one less than the sample size ($\nu = n-1$). Also, we have divided $\alpha$ by 2 because we are wanting upper and lower bounds to the confidence interval. It might be that we only need an upper bound, as we considered when we were estimating squirrel numbers earlier. In this case we would just quote $\overline{x} + t_{\alpha,n-1}\hat{\sigma}_{\overline{X}}$. This is still a $100(1-\alpha)$ confidence interval, since the interval from $-\infty$ to the upper bound contains $100(1-\alpha)$ of the area under the $t$-distribution.

To continue the squirrel example, suppose we want to find a 95% confidence interval for the weight. The 95% confidence interval implies $\alpha = 0.05$ and $\nu = n - 1 = 31$. We would then look up the $t_{0.025,31} = 2.040$ and substitute it into Equation 4 along with the sample mean and estimated SEM. and then use this to generate the confidence interval, which we could quote as $\hat{\mu} = 341.0 \pm 8.0$g

(95% confidence interval, $n = 32$). This is a bit wider than the interval we would obtain using the corresponding critical value of a normal distribution $z_{0.025} = 1.96$.

# References

Wauters, L. A. and Dhondt, A. A. (1989). ‘Variation in length and body weight of the red squirrel (*Sciurus vulgaris*) in two different habitats’. *Journal of Zoology* **217**:93–106