Online Resource

# Chapter 5: Uncertainty

## Central limit theorem

It is no exaggeration to say that without the Central Limit Theorem, there would be no statistical inference. Statistics can do its first task (describe patterns in the data) without the Central Limit Theorem, but it cannot do its second task of estimating uncertainty. Actually, like all maths, the Central Limit Theorem is a work in progress – it is regularly being improved.

The basic Central Limit Theorem is the reason why we can say that a sample can simultaneously give us both the best estimate of a population and an estimate of our uncertainty. This is the magic moment in statistics. One sample tells you what all the other samples could look like.

At its heart, the Central Limit Theorem is an extension of the very long-standing law of large numbers: the results of any uncertain procedure converge on the expected value with a large enough sample size. This is a guarantee of stability that nearly always applies.

The basic Central Limit Theorem states specifically that if we are estimating the mean of a population, then the distribution of sample estimates will converge on (i.e., become, when there are enough of them) a normal distribution. That sampling distribution has a mean that is the same as the mean of the population those samples come from. It has a standard deviation that is the standard deviation of the population itself, divided by the square root of the sample size.

This is quite remarkable. For all practical situations that we may encounter, it says that if we knew the mean and standard deviation of the population and we know the size of our sample, we can also know the sampling distribution. Since any sample gives us an estimate of the population mean and standard deviation, and we must know our sample size, it allows us to use the sample to find the standard error.

Equally remarkable is that all of this applies almost regardless of the shape of the population distribution. That distribution can be skewed, can have a high or very low kurtosis, but the sampling distribution will still be a normal distribution. There are some very rare situations where Central Limit Theorem doesn’t apply, but these have no practical importance. Specifically, any distribution that can produce values that are infinite is going to cause problems. The Cauchy distribution is one widely cited example, which you are unlikely to encounter.

In this next figure, we show some types of populations (left) and the distribution of sample means that they give rise to. The populations are:

1. Normal distribution

2. High skew

3. Very low kurtosis

4. Cauchy distribution

Notice how, even for the last, the sampling distributions are all similar. In theory, the Cauchy distribution does not obey the Central Limit Theorem, but in practice it comes very close.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Distribution | 10 samples | 100 samples | 10000 samples |
| 1 |  |  |  |  |
| 2 |  |  |  |  |
| 3 |  |  |  |  |
| 4 |  |  |  |  |

## Standard error

The standard error is the standard deviation of the sampling distribution. It is a measure of how variable sample effect sizes will be given a population and a design. That is their fundamental usefulness when we use statistics in psychology.

In this further commentary, we cannot resist adding in a little extra feature of them that is not widely known. This extra feature requires a little maths but even if you don’t follow maths, there is something here to be seen.

Suppose that we are interested in measuring the mean of a distribution and obtain a sample to get an estimated mean and standard error. For some reason, we want to change the scale we measured the distribution on. We’ll look at 3 cases and then explain it all:

1. Add a constant number. We have measured risk taking on a scale that goes from 0 to 5; someone else has measured it on a scale of 1 to 6. We want to compare.

It is easy to convert our mean value to theirs, by adding 1.

It is also easy to convert our standard error: do nothing.

If we add a constant to a scale, then any standard error isn’t changed.

2. Multiply by a constant number. We have measured risk taking on a scale of 0–5, someone else has measured it on a scale of 0–10. We want to compare.

It is easy to convert our mean value to theirs, by multiplying by 2.

It is also easy to convert our standard error: multiply by 2.

If we multiply a scale by a constant, then any standard error is also multiplied by that constant.

3. Convert scale to a logarithmic scale.

It is easy to convert our mean value to theirs, by taking the log of it.

It is also easy to convert our standard error: divide it by the mean value.

If replace a scale by its logarithm, then any standard error is divided by the mean.

Explanation (this bit is mathematical, sorry. Read on if you are interested!). The standard error can be treated as a gradient on the measurement scale. Technically it behaves as a partial derivative.

If we add a constant number to a scale, we don’t change the gradient of the scale, so we don’t have to change the standard error.

If we multiply a scale by a constant, we have also multiplied the gradient by that constant, so we multiply the standard error by the constant.

If we replace a scale by its logarithm, then the gradient is changed differently everywhere. Here, calculus makes life easy:

We have:



We can differentiate:



Treating the partial operators (*dx*) as standard errors (*se*(*x*)), we can then see that:



For the specific example:



Since





## Maximum likelihood

It will seem quite obvious to some readers that the sample mean or sample effect size is the best estimate of the population mean or population effect size. From a more rigorous, mathematical standpoint, this is something that requires proof. We will not supply the proof here, but instead explain how that proof works so that the non-mathematical reader can still get a sense of what it all means.

Think about this. I measure the IQ of 4 students and the scores I get are 101, 107, 96 & 104. The mean of these is 102. My sample mean is 102 and I am going to believe that this is the best estimate I have of the mean of the distribution of student IQs. Now we can ask what is the likelihood that a population of mean 102 (and sd = 15) would give me a single participant with an IQ of 101 (the first of my measured IQs). It will be some small number. We can do the same thing for each of the measured values.

From those individual likelihoods, we can calculate the overall likelihood of a distribution of mean 102 giving us the 4 values we started with. This is the product of the individual likelihoods. We could exhaustively try all possible populations and find the one that has the highest likelihood (the maximum likelihood).

However, the formula for a normal distribution allows us to do something more useful. The likelihood of getting a value of *x* from a normal distribution with mean *M* and standard deviation *S* is this:



The likelihood of getting 2 values, *x*1 and *x*2 is given by this:



or, more usefully, the log likelihood is:



and our task is to find out what value of *M* sets this quantity to its minimum.

Mathematically, we equate the derivative of this to zero, to find the value of *M* that gives the highest log likelihood. That leads to solve this equation:



From this we have 

or 

You will be able to see how to extend this to more than two values:



## Using Fisher *Z* to calculate confidence limits for normalized effect sizes

Confidence limits for some quantities (such as a mean) are easily calculated as being the mean itself plus and minus 1.96 times the standard error. For normalized effect sizes, the calculation is more complex.

The broad logic is that the sampling distribution and therefore the likelihood function for sample means is symmetric and normally distributed. For normalized effect sizes, they are not. But there exists a way of making them symmetric and normal, called the Fisher *Z*-transform. The Fisher *Z*-transform has the additional nice property that the standard deviation of the normal distribution it produces is determined solely by the sample size, *n*.

So, we take the estimated effect size and sample size, apply the Fisher *Z*-transform. Then we calculate the confidence limits in that transformed scale. Then we reverse the transform to bring our result back to the scale we want (the original normalized effect size scale).

**Step 1: Transform to *z***

We start by converting the normalized effect size, *r*, to a (yet another) new measure of effect-size (we will call it *z* here): this conversion is called the **Fisher *Z*-transformation**:

*z* = *a* tan *h*(*r*)

The transformation is useful because it produces sampling distributions that are very, very nearly normal. Even better, the standard error (the standard deviation of the sampling distribution) is very, very nearly:

se(*z*) = 1/sqrt(*n*–3)

where *n* = sample size.

So, if we have a normalized effect size of 0.3 and a sample size of 42, then we can do this:

*z* = *a* tan *h*(0.3) = 0.3095

se(*z*) = 1/sqrt(39) = 0.1601

**Step 2: Calculate confidence limits**

So, keeping to *z* for a minute, the 95% confidence limits are given by *z* plus or minus the se(*z*) times 1.96:

0.3095 – 1.96 \* 0.1601 to 0.3095 + 1.96 \* 0.1601

–0.0043 to 0.623

**Step 3: Transform back to *r***

Finally, we can convert from *z* back to *r*. The inverse transformation for this is:

*r* = tan*h*(*z*)

So, the 95% confidence interval becomes:

tan*h*(–0.0043) to tan*h*(0.623)

–0.004 to 0.55

You can now see that the two confidence limits are not at equal distances from the effect size of 0.3.