# Research Methods 2

# Week 6: Document 1

# How good is my estimate? The meaning of Standard Errors

### *Recapitulation*

In Document 1 of Week 3, you met a sample that had been obtained from the population of heights of 5-year-old boys. If we assume that this variable has a Normal distribution (an assumption that is, in fact, entirely reasonable) then it will have a population mean, $m$ For many purposes, such as assessing the height of a five-year-old in the follow-up clinic, the value of $m$ is of interest. As this is a population parameter we will never know its true value, because we will never have a complete enumeration of the population. As we learnt in Week 4 we have to content ourselves with whatever information about $m$ that can be gleaned from a random sample drawn from the population.

As was learnt in Week 4, the natural way to estimate $m$ is to compute the mean, $m$, of the sample and say that this value is our *estimate* of $m$ The mean of the sample of 99 heights is 108.34 cm. Had we only measured the heights of the first ten boys in this sample the value obtained would have been 107.77 cm. If the first 20 boys had been measured then the value would have been 107.68 cm. The three means are summarised below.

| Sample size | Sample mean (cm.) |
|-------------|-------------------|
| 10 | 107.77 |
| 20 | 107.68 |
| 99 | 108.34 |

Each of these sample means is a legitimate estimate of $m$ Indeed, even a single height measurement, such as the first measurement, 117.9 cm, is a *legitimate* estimate of $m$ If we only had a single measurement and we were asked what was our best estimate of $m$ then we could give this one value (it can be thought of as the mean of a sample of size 1). However, in this instance we have a range of alternatives, so which one should we use to estimate $m$ and, more importantly, why?

Most investigators would intuitively say that the mean of the sample of size 99 was the best one. This intuition would be based on the notion that by using data from 99 boys the sample mean was based on more information and so must be 'better' in some sense. In this instance intuition turns out to be a reliable guide and the basic idea that a sample mean based on a larger sample is 'better' than one based on a smaller sample is correct. However, it is useful to be more precise about what we mean by 'better'.

It will also be helpful to try to be more quantitative about the relative merits of samples of different sizes. The larger sample provides a 'better' estimate only in a narrow statistical sense. It may well be that a larger sample is more expensive to collect and it is likely to be important to know *how much* improvement will be obtained for a given increase in sample size.

### Principles behind the Standard Error

The mean of a larger sample will tend to be closer to the mean of the population than the mean of a smaller sample. This observation is an important step to a more precise understanding of why the mean of a larger sample is 'better' than the mean of a smaller sample. It is more precise because we are now focussing on a specific quantity, the difference between the population mean and the sample mean. However, it remains unsatisfactory from a practical point of view because it has not been made clear what is meant by the phrase '…will *tend* to be closer to …' used above.

Of course, as the population mean is unknown we cannot evaluate the difference between *m* and any particular sample mean. So what basis is there for the assertion that the mean of a larger sample 'tends' to be closer to *m* and does this allows us to be more specific about what 'tends' means?

This is most readily approached by considering the following. In practice we have a single sample from a population and must base our inferences on this one sample. However, suppose for the moment that we have available not one but many samples from the same population. For each sample we can compute the sample mean and then we can look at how the *sample means* are distributed, for example by drawing the histogram of these means. If the sample mean is a useful estimator of the population mean then the distribution of sample means will cluster about the population mean. Moreover, if the sample mean 'tends' to be closer to the population mean for larger samples then the histogram of sample means for samples of size 1000, say, will be more tightly clustered around the central value, whatever it is, then the distribution of means of samples of size 10.

This can be demonstrated using the technique you have met in the exercises in weeks 4 and 5, namely asking Minitab to generate data from specific populations. Minitab can be asked to generate a sample of size 10 from a Normal population with population mean of 108 cm and population SD of 4.7 cm. It can be asked to do this repeatedly. If this is done 500 times, and the mean of each sample is computed then you will have a collection of 500 *means*. A histogram of 500 such numbers is shown in figure 1. As a comparison, figure 2 shows a histogram of 500 individual heights values from the same population.

There are three things to note about figures 1 and 2.

1. Like the individual heights, the sample means have a Normal distribution.
2. Like the individual heights, the mean of the distribution of sample means seems to be close to 108 cm (evaluating the mean of the 500 means gives 107.96 cm).
3. While the heights are spread between about 95 cm and over 120 cm, the spread of the sample means is much less, being between about 104 and 114 cm.
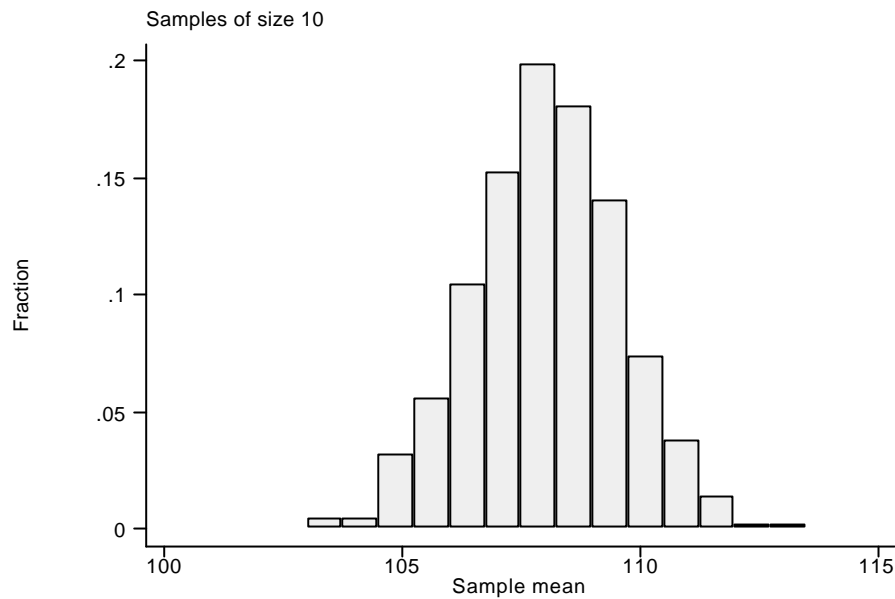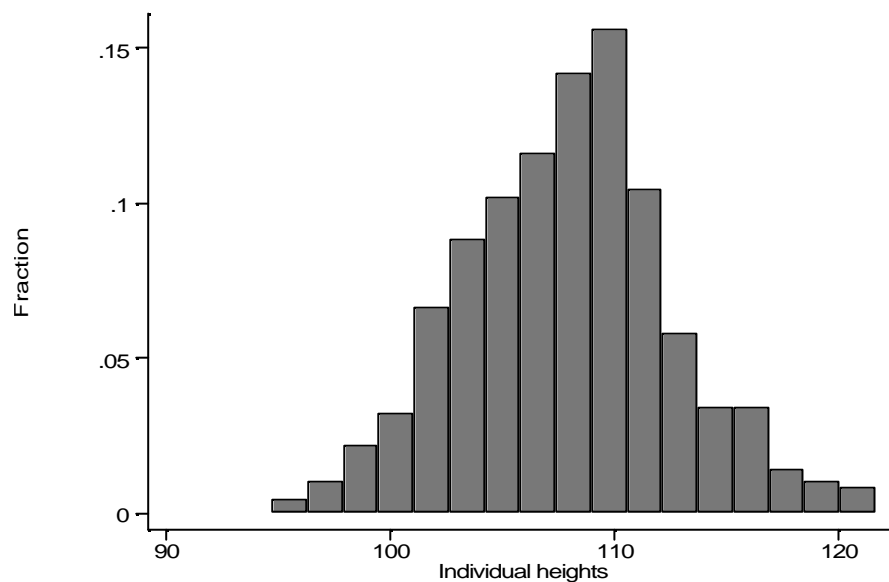
Samples of size 10



Figure 1



Figure 2

The distribution of the sample means is very similar to that of the population – it appears to be Normal, with a mean which is the same as that of the population, but which is more tightly clustered around the mean. We can measure spread and computing the sample SD of the 500 sample means of figure 1 gives 1.50 cm, whereas the corresponding quantity for the individual heights is 4.66 cm[†], confirming that the distribution of sample means is less spread out than the distribution of individual heights.

This is why we take means of samples – while both the mean of a sample and an individual height will cluster around the population mean, the sample mean will stray less far from the population mean. This effect gets more marked as the size of the sample increases. Figure 3 shows the result of repeating the exercise for figure 1 but

[†] note that this is what we would expect for a sample from a population known to have an SD of 4.7 cm

3

now with samples of size 1000 rather than 10, and this is much more concentrated about its central value than that in figure 1, with no mean smaller than 107.4 and none greater than 108.6 cm. Again the distribution is Normal and the mean of these means is 107.99 cm, in line with previous values. The SD of these means is 0.15 cm, indicating that the mean of a sample of size 1000 will not stray very far at all from the population mean.

The key quantity we have used to summarise the differences between these histograms is the standard deviation of the distribution of sample means. This is such an important quantity in statistical inference that it is given its own name: it is called the *standard error* (*SE*) or sometimes the *standard error of the mean* (*SEM*). Some points about nomenclature are given in the information bar at the end of the document.
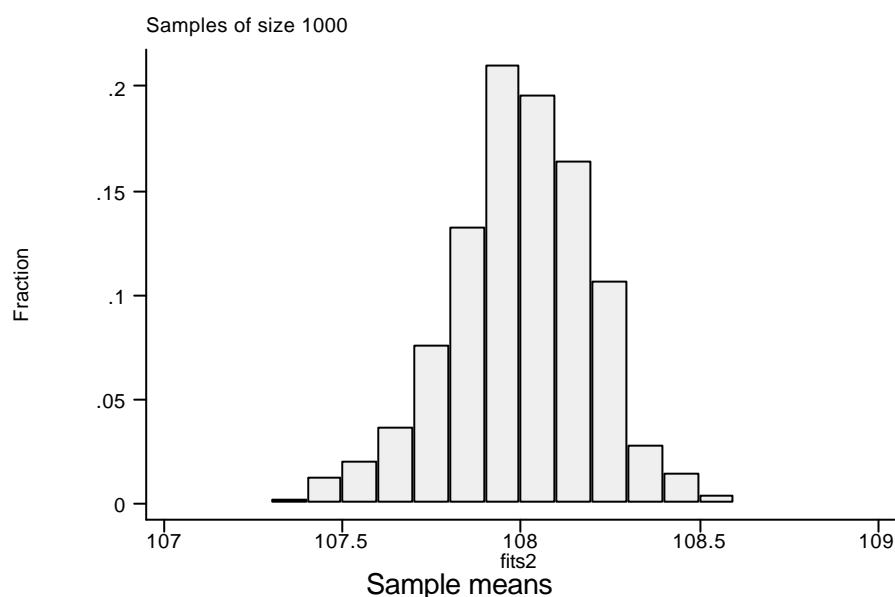


Figure 3

In practice we have just a single sample, for example one obtained in the course of some investigation. We can compute its mean and the foregoing analysis shows that it will tend to be closer to the population mean if it is based on a larger sample.

So far we have been relative in our claims for the effect of sample sizes. We have said that means from larger samples are less dispersed than those from smaller samples. Can we be more quantitative about the mean of a single sample of a given size? Can we say something about how far a sample mean is likely to stray from the population mean?

The answer to both these questions is yes, but both answers require us to evaluate the standard error and this presents a problem. We were only able to estimate SEs in the above examples because we used the unrealistic device of assuming that we had access to arbitrarily many samples, not to just one sample. Fortunately there is a way round this difficulty and this will now be described.

### *Calculating the Standard Error from a single sample*

The solution to the problem is to use a theoretical result of great importance in statistics. This result states that:

> *If a variable has population standard deviation $s$ then the standard error of the mean of a sample of size n from this population is $\dfrac{s}{\sqrt{n}}$.*

In practice we do not know $s$ so we must estimate it using the sample SD, *s*, and hence the estimate of the SE of the mean of a sample of size *n* is $s/\sqrt{n}$. All of these quantities are available from a single sample.

So, e.g. the sample of 500 heights shown in figure 2 has SD 4.66 cm and the SE of the mean of the sample is computed as $4.66/\sqrt{500} = 0.21$ cm.

As another example , generate a sample of size 10 from the same Normal population. This has a sample SD of 3.89 cm, giving an estimate of the SE of the mean of this sample as $3.89/\sqrt{10} = 1.23$ cm. This is quite close to the value of 1.50 cm we obtained as the SD of the sample of means plotted in figure 1. {The two should not be expected to be exactly equal because of sampling error}.

As a third example a sample of size 1000 is generated from the Normal population underlying figure 3. This has SD 4.78 cm, giving an SE for the mean of that sample of $4.78/\sqrt{1000} = 0.15$ cm, which is (to two d.p.s) the same as that obtained as the SD of the 500 sample means in figure 3. {The agreement is much closer than in the previous example because sampling error is much reduced in larger samples}

The formula shown above is worth thinking about for a moment or two (not least because it is one of very few you will encounter in this course!).

- It shows that as the sample size gets bigger, the SE must get smaller (the *n* in the denominator gets large, so the ratio gets small).
- Indeed, the SE can be made as small as we like provided we collect a large enough sample (and provided it is collected appropriately).
- The square root sign in the denominator means that the SE does not get small quite as quickly as we would like when we increase the sample size. For example, in order to halve the SE we need to quadruple the size of our sample.

---

## Nomenclature

It might be asked, when the SE is simply the standard deviation of the distribution of sample means, why a term other than standard deviation is necessary. While a reasonably cogent argument can be made for abandoning the term *standard error* there are at least three reasons why *standard error* is used and should be retained.

i)      While the standard error is a form of standard deviation, it is a very special form. It is the standard deviation of a hypothetical distribution that is never actually observed. Values for the standard error usually need to be computed using a theoretically derived formula.

ii) Standard errors and standard deviations are put to different uses.  Standard deviations are descriptive tools that indicate the dispersion in a sample.  A standard error is an inferential tool, which measures the precision of estimates of population parameters.

iii) *Standard error* is the term that has been widely used for the standard deviation of the distribution of sample means and to change nomenclature now would cause great confusion.

### *Practical use of the Standard Error*

So far we have learnt what the Standard Error of the mean is, and how it can be calculated.  What we have not covered is how it can be used in practice to improve our inferences.  This will be covered next week.