Clinical Biostatistics

Estimation

Martin Bland

Professor of Health Statistics

University of York

http://martinbland.co.uk/

The Normal distribution

Many statistical methods are only valid if we can assume that our data follow a distribution of a particular type, the Normal distribution. This is a continuous, symmetrical, unimodal distribution described by a mathematical equation, which we shall omit.
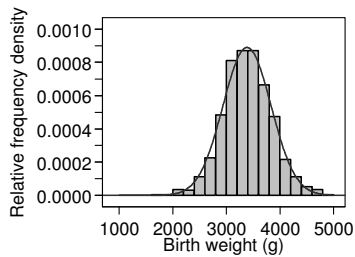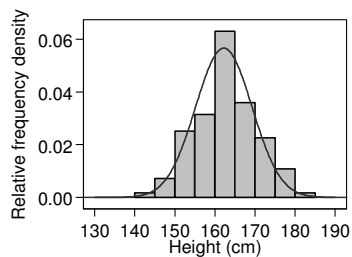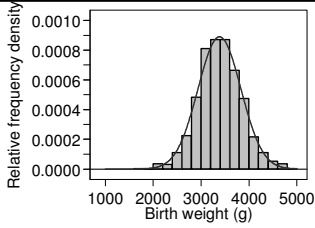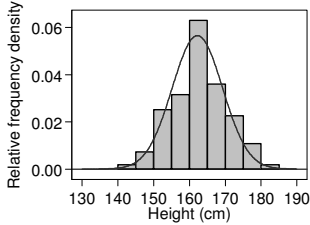

The Normal distribution

Many statistical methods are only valid if we can assume that our data follow a distribution of a particular type, the Normal distribution. This is a continuous, symmetrical, unimodal distribution described by a mathematical equation, which we shall omit.

Mean = 3384 g
Variance = 201164 g2
SD = 449 g

Mean = 162.2 cm
Variance = 49.7 cm2
SD = 7.1 cm

The Normal distribution is not just one distribution, but a family of distributions.

The particular member of the family that we have is defined by two numbers, called parameters.

Parameter is a mathematical term meaning a number which defines a member of a class of things.

The parameters of a Normal distribution happen to be equal to the mean and variance.

These two numbers tell us which member of the Normal family we have.

The parameters (mean and variance) of a Normal distribution happen to be equal to the mean and variance.

These two numbers tell us which member of the Normal family we have.
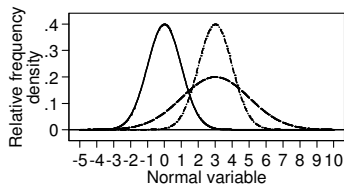
Mean=0, variance=1 is called the **Standard Normal distribution**.

— Mn=0, Var=1 --- Mn=3, Var=1
–– Mn=3, Var=4

The parameters (mean and variance) of a Normal distribution happen to be equal to the mean and variance.

These two numbers tell us which member of the Normal family we have.
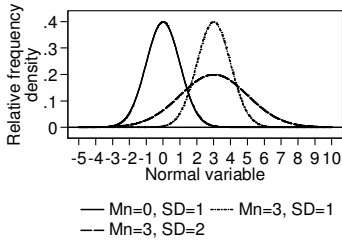

The distributions are the same in terms of standard deviations from the mean.

The Normal distribution is important for two reasons.

1. Many natural variables follow it quite closely, certainly sufficiently closely for us to use statistical methods which require this.

2. Even when we have a variable which does not follow a Normal distribution, if we the take the mean of a sample of observations, such means will follow a Normal distribution.

There is no simple formula linking the variable and the area under the curve.

Hence we cannot find a formula to calculate the frequency between two chosen values of the variable, nor the value which would be exceeded for a given proportion of observations.

Numerical methods for calculating these things with acceptable accuracy were used to produce extensive tables of the Normal distribution.

These numerical methods for calculating Normal frequencies have been built into statistical computer programs and computers can estimate them whenever they are needed.

Two numbers from tables of the Normal distribution:

1. we expect 68% of observations to lie within one standard deviation from the mean,

2. we expect 95% of observations to lie within 1.96 standard deviations from the mean.

This is true for all Normal distributions, whatever the mean, variance, and standard deviation.

Sampling

Most research data come from subjects we think of as samples drawn from a larger population.

The sample tells us something about the population.

E.g. blood sample to estimate blood glucose.

> One drop of blood to represent all the blood in the body.
>
> Three measurements: 6.0, 5.9, and 5.8.
>
> Which is correct?
>
> None are; they are all estimates of the same quantity.
>
> We do not know which is closest.

Sampling

Most research data come from subjects we think of as samples drawn from a larger population.

The sample tells us something about the population.

E.g. three trials of elastic versus inelastic multilayer compression bandaging for venous ulcers.

> Differences in percentage of patients achieving complete healing: 13, 25, and 20 percentage points in favour of elastic bandaging (Fletcher *et al.*, 1997).
>
> All estimates of the advantage to elastic bandaging.
>
> Natural random variation between samples.

Fletcher A, Cullum N, Sheldon TA. (1997) A systematic review of compression treatment for venous leg ulcers. *BMJ* **315**: 576-580.

Sampling distributions

The estimates from all the possible samples drawn in the same way have a distribution.

We call this the **sampling distribution**.

Sampling distributions

Example: an ordinary, six-sided die.

Rolls of the die will produce a score that will act as the original measurements that we make.

Rolling a die can produce: 1, 2, 3, 4, 5, or 6.

Each will happen in the same proportion of rolls, 1/6.

Proportions of rolls which would give each possible score, all equal to 1/6 or 0.167.

Sampling distributions

Example: an ordinary, six-sided die.

Rolls of the die will produce a score that will act as the original measurements that we make.

Rolling a die can produce: 1, 2, 3, 4, 5, or 6.

Each will happen in the same proportion of rolls, 1/6.

Let us put ourselves in the position of not knowing what the average score would be.

We take a sample of dice rolls to enable us to estimate the mean.

Roll two dice and calculate the mean (or average) of the two scores to provide the estimate of the mean for the population of all dice rolls.


The mean is 3.5, as before, but the standard deviation is not the same. It is 1.21, which is less than the 1.71 for a single die.

The more dice we roll, the more accurate we might expect our sample mean to be as an estimate for the population mean.
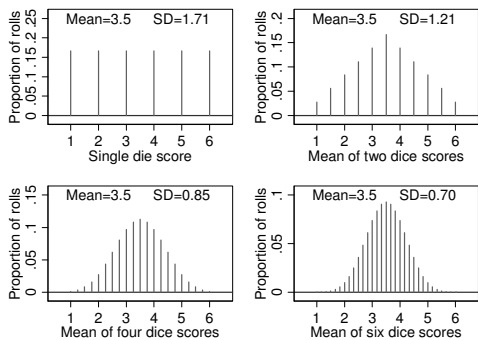
1. The mean of the distribution is always the same, 3.5. When we take several observations and average them, the mean of many such averages is the same as the mean for the distribution of single observations.

2. The distributions become less variable as the number of dice increases. There are much smaller proportions of rolls producing means close to 1.0 or 6.0 and the standard deviations get smaller.

3. The shape of the distribution changes as the number of dice increases. Becomes similar to the Normal distribution curve.

Superimposed Normal distribution of same mean and variance:
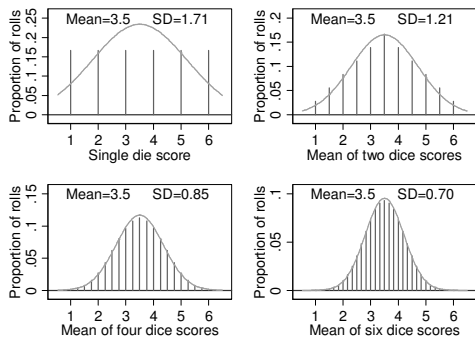
For almost any observations we care to make, if we take a sample of several observations and find their mean, whatever the distribution of the original variable was like:

1. Such sample means will have a distribution which has the same mean as the whole population.

2. These sample means will have a smaller standard deviation than the whole population, and the bigger we make the sample the smaller the standard deviations of the sample means will be.

3. The shape of the distribution gets closer to a Normal distribution as the number in the sample increases.

Any statistic which is calculated from a sample, such as a mean, proportion, median, or standard deviation, will have a sampling distribution.

Standard error

We can use standard error to describe how good our estimate is.

The standard error comes from the sampling distribution.

The standard deviation of the sampling distribution tells us how good our sample statistic is as an estimate of the population value.

We call this standard deviation the **standard error** of the estimate.

Hence the standard error of the mean of six dice scores is 0.70.

Standard error

People find the terms 'standard error' and 'standard deviation' confusing.

This is not surprising, as a standard error is a type of standard deviation.

We use the term 'standard deviation' when we are talking about distributions, either of a sample or a population.

We use the term 'standard error' when we are talking about an estimate found from a sample.

Standard error

In the dice example, we know exactly what the distribution of the original variable is because it comes from a very simple randomising device (the die).

In most practical situations, we do not know this.

Elastic vs. inelastic bandages, first trial:

Difference between the two percentages whose ulcers were completely healed = 13.

This is an estimate of the difference in the population of all venous ulcer patients.

What is its standard error?

Standard error

Elastic vs. inelastic bandages, first trial.

Difference between the two percentages whose ulcers were completely healed = 13.

What is its standard error?

There were 31 healed out of 49 patients in the elastic bandage group and 26 out of 52 in the inelastic bandage group.

On theoretical grounds, we know the family of distributions which the difference will be from. In this case, it is approximately Normal.

Standard error

Elastic vs. inelastic bandages, first trial.

Which member of the Normal distribution family we have depends on the proportion of the whole patient population who would heal if given elastic bandages and the proportion of the whole patient population who would heal if given inelastic bandages.

We do not know these and there is no way that we could know them.

However, we can estimate them from the data, using the sample percentages 63% (31/49) and 50% (26/52).

We then calculate what the standard error would be if the unknown population percentages were, in fact, equal to these sample percentages.

Standard error

Elastic vs. inelastic bandages, first trial.

We calculate what the standard error would be if the unknown population percentages were, in fact, equal to these sample percentages.

This estimated standard error can then be used to assess the precision or the estimate of the difference.

Now things get very confusing, because we call this estimated standard error the 'standard error' also.

For our difference between the two percentages with healed ulcers, which was 13, the standard error is 10 percentage points.

Standard error

The standard error of an estimate tells us how variable estimates would be if obtained from other samples drawn in the same way as one being described.

Even more often, research papers include confidence intervals and P values derived using them.

Estimated standard errors can be found for many of the statistics we want to calculate from data and use to estimate things about the population from which the sample is drawn.

Standard error

Plus or minus notation

We often see standard errors written as 'estimate ± SE'.

E.g. Elastic vs. inelastic bandages, first trial,
difference = 13 ± 10.

A bit misleading as many samples will give estimates more than one standard error from the population value.

Confidence intervals

Confidence intervals are another way to think about the closeness of estimates from samples to the quantity we wish to estimate.

Some, but not all, confidence intervals are calculated from standard errors.

Confidence intervals are called '**interval estimates**', because we estimate a lower and an upper limit which we hope will contain the true values.

An estimate which is a single number, such as the difference we observed from the trial, is called a **point estimate**.

Confidence intervals

It is not possible to calculate useful interval estimates which always contain the unknown population value.

There is always a very small probability that a sample will be very extreme and contain a lot of either very small or very large observations, or have two groups which differ greatly before treatment is applied.

We calculate our interval so that most of the intervals we calculate will contain the population value we want to estimate.

Confidence intervals

Often we calculate a **confidence interval:** a range of values calculated from a sample so that a given proportion of intervals thus calculated from such samples would contain the true population value.

For example, a **95% confidence interval** calculated so that 95% of intervals thus calculated from such samples would contain the true population value.

Confidence intervals

E.g., for the venous ulcer bandage study we have an estimated difference of 13 and a standard error of 10.

The sampling distribution is approximately Normal, with mean equal to the unknown population difference and standard deviation equal to the standard error, estimated to be 10.

We know that 95% of observations from a Normal distribution are closer than 1.96 standard deviations to the mean.

Hence 95% of possible samples will have the difference closer to the unknown population mean than 1.96×10 percentage points.
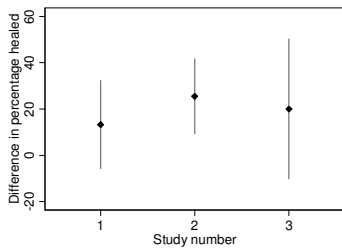
Confidence intervals

If we estimate our unknown population value to be between the observed sample value minus 1.96 standard errors and the observed sample value plus 1.96 standard errors, that range of values would include the population value for 95% of possible samples.

Thus the 95% confidence interval is 13 – 1.96×10 = –7 to 13 + 1.96×10 = 33 percentage points.

Hence we estimate that the true difference in the population lies between –7 and +33 percentage points. .
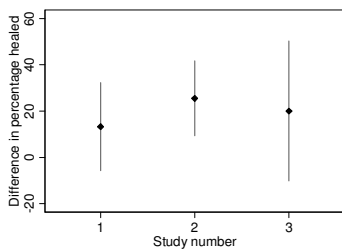

Venous ulcer trials: estimates from the other samples, with 95% confidence intervals:

A **forest plot**.

The width of the confidence interval depends on how many observations there were and the third study was smaller than the others.

Venous ulcer trials: estimates from the other samples, with 95% confidence intervals:

These confidence intervals all overlap, so they are quite consistent with the same unknown true value, which could lie within all of them.
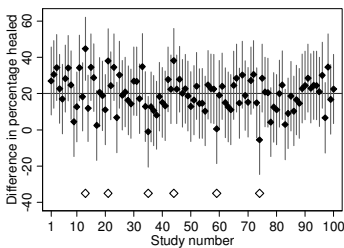
Confidence intervals do not always include the population value.

If 95% of 95% confidence intervals include it, it follows that 5% must exclude it.

In practice, we cannot tell whether our confidence interval is one of the 95% or the 5%.

Computer simulation of a trial of elastic bandaging.

Sample sizes in trial 2, 49 and 52 patients in the elastic and inelastic bandage groups, and assumed that in whole population of patients, the proportion of patients experiencing total healing would be 57% in the elastic bandage group and 37% in the inelastic bandage group.

Of the 100 trials, six 95% confidence intervals do not include the population difference (20).

Expect to see 5% of the intervals having the population value outside the interval.

Examples of confidence intervals:

```
Serum cholesterol (mmol/L) measured on a sample of 86
stroke patients (data of Markus et al., 1995)
  3.7   4.8   5.4   5.6   6.1   6.4   7.0   7.6   8.7
  3.8   4.9   5.4   5.6   6.1   6.5   7.0   7.6   8.9
  3.8   4.9   5.5   5.7   6.1   6.5   7.1   7.6   9.3
  4.4   4.9   5.5   5.7   6.2   6.6   7.1   7.7   9.5
  4.5   5.0   5.5   5.7   6.3   6.7   7.2   7.8  10.2
  4.5   5.1   5.6   5.8   6.3   6.7   7.3   7.8  10.4
  4.5   5.1   5.6   5.8   6.4   6.8   7.4   7.8
  4.7   5.2   5.6   5.9   6.4   6.8   7.4   8.2
  4.7   5.3   5.6   6.0   6.4   7.0   7.5   8.3
  4.8   5.3   5.6   6.1   6.4   7.0   7.5   8.6
```

Statistic	Point	95% confidence interval
Mean	6.34	6.04 to 6.64 (SE method)
Median	6.15	5.7 to 6.5 (Binomial method)

Note that mean is better estimated than median.

Examples of confidence intervals:

```
Serum cholesterol (mmol/L) measured on a sample of 86
stroke patients (data of Markus et al., 1995)
  3.7   4.8   5.4   5.6   6.1   6.4   7.0   7.6   8.7
  3.8   4.9   5.4   5.6   6.1   6.5   7.0   7.6   8.9
  3.8   4.9   5.5   5.7   6.1   6.5   7.1   7.6   9.3
  4.4   4.9   5.5   5.7   6.2   6.6   7.1   7.7   9.5
  4.5   5.0   5.5   5.7   6.3   6.7   7.2   7.8  10.2
  4.5   5.1   5.6   5.8   6.3   6.7   7.3   7.8  10.4
  4.5   5.1   5.6   5.8   6.4   6.8   7.4   7.8
  4.7   5.2   5.6   5.9   6.4   6.8   7.4   8.2
  4.7   5.3   5.6   6.0   6.4   6.4   7.0   7.5   8.3
  4.8   5.3   5.6   6.1   6.4   7.0   7.5   8.6
```

Statistic	Point	95% confidence interval
SD	1.3998	1.1894 to 1.6102 (SE method)
SD	1.3998	1.1896 to 1.6096 (chi-squared method)

Note that there may be more than one way to calculate a CI.

Do we always use 95% confidence intervals?

So far we have looked at 95% confidence intervals, chosen so that 95% of intervals will include the population value.

The choice of 95% is just that, a choice.
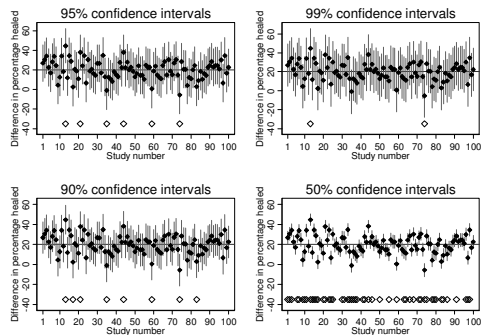
There is no reason why we have to use it.

We could use some other percentage, such as 99% or 90% confidence intervals.

If 99% of intervals include the population value, they must be wider than 95% confidence intervals.

90% intervals are narrower.

Do we always use 95% confidence intervals?

Elastic bandage trial simulation with 95%, 99%, 90% and 50% intervals.

Do we always use 95% confidence intervals?

99% confidence intervals are wider.

90% confidence intervals are narrower.

The 99% confidence interval is 19% wider than the 95% confidence interval, but more intervals are 'correct', in that the interval estimate includes the population value.

We do not gain a lot by widening the interval, that extra 19% of width includes the population value in only 4% (99% – 95%) of 99% confidence intervals.

The 90% confidence interval is 16% narrower than the 95% confidence interval, but fewer intervals are 'correct'.

Do we always use 95% confidence intervals?

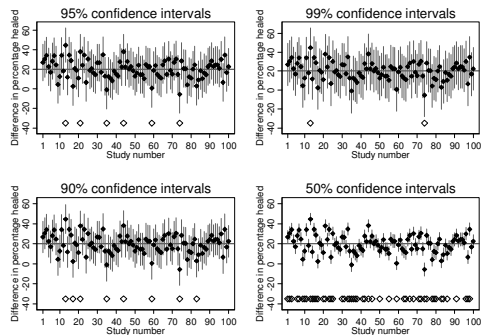99% confidence intervals are wider.

90% confidence intervals are narrower.

The choice of 95% is a compromise between the desire to have a confidence interval which includes the population value and one which is narrow enough to provide some useful information.

Occasionally researchers may want a different compromise and use 90% or 99% intervals, often without giving any indication in their reports as to why they did this.
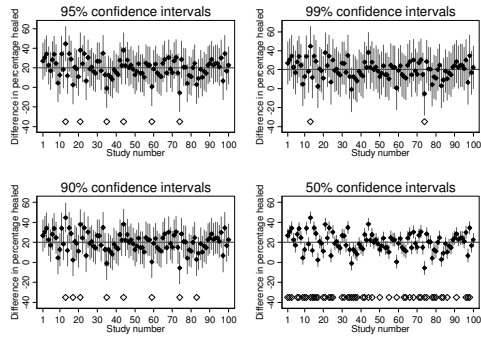

Do we always use 95% confidence intervals?

50% confidence intervals are much narrower than 95% intervals, forming the central third of the 95% interval.
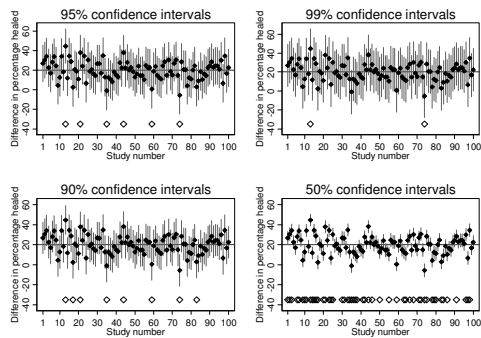
Do we always use 95% confidence intervals?

Only half of the 50% confidence intervals will include the population value, so they are not much use.
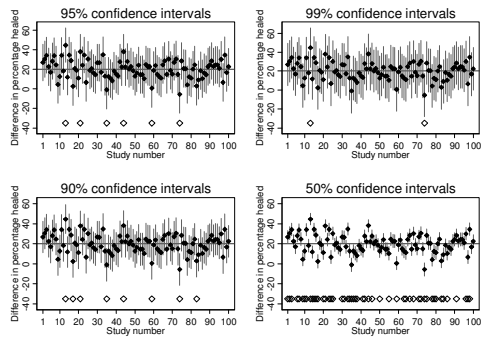
Do we always use 95% confidence intervals?

A method of estimation which was wrong on half the occasions when it was used would be of limited value.

Do we always use 95% confidence intervals?

The central third of a 95% confidence interval contains the population value for half of the intervals calculated.

Pitfalls in using and interpreting confidence intervals

The most common pitfall is not quoting confidence intervals at all.

Most of the leading journals specify in their instructions to authors that results should be given in the form of confidence intervals, yet authors persist in giving only P values.

For example, the *Lancet's* instructions to authors state "When possible, quantify findings and present them with appropriate indicators of measurement error or uncertainty (such as confidence intervals). Avoid relying solely on statistical hypothesis testing, such as the use of P values, which fails to convey important quantitative information."

Pitfalls in using and interpreting confidence intervals

A second pitfall is to quote a confidence interval which does not actually answer the research question directly.

One example is to give confidence intervals for estimates from treatment groups separately, rather than for the difference between the groups.

For example, in the study of venous ulcers by Northeast *et al.* (1990), the confidence interval for percentage healed in the elastic bandage group was 50% to 77% and for the inelastic bandage group it was 36% to 64%.

Quoting these would not show how well we had estimated the difference, which is what we are really interested in. The confidence interval for this difference in percentage healed, –7 to +33, addresses the question directly.

Pitfalls in using and interpreting confidence intervals

A third pitfall is to calculate the confidence interval for an estimate obtained from a small sample using a method designed for large samples.

For example, in a study of the prevalence of HIV in ex-prisoners (Turnbull *et al.*, 1992), of 29 women who did not inject drugs one was HIV positive.

The authors reported this to be 3.4%, with a 95% confidence interval –3.1% to 9.9%. The lower limit of –3.1%, obtained from the observed proportion minus 1.96 standard errors (a large sample method), is impossible.

A correct small sample method gives 0.1% to 17.8%.

Turnbull, P.J., Stimson, G.V., and Dolan, K.A. (1992) Prevalence of HIV infection among ex-prisoners. *BMJ* **304**, 90-1.