

## Standard Error and Confidence Intervals

Martin Bland

Professor of Health Statistics

University of York

<http://www-users.york.ac.uk/~mb55/>

---

---

---

---

---

---

---

---

### 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. randomised controlled trial comparing two obstetric regimes, active management of labour and routine management.

Proportion of women in the active management of labour group who had a Caesarean section was 0.97 times the proportion of women in the routine management group who had sections (Sadler *et al.*, 2000). (Relative risk.)

Sadler LC, Davison T, McCowan LM. (2000) A randomised controlled trial and meta-analysis of active management of labour. *British Journal of Obstetrics and Gynaecology* **107**, 909-15.

---

---

---

---

---

---

---

---

### 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.

Four randomised controlled trial comparing active management of labour and routine management.

Relative risks of Caesarean section:

0.97, 0.75, 1.01, and 0.64.

They are not all the same.

Estimates vary from sample to sample.

---

---

---

---

---

---

---

---

### Sampling distributions

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

We call this the **sampling distribution**.

---

---

---

---

---

---

---

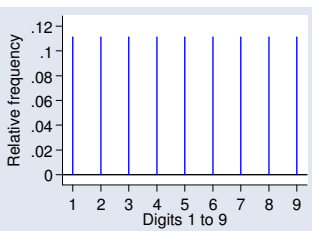
---

### Sampling distributions

Example: drawing lots.

Put lots numbered 1 to 9 into a hat and sample by drawing one out, replacing it, drawing another out, and so on.

Each number would have the same chance of being chosen each time and the sampling distribution would be:



---

---

---

---

---

---

---

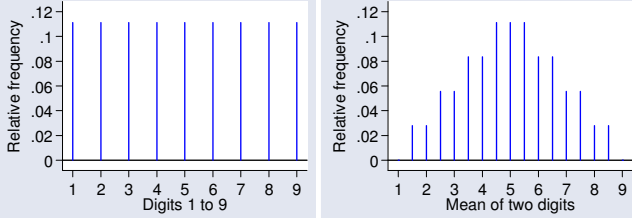
---

### Sampling distributions

Example: drawing lots.

Now we change the procedure, draw out two lots at a time and calculate the average.

There are 36 possible pairs, and some pairs will have the same average. The sampling distribution would be:



---

---

---

---

---

---

---

---

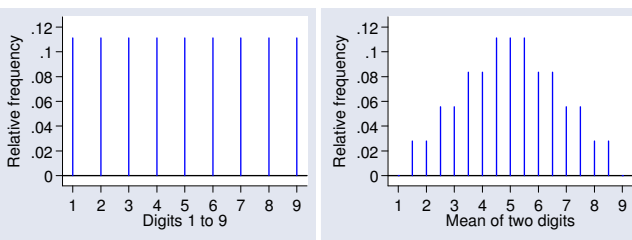
---

---

### Sampling distributions

Example: drawing lots.

The mean of the distribution remains the same, 5.



---

---

---

---

---

---

---

---

---

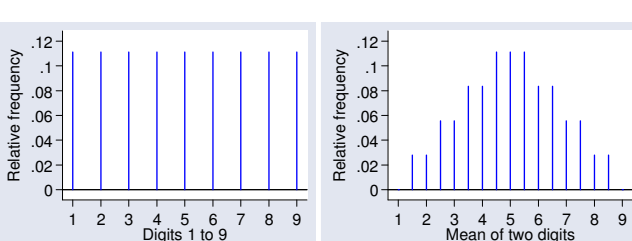
---

### Sampling distributions

Example: drawing lots.

The sampling distribution of the mean is not so widely spread as the parent distribution.

It has a smaller variance and standard deviation.



---

---

---

---

---

---

---

---

---

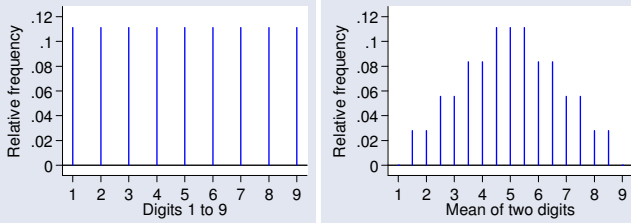
---

## Sampling distributions

Example: drawing lots.

The sampling distribution of the average has a different shape to the distribution of a single draw.

It looks closer to a Normal distribution than does the distribution of the observations themselves.



---

---

---

---

---

---

---

---

---

---

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.

---

---

---

---

---

---

---

---

---

---

**Standard error**

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

In most practical situations, we do not know this.

---

---

---

---

---

---

---

---

**Standard error**

FEV1 (litres) of 57 male medical students

2.85 3.19 3.50 3.69 3.90 4.14 4.32 4.50 4.80 5.20  
2.85 3.20 3.54 3.70 3.96 4.16 4.44 4.56 4.80 5.30  
2.98 3.30 3.54 3.70 4.05 4.20 4.47 4.68 4.90 5.43  
3.04 3.39 3.57 3.75 4.08 4.20 4.47 4.70 5.00  
3.10 3.42 3.60 3.78 4.10 4.30 4.47 4.71 5.10  
3.10 3.48 3.60 3.83 4.14 4.30 4.50 4.78 5.10

Mean = 4.062, SD  $s = 0.67$  litres.

What is the standard error of the mean?

For a sample mean, the standard error is =

$$\text{Standard deviation} / \text{square root sample size} \\ = 0.67 / \sqrt{57} = 0.089 \text{ litres.}$$

---

---

---

---

---

---

---

---

**Standard error**

FEV1 (litres) of 57 male medical students

Mean = 4.062, SD  $s = 0.67$  litres.

What is the standard error of the mean?

For a sample mean, the standard error is =

$$\text{Standard deviation} / \text{square root sample size} \\ = 0.67 / \sqrt{57} = 0.089 \text{ litres.}$$

In general, standard errors are proportional to one over the square root of the sample size, approximately.

To half the standard error we must quadruple the sample size.

---

---

---

---

---

---

---

---

**Standard error**

Plus or minus notation

We often see standard errors written as 'estimate  $\pm$  SE'.

E.g. mean FEV1 = 4.062  $\pm$  0.089.

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

---

---

---

---

---

---

---

---

**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.

If we want to say how good our estimate of the mean FEV1 measurement is, we quote the standard error of the mean. If we want to say how widely scattered the FEV1 measurements are, we quote the standard deviation, *s*.

---

---

---

---

---

---

---

---

**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.

---

---

---

---

---

---

---

---

**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 mean FEV1 observed from the sample, 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**

In the FEV1 example, we have a fairly large sample, and so we can assume that the observed mean is from a Normal Distribution.

For this illustration we shall also assume that the standard error is a good estimate of the standard deviation of this Normal distribution. (We shall return to this in Week 5.)

We therefore expect about 95% of such means to be within 1.96 standard errors of the population mean.

Hence, for about 95% of all possible samples, the population mean must be greater than the sample mean minus 1.96 standard errors and less than the sample mean plus 1.96 standard errors.

---

---

---

---

---

---

---

---

**Confidence intervals**

If we calculate mean minus 1.96 standard errors and mean plus 1.96 standard errors for all possible samples, 95% of such intervals would contain the population mean.

For the FEV1 sample, these limits are

$$4.062 - 1.96 \times 0.089 \text{ to } 4.062 + 1.96 \times 0.089$$

which gives 3.89 to 4.24, or 3.9 to 4.2 litres.

3.9 and 4.2 are called the **95% confidence limits** for the estimate, and the set of values between 3.9 and 4.2 is called the **95% confidence interval**.

The confidence limits are the ends of the confidence interval.

---

---

---

---

---

---

---

---

**Confidence intervals**

It is incorrect to say that there is a probability of 0.95 that the population mean lies between 3.9 and 4.2.

The population mean is a number, not a random variable, and has no probability.

We sometimes say that we are 95% confident that the mean lies between these limits, but this doesn't help us understand what a confidence interval is.

The important thing is: we use a sample to estimate something about a population. The 95% confidence interval is chosen so that 95% of such intervals will include the population value.

---

---

---

---

---

---

---

---



### Confidence intervals

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%.

---

---

---

---

---

---

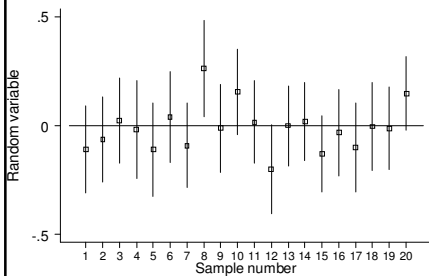
---

---

### Confidence intervals

Confidence intervals for the mean for 20 random samples of 100 observations from the Standard Normal Distribution.

Population mean is = 0.0, SD = 1.0, SE = 0.10.



The population mean is contained by 19 of the 20 confidence intervals.

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

---

---

---

---

---

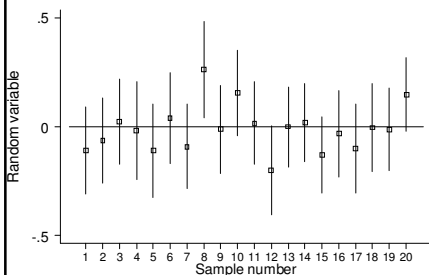
---

---

---

### Confidence intervals

We expect to see 5% of the intervals having the population value outside the interval and 95% having the population value inside the interval.



Note that this is not the same as saying that 95% of further samples will have estimates within the interval.

---

---

---

---

---

---

---

---

**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.

---

---

---

---

---

---

---

---

**Confidence intervals**

For the trial comparing active management of labour with routine management (Sadler *et al.*, 2000), the relative risk for Caesarean section was 0.97.

Sadler *et al.* quoted the 95% confidence interval for the relative risk as 0.60 to 1.56.

Hence we estimate that in the population which these subjects represent, the proportion of women undergoing Caesarean section when undergoing active management of labour is between 0.60 and 1.56 times the proportion who would have Caesarean section with routine management.

---

---

---

---

---

---

---

---

**CI for the mean of a small sample**

In the FEV1 example, we assumed that the observed mean is from a Normal Distribution and that the standard error is a good estimate of the standard deviation of this Normal distribution.

This assumption may not be valid.

Usually need 100 or more observations to make it.

We should not use the large sample Normal method because the sample is too small.

---

---

---

---

---

---

---

---

### CI for the mean of a small sample

The standard error will not be sufficiently well estimated.

The distribution of the standard error estimate depends on the distribution of the observations themselves.

There are several ways to estimate a confidence interval for the mean.

The most frequently used is the t Distribution method.

We must make an assumption about the data:

that the observations come from a Normal distribution.

---

---

---

---

---

---

---

---

### CI for the mean of a small sample

Large sample Normal method: calculate mean minus 1.96 standard errors and mean plus 1.96 standard errors.

For the FEV1 sample, these limits are

$$4.062 - 1.96 \times 0.089 \text{ to } 4.062 + 1.96 \times 0.089$$

which gives 3.89 to 4.24, or 3.9 to 4.2 litres.

For a small sample from a Normal Distribution, we replace 1.96 by a number from the t distribution, also known as Student's t distribution.

---

---

---

---

---

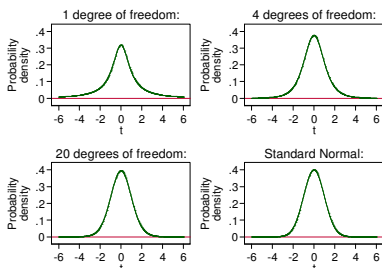
---

---

---

### Two sample t method

The t distribution family has one parameter, the degrees of freedom.



As the degrees of freedom increases, the t distribution becomes closer to the Standard Normal distribution.

---

---

---

---

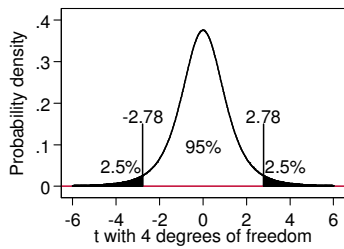
---

---

---

---

Two tailed probability point of the t Distribution




---

---

---

---

---

---

---

---

Two tailed probability points of the t Distribution

D.f.	Probability				D.f.	Probability			
	0.10 (10%)	0.05 (5%)	0.01 (1%)	0.001 (0.1%)		0.10 (10%)	0.05 (5%)	0.01 (1%)	0.001 (0.1%)
1	6.31	12.70	63.66	636.62	16	1.75	2.12	2.92	4.02
2	2.92	4.30	9.93	31.60	17	1.74	2.11	2.90	3.97
3	2.35	3.18	5.84	12.92	18	1.73	2.10	2.88	3.92
4	2.13	2.78	4.60	8.61	19	1.73	2.09	2.86	3.88
5	2.02	2.57	4.03	6.87	20	1.73	2.09	2.85	3.85
6	1.94	2.45	3.71	5.96	21	1.72	2.08	2.83	3.82
7	1.90	2.36	3.50	5.41	22	1.72	2.07	2.82	3.79
8	1.86	2.31	3.36	5.04	23	1.71	2.07	2.81	3.77
9	1.83	2.26	3.25	4.78	24	1.71	2.06	2.80	3.75
10	1.81	2.23	3.17	4.59	25	1.71	2.06	2.79	3.73
11	1.80	2.20	3.11	4.44	30	1.70	2.04	2.75	3.65
12	1.78	2.18	3.06	4.32	40	1.68	2.02	2.70	3.55
13	1.77	2.16	3.01	4.22	60	1.67	2.00	2.66	3.46
14	1.76	2.15	2.98	4.14	120	1.66	1.98	2.62	3.37
15	1.75	2.13	2.95	4.07	∞	1.65	1.96	2.58	3.29

D.f. = Degrees of freedom  
 ∞ = infinity, same as the Standard Normal Distribution

---

---

---

---

---

---

---

---

**CI for the mean of a small sample**

For a small sample from a Normal Distribution, we replace 1.96 by a number from the t distribution, also known as Student's t distribution.

The degrees of freedom we use are the degrees of freedom for the standard deviation used to calculate the standard error:  $n - 1 = 57 - 1 = 56$ .

---

---

---

---

---

---

---

---

