Standard Errors and Confidence Intervals

- Population of five-year-old boys, has mean height μ.
- Mean of sample of 99 heights is 108.3 cm

  10 boys were sampled, mean would have been 107.8 cm

  20 boys, mean would have been 107.7 cm

  1000 boys mean would have been 108.0 cm

- Each mean is a legitimate estimator of μ

  which one should we choose?

Choice of Mean

- Which should we choose?

- Assume resources not an issue (see later)

- Intuition suggests using largest sample, as it contains ‘more information’
  Aliter: a larger sample should be ‘more representative’ so its mean should be closer to μ

- Idea that the mean of a larger sample will be closer to μ is central

The Idea of the Standard Error

- To demonstrate this important idea it is useful to have access not to one but to very many samples

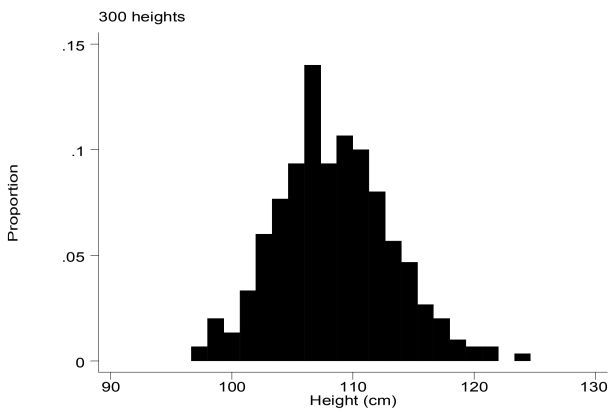
- To this end we will use the computer to generate artificial samples of ‘heights’ of five-year-old boys, assuming population values of μ = 108 cm and σ = 4.7 cm (these are quite sensible values)

- This is not the position in practice, but let’s worry about that later.

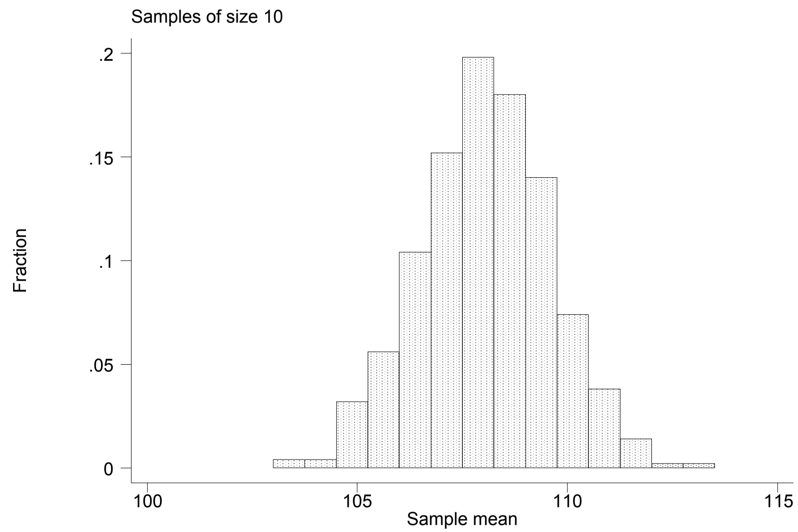
- Suppose we generate 500 samples, each of size 10 from the above population

Distribution of Means, samples of 10

Histogram of individual heights:



Histogram of 500 samples of size 10



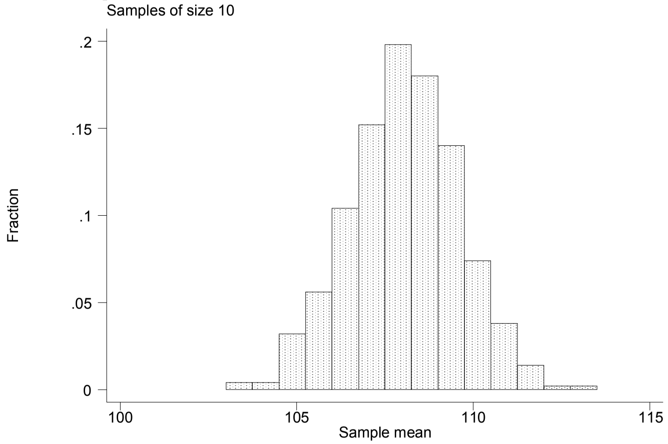
Histogram of sample means is less dispersed than original heights.

(Centred on 108 cm, Normally distributed)

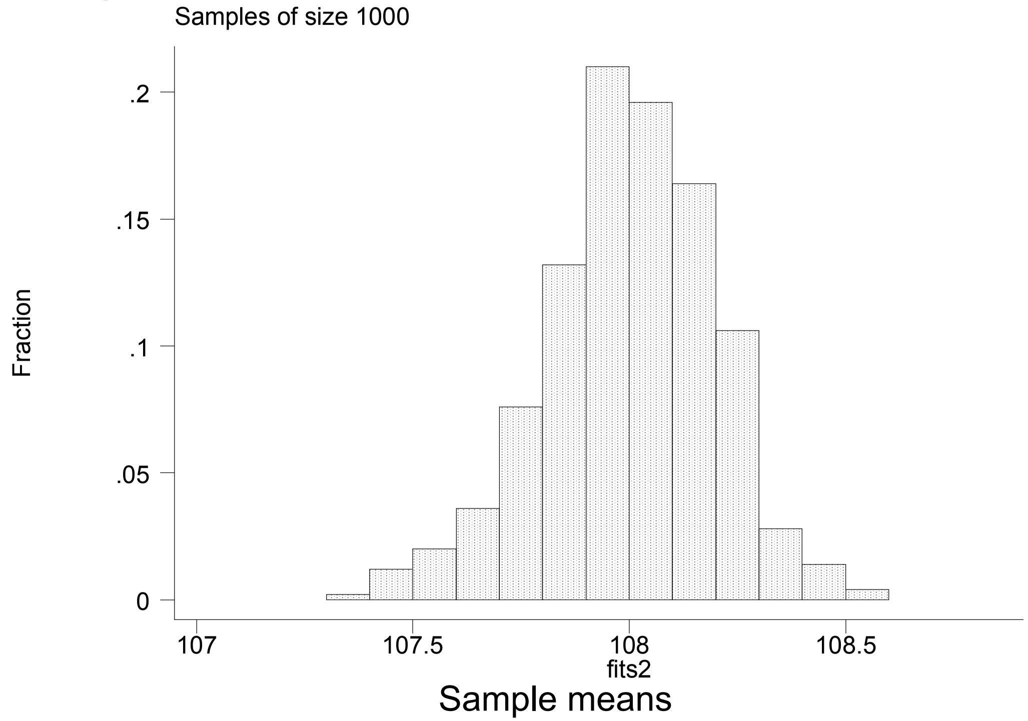
Distribution of Means, samples of 1000

Repeating above with samples of 1000

Samples size 10



Samples size 1000



Dispersion about μ now even less

Definition of Standard Error

- These histograms demonstrate that means of larger samples are better because they depart from the population mean by less.

- We can be quantitative about this: the standard deviation of the distribution of sample means measures the spread about μ

  samples size 10   1.54 cm

  samples size 1000   0.20 cm

- This very special form of standard deviation is given a special name: *the standard error of the mean* (SEM) or simply *the standard error* (SE).

Calculating SE from a Single Sample

- Computing the standard deviation of 500 samples is not a practical way of finding the SE.

- In practice there is only one sample available

- Statistical theory can be used to derive the fact that the standard error is determined by the standard deviation of the height of an individual, σ, and the size of the sample, n viz.

$$\text{SE}= \frac{\sigma}{\sqrt{n}}$$

Formula for SE

Thus, for samples of size 10, SE = $\sigma/\sqrt{10} = 0.316\sigma$

Samples of size 1000 it is a tenth of this, $\sigma/\sqrt{1000} = 0.0316\sigma$

Sample size	From multiple samples	From formula
10	1.54	1.49
1000	0.20	0.15

Using the Standard Error I

In sample of 99 heights, mean is 108.34 cm, SE 0.52 cm

How should we use this measure of 'error'?

Do we quote mean as 108.34 ± 0.52 cm?

Encouraged by all the 'Mean (±SE)' headings in papers and plots with SE bars

Should be discouraged, but why?

Invites reader to think mean is between 107.82 & 108.86 cm, which is false.

Using the Standard Error II

Recall that distribution of sample means is Normal

Recall that approx. 95% of a Normal sample is within two SDs of the mean.

Applying this to distribution of *sample means* indicates that 95% of the time, m is between $\mu \pm 2\text{SE}$, amounts to

95% confident that μ is between $m \pm 2\text{SE}$.

Confidence Intervals

95% Confidence interval, alternatively *interval estimate*:

$$\left(m - 2\frac{s}{\sqrt{n}}, m + 2\frac{s}{\sqrt{n}} \right)$$

Here *s* has been used in place of σ - has some implications for a precise definition – see appendix.

Use of Confidence Intervals

Suppose a trial of two treatments, A and B, results in a mean blood pressure 1 mmHg lower on A than B.

How should this be interpreted?

It can't

If 95% confidence interval for difference is (-3, 5) mmHg then there is no material difference between the treatments.

If 95% confidence interval for difference is (-30, 32) mmHg then there could be an important difference between the treatments.

Only confidence intervals can do this.

Appropriate Precision

The sample standard deviation does not systematically get smaller or larger as the sample size increases

The standard error gets smaller as the sample size increases

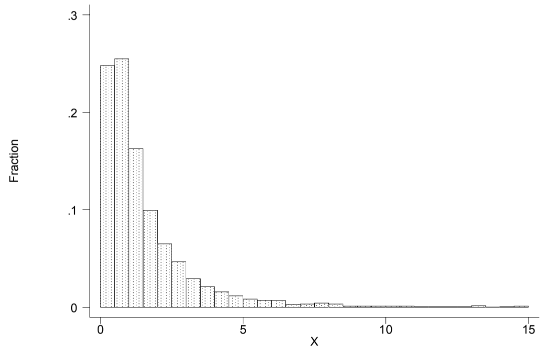
Because of $\sqrt{n}$ in denominator

Because of $\sqrt{}$ does not get small all that quickly

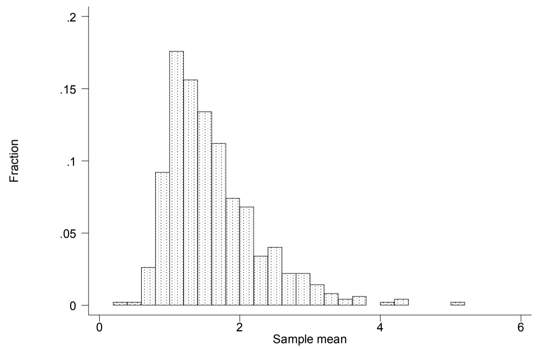
Can make precision arbitrarily high by making sample arbitrarily large but this is likely to be wasteful. Should aim for appropriate precision

Distribution of Sample Mean

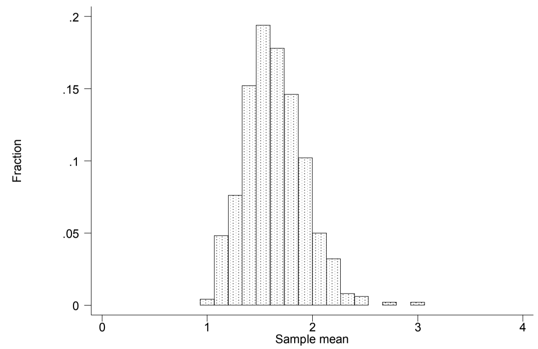
Sample means tend to be Normally distributed, at least approximately



(a) original variable



(b) means of samples of size 10



(c) means of samples of size 50

Figure 3