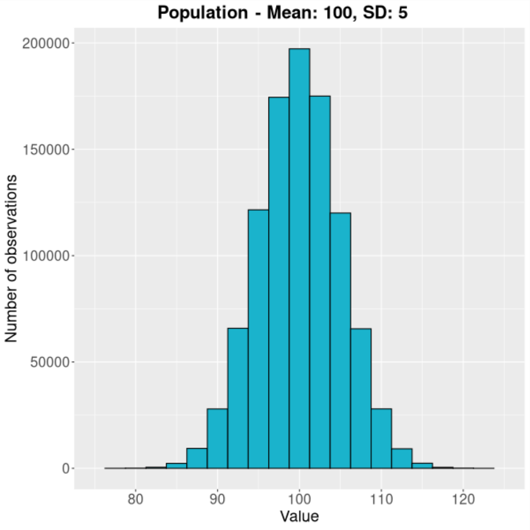
Sampling Distribution & The Central Limit Theorem

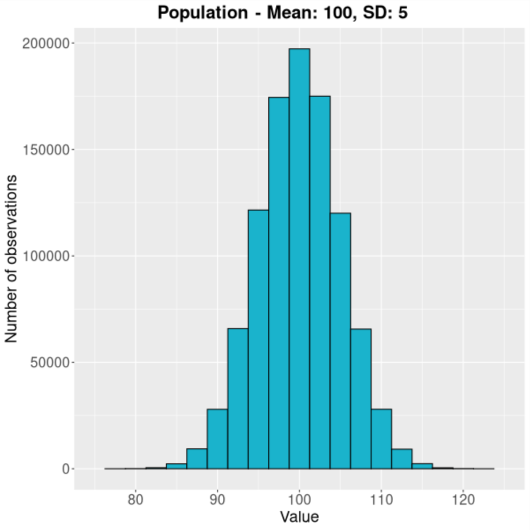
Nicola Romanò - nicola.romano@ed.ac.uk
Based on slides by Duncan McGregor

What’s up with normal distributions?



We talk about normal distributions a lot.

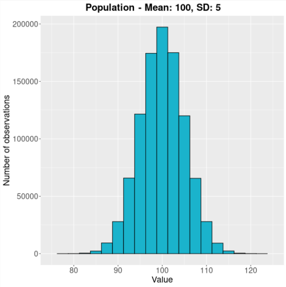
What's up with normal distributions?

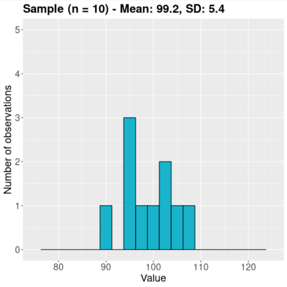


We talk about normal distributions a lot.**But why, actually?**

This lecture is about...

- Properties of sampling distributions.
- The normal distribution...
- ...its properties...
- ...and why it is special.

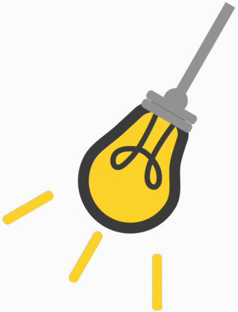




Learning objectives

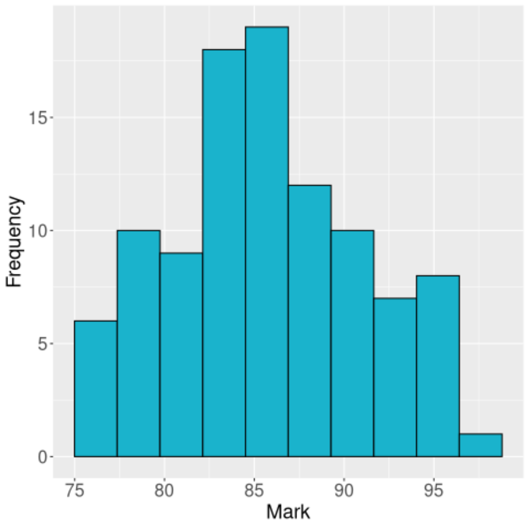
At the end of this lecture, you should be able to:

- Define the standard error of the mean
- Compare sampling distributions and underlying population distributions
- Describe a normal distribution and explain its importance
- Explain the Central Limit Theorem



Examining normal distributions

Create a "virtual class" of 100 exam grades with a mean of 86 and standard deviation of 5.

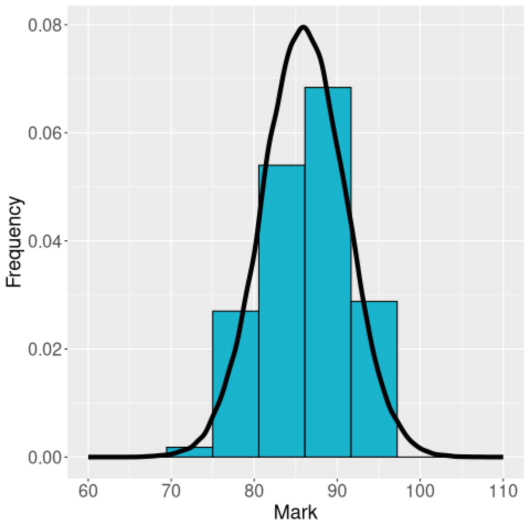


Notation

How would you read this?

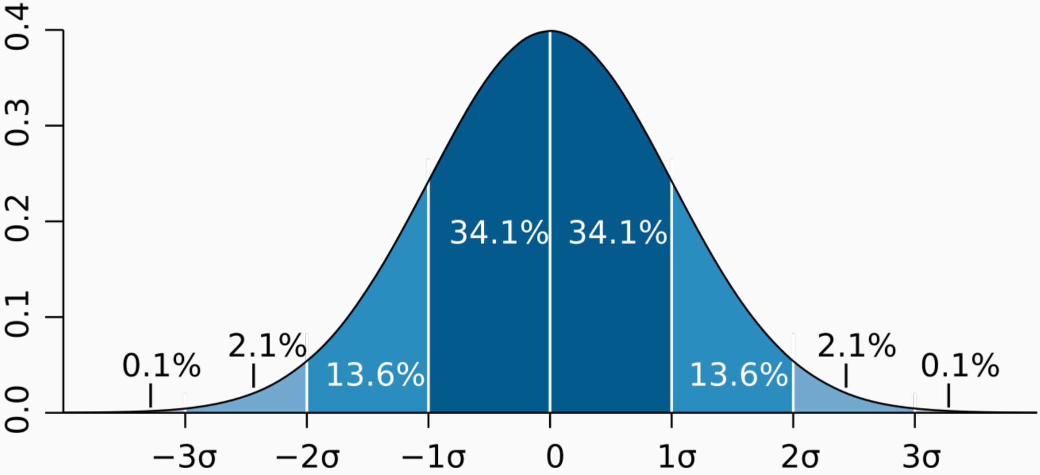
$$X \sim \mathcal{N}(\mu, \sigma^2)$$

What are the properties of a normal distribution?



???

The 68 - 95 - 99.7 Rule



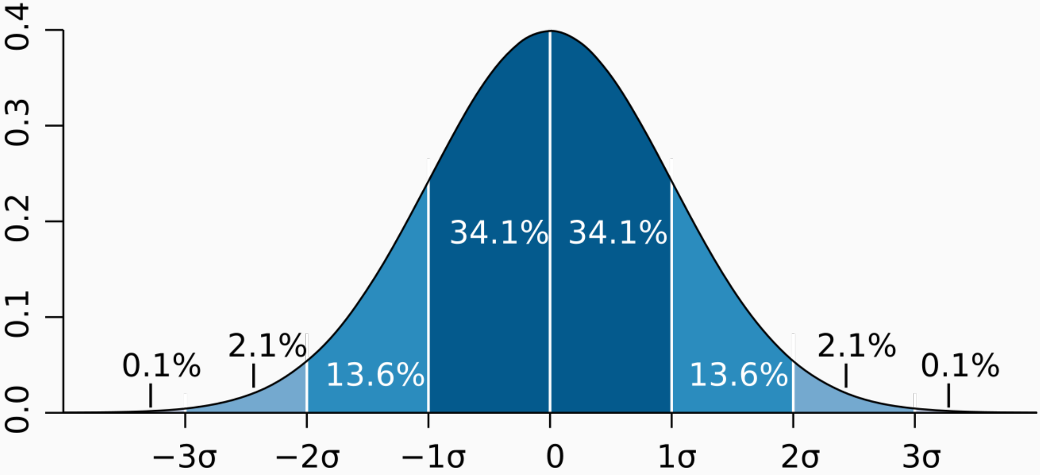
Sampling distributions

From Problem Set 3 (Sampling)

"Take samples of size 5 from a normal distribution, record mean and standard deviation."

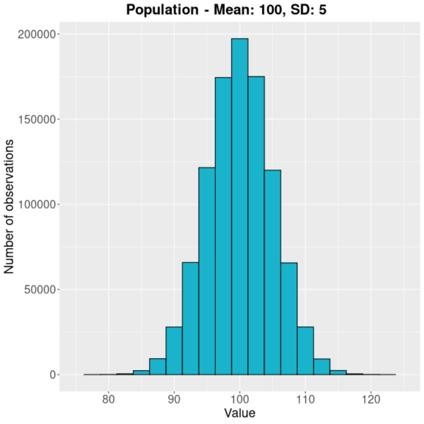
- Is our sample likely to have a higher or lower standard deviation than the population?
- Why?
- How does this relate to sample size?

We are unlikely to sample from the "edges"



From Problem Set 3 (Sampling)

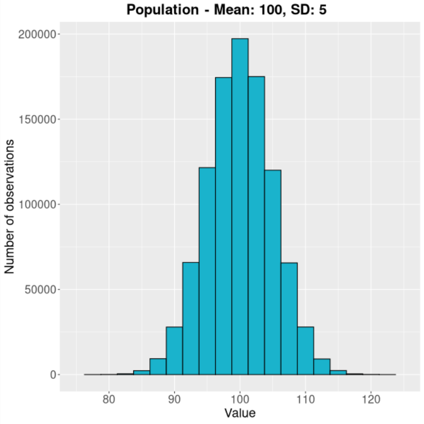
Take samples of size 5 from a normal distribution, record the **mean**



If you do this repeatedly, the distribution of sample means is called the **sampling distribution**.

Where is the sampling distribution centred? How much spread is there?

Population



Sampling distribution (n=5)

?

Fine, but all we (usually) have is a few samples…

How do we know how good a guess our sample mean is for the true population mean?

The **Standard Error of the Mean** (SEM) is a measure of how well your sample mean estimates the true population mean.

$$SEM = \frac{\sigma}{\sqrt{n}}$$

Where

σ is the standard deviation of the population
n is the sample size

Fine, but all we (usually) have is a few samples...

How do we know how good a guess our sample mean is for the true population mean?

The **Standard Error of the Mean** (SEM) is a measure of how well your sample mean estimates the true population mean.

$$SEM = \frac{\sigma}{\sqrt{n}}$$

Where

σ is the standard deviation of the population
n is the sample size

What happens if n increases?
What happens if σ increases?

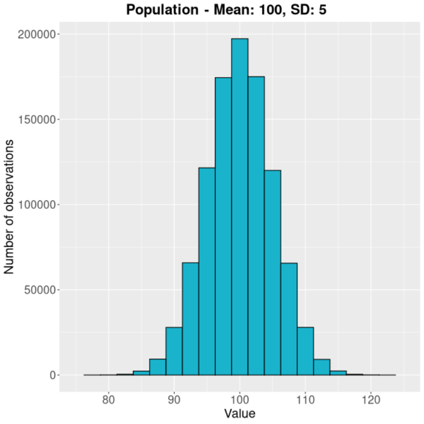
What is the difference between Standard Error of the Mean and Standard Deviation?

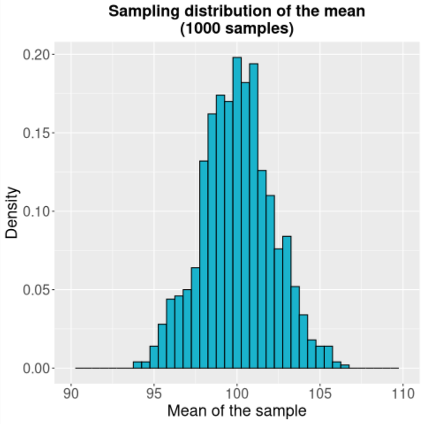
?

Where do normal distributions come from?

Why do we like normal distributions so much?

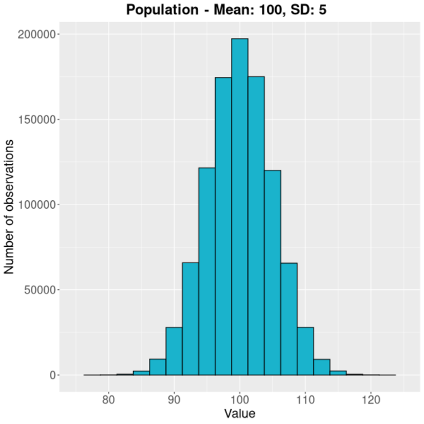
Let's recap:

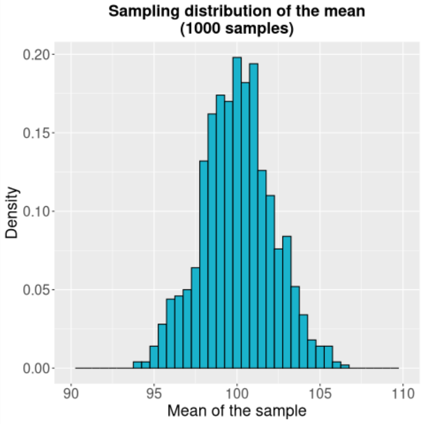




Why do we like normal distributions so much?

Let's recap:





But what happens if we are not sampling from a normal distribution?

The central limit theorem

Source: creaturecast.org

For sample means
Even if a population is not normally distributed, the sampling distribution (for large enough samples) will tend to be normal

The central limit theorem

For sample means
Even if a population is not normally distributed, the sampling distribution (for large enough samples) will tend to be normal

More in general
If we take n independent random variables from any distribution, and take their (normalised) sum, then that sum will tend towards a normal distribution with increasing n.

For sample means
Even if a population is not normally distributed, the sampling distribution (for large enough samples) will tend to be normal
More in general
If we take n independent random variables from any distribution, and take their (normalised) sum, then that sum will tend towards a normal distribution with increasing n.

Maybe you have seen this in real life before?

For sample means
Even if a population is not normally distributed, the sampling distribution (for large enough samples) will tend to be normal
More in general
If we take n independent random variables from any distribution, and take their (normalised) sum, then that sum will tend towards a normal distribution with increasing n.

Maybe you have seen this in real life before?

So…why do we love normal distributions so much?

Because it comes up all the time. Even when things are not normally distributed, a normal distribution often “comes out” of parameter combinations, such as taking the mean.

Learning objectives

Now you should be able to:

- Define the standard error of the mean
- Compare sampling distributions and underlying population distributions
- Describe a normal distribution and explain its importance
- Explain the Central Limit Theorem