浙江大学爱丁堡大学联合学院

ZJU-UoE Institute

Sampling distributions and The Central Limit Theorem

ADS 2, Lecture 3

Melanie Stefan - melanie.stefan@ed.ac.uk

Semester 1, 2019/20

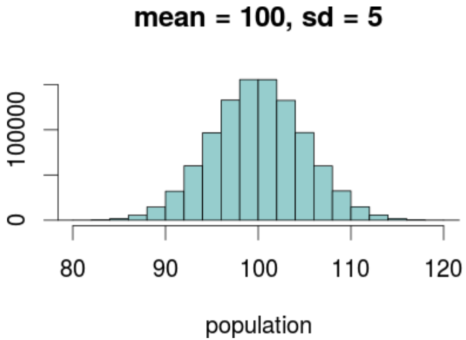
Before we start . . .

Please congratulate our class reps:

Before we start . . .

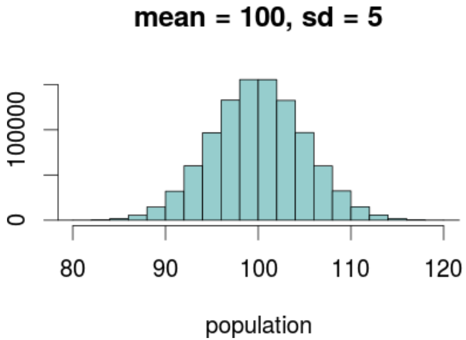
Please congratulate our class reps: Adele, Alana, and Jeff

What's up with normal distributions?



We talk about normal distributions a lot.

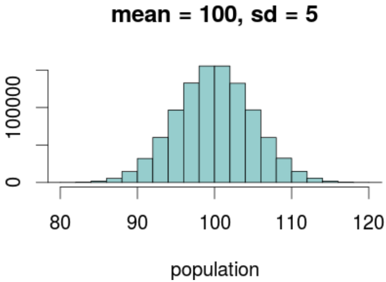
What's up with normal distributions?

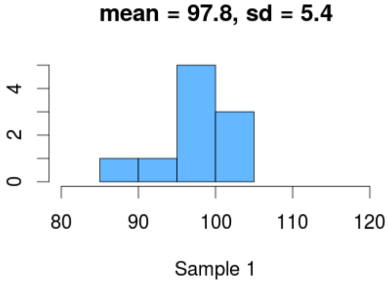


We talk about normal distributions a lot. But why, actually?

This lecture is about . . .

Properties of sampling distributions;
the normal distribution, its properties and why it is special.





Learning Objectives

After 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

Outline

From Problem Set 1

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

Notation

How would you read this?

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

Notation

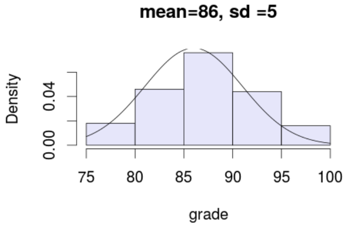
How would you read this?

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

“X follows a normal distribution with mean μ (mu) and standard deviation σ (sigma).”

“X is normally distributed with mean μ and variance σ^2 (sigma square).”

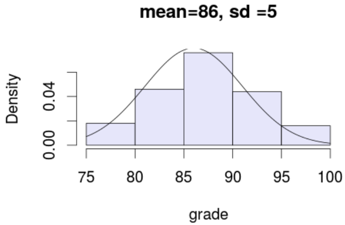
Features of a normal distribution



Properties of a normal distribution

- "Bell-shaped" curve
- Mean, median and mode are the same
- Distribution is symmetric around the mean

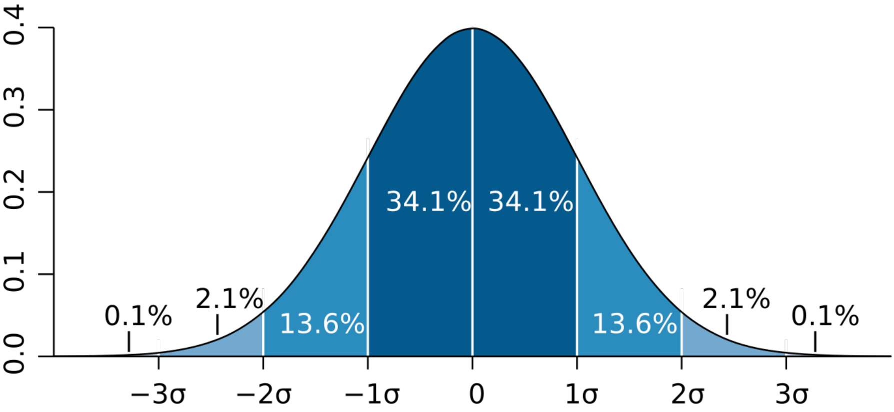
Features of a normal distribution



In problem set 1, we asked:

- How many students are more than one standard deviation away from the mean (less than 81 or more than 91)?
- How many students are more than 2 standard deviations away?

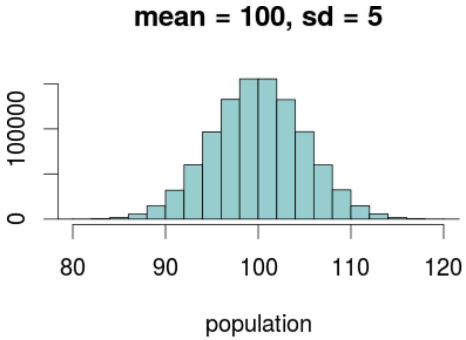
The 68 - 95 - 99.7 Rule



Outline

From Problem Set 2

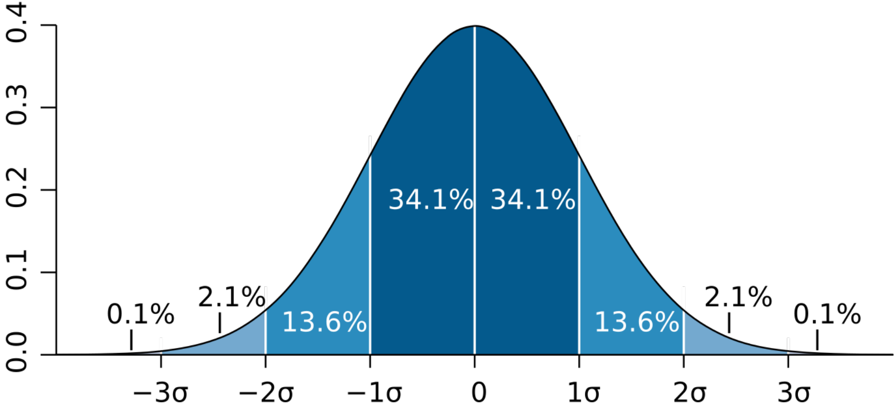
Take samples of size 5 from a normal distribution, record mean and standard deviation.



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

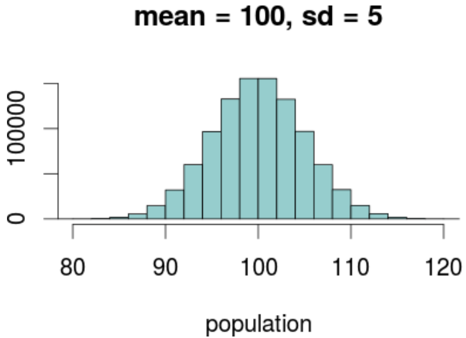
We are unlikely to sample from "the edges"

Small sample sizes will underestimate the population variance.



From Problem Set 2

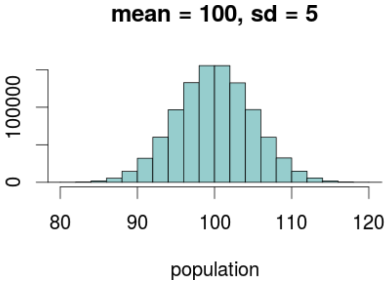
Take samples of size 5 from a normal distribution, record mean



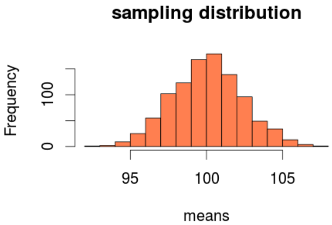
If you do this repeatedly, the distribution of sample means is called the sampling distribution

Sampling distribution

Population:



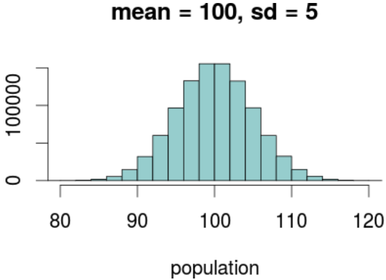
Sampling distribution (n=5)



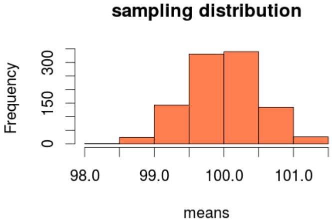
Where is the sampling distribution centred? How much spread is there?

Sampling distribution

Population:



Sampling distribution (n=100)



Fine, but all we (usually) have is one sample . . .

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

Fine, but all we (usually) have is one sample ...

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.

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

sd ... standard deviation
n ... sample size

What happens if n increases? What happens if sd increases?

What's the difference between Standard Error of the Mean and Standard Deviation?

What's the difference between Standard Error of the Mean and Standard Deviation?

- Different number:

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

What's the difference between Standard Error of the Mean and Standard Deviation?

- Different number:

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

- *More importantly:* Different concept

  Standard deviation measures variability in a dataset
  Standard error of the mean measures how good your estimate of the population mean is

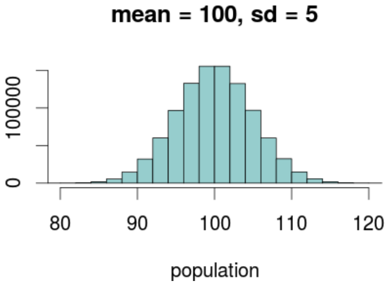
Outline

Why do we like normal distributions so much?

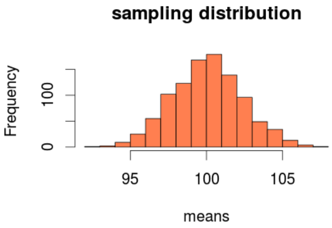
Why do we like normal distributions so much?

Let's recap:

Population:



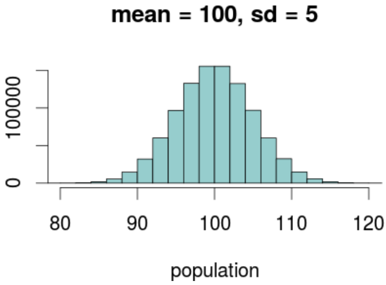
Sampling distribution (n=5)



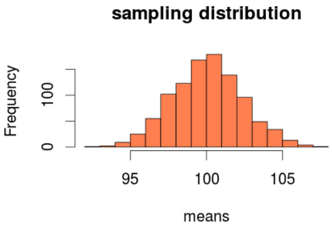
Why do we like normal distributions so much?

Let's recap:

Population:



Sampling distribution (n=5)



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

Central Limit Theorem

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 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?

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 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?

Why do we like normal distributions so much?

Why do we like 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.

What questions do you have?

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

Image credits

- Bunnies, Dragons, and the Central Limit Theorem. By ShuYi Chiou (animation), `creaturecast.org`
- Dice. By Diacritica - Own work, CC BY-SA 3.0, `https://commons.wikimedia.org/w/index.php?curid=12242533`, via Wikimedia Commons
- Fireworks. By Frank Vincentz - Own work, CC BY-SA 3.0, `https://commons.wikimedia.org/w/index.php?curid=37101006`, via Wikimedia Commons
- Histograms of populations, randomly drawn samples and sampling distributions. My own work using RStudio, 2019. CC BY-SA 3.0.
- The 68 - 95 - 99.7 rule. By M. W. Toews - Own work, based (in concept) on figure by Jeremy Kemp, on 2005-02-09, CC BY 2.5, `https://commons.wikimedia.org/w/index.php?curid=1903871`, via Wikimedia Commons