Practical 5: Sampling distributions

ADS2

Semester 1, 2022/23

Work through this guide alone or in groups. Facilitators are here to help. The time it takes to complete this practical can vary between individuals - this is OK. Do not worry if you do not finish within the session.

Learning Objectives

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

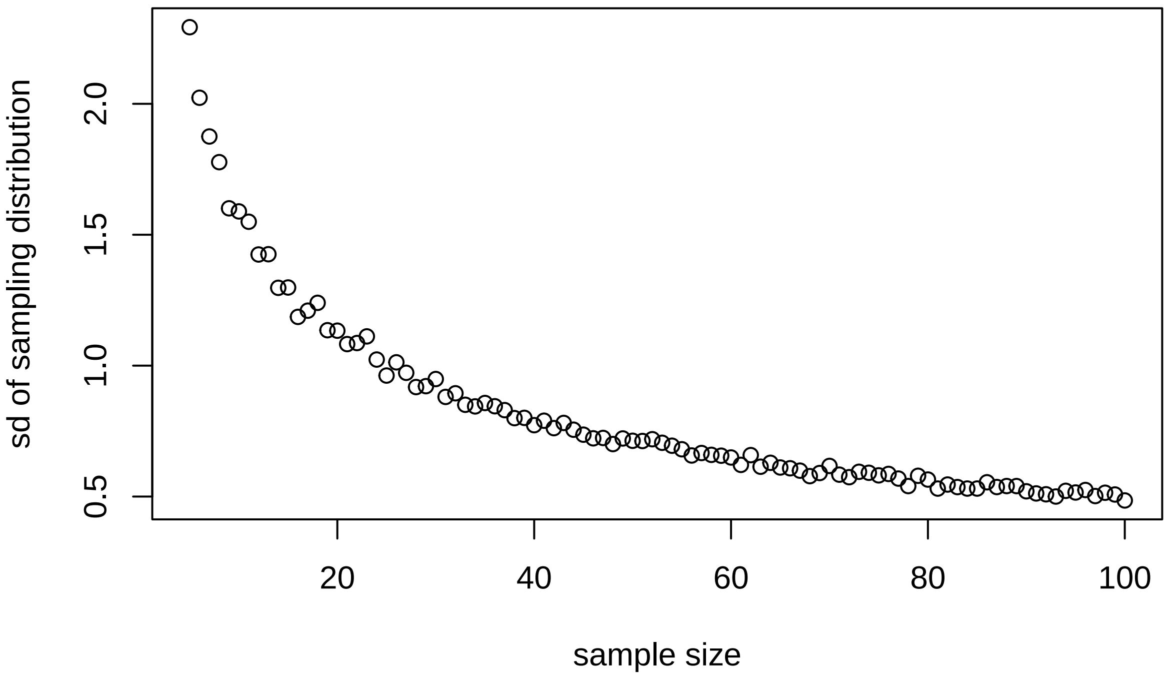
Width of the sampling distribution and sample size

In problem set 3, you were asked to sample from the following population:

```r
population <- rnorm(1e6,100,5)
```

You were asked to take 1000 samples of size 5 and record the mean - i.e. create a sampling distribution for samples of size 5. You were then asked to repeat this for samples of size 100.

This time, we want to be systematic about it. We would like to know what's the **standard deviation of the sampling distribution** as a function of sample size (for sample sizes between 5 and 100). We are looking for a plot like this:



- Create this plot (running the code may take a few minutes)
- What do you notice? What other name is there for the standard deviation of the sampling distribution?

Rolling dice

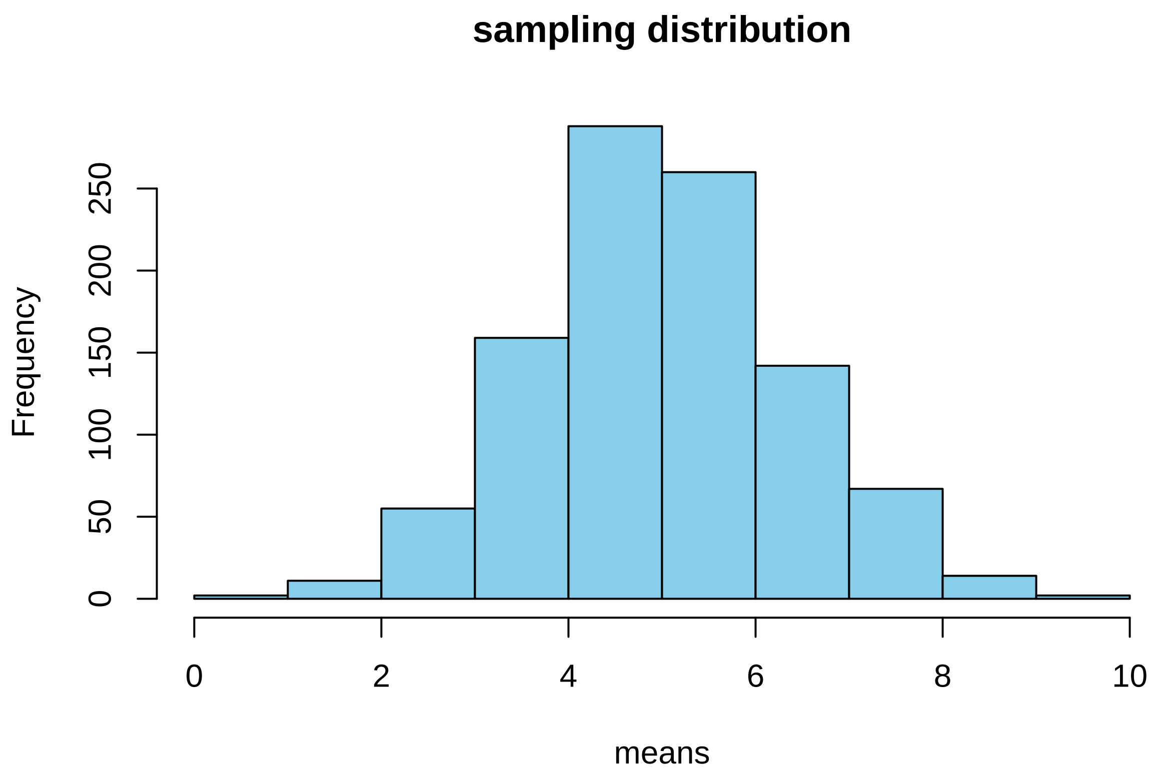
At first, it may seem weird that the Central Limit Theorem holds even when the underlying distribution is not normal. But here is an example that may be familiar to you from your lived experience.

- Roll a six-faced dice in R
- This should be a uniform distribution - every number from 1 to 6 is equally likely to appear. Roll your dice 1000 times and visualise the outcome to convince yourself that this is true
- If you drew a histogram and it looks a bit weird, this is because of the way that R decides where to break between each of the bins. Try adding breaks=0.5:6.5 to your hist command to manually set break points.
- Everything changes though when you roll 2 dice and add up their numbers. You may know from your own experience what number is the most frequent. Why?
- Let's do this 1000 times and see what we get
- Does this look normal to you? Why or why not?
- OPTIONAL (if you have time!): What about using 3 or 4 dice?

Dragon wingspans

Of course, we have to look at dragon wingspans in order to check the Central Limit Theorem for ourselves!

- The file `dragons.csv` contains wing spans for a random sample of 500 dragons. Import the file and visualise the data to convince yourself that this is indeed a non-normal distribution.
- Create a sampling distribution for a sample size of your choice. Here is one we did with n=10



Last update by DJ MacGregor in 2022