Problem Set 5: Sampling Distributions and the Central Limit Theorem

ADS2

Semester 1 2022/23

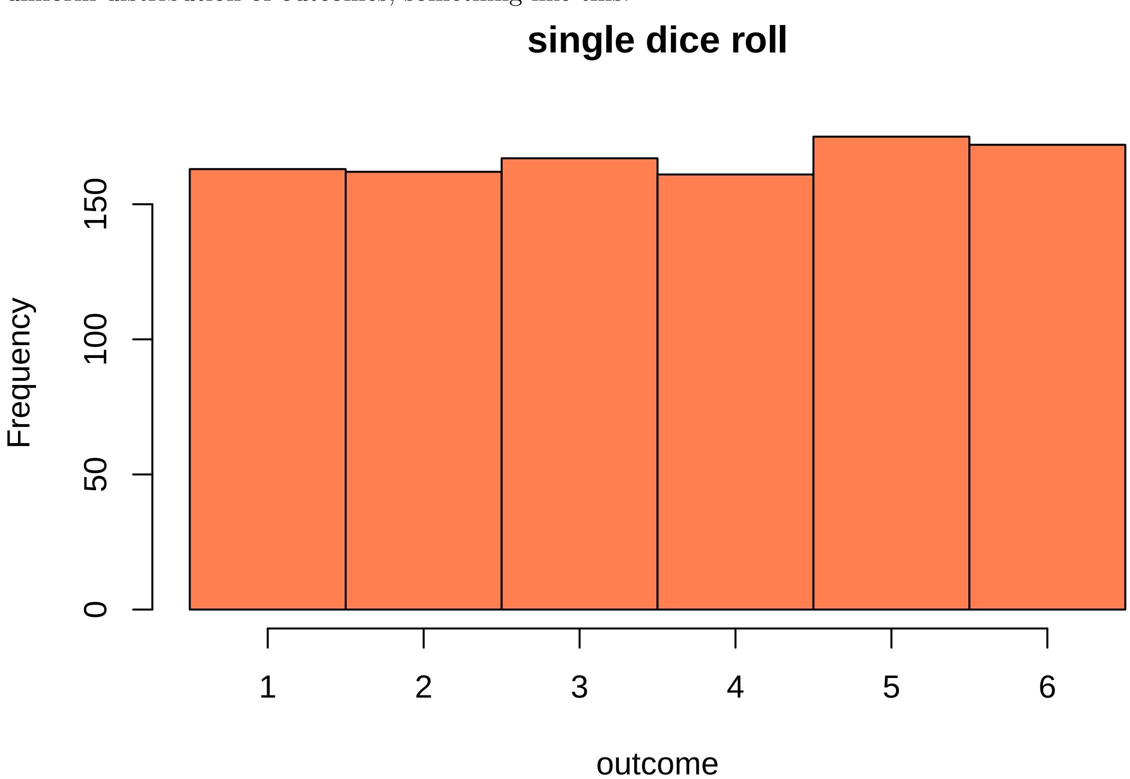
We expect this problem set to take around an hour to complete. But professors are sometimes wrong![citation missing] If this or future problem sets are too long, please let us know, so we can adjust and plan accordingly.

Learning Objectives

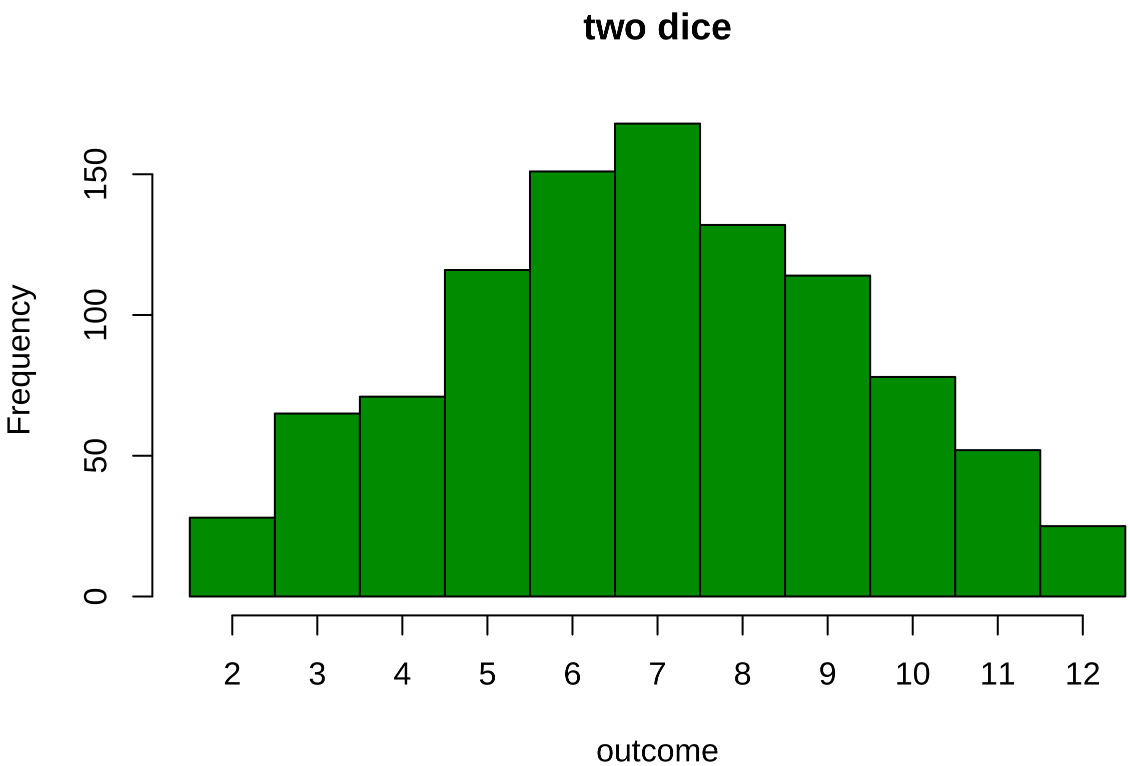
- Gain additional practice around sampling distributions and the Central Limit Theorem

Rolling dice

In the practical, you simulated a dice rolled once, and then two dice. You saw that rolling one dice gives a uniform distribution of outcomes, something like this:

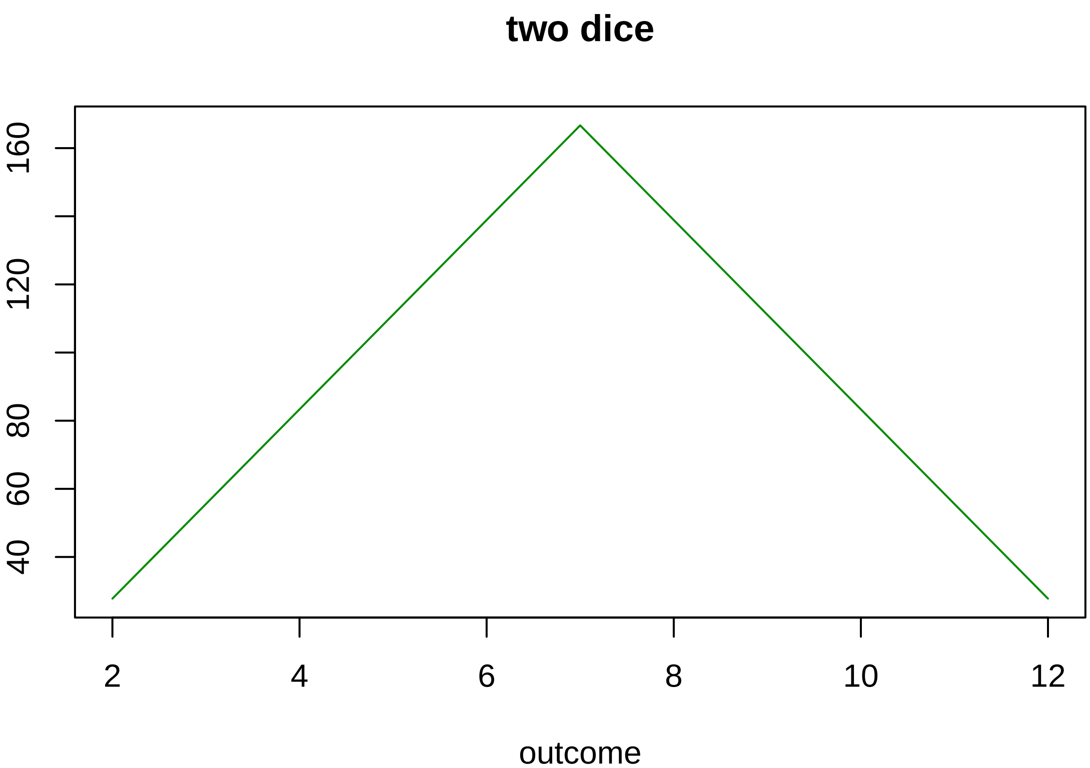


But rolling two dice gives a different distribution:



This second distribution already has features similar to a normal distribution: It is symmetric, with a peak in the middle, which is the mean, the median and the mode.

But it's not exactly a normal distribution. Actually, if you roll two dies, the resulting (theoretical) distribution is this:



- Why? Can you explain this mathematically?
- Remember the Central Limit Theorem says that as the sample size increases, the sums of samples become "more normal" and narrower. Run simulations and draw the distribution of sums for rolling 3, 4, and 5 dice to see what happens.
- How would you convince yourself that a distribution is "normal enough"?

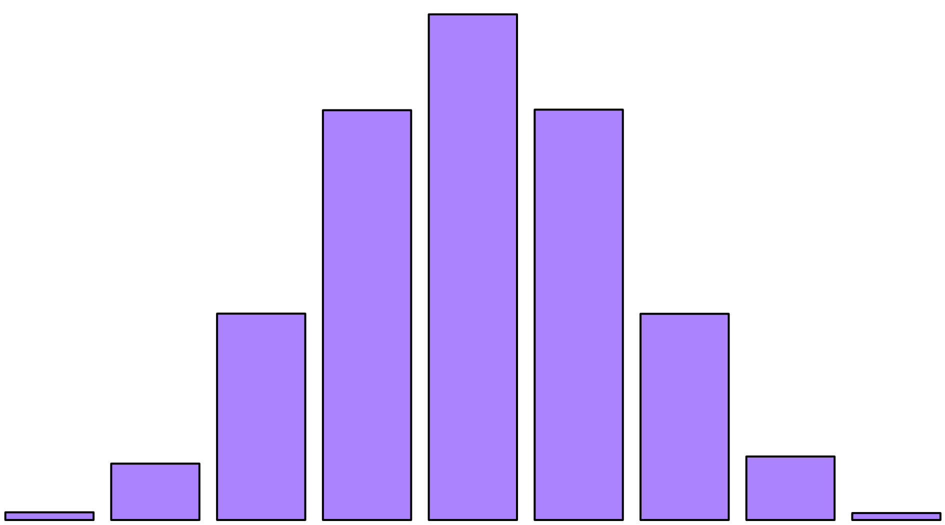
Bean machines

One way in which normal distributions arise quite naturally is as the result of many, many binary choices. To be more precise, a sequence of n binary choices gives a binomial distribution, which for large enough n can be well approximated by a normal distribution.

A Bean Machine (sometimes also called a quincunx, or a Galton Board) is a vertical board with several rows of round pegs. A bead that is released at the top of the board will hit a peg in the first row and will be deflected either to the left or to the right with a 50:50 probability. It will then hit a peg in the second row, where again there is a left-right decision, and so on. At the bottom, there are bins in which beads are collected, so that they form a "physical histogram".

A demonstration of this can be found in the following video (by Matemateca (IME/USP)/Rodrigo Tetsuo Argenton, 2016, CC-BY-SA 4.0, via Wikimedia commons): https://en.wikipedia.org/wiki/File:Galton_box.webm

- Let's make our own bean machine! Let's assume there are 8 layers of pegs. We don't need to simulate the beads physically going through the machine. All we need to know is that each bead has to make 8 left/right decisions and that we want to record the position of the bead at the end of those decisions. At the end, we want something like this:



- Why does the outcome of a bean machine look like a normal distribution if there are enough layers of pegs?
- What would happen if the probability to go left or right at each peg was not 50:50? For instance, what if there was an 80% probability to go left and only a 20% probability to go right at each peg?

Class grades

In a previous problem set, we modelled the grades that students received on a multiple-choice test as a normal distribution. We added as a caveat that this is an assumption, and that the distribution could have looked different.

- From what you have learned this week, do you think test scores can be modelled as normally distributed? Why or why not? Think about how a test score is computed, and work from there.

Last update by DJ MacGregor in 2022