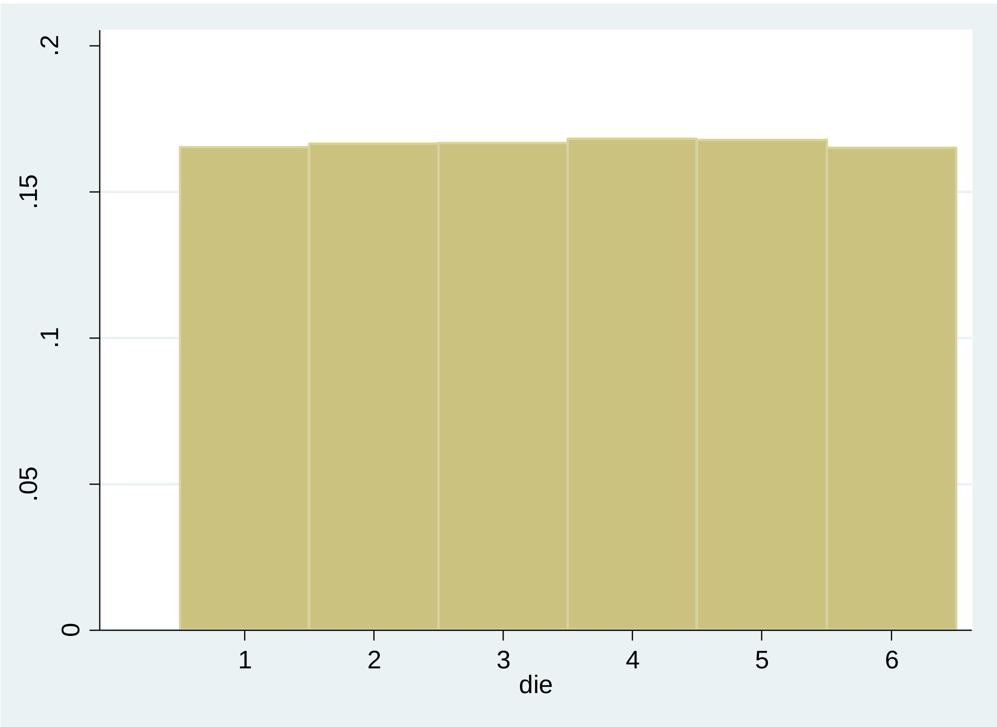
Class exercise on Central Limit Theorem

Students often struggle to understand the difference between a sample distribution and a sampling distribution. The aims of this exercise are:

1) To understand what a sampling distribution is
2) To see that as the central limit theorem states, the sampling distribution of the mean of a variable is approximately normal, even if the population distribution is not.

The exercise builds on variable familiar to students: the number of dots on dice.
I first show a distribution of a (very) large sample.

With a fair die each value should come up 1/6 = .17
Distribution of 50,000 rolls of a die (raw scores)



Then I challenge students to estimate the sample mean:

If you would do a set of 6 dice rolls, what do you expect the mean number of dots to be?

Given that each number of dots has an equal probability, the best estimate for the mean is that each number shows up once:

$$\frac{1+2+3+4+5+6}{6} = 3.5$$

The best estimate of the mean number of dots across 6 dice rolls is 3.5
Once this prediction is accepted by all students, we move on to an empirical test.

Students are given a die and asked to draw three samples of 6 rolls, calculate the mean of each sample and write it down on the post-its (each mean on one post-it). To help them calculate the means, they are given a form.

Exercise:

- 3 sets (samples) of 6 dice rolls
- Write down the number of eyes for each roll
- Write down the sum & mean for each 'sample' of 3 rolls
- Write down the mean of each 'sample' on a post-it
- Paste the post-it on the axis on the whiteboard

sum	mean	sum	mean
6	1.00	22	3.67
7	1.17	23	3.83
8	1.33	24	4.00
9	1.50	25	4.17
10	1.67	26	4.33
11	1.83	27	4.50
12	2.00	28	4.67
13	2.17	29	4.83
14	2.33	30	5.00
15	2.50	31	5.17
16	2.67	32	5.33
17	2.83	33	5.50
18	3.00	34	5.67
19	3.17	35	5.83
20	3.33	36	6.00
21	3.50		

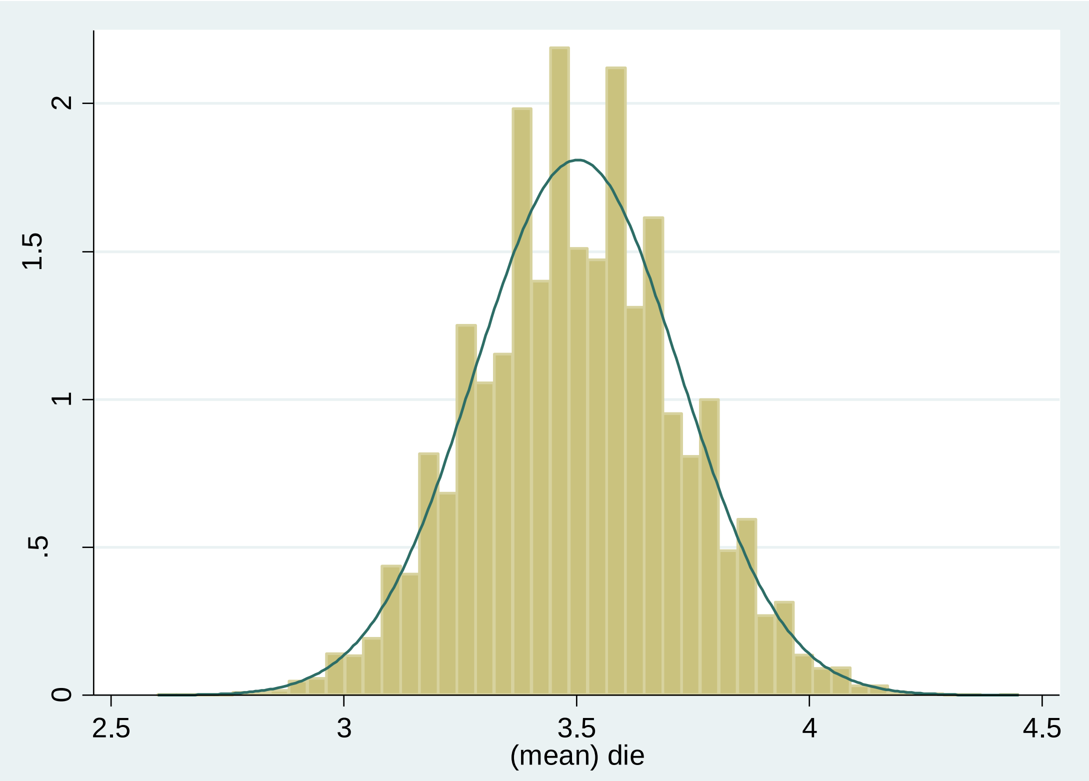
	round 1	round 2	round 3
1			
2			
3			
4			
5			
6			
sum			
mean			

The result will look something like this:

The central limit theorem works when the sample is 'sufficiently large' – generally 30 or more. However the samples used in this exercise are smaller (N=6) to remain closer to students' intuition and keep the exercise manageable (the average of 30 dice rolls is more difficult to calculate and takes a lot of time to generate by hand). Even with a sample size of 6 the sampling distribution takes on a shape resembling the normal distribution and thus provides a good illustration of how the central limit theorem works.

Distribution of mean scores of 50,000 <u>series of 60 rolls</u> (i.e. samples of n=60) of a die (sample means)

Population mean for a fair die = 3.5



Central limit theorem: <u>sampling distribution</u> of means will approximate a <u>normal distribution</u>, even if the distribution of scores is not normal, as long as the sample size n ≥30

But…only if the sample is random!

Mean of the sampling distribution = the population mean

SD of the sampling distribution = <u>*standard error (se)*</u>

I have done this exercise in classes of up to 31 students. For large groups it might not be feasible to hand out dice and ask students to all walk up to the board. Instead students could use coin flips and use an online survey and interactive lecture tool such as Shakespeak to send their results to the lecturer.