Sampling distributions

Some illustrative examples

Sonja Petrović
Created for Math 563

Jan 21, 2020.

Context

Statistics:

- Functions of random variables
- Therefore, are random variables themselves.
  - In particular, they have their own distributions, called sampling distributions.

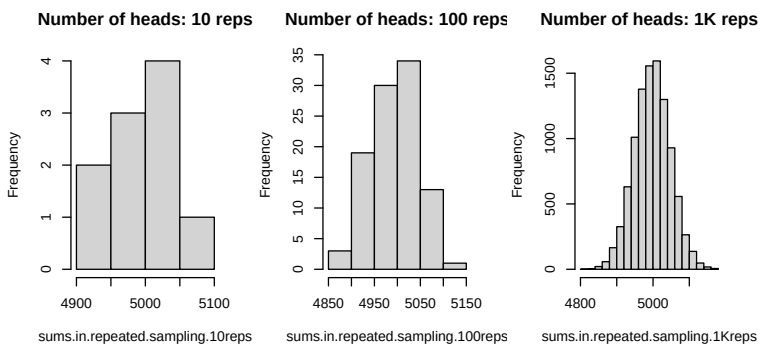
Context

Statistics:

- Functions of random variables
- Therefore, are random variables themselves.
  - In particular, they have their own distributions, called sampling distributions.

Simulation of a coin toss

- Let x be the random variable recording the outcome of a coin toss:
  - $x_i = 0$ if we see Tail on the i-th trial (toss),
  - $x_i = 1$ if we see heads on the i-th trial.
- Fix $n = 10000$.
- $Y =$ the number of heads.
  - Is the number of heads supposed to be $n/2$? How far off is it? Does it vary? What does this mean?



- Sampling distribution of Y appears to have a mean around *the expected number of heads when a fair coin is tossed*, which is about $n/2$.
- The more times we repeat the experiment of n coin tosses, the closer Y gets to its expected value – this can be measured by looking at both the mean and the variance of Y.

Means of Y
4987.600
4995.380
4998.437

Vars of Y
2618.056
2618.056
2445.156

Question:

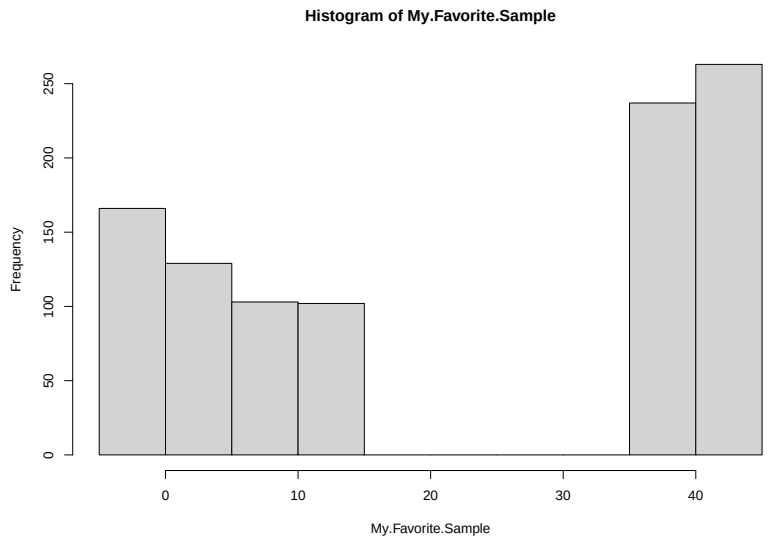
is it possible that something similar to this always happens?

- As we will see, the sampling distribution of Y is approximately *normal* with mean equal to the expected value of X.
- In other words, the example above illustrates a known result–**the Central Limit Theorem**, one of the cornerstone results used in inference.
- You should already be familiar with it from your probability class.

Importance of sampling distributions

- Sampling distributions tell a story about the model behind the data (i.e., the probability distribution or population from which the data was sampled);
- they give a glimpse into how it was generated.

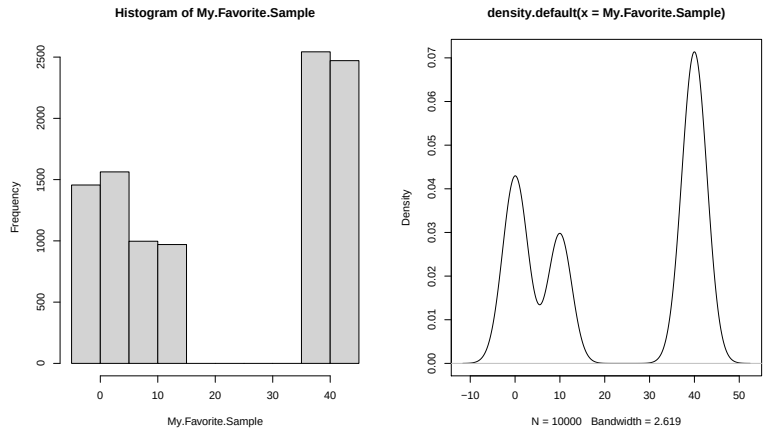
Example




```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-2.5415  0.8506 24.1903 22.0272 40.0701 42.6451
```

Hmm...

- Is it strange to see "two bumps" in the histogram instead of one, as usual?
- Maybe the sample size is too small, we need to simulate more data?

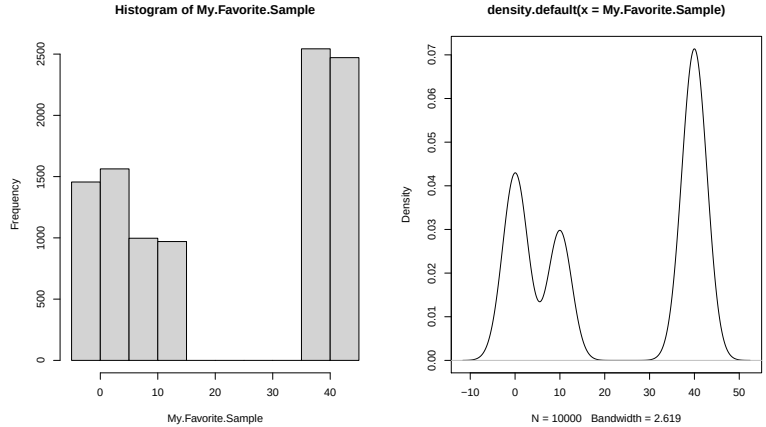


AhaMoment!

What do you see? → ahaslides.com quiz ←

Hmm...

- Is it strange to see "two bumps" in the histogram instead of one, as usual?
- Maybe the sample size is too small, we need to simulate more data?



AhaMoment!

What do you see? → ahaslides.com quiz ←

- This data is *not* being drawn from anything like a normal distribution.
- Consequently, knowing simply the mean and the variance ... is not enough to understand the data, that is, the data-generating mechanism behind it.

. . . Wait, what was that?!

This was an example of a **mixture** of normal distributions.

Check out the handout on Campuswire to see more, and ponder some important questions.

License

This document is created for Math 563, Spring 2021, at Illinois Tech.