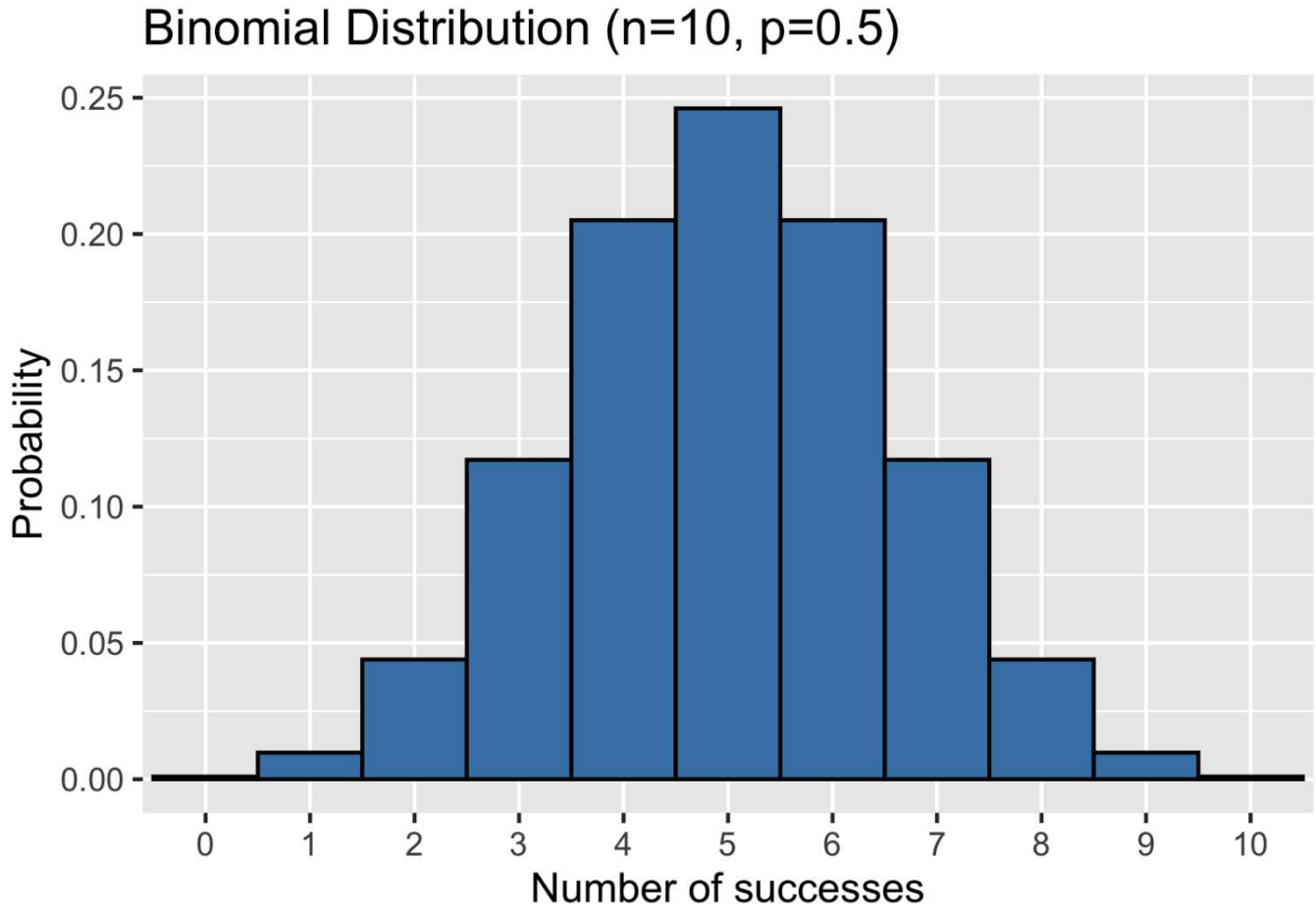
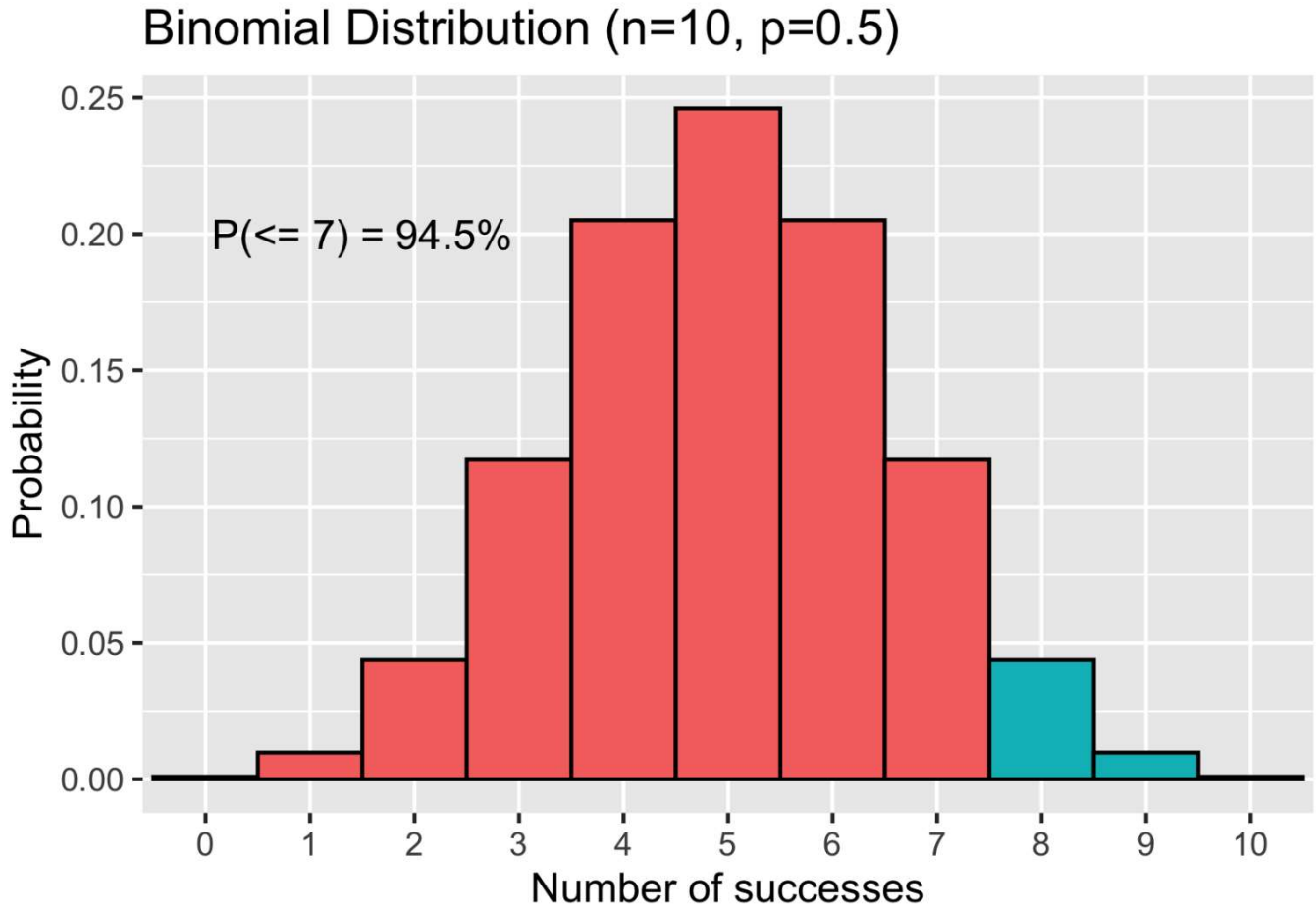
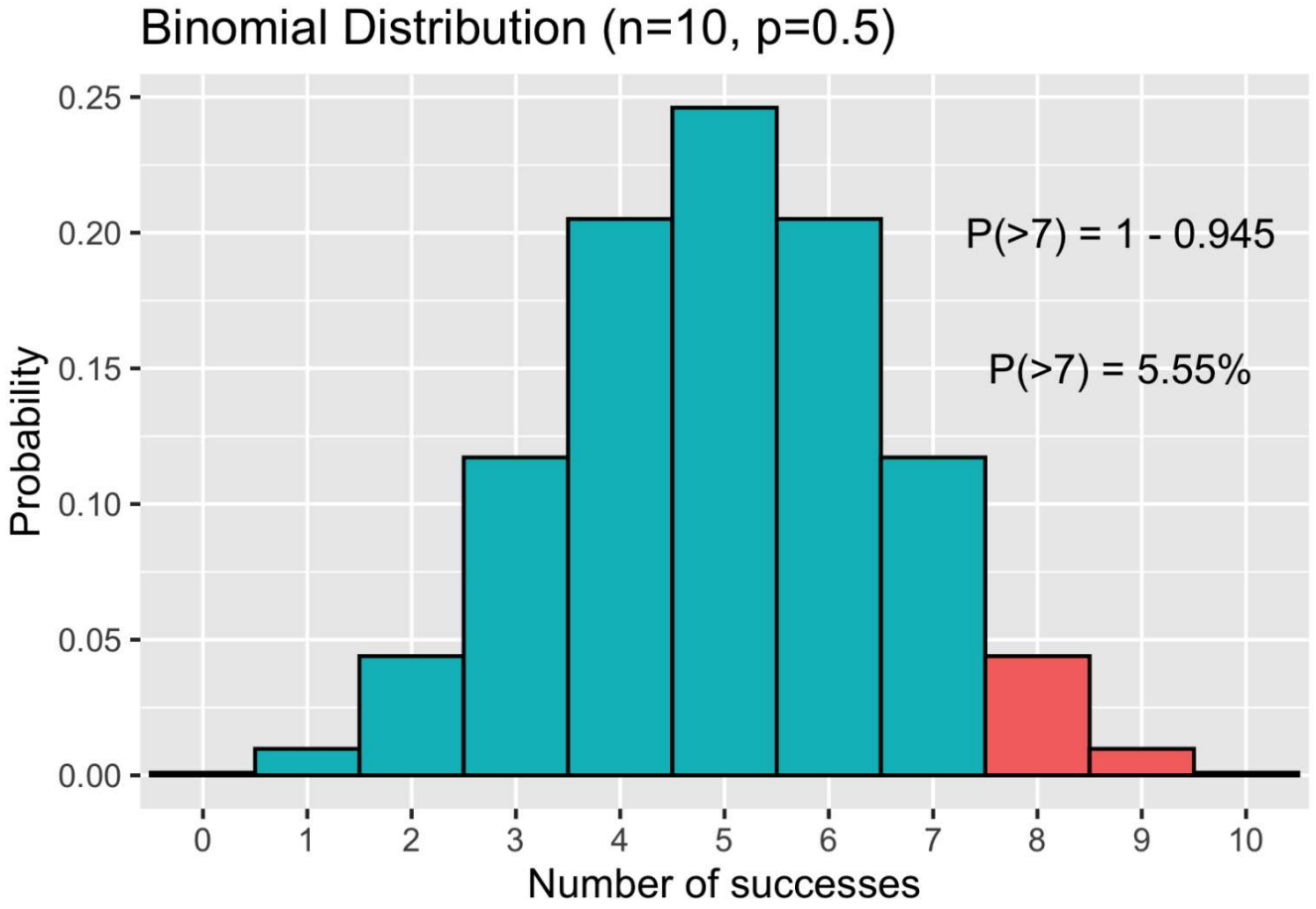
Binomial distribution



Probability of 7 or fewer heads



Probability of 8 or more heads



Expected value

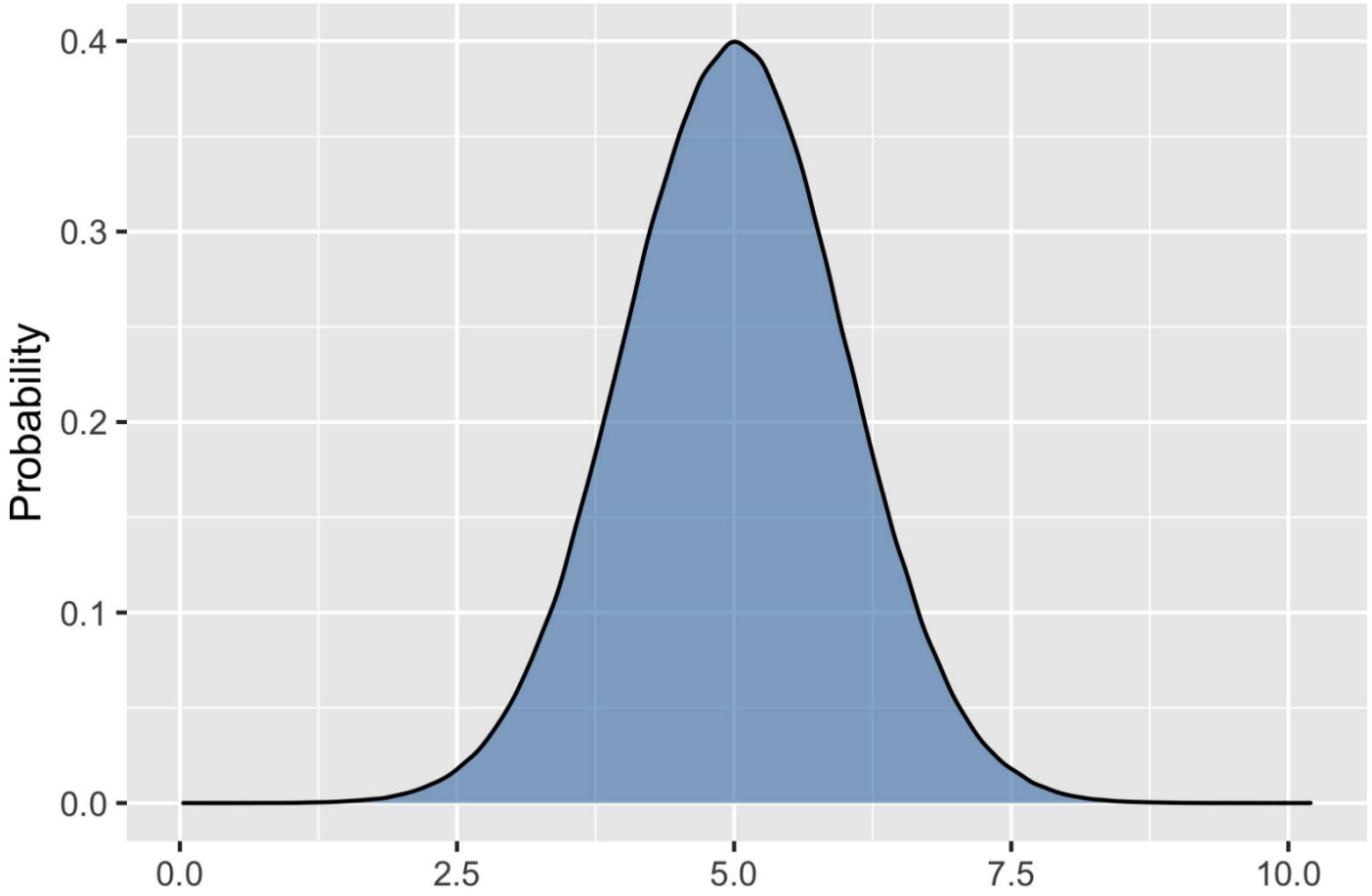
$$Expected\ value = n \times p$$

Expected number of heads out of 10 flips $= 10 \times 0.5 = 5$

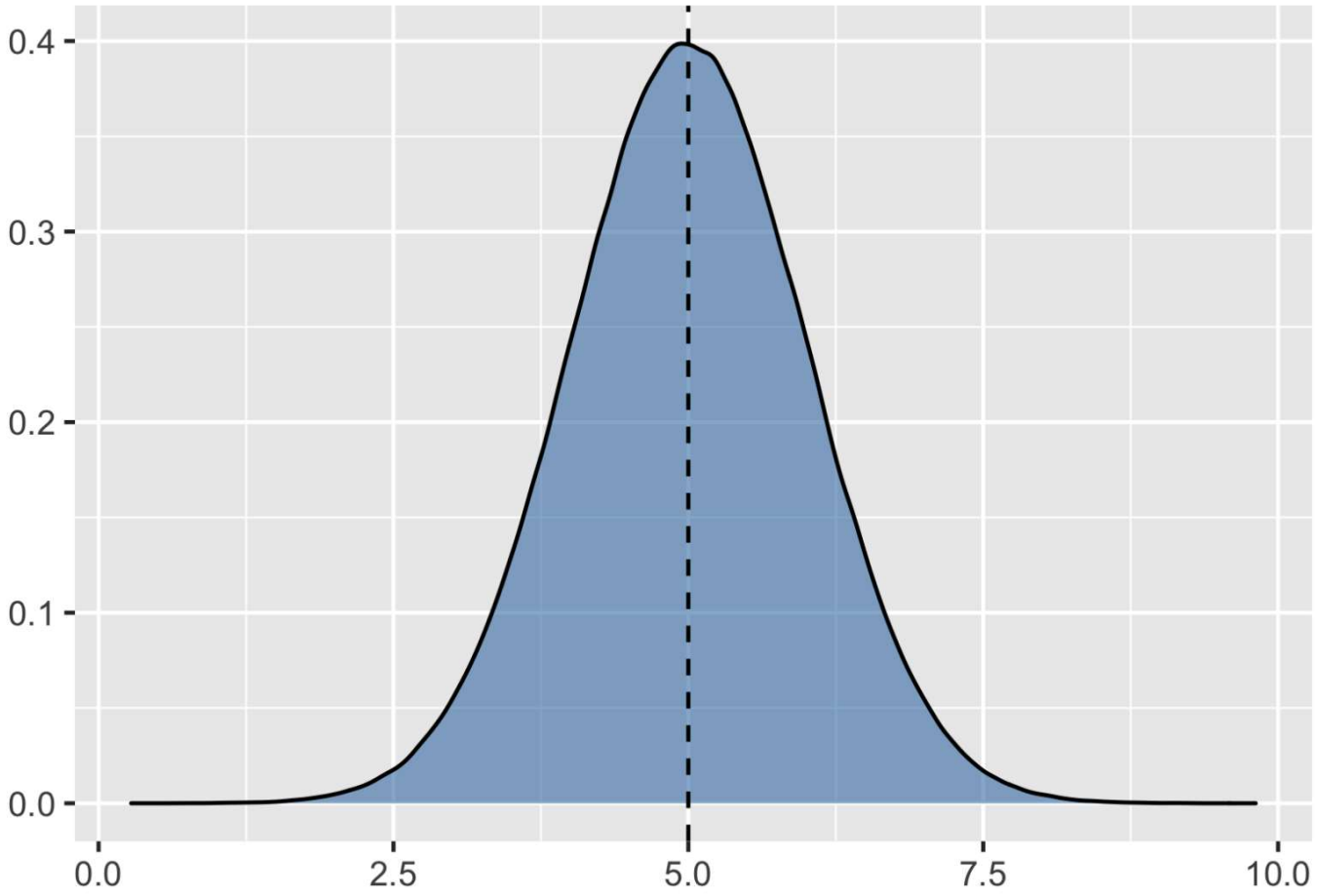
If we don't know p, but know n and the expected value:

$$p = \frac{expected\ value}{n}$$

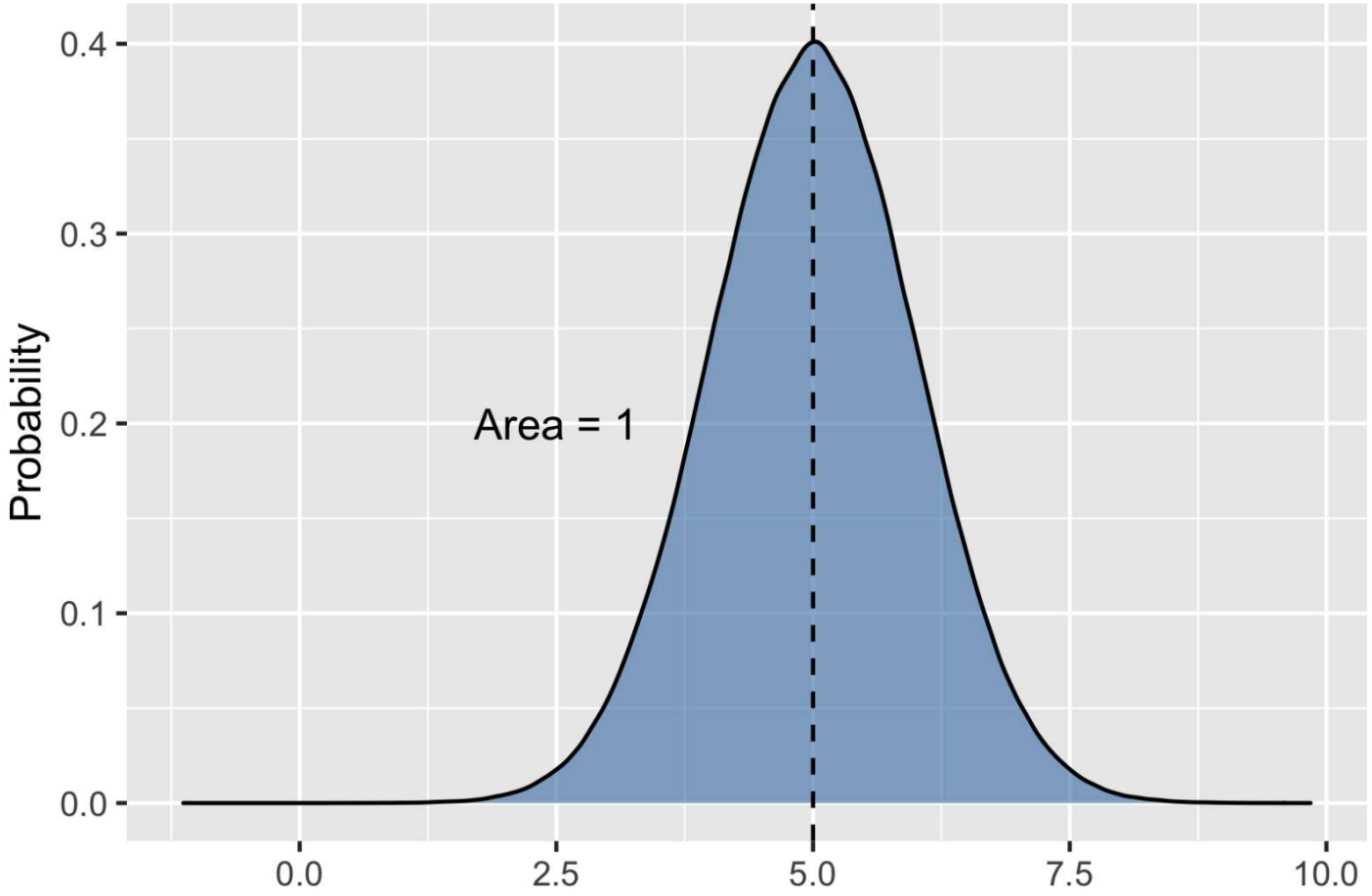
What is the normal distribution?



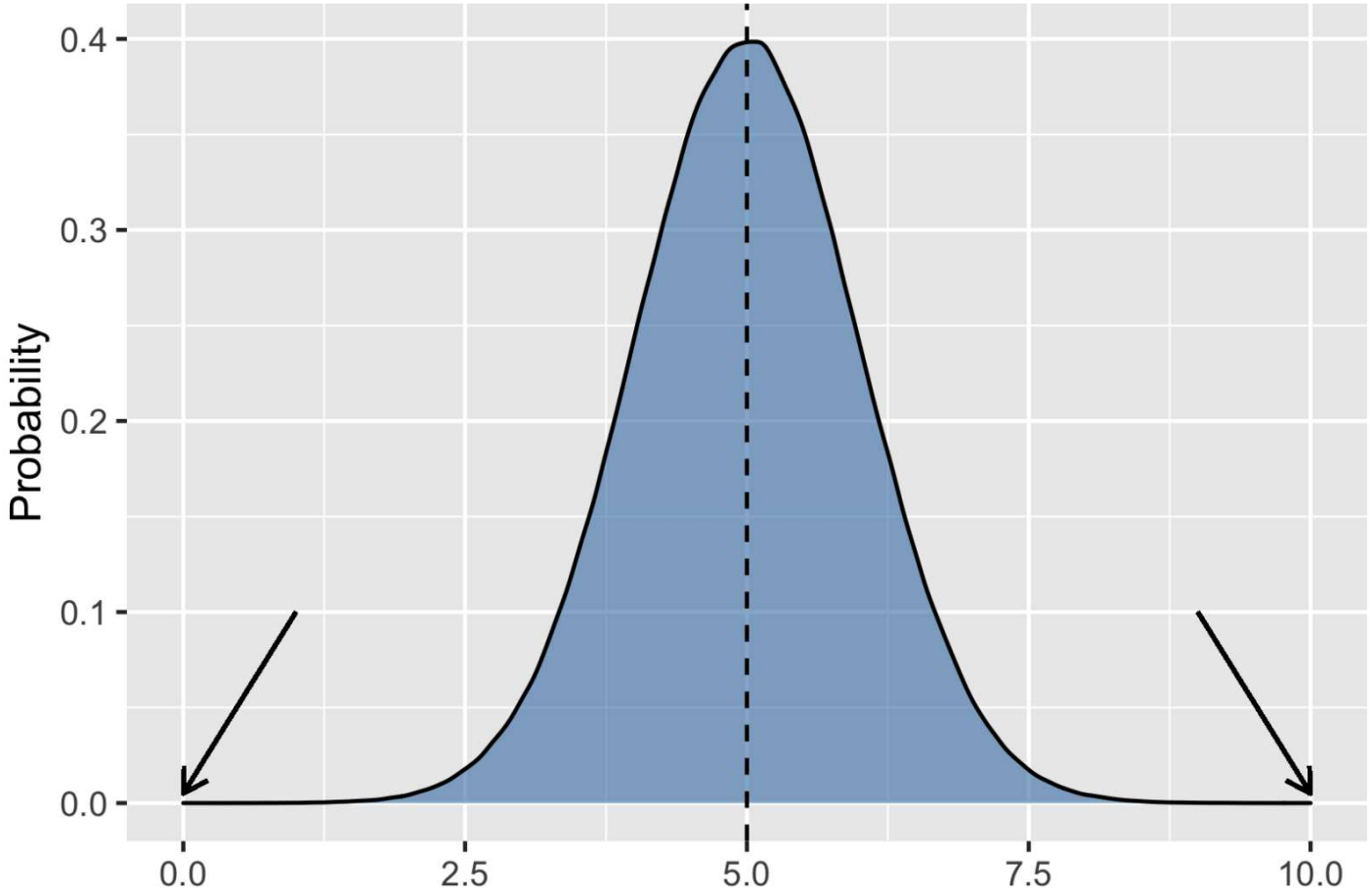
Symmetrical



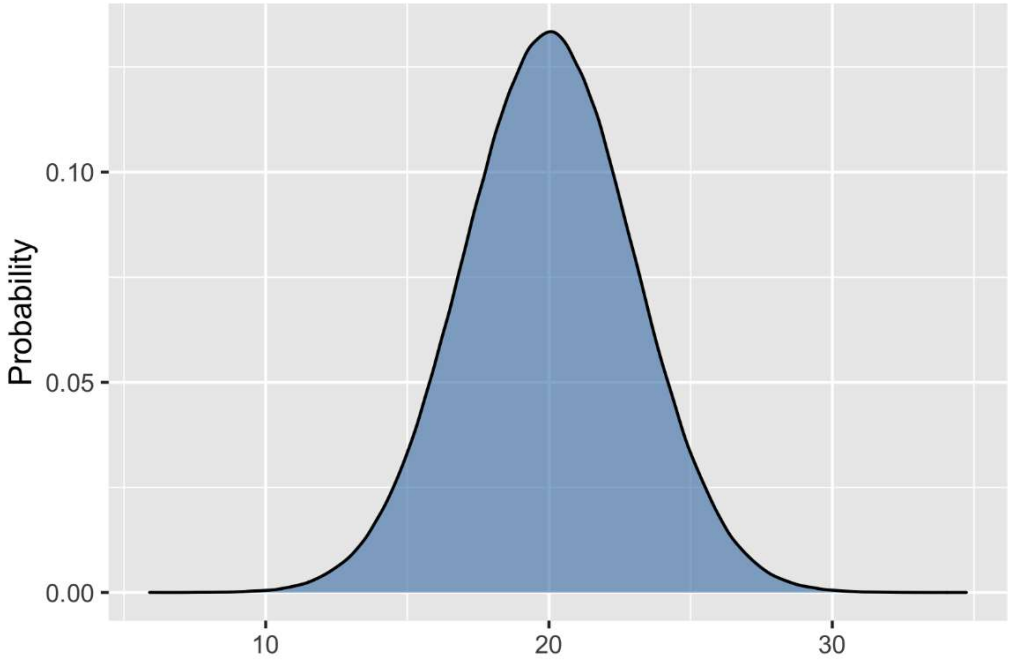
Area = 1

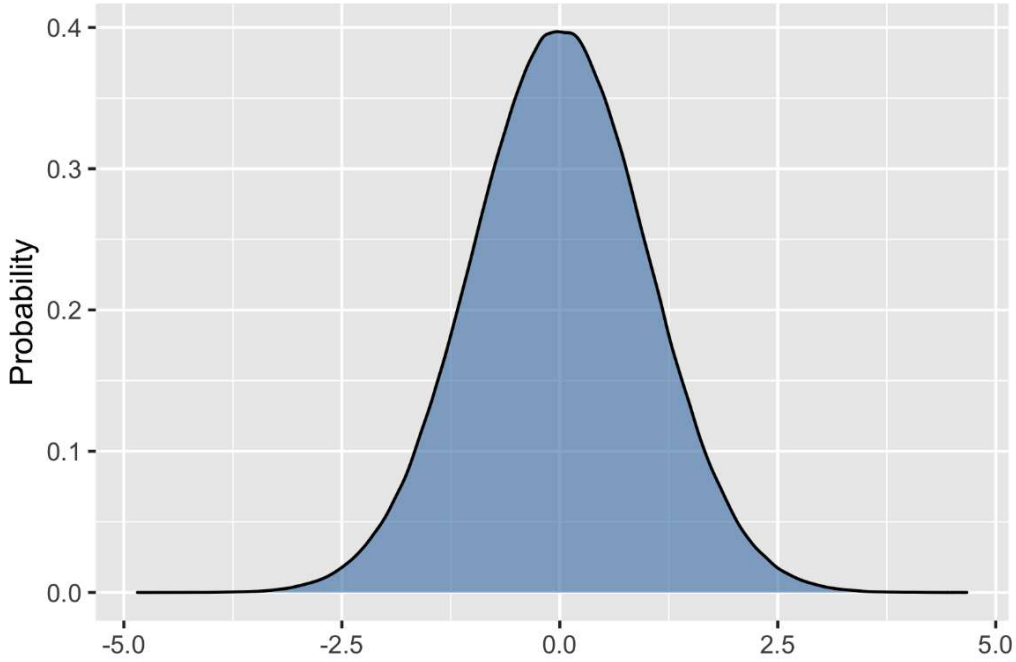


Curve never hits 0



Described by mean and standard deviation





$Mean = 20$

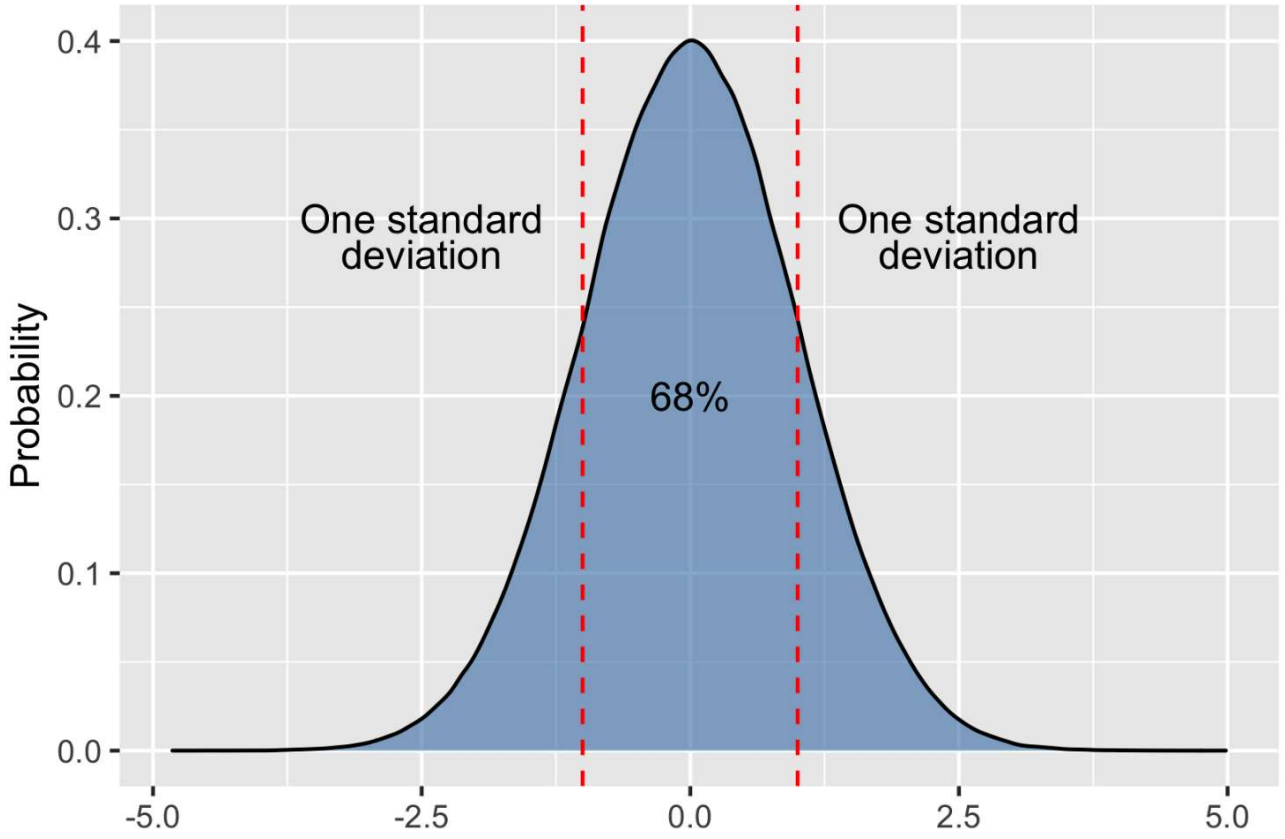
$Standard\ Deviation = 3$

$Mean = 0$

$Standard\ Deviation = 1$

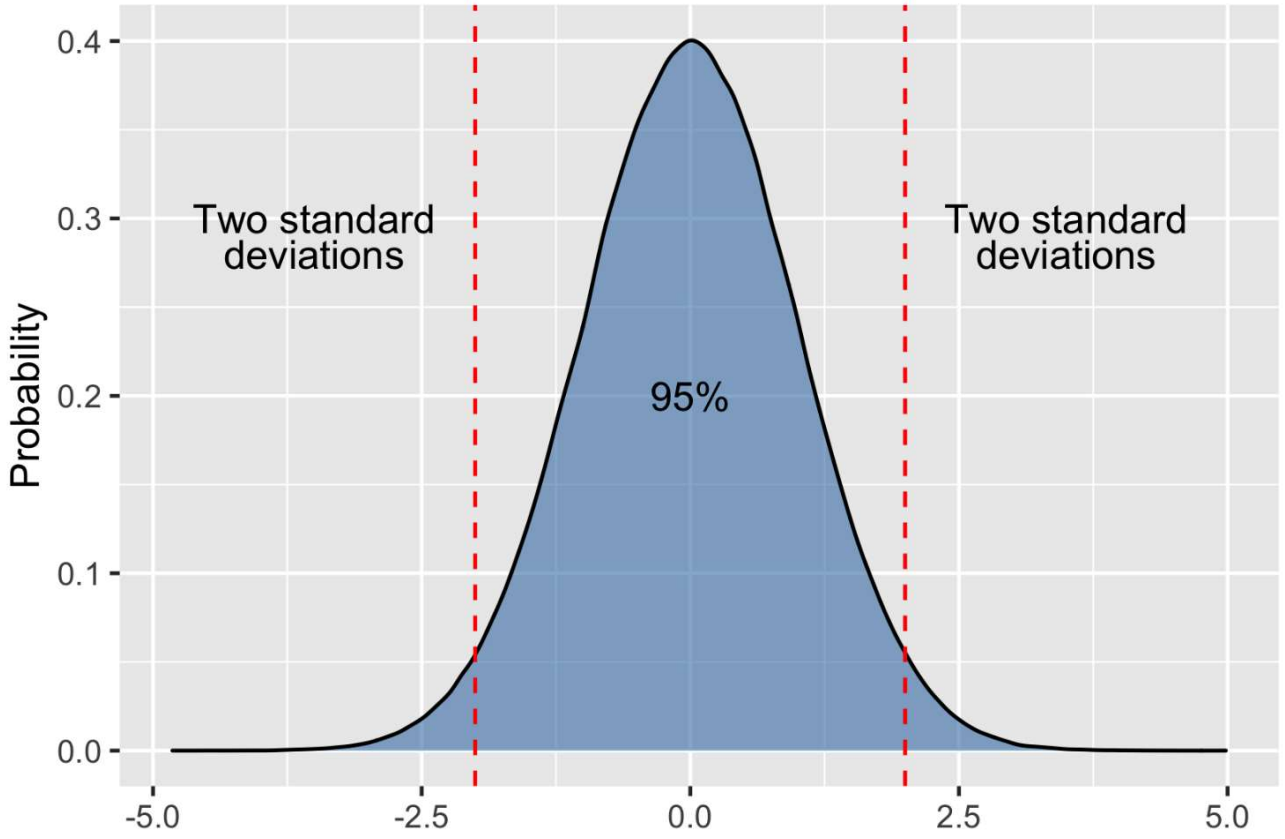
Areas under the normal distribution

68% falls within one standard deviation



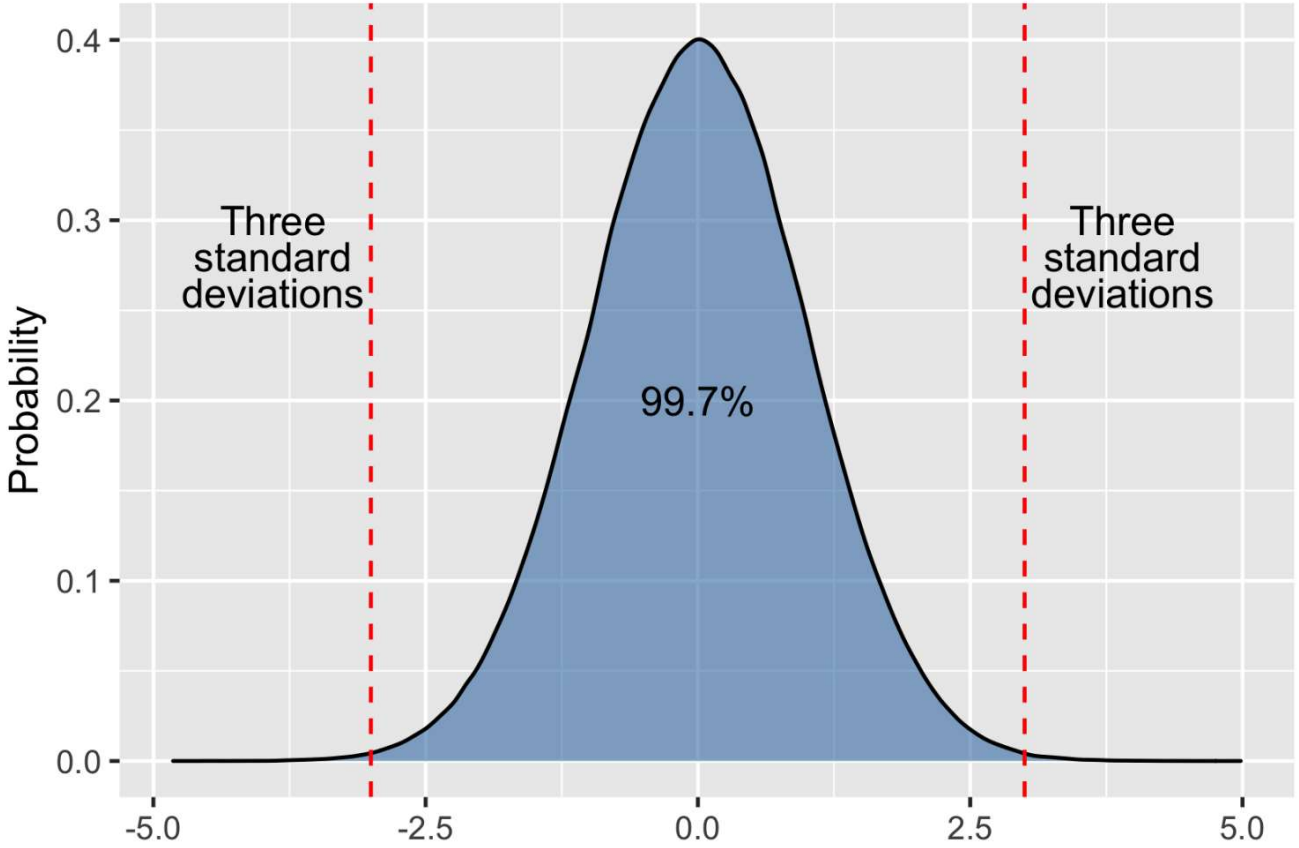
Areas under the normal distribution

95% falls within two standard deviations



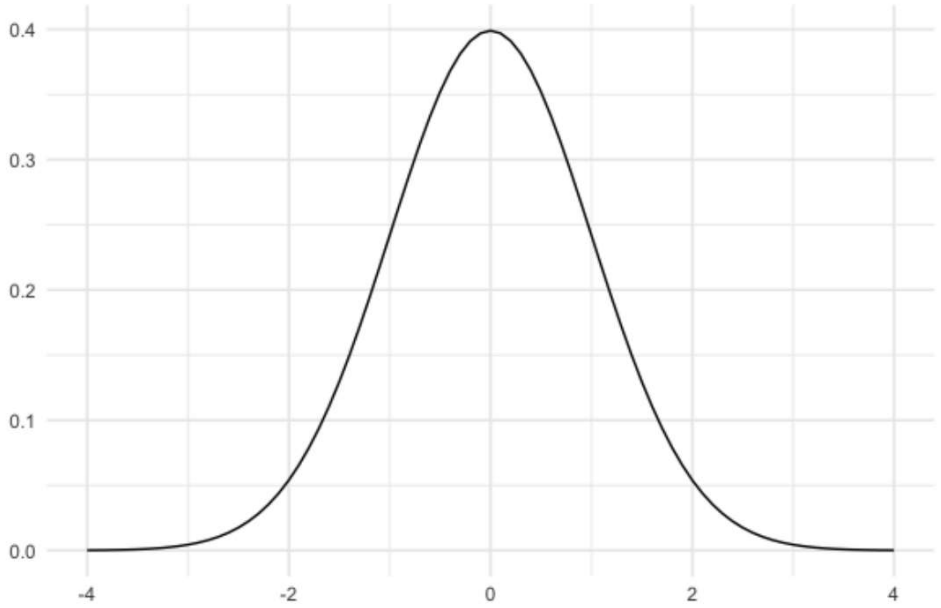
Areas under the normal distribution

99.7% falls within three standard deviations

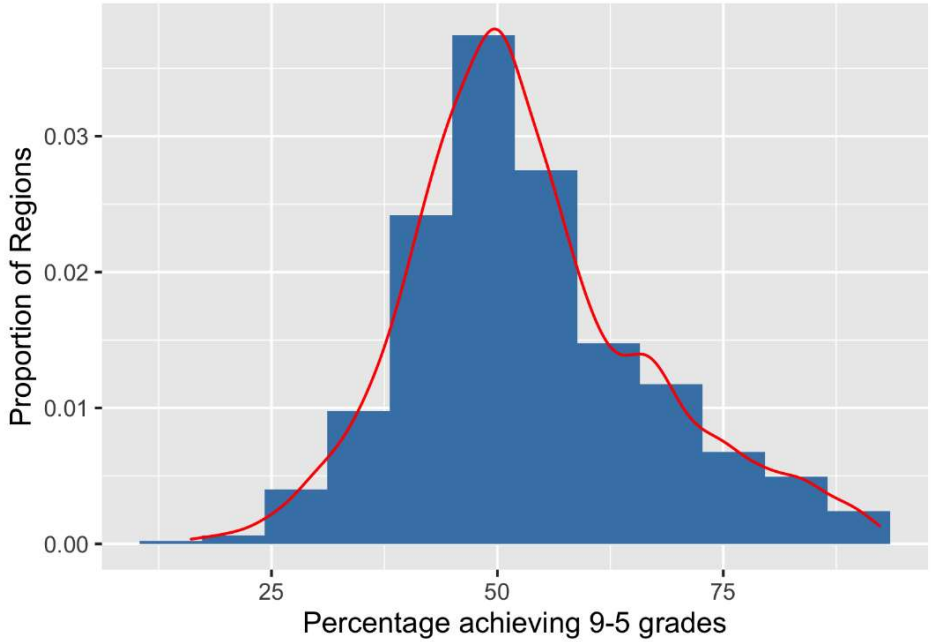


Why is the normal distribution important?

- Lots of real-world data resembles a normal distribution.

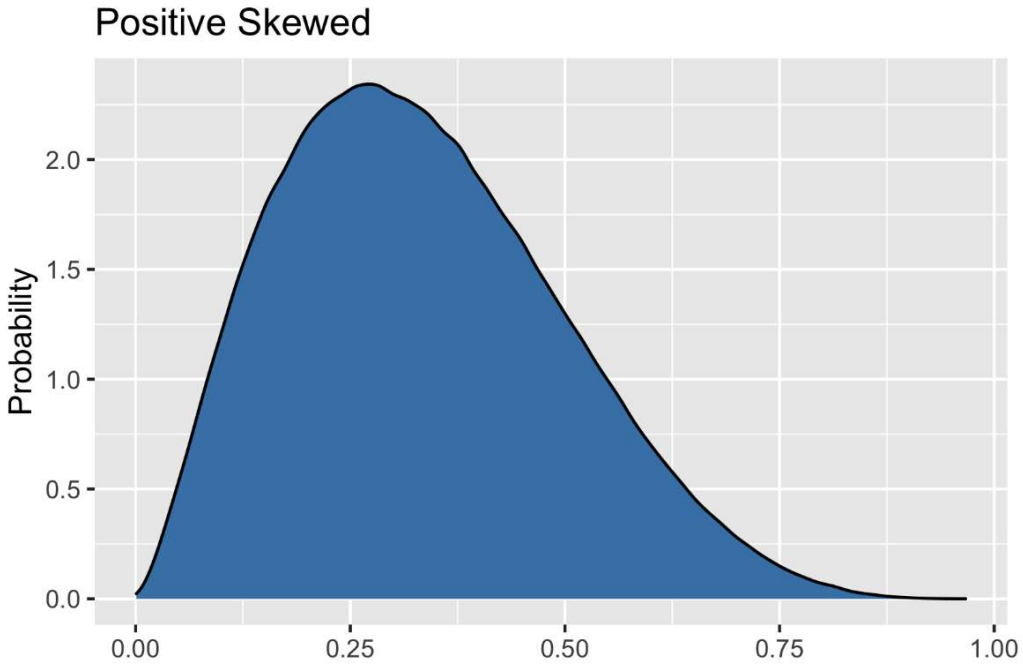


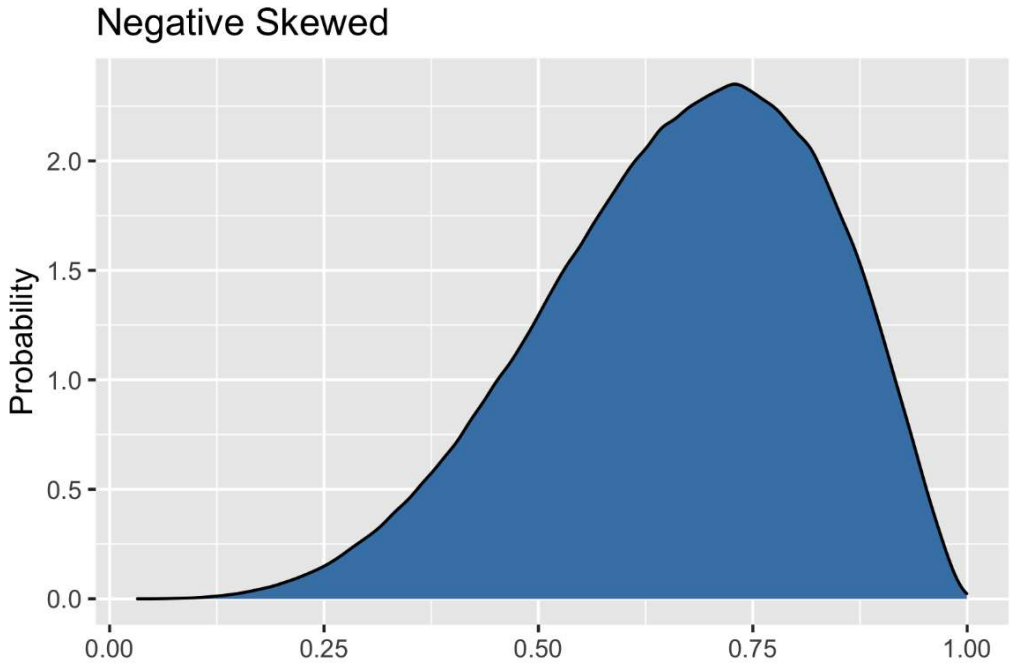
- A normal distribution is required for many statistical tests.



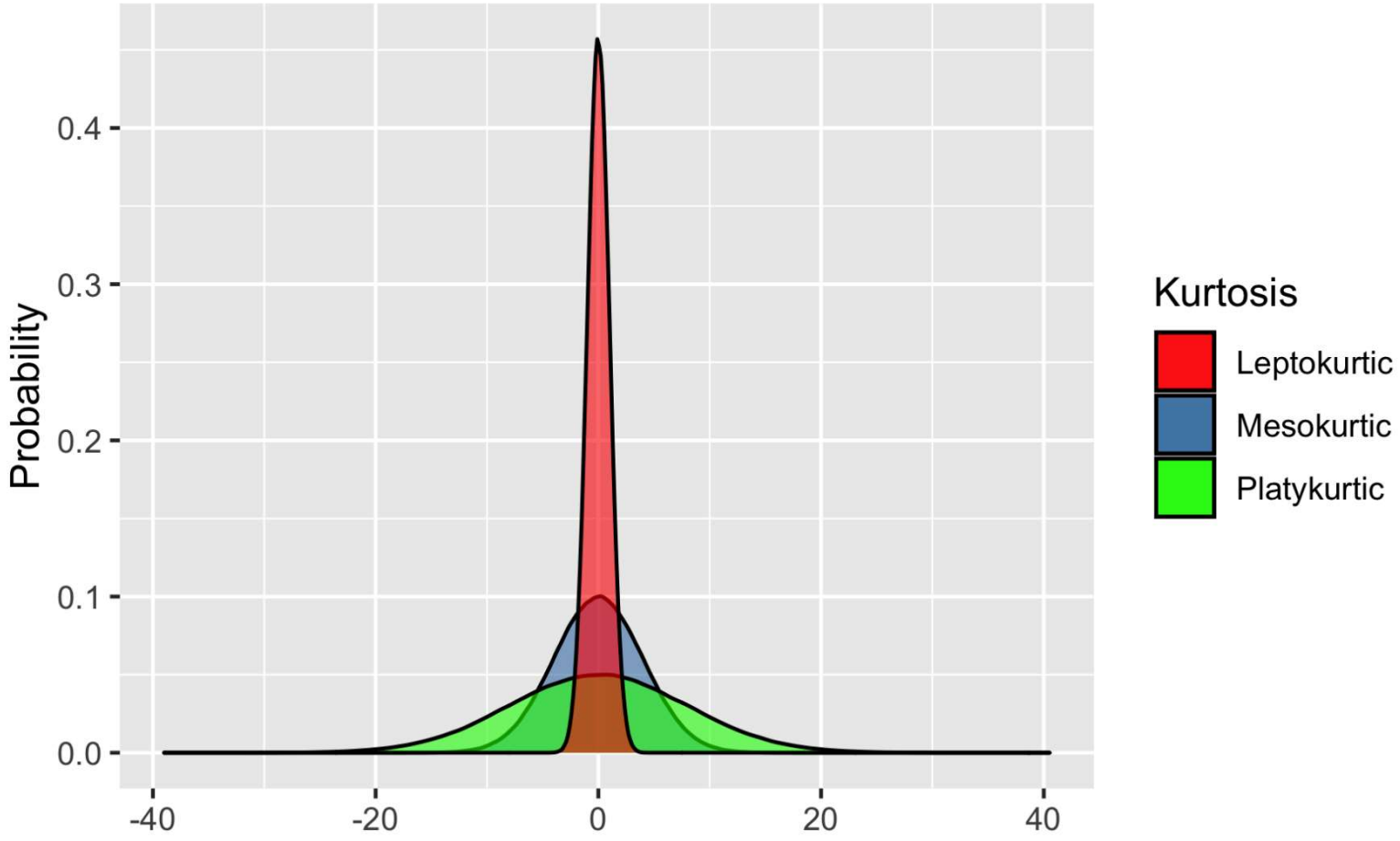
[1] Data source: https://data.gov.uk/dataset/ec1efd76-d6ad-4594-9b4d-944aa4170e63/gcse-english-and-maths-results-by-ethnicity

Skewness


Positive Skewed
Probability
2.0
1.5
1.0
0.5
0.0
0.00
0.25
0.50
0.75
1.00

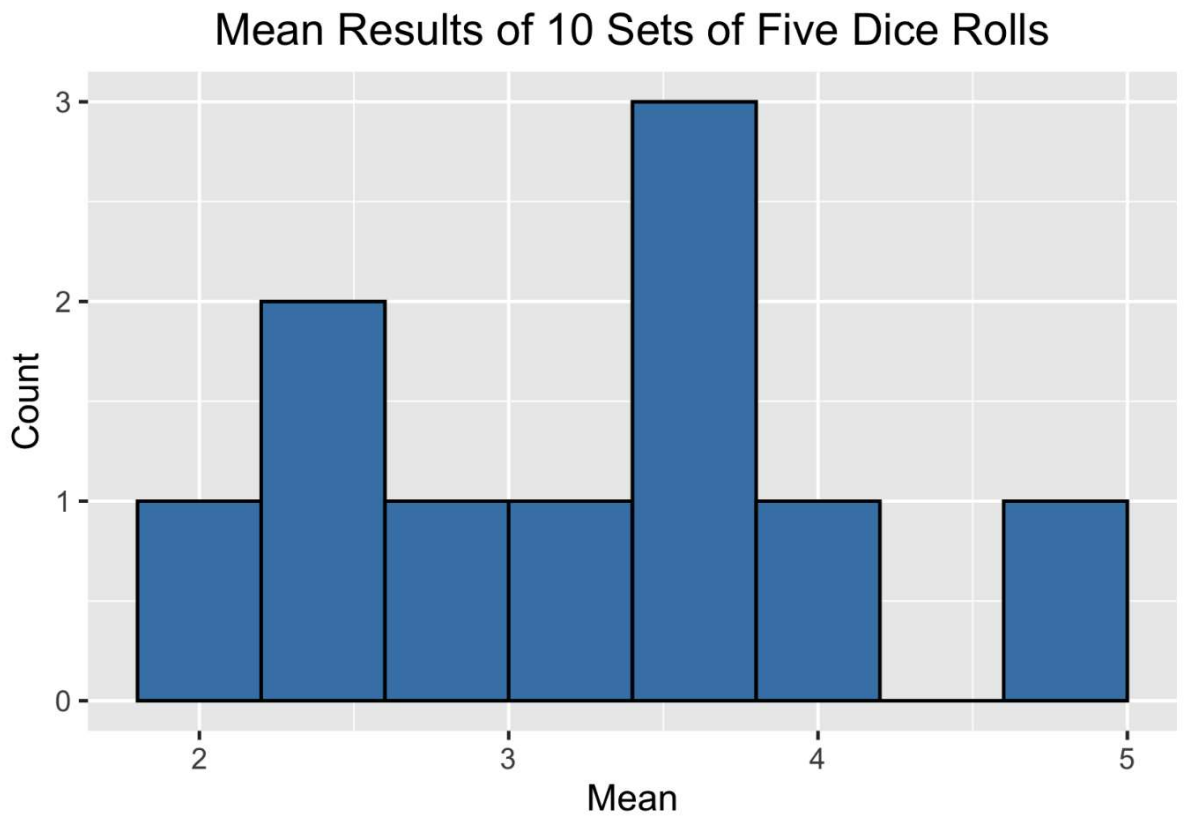

Negative Skewed
Probability
2.0
1.5
1.0
0.5
0.0
0.00
0.25
0.50
0.75
1.00

Kurtosis

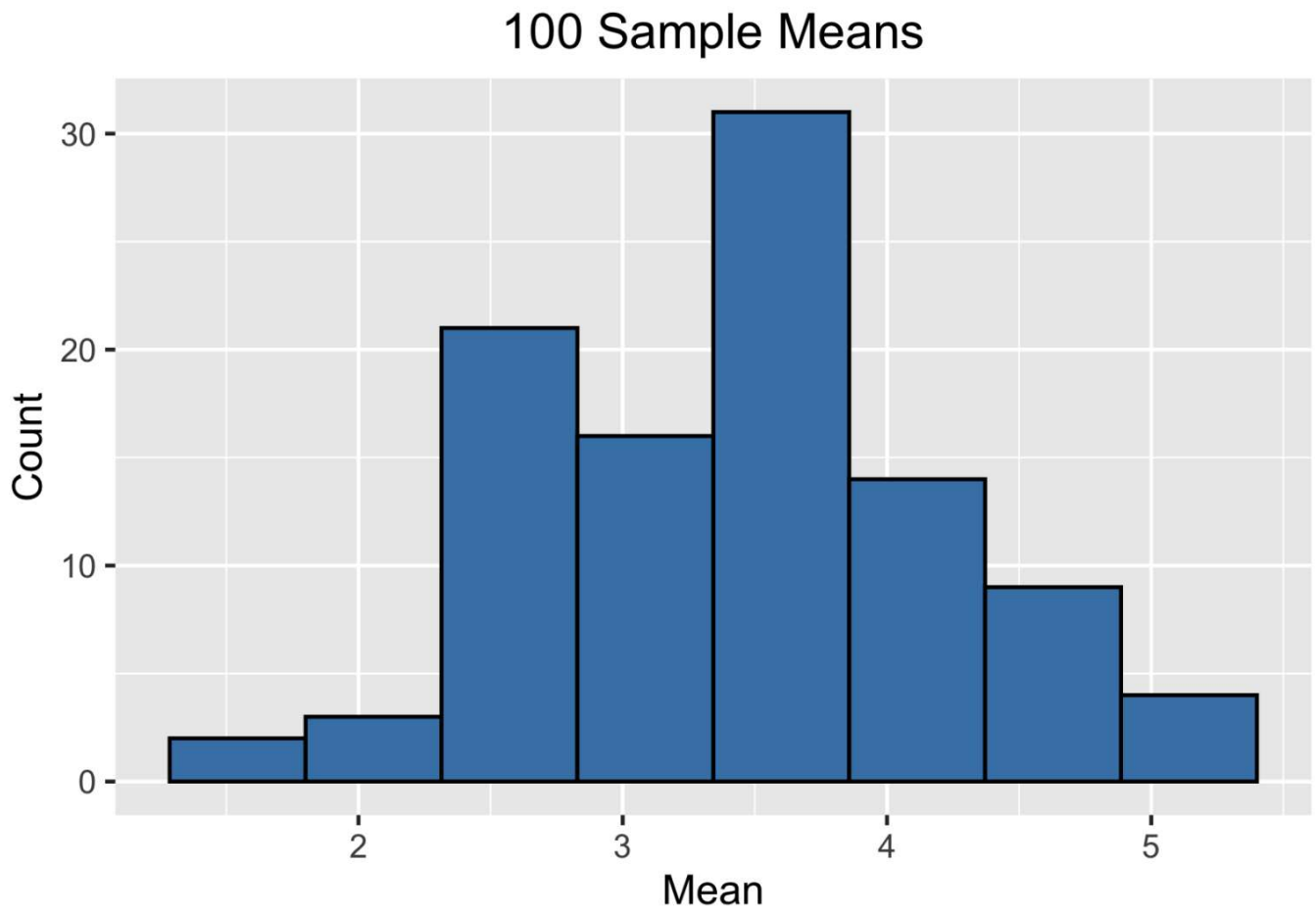

Probability
0.4
0.3
0.2
0.1
0.0
-40
-20
0
20
40
Kurtosis
Leptokurtic
Mesokurtic
Platykurtic

Sampling distributions

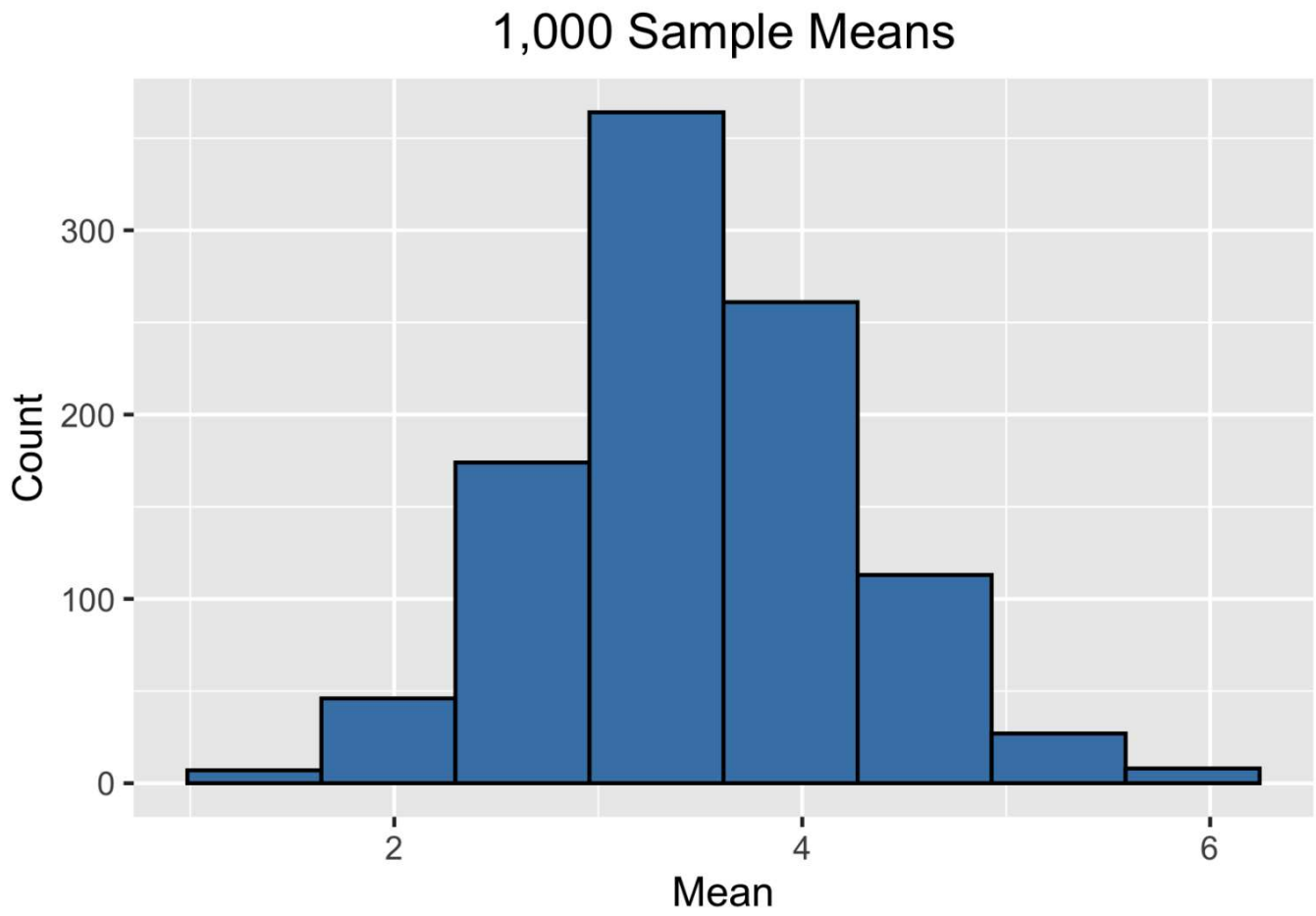
Sampling distribution of the sample mean



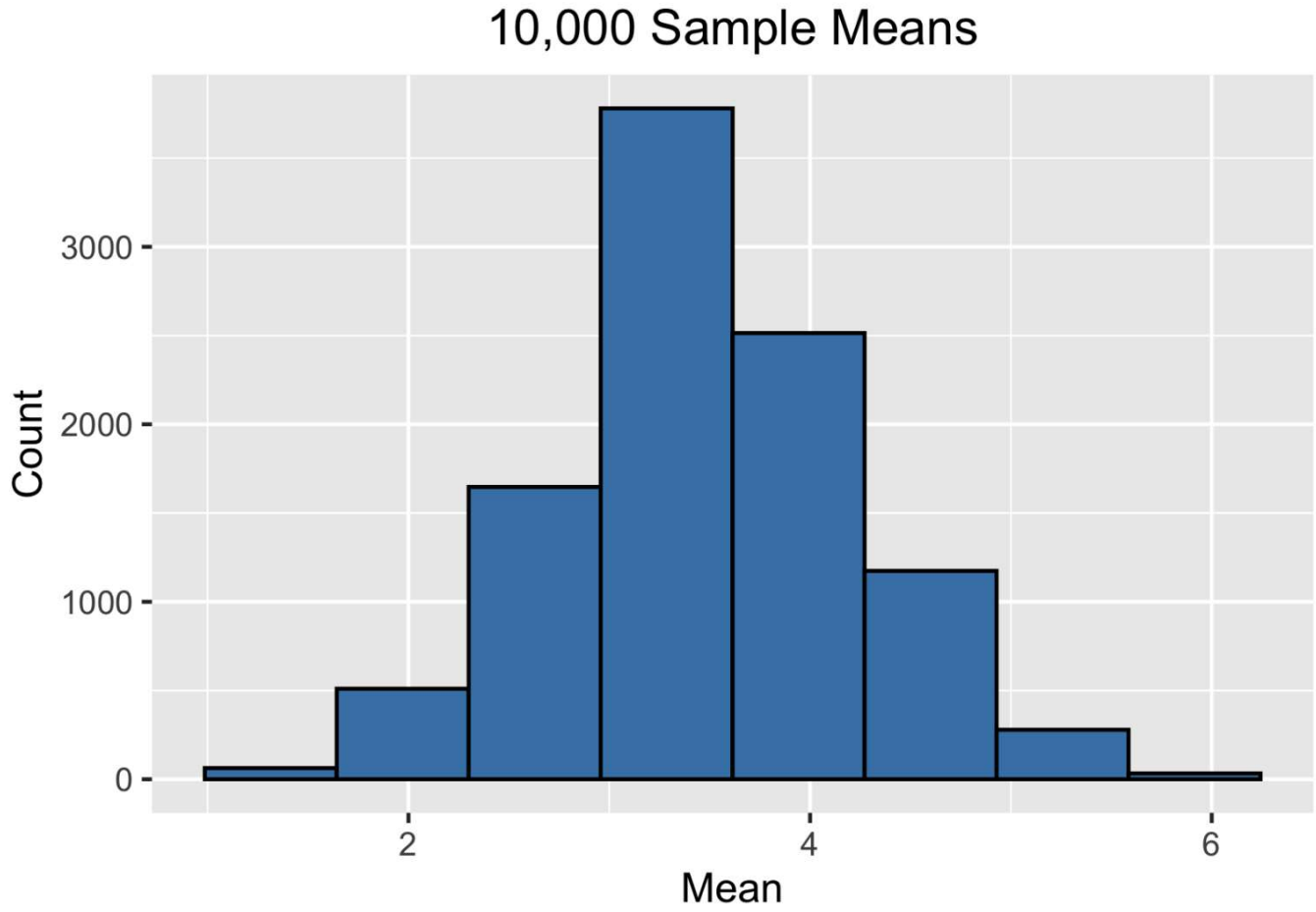
100 sample means



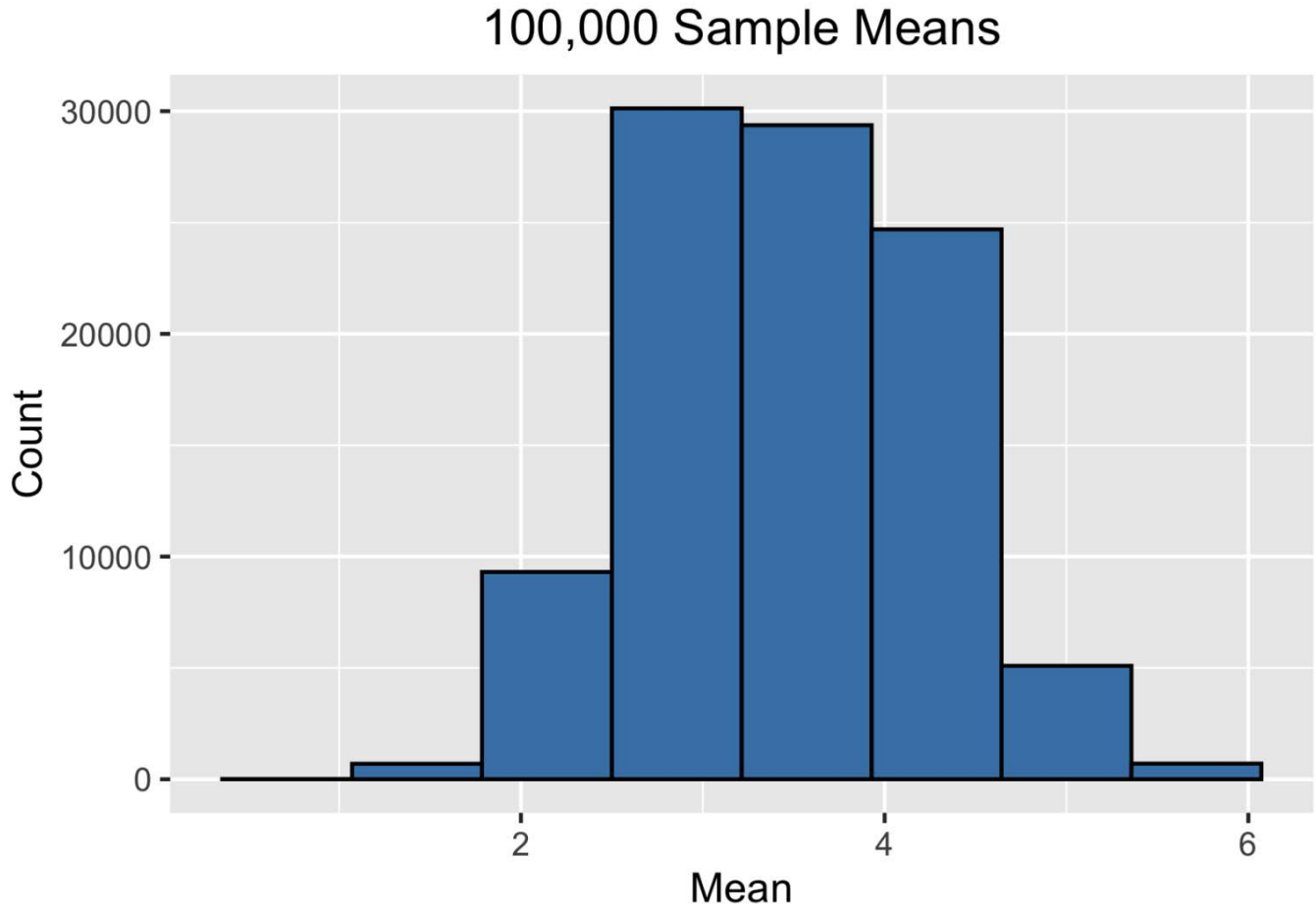
1000 sample means



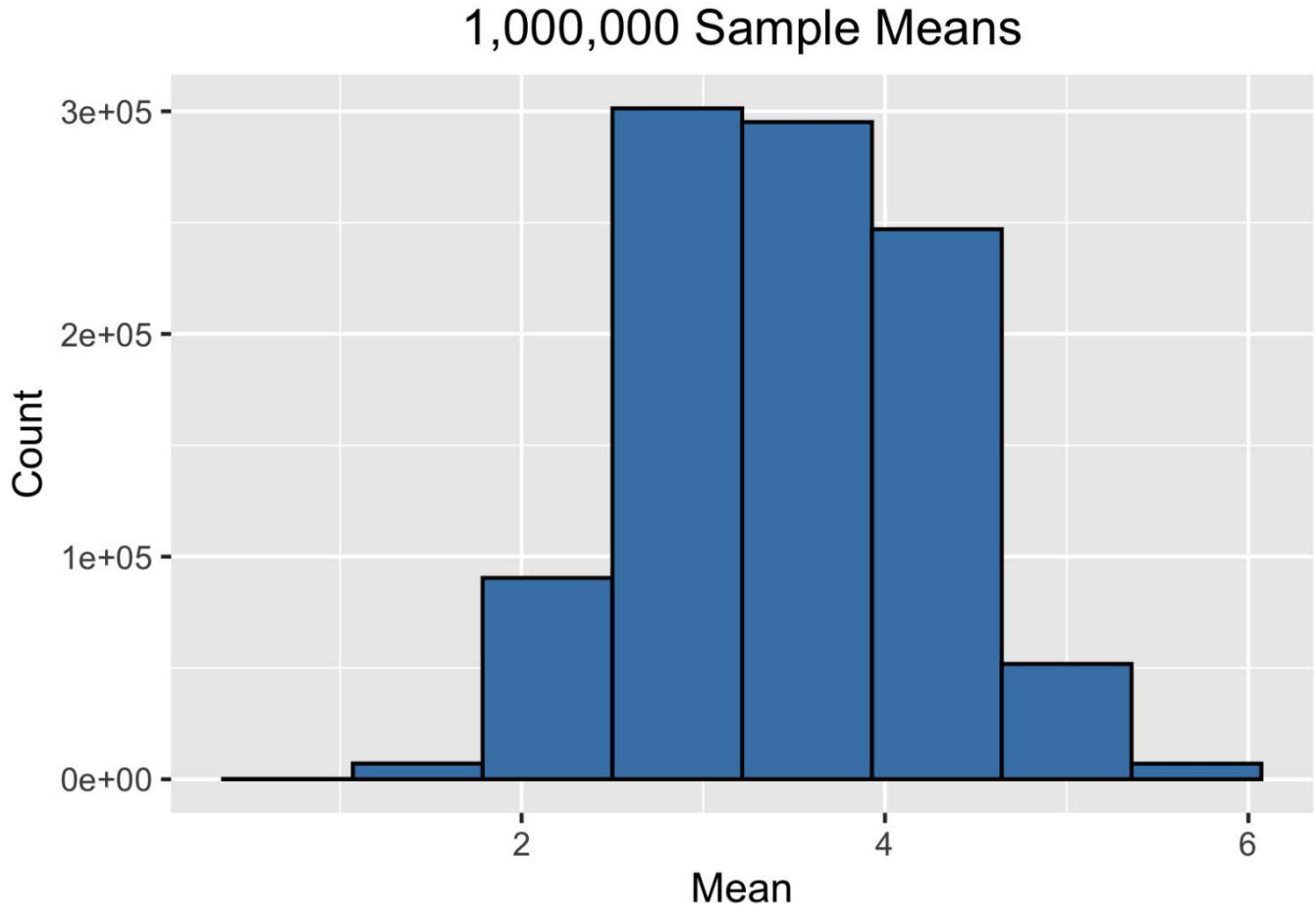
10000 sample means



100000 sample means

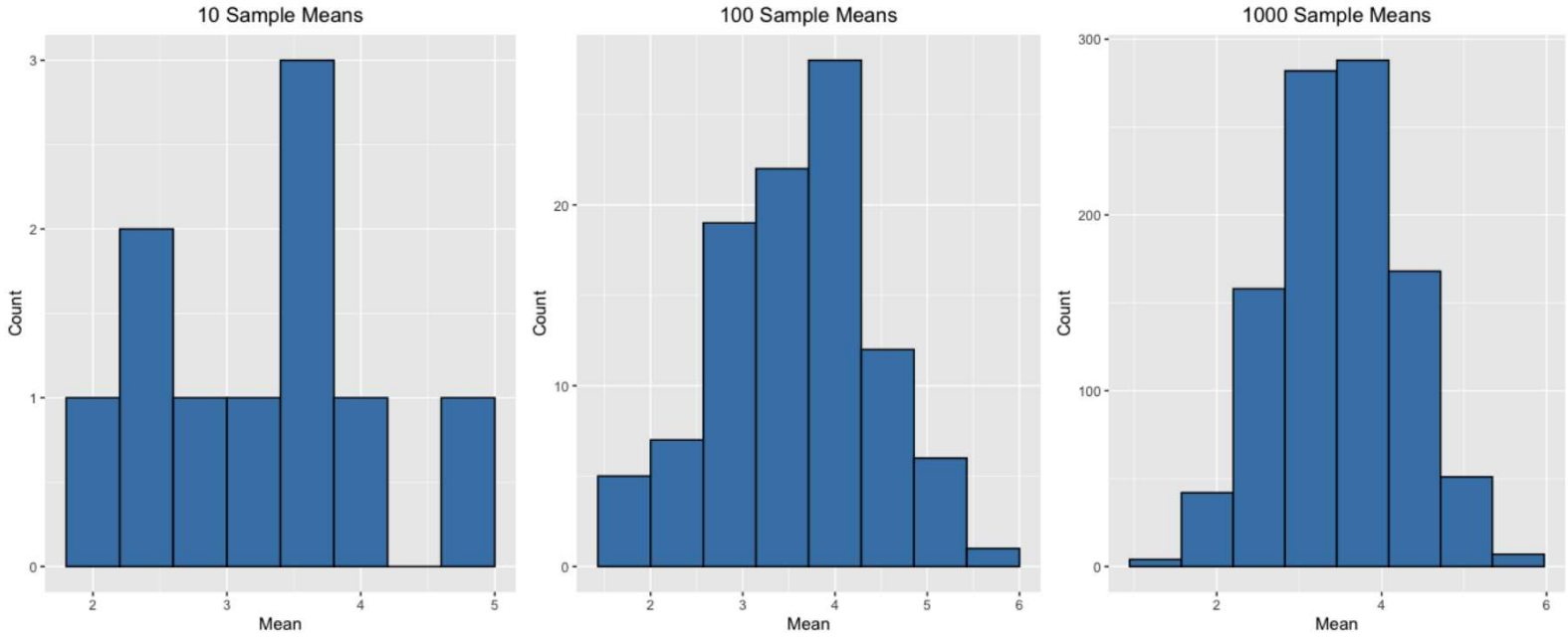


One million sample means



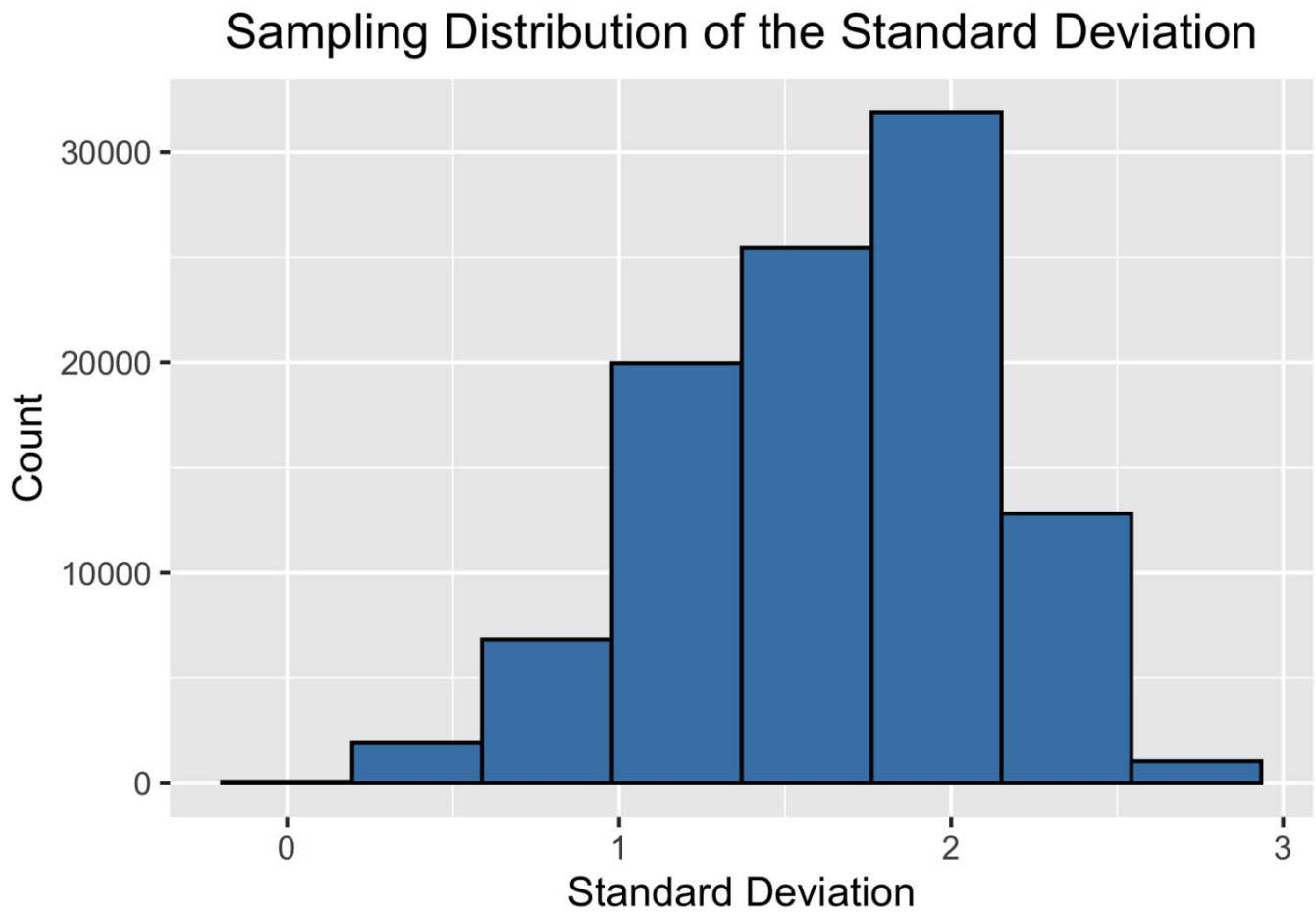
Central limit theorem

The sampling distribution of a statistic becomes closer to the normal distribution as the size of the sample increases.



** Samples should be random and independent*

Standard deviation and the CLT



Proportions and the CLT

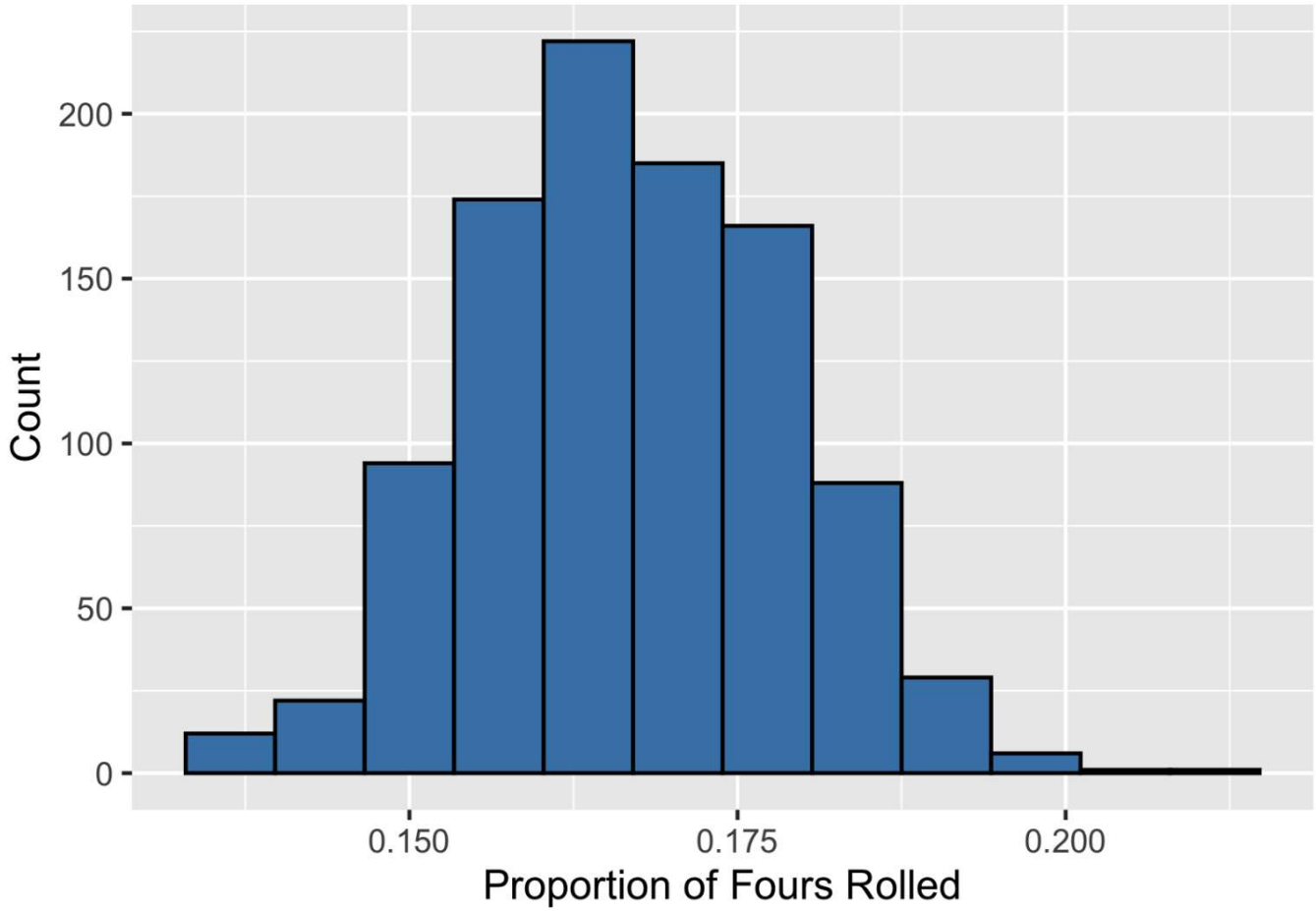
Roll	Result
1	2
2	1
3	4
4	2
5	6

- $\frac{1}{5}$ or 20% are a 4

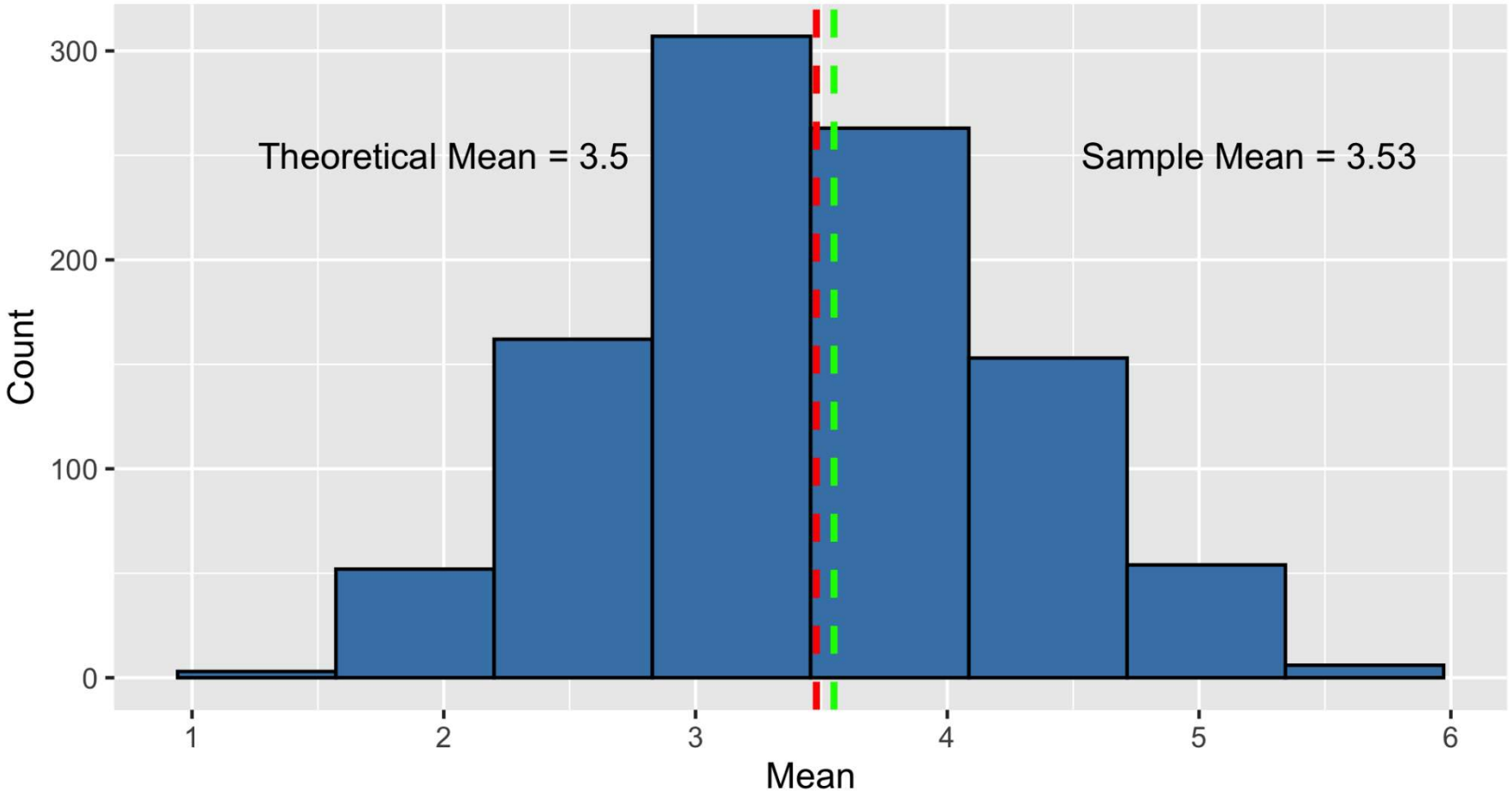
Set	Mean
1	4
2	4
3	1
4	4
5	3

- $\frac{3}{5}$ or 60% are a 4

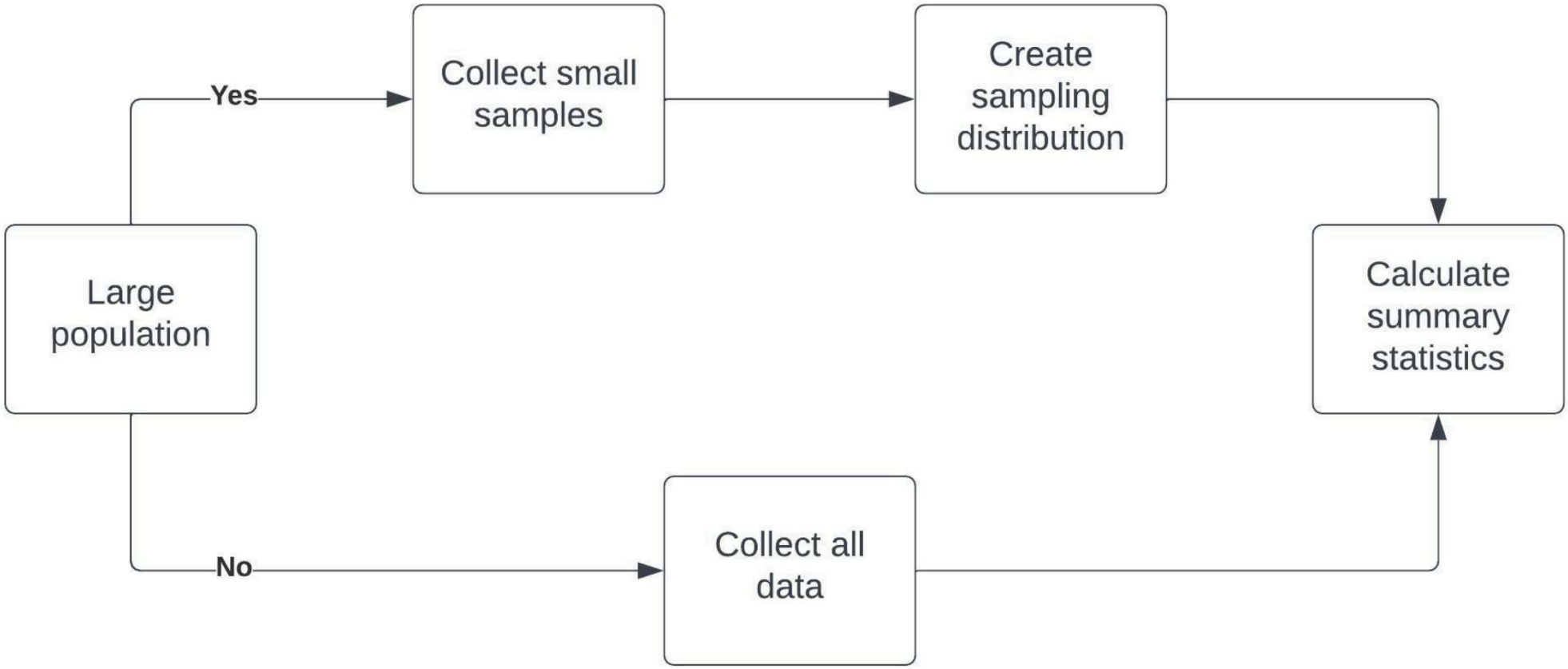
Sampling distribution of proportion



Mean of the sampling distribution

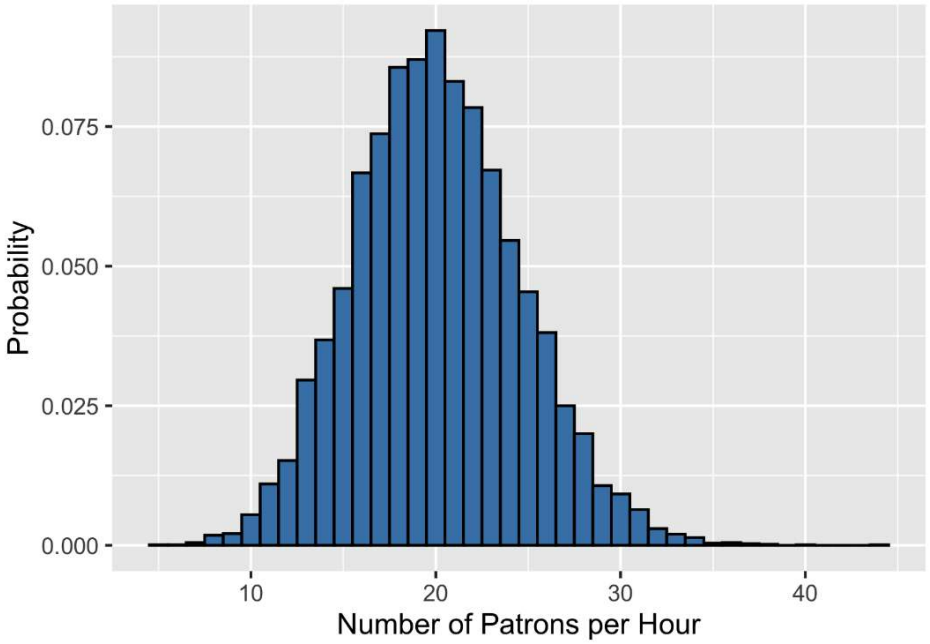


Benefits of the central limit theorem



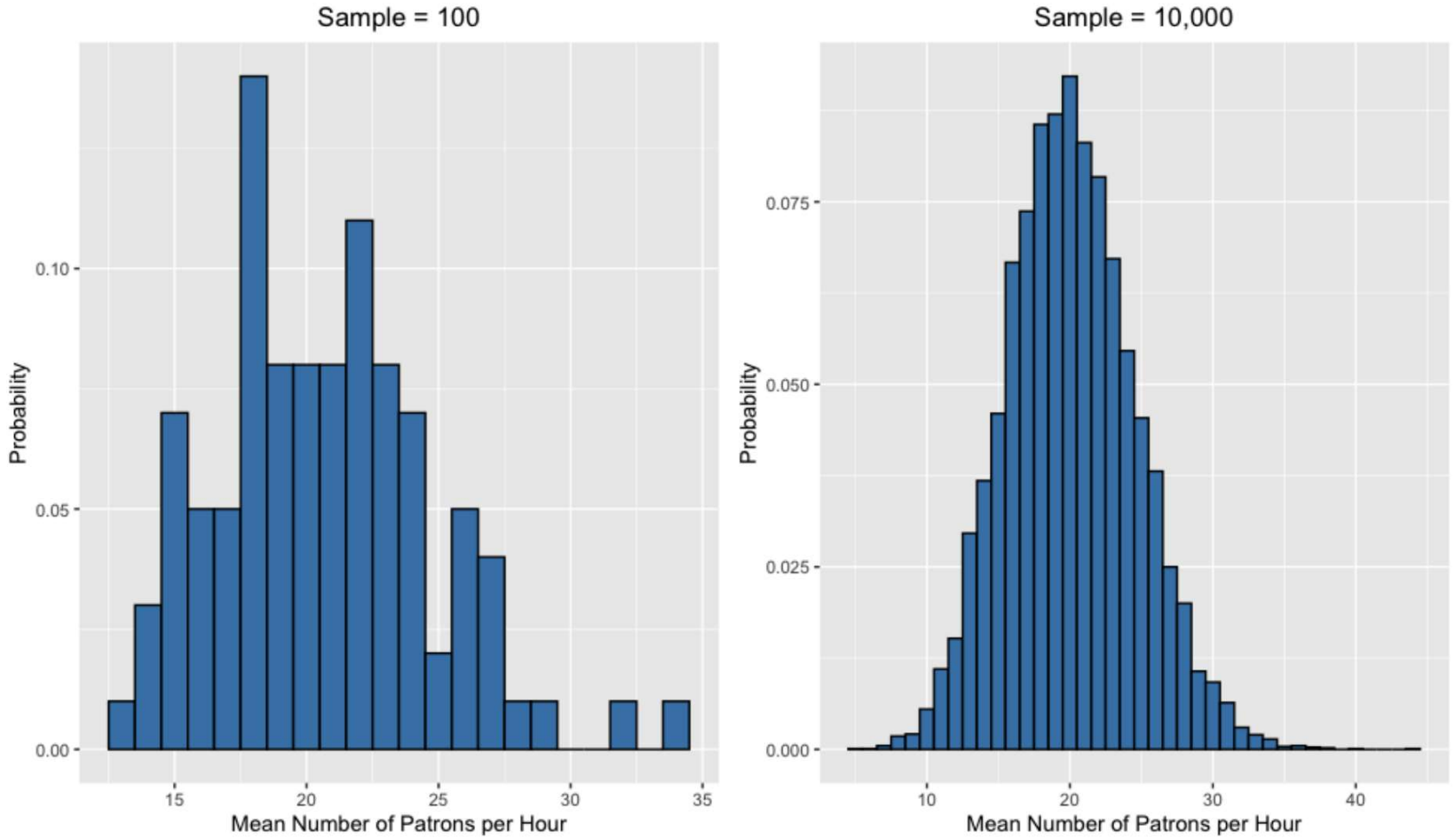
Lambda (λ)

- λ = average number of events per time interval
  - Average number of patrons per hour = 20
  - λ = the expected value of the distribution!

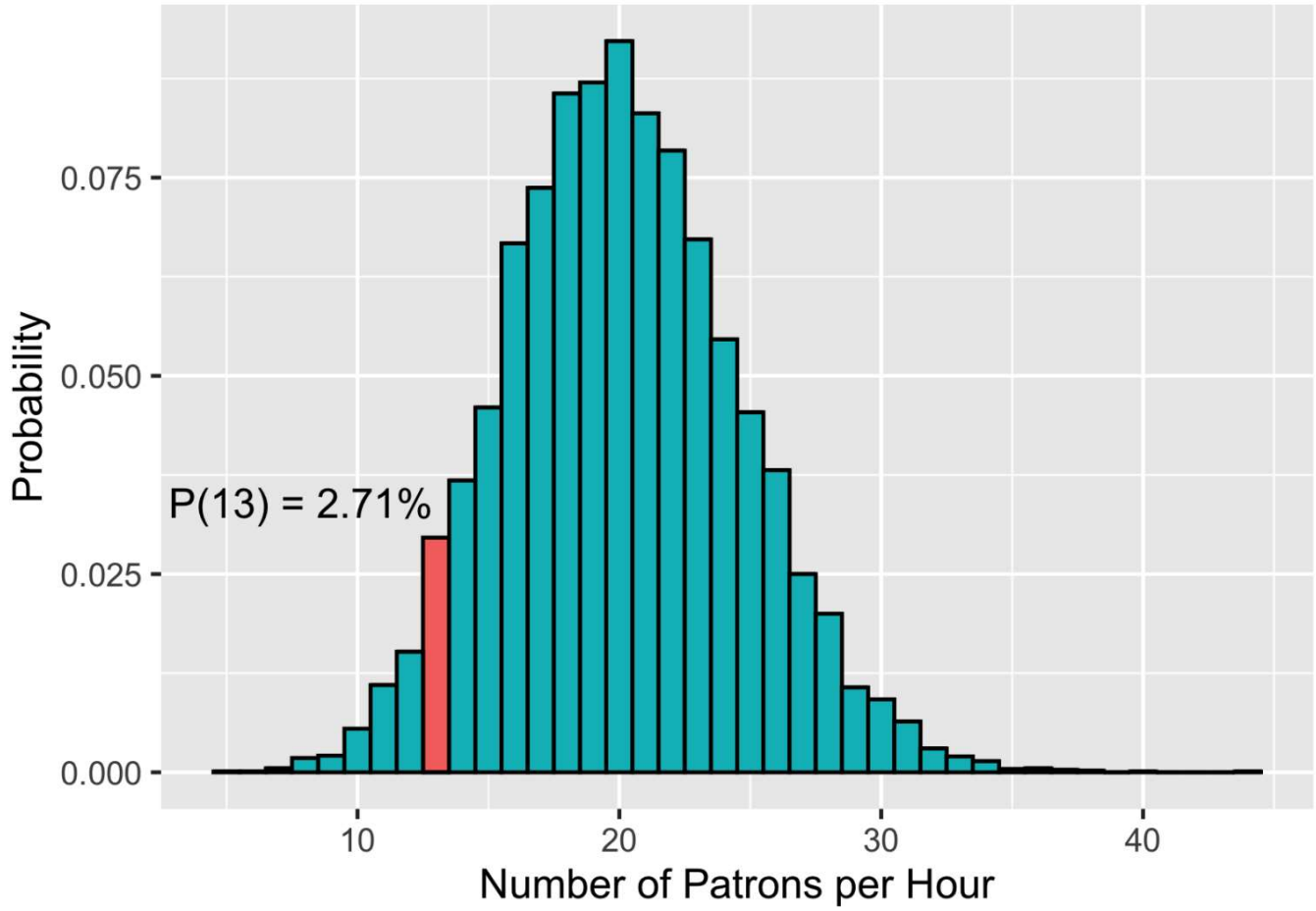


Lambda is the distribution's peak

Central limit theorem still applies!



Probability of 13 patrons in an hour



Probability of 25 or more patrons