6.100B: Recitation 4

Distributions, Simulation, and Significance Testing

April 28th, 2023

Actionables

PS5 is released on **Monday, 5/1**

PS4 is due on **Thursday, 5/4**

PS3 Checkoff is due on **Friday, 5/5** by **5pm**

Microquiz 3 on **Wednesday, 5/3** in-class

Agenda

Distributions

Uniform vs. Gaussian

Discrete Distributions

Expected Value

Simulations

The Dice Problem

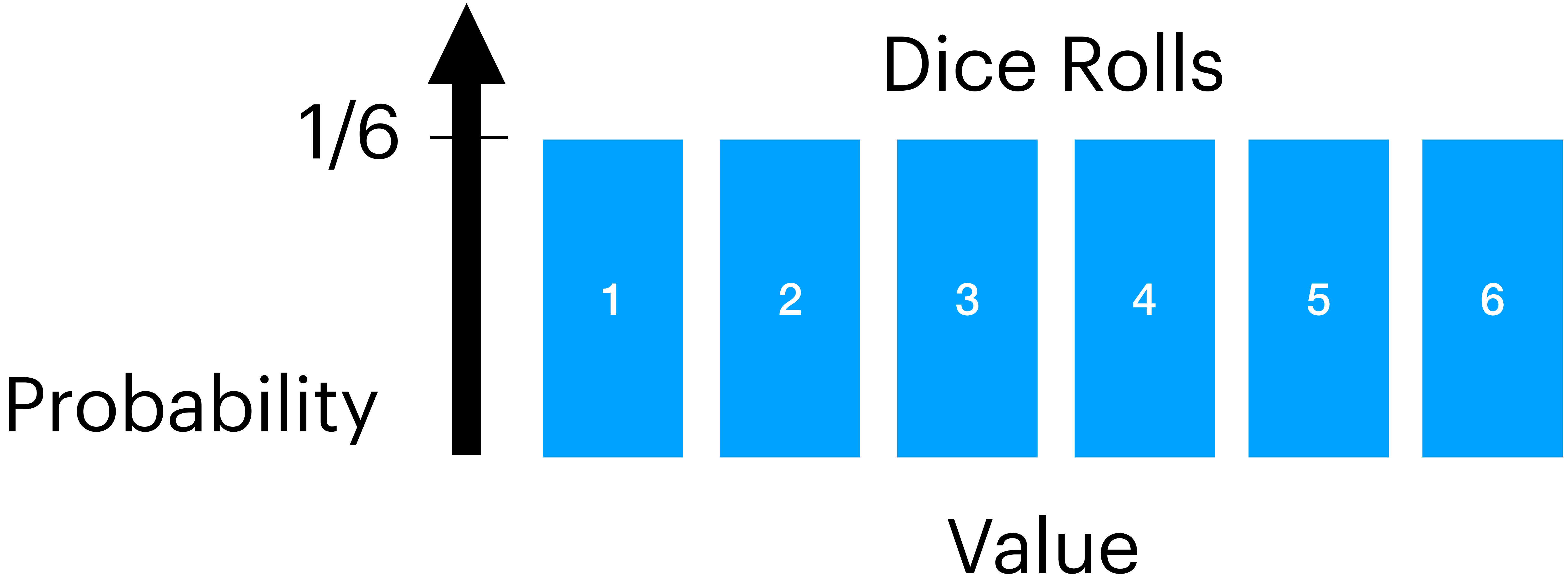
Let’s get started!

https://mattfeng.tech/teaching/6.100B/

Distributions | Uniform Distributions

We need to be able to model **randomness** and its effects on problems we simulate

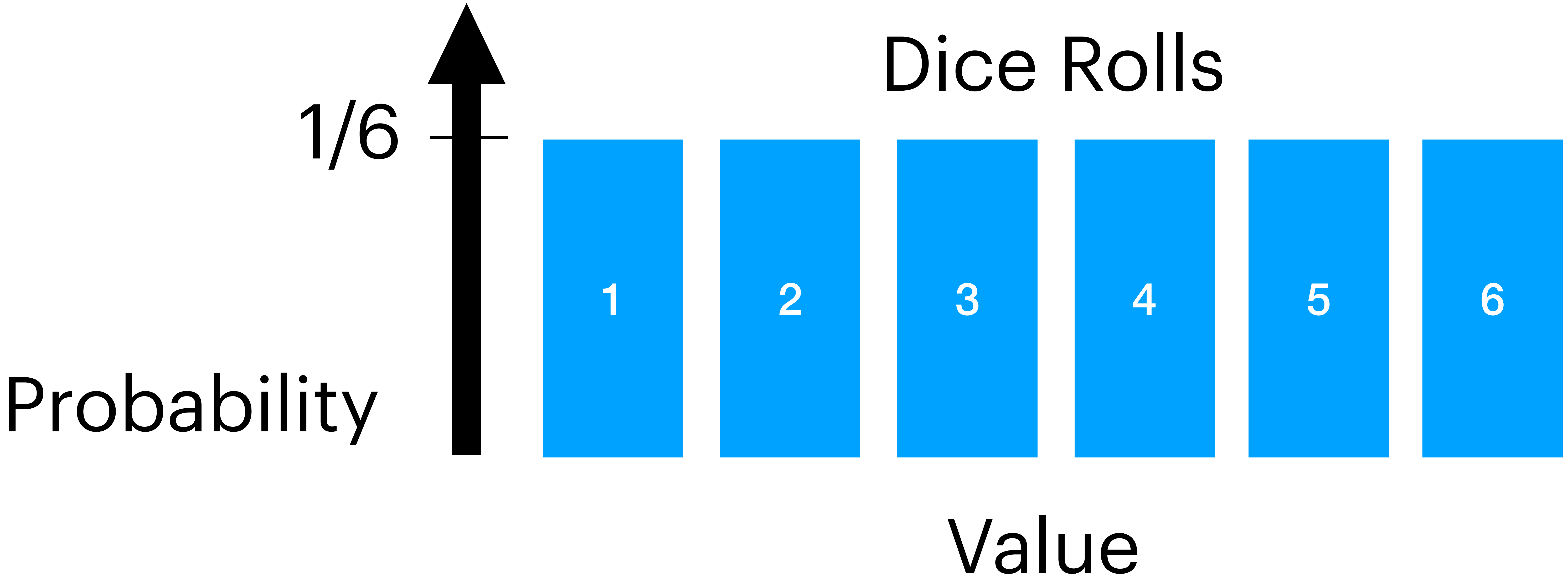
A common case that we may see in problems is **uniform probability distributions**



Distributions | Uniform Distributions

For **Discrete Probability Distributions**, we have finite number of values in our domain (we can't roll a 4.2 or 5.7)

A uniform probability density considers all outcomes to be of **equal likelihood**



Distributions | Discrete Distributions

Consider we have **two** dice, we'd like to determine the distribution for each **sum** of the two dice rolls

For each outcome, we need to determine the number of **ways the outcome can occur**, and find the ratio between that and the **total possible outcomes**

Rolling a 1
Impossible

Distributions | Discrete Distributions

Consider we have **two** dice, we'd like to determine the distribution for each **sum** of the two dice rolls

For each outcome, we need to determine the number of **ways the outcome can occur**, and find the ratio between that and the **total possible outcomes**

Rolling a 2

One possibility

Distributions | Discrete Distributions

Consider we have **two** dice, we'd like to determine the distribution for each **sum** of the two dice rolls

For each outcome, we need to determine the number of **ways the outcome can occur**, and find the ratio between that and the **total possible outcomes**

Rolling a 3
Two possibilities

Distributions | Discrete Distributions

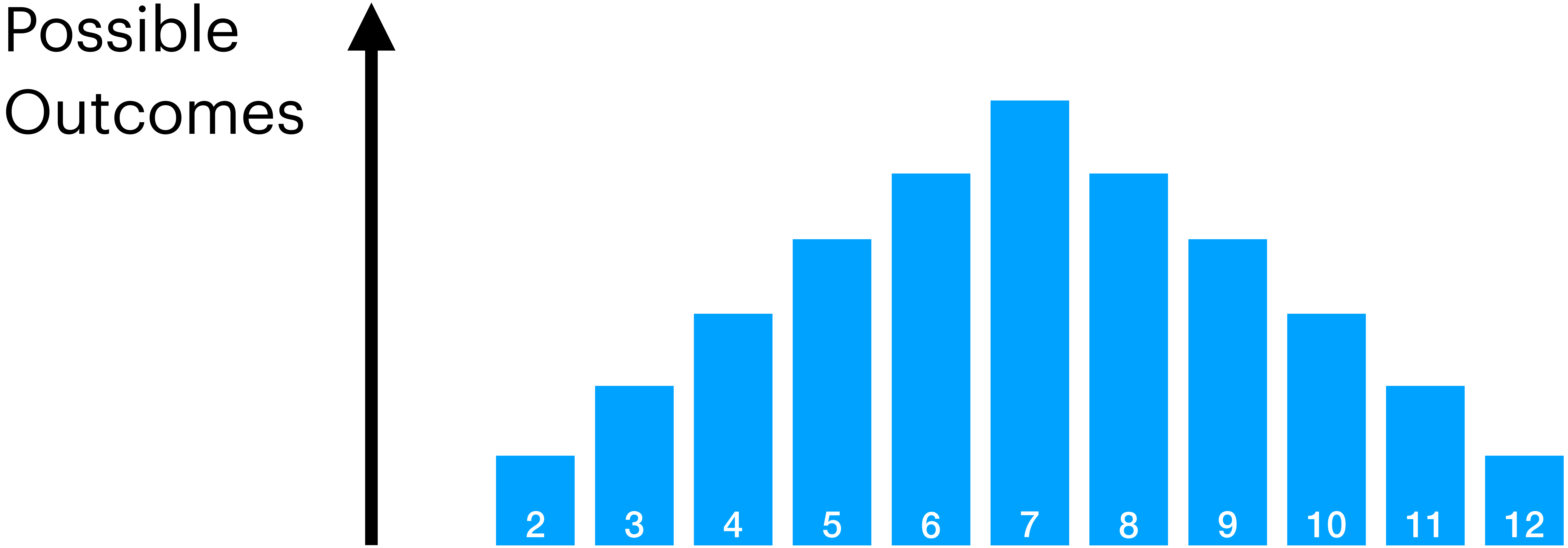
Consider we have **two** dice, we'd like to determine the distribution for each **sum** of the two dice rolls

For each outcome, we need to determine the number of **ways the outcome can occur**, and find the ratio between that and the **total possible outcomes**

Rolling a 4
Three possibilities

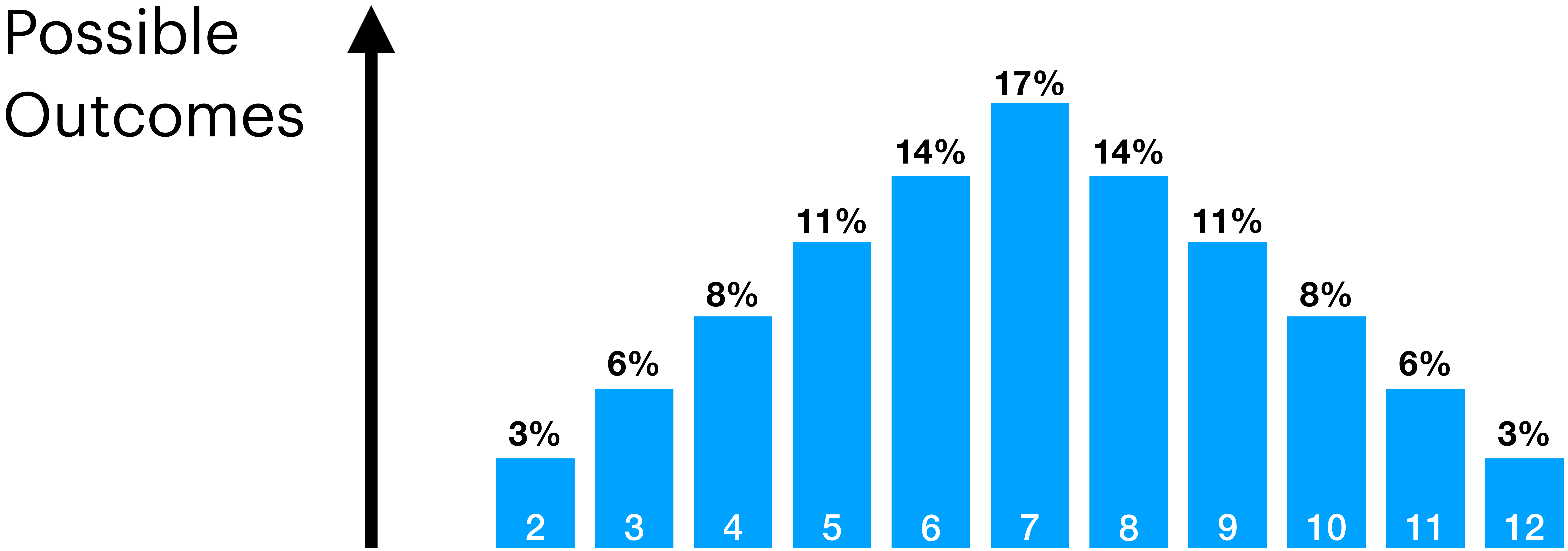
Distributions | Discrete Distributions

Here we see the **number of possibilities** for two-dice sum rolls:



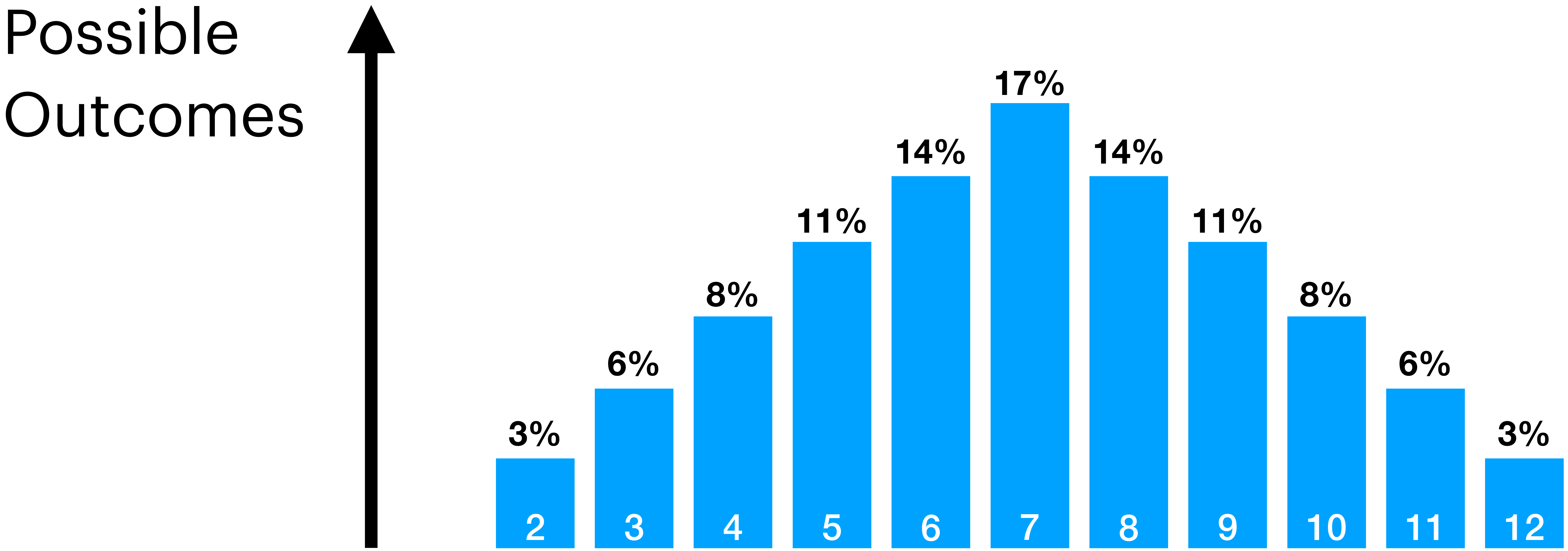
Distributions | Probability Distributions

Given the **36** possible outcomes, we can turn this into a **probability distribution** by dividing the number of possible ways the event can happen by the **total outcomes**



Distributions | Probability Distributions

What is the **expected value** of this roll?



Distributions | Expected Value

Expected Value is the outcome we can **expect** to be the average outcome of a system (like rolling two dice)

$$E[X] = \sum_i p_i v_i$$

In which **p** is the probability of outcome **i** of event **X**, and **v** is the value associated with that outcome

Distributions | Expected Value

The expected value for a single die roll is:

$$E[\text{Die Roll}] = \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6}$$

$$E[\text{Die Roll}] = \frac{21}{6} = 3.5$$

Distributions | Expected Value

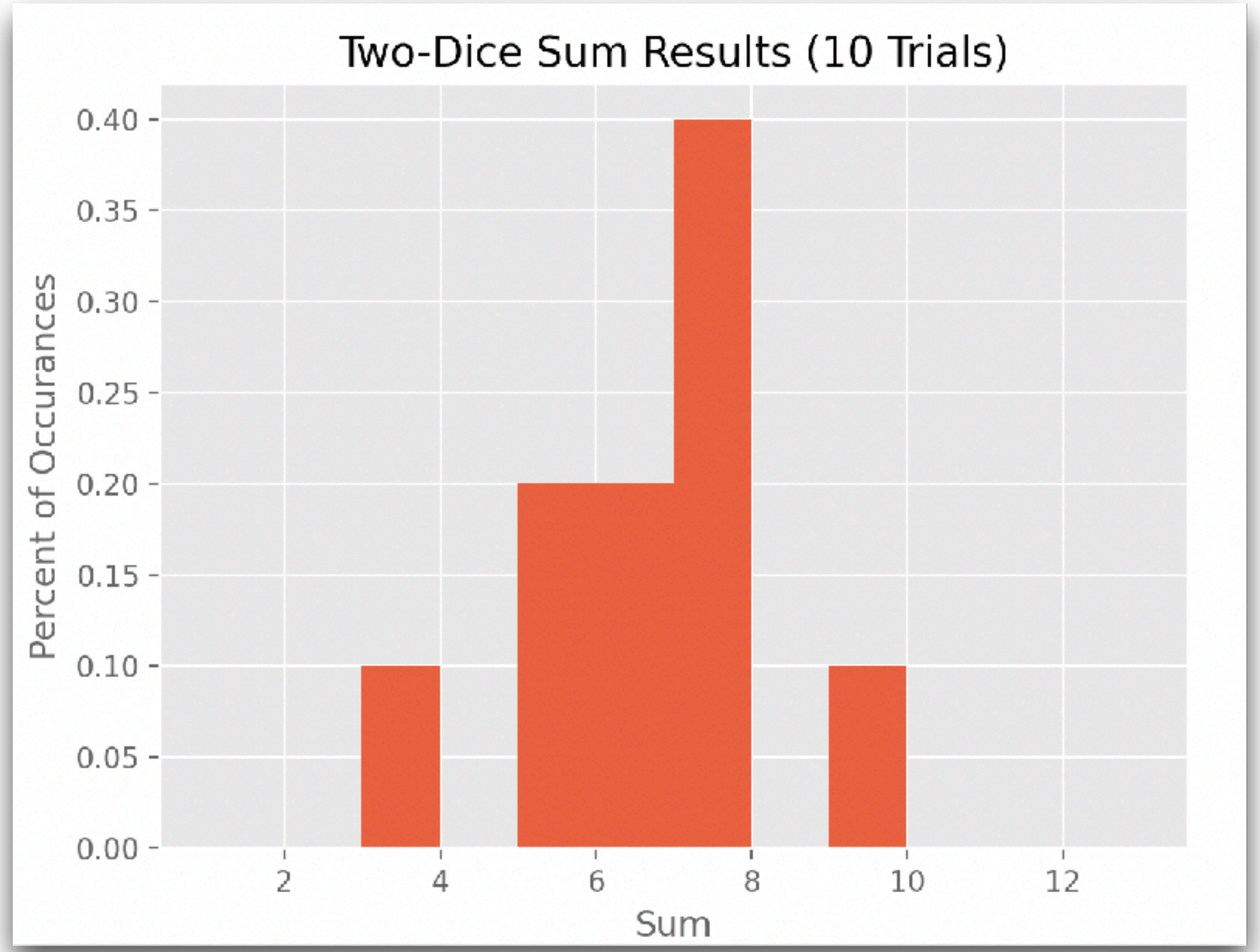
Expected Value is like a **weighted average** based on the likelihood of outcomes to happen

$$E[\text{2 Die Roll}] = 1\frac{2}{36} + 2\frac{3}{36} + 3\frac{4}{36} + 4\frac{5}{36} + 5\frac{6}{36} + 6\frac{7}{36} + 5\frac{8}{36} + 4\frac{9}{36} + 3\frac{10}{36} + 2\frac{11}{36} + 1\frac{12}{36}$$

$$E[\text{2 Die Roll}] = 7$$

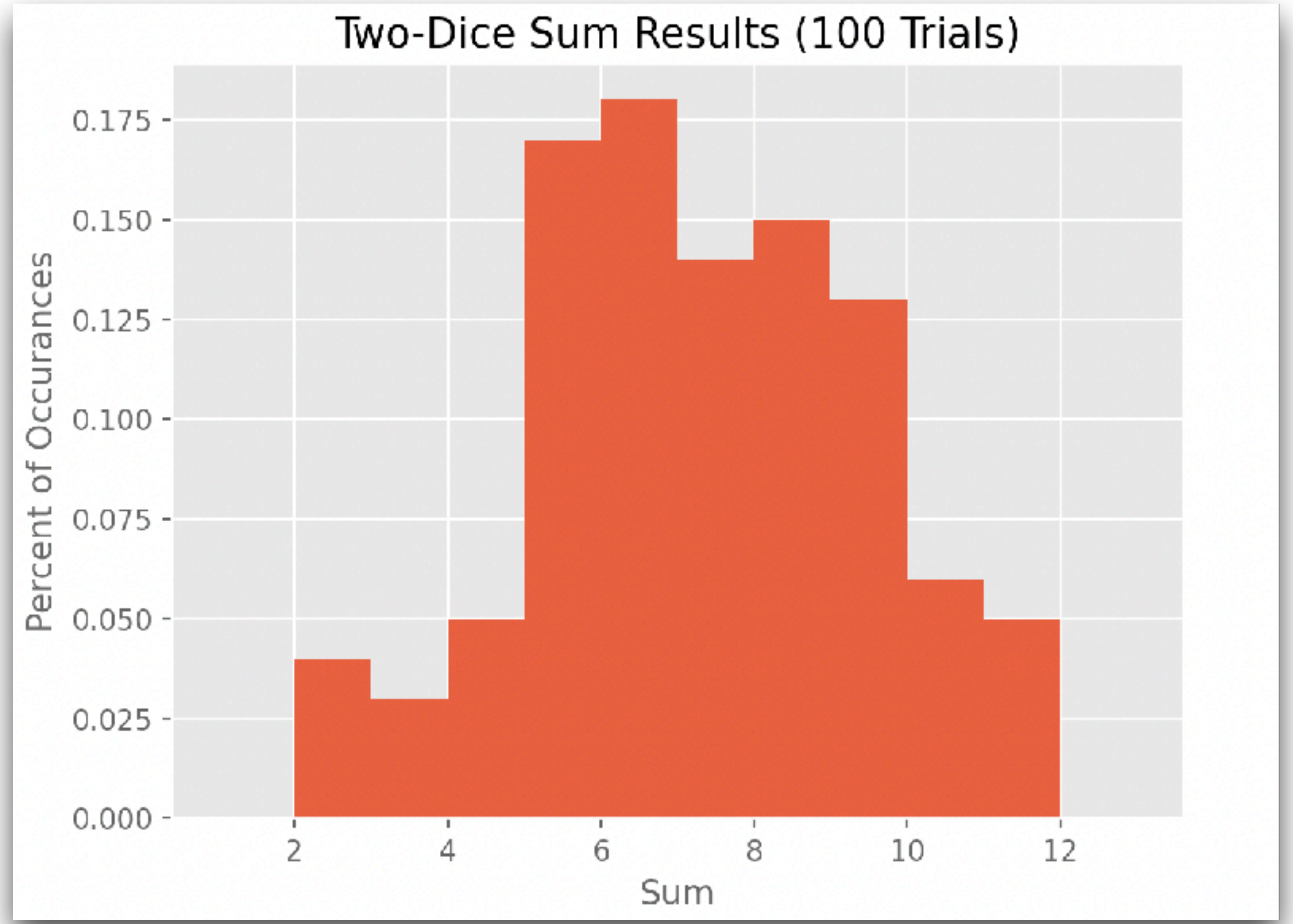
Distributions | Expected Value

Now if we roll two dice and add the sum, what's the **distribution** we get in practice?



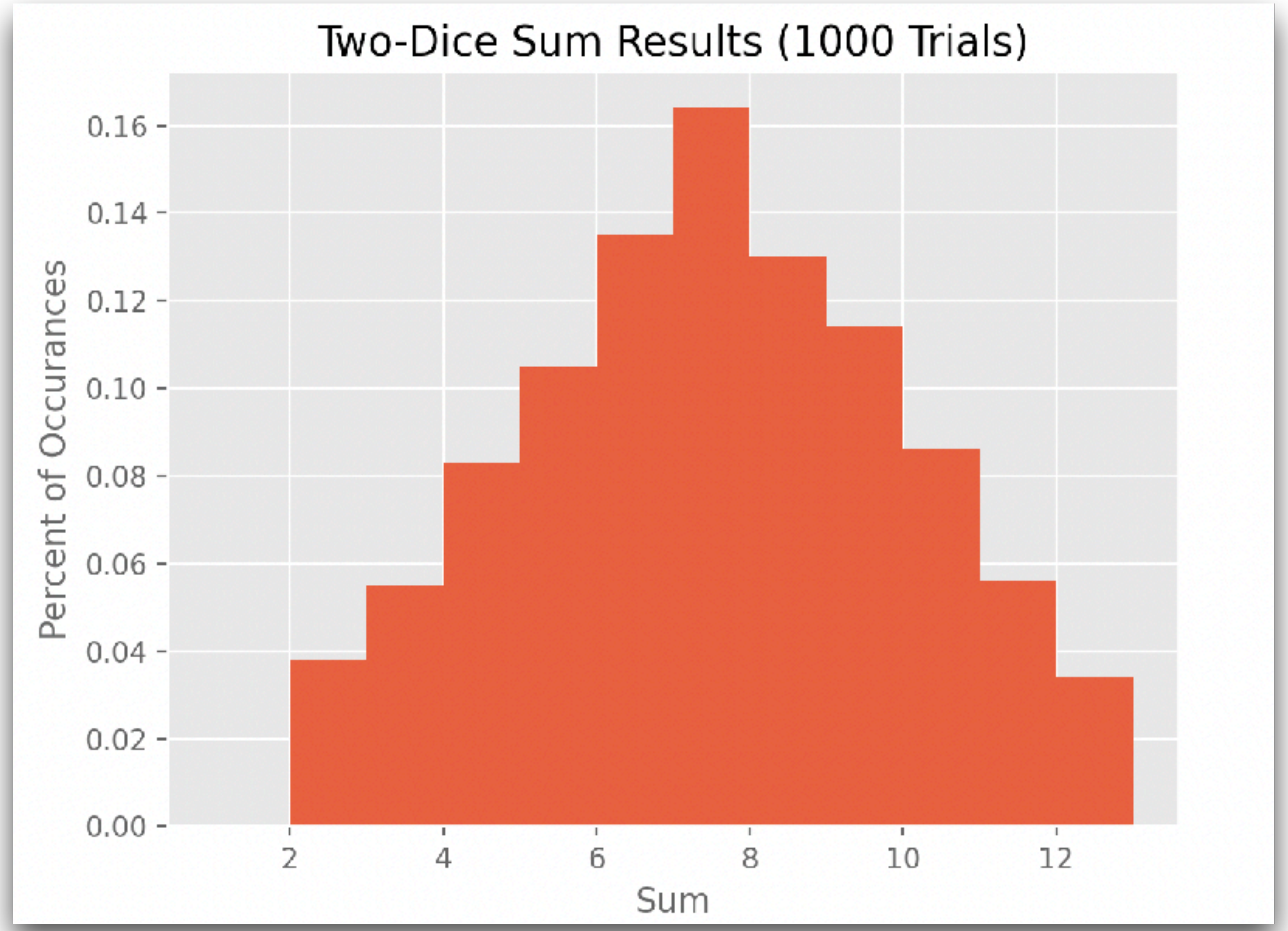
Distributions | Expected Value

Now if we roll two dice and add the sum, what's the **distribution** we get in practice?



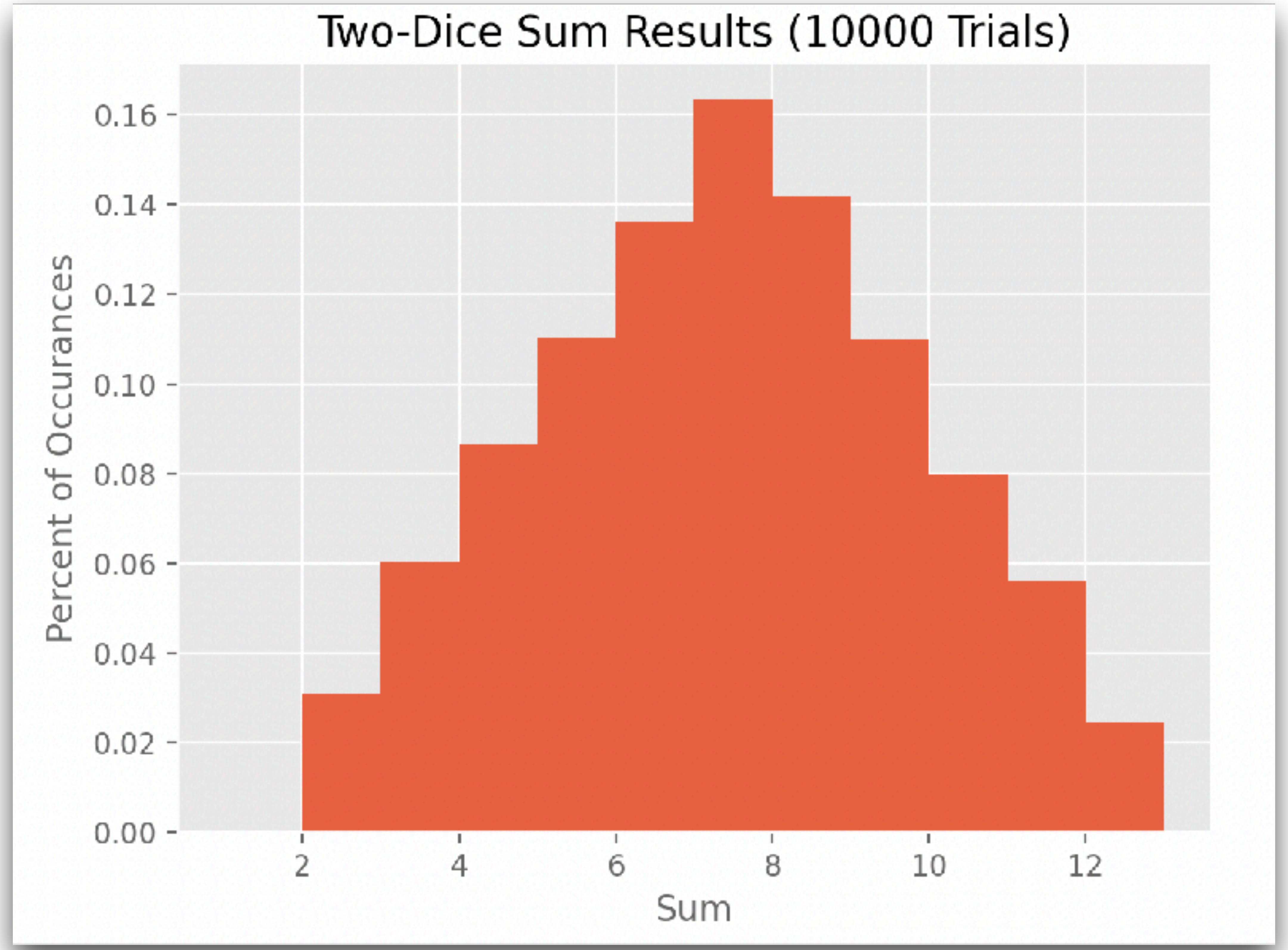
Distributions | Expected Value

Now if we roll two dice and add the sum, what's the **distribution** we get in practice?



Distributions | Expected Value

Now if we roll two dice and add the sum, what's the **distribution** we get in practice?



Simulations | Dice Problem

Okay, we collected a bunch of data – now what?

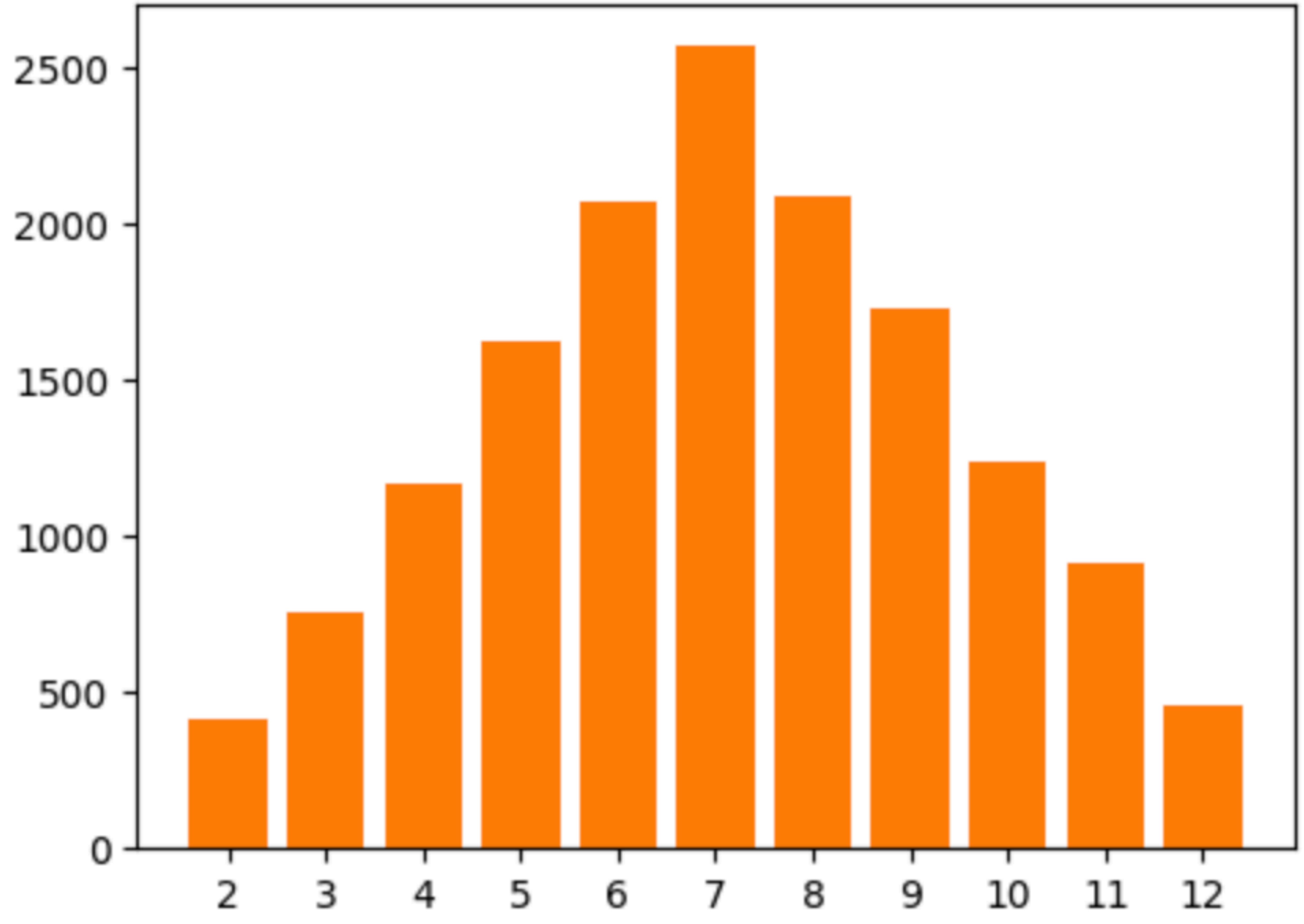
We can use **statistical tests** to determine whether or not our results are **significant** (i.e. likely signal and not just noise).

For example, could we figure out if another player is using **weighted dice**?

Simulations | Dice Problem

Fair or unfair?

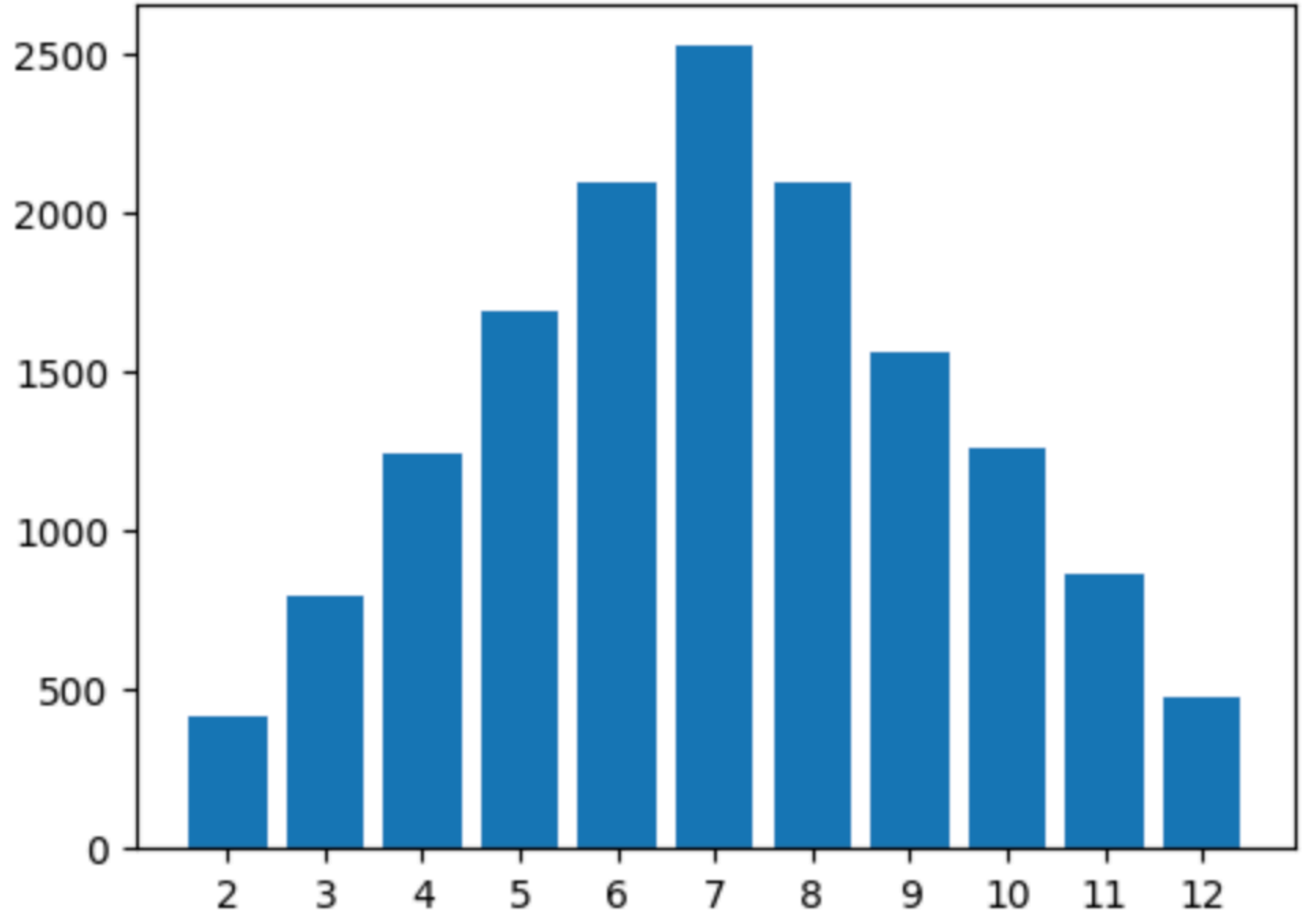
Unfair



Simulations | Dice Problem

Fair or unfair?

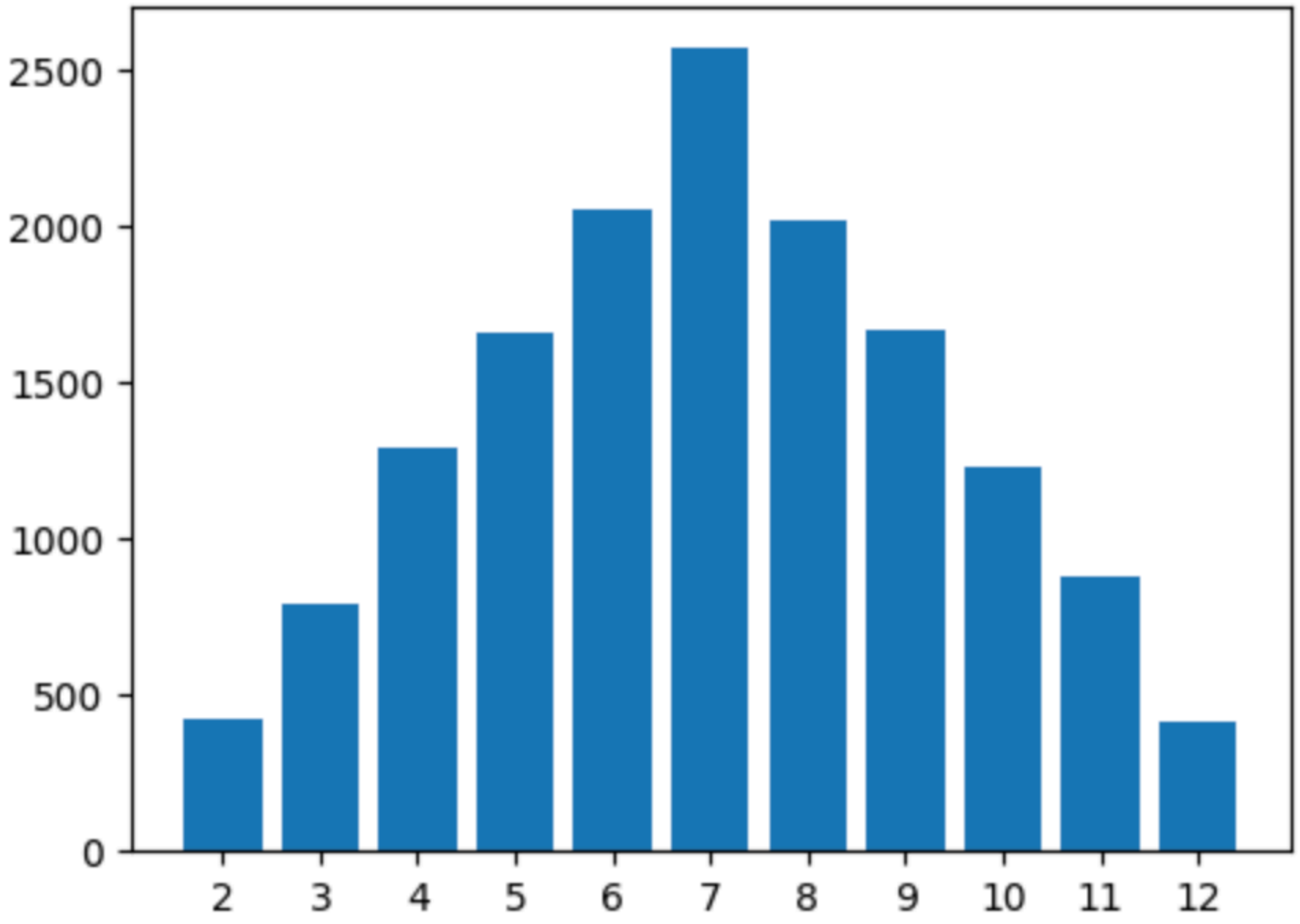
Unfair



Simulations | Dice Problem

Fair or unfair?

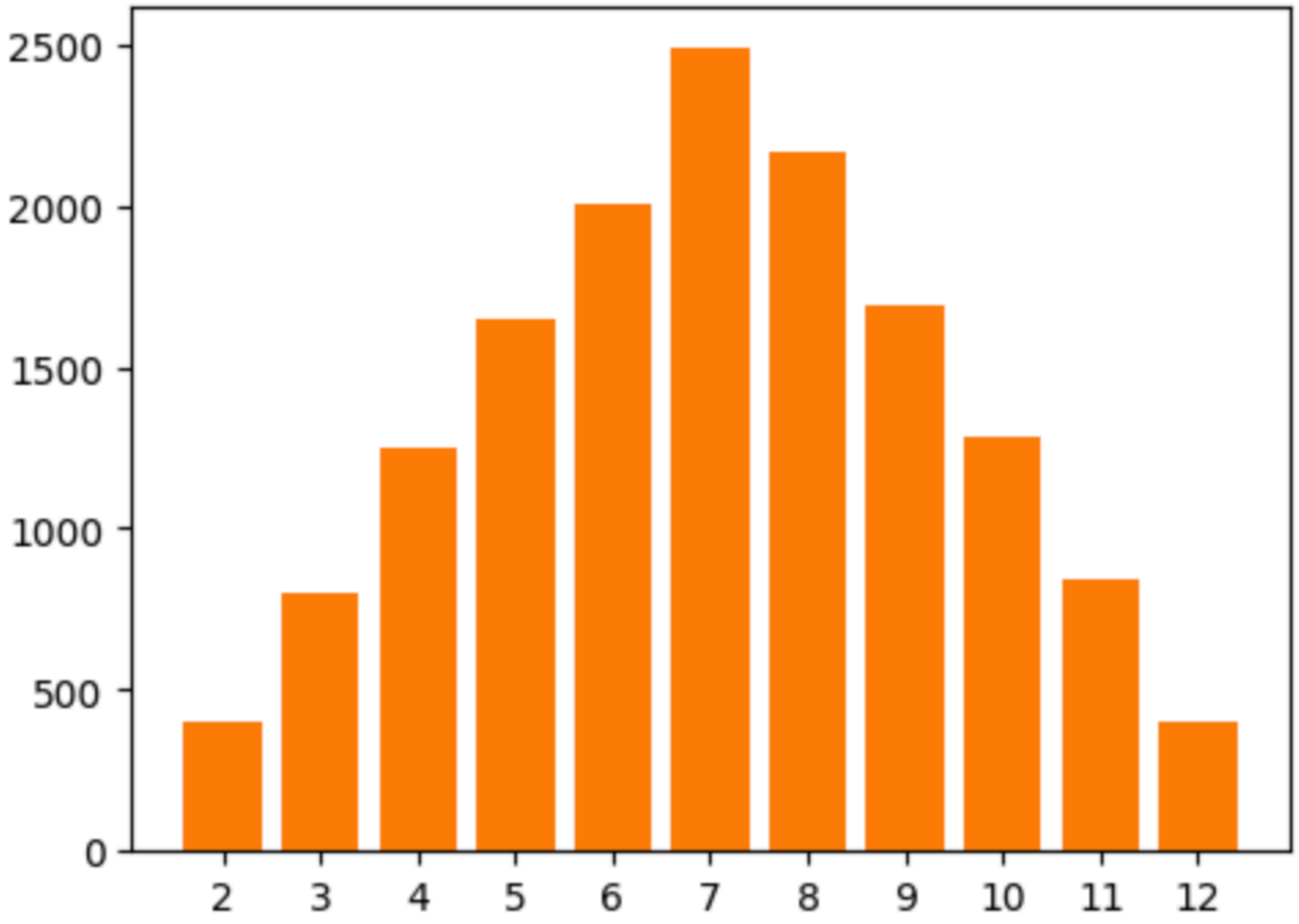
Fair



Simulations | Dice Problem

Fair or unfair?

Fair



Simulations | Dice Problem

If you're going to accuse someone of cheating, you need to be able to back up your claims

Simulations | Dice Problem

How do we tell them apart? Namely with **95% confidence?**

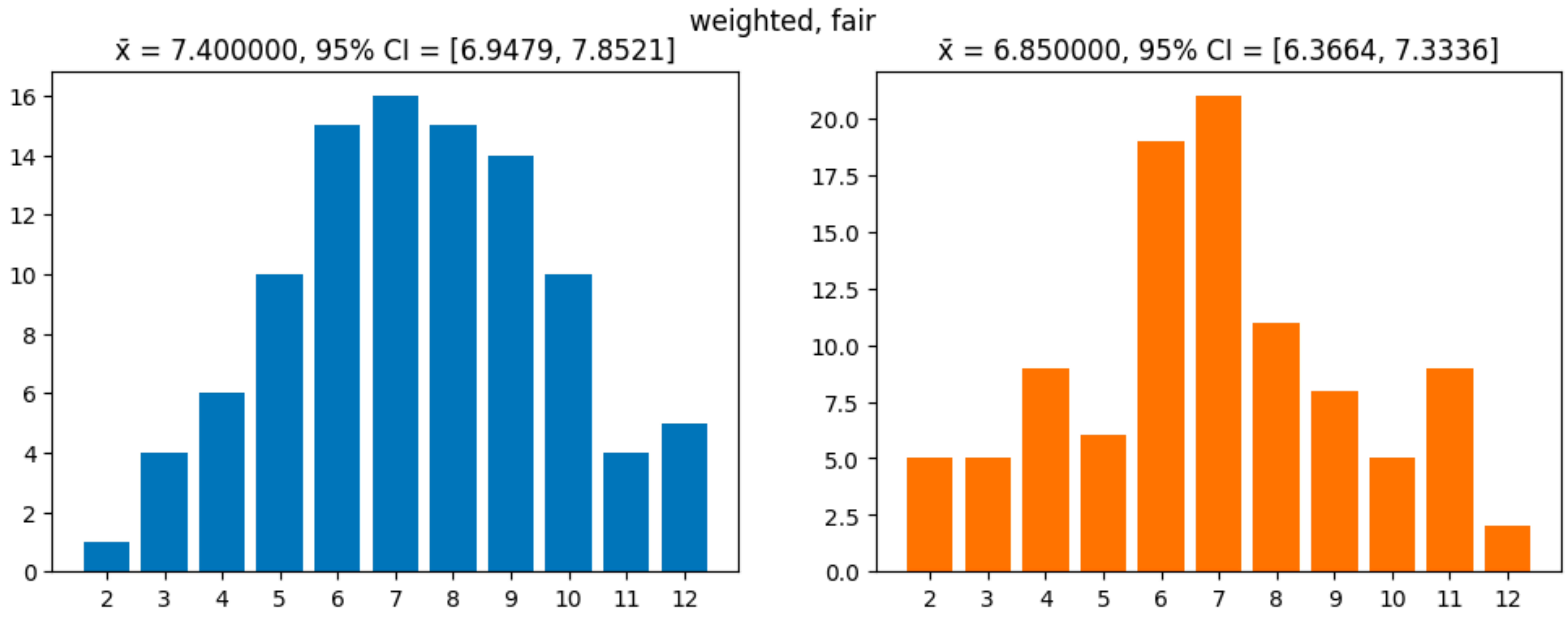
We can perform a statistical test (called a z-test) to evaluate whether the sample mean of the fair dice differs the from the sample mean of the weighted dice.

The z-test let's us determine whether or not we have enough data to claim with 95% confidence that the means of two **normal distributions** are different

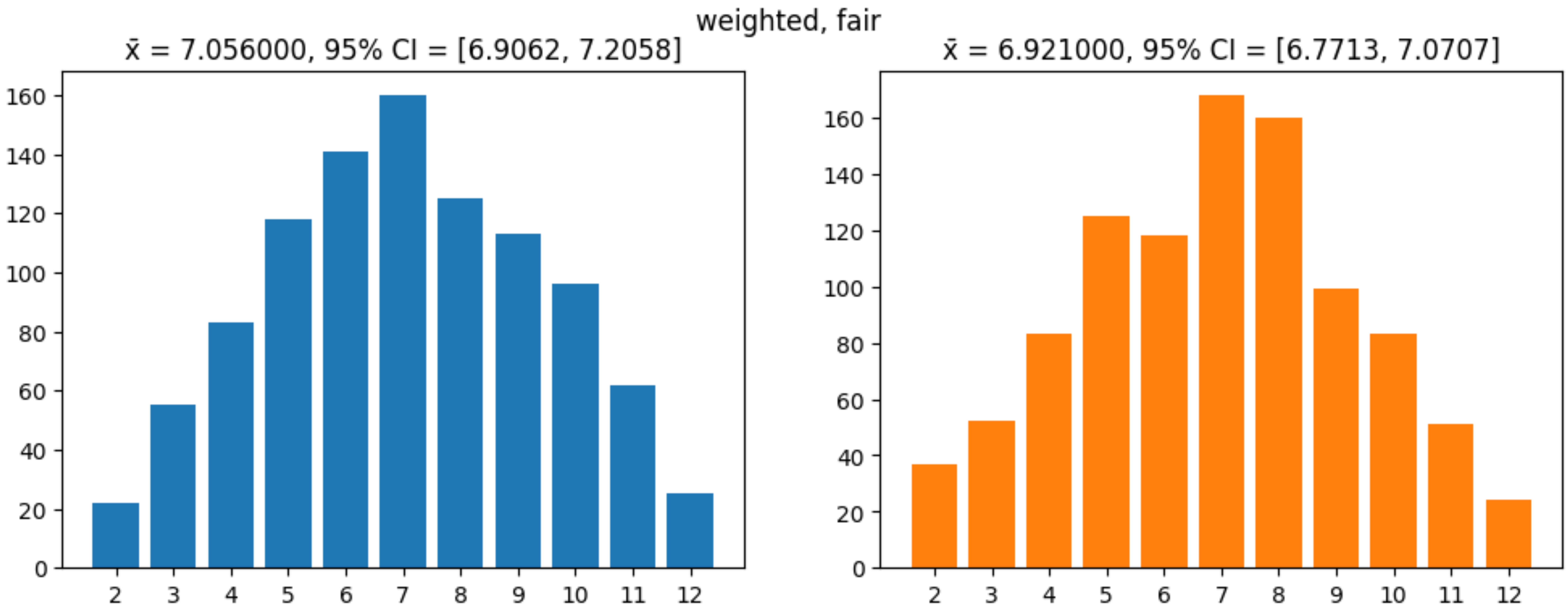
Because of the Central Limit Theorem, the distribution of the **sample means** of our dice roles approaches a normal distribution

Confidence intervals give the ranges we believe the true means of the dice lie.

Simulations | Dice Problem



Simulations | Dice Problem



Simulations | Dice Problem