# 6.10: Another Look at Probability (2 of 2)

A link to an interactive elements can be found at the bottom of this page.

A link to an interactive elements can be found at the bottom of this page.

Source: GeoGebra, license: CC BY SA

Note that by clicking the GeoGebra link above you can launch a new window with this simulation in it if you would like to position it closer to the questions you'll be answering below to avoid scrolling so much.

1. Make sure **Coins = 1** and **P(heads) = 0.2**.
2. Click the **Auto** button and watch the count of heads and tails change.
3. Click the **Pause (II)** button once **Total Flips** is over 100 or so.
4. Record the total number of Heads (1's) and the total number of flips.
5. Calculate P(H) (Number of heads / Total Flips) when Total Flips is about 100.
6. Click the **Auto** button again to continue the flips.
7. Click the **Pause (II)** once **Total Flips** is over 1,000 or so.
8. Record the total number of Heads (1's) and the total number of flips.
9. Calculate P(H) (Number of heads / Total Flips) when Total Flips is about 1,000.

https://assessments.lumenlearning.co...sessments/3887

https://assessments.lumenlearning.co...sessments/3554

---

Let's summarize what we have learned from these activities:

- The empirical probability will approach the theoretical probability after a large number of repetitions. In some situations, such as in flipping an unfair coin, we cannot calculate the theoretical probability. In these cases, we have to depend on data.
- There is less variability in a large number of repetitions. This means that in the long run, we will see a pattern, so we are more confident about estimating the probability of an event using empirical probability with a large number of repetitions.

## What Do We Mean When We Say an Event Is Random or Due to Chance?

In the discussion of the role of probability in the Big Picture of Statistics, we said that probability is the machinery that allows us to draw conclusions about a population on the basis of a random sample. To understand why we can trust random selection in an observational study and random assignment in an experiment, we need to look more closely at what we mean by random or chance behavior.

When we say that an event is random or due to chance, we mean that the event is unpredictable in the short run but has a regular and predictable behavior in the long run. This is obviously true for the coin-tossing activity. We cannot predict whether an individual toss will be heads, but in the long run, the outcomes have a predictable pattern. The relative frequency of heads is very close to 0.5 for a fair coin.

*We can make probability statements only about random events.*

## What Is the Connection between the Coin-Flipping Activities and the Discussion of Probability in the Previous Module?

Let's look at two probability questions that we might answer using the familiar data set from *Relationships in Categorical Data with Intro to Probability*. Recall that 6,198 of the 12,000 students at a West Coast community college are female. Previously, we calculated $P(\text{female}) = 6{,}198 / 12{,}000 = 0.5165$. What is the random event in this case? Let's be very specific about the question this calculation is meant to answer.

*What is the probability that a student at the West Coast community college is a female?*

- In this case, the relative frequency 6,198 / 12,000 is the actual proportion of females at the college. This is like the fair coin situation. Because we know the gender distribution at the college, we can think of 0.5165 as the theoretical probability that a *randomly selected student at this particular college* is a female. Tossing the fair coin in the simulation is like randomly selecting a student from the spreadsheet of data. We do not know if a randomly selected student will be female. But if we repeat this process many, many times, in the long run, the relative frequency of females will have a predictable pattern. The relative frequency will be very close to the proportion of females in the data set.

*What is the probability that a community college student in the United States is female?*

- In this case, we are using the data from the 12,000 West Coast community college students to represent students at all community colleges in the United States. The relative frequency is an estimate for the chance that a *randomly selected U.S. student* is female. This is like tossing the unfair coin 12,000 times and using the relative frequency of heads as an estimate of $P$(head). We do not know $P$(female) for all community colleges, just as we did not know the $P$(heads) with an unfair coin. But if the sample is random, we can use the relative frequency of females in the sample as an estimate of $P$(female) in all community colleges.

The main points are these:

- We can make probability statements only about random events.
- Probability of an event A is the relative frequency with which that event occurs in a long series of repetitions.

## Contributors and Attributions