

Probability –the likelihood that a random event will occur. In other words, how often an event will occur if an experiment is repeated a large number of times. If we were in a traditional probability class we would spend 3 weeks learning about combinatorics, conditional probabilities, etc. in order to calculate theoretical probabilities. But in this class we are going to learn how to use Excel to estimate theoretical probabilities with empirical probabilities. You will still need to know these rules:

Probability of an Event – the measure of the degree of certainty of an event occurring. The probability of an event, E is defined by

$$P(E) = \frac{\text{number of ways for E to happen}}{\text{total number of **equally likely** outcomes}}$$

$$0 \leq P(E) \leq 1 \quad (\text{or } 0\% \leq P(E) \leq 100\%)$$

Probability Rule #1: If two events E_1 and E_2 are **disjoint** (non-overlapping), then

$$P(E_1 \text{ or } E_2) = P(E_1) + P(E_2)$$

Probability Rule #2: If two events E_1 and E_2 are **independent** (occurrence of the first doesn't affect the second), then

$$P(E_1 \text{ and } E_2) = P(E_1) \cdot P(E_2)$$

Complementary Events:

$$P(\text{event occurs}) = 1 - P(\text{event does not occur})$$

Empirical probabilities are probabilities based on observation. Our genetics example from the last module was based on data gathered from a study of 1000 people:

	FEMALE (F)	MALE (M)	Totals
Color blind (C)	2	24	26
Normal (N)	518	456	974
<i>Totals</i>	520	480	1000

We don't know the theoretical probability of being color blind, but if a large enough study is conducted we can get a pretty good estimate. How large is large enough? That is a complicated topic. For this class we will be running simulations of anywhere from 100 to 5000 iterations to collect our data.

Say we were to take a fair die, drill a hole in one of the spots, fill the hole with lead and then paint over the lead to make the die appear untampered with. For a fair die the probability of any one side appearing is $1/6$, but we do not know the probabilities for our tampered die. If we were to roll a fair die 600 times we would expect to observe each side about 100 times. With our tampered die some sides may appear more or less frequently and we could estimate the probability of each side by $P(E) = \frac{\text{frequency of } E}{\text{number of trials}}$. If the side with three spots was observed 55 times in our 600 rolls then we would estimate $P(3) = \frac{55}{600} \approx 0.0917$. Of course, if we were to roll the die another 600 times we would likely get a slightly different result but as long as we are running hundreds–or thousands–of iterations we are going to get some pretty good estimates.

To get started, we are going to need some new functions in Excel. You will want to watch the video, but here is a summary for some of the tools we will use.

Excel functions:

=rand() will return a random number between 0 and 1

Why would that be useful? All probabilities are between 0 and 1—such as $P(\text{Heads}) = 0.5$ —then we can simulate flipping a fair coin and recording whether we get Heads by using IF statements along the lines of “if a random number is less than $1/2$, then we will call that Heads (success) otherwise it is Tails (failure)”. For convenience, let 1 =success and 0 = failure. You will see why that is convenient soon enough.

=randbetween(bottom, top) will return an integer value from the bottom value through the top value. This is a useful tool for simulating rolling a die, using =randbetween(1, 6)

=if(test, value if true, value if false) is a familiar function. One way we will use it in probability is to return values of 1 for success and 0 for false

example: =IF(A3<2, 1,0) For this IF statement, if the value in cell A2 is less than 2 the result will be 1, otherwise the result is a 0. The reason that using 1 and 0 to indicate “success” and “failure” is convenient? If every “success” is 1 and failure “0” summing up the column of values will count the number of successes.

Other tools

F9 is a function key on your keyboard. If you have a full sized keyboard it will be in the row of keys above the row of numbers on your keyboard. On compact keyboards you may need to first press a function (Fn) key down near your space bar, then the F9 key at the top. Pressing the F9 key will cause all of your rand() and randbetween() function to regenerate...which you will find disturbing at first. You’ll get used to it.