Learning Goal: Use probability distributions to identify typical and unusual events and, when appropriate, to determine an expected value.

Introduction:

In our previous discussions of probability, we focused on determining the probability of one outcome at a time, such as estimating the probability of a head in the toss of a weighted coin. Now we shift our focus to describing the probabilities of all possible outcomes instead of the probability of just one outcome.

Statisticians describe outcomes as variable values. Each variable value has a probability. All possible variable values together with their probabilities are a *probability distribution*.

We can generate a probability distribution from theoretical probabilities (such as a fair coin or fair die) or from empirical probabilities (such as using relative frequency to find probabilities for a weighted coin or weighted die).

Examples of probability distributions:

Probability distributions for a categorical variable

The variable is *coin toss.* This is a categorical variable with variable values: Head, Tail. The probability distribution assigns each variable value (H or T) a probability.

Here is a probability distribution for a fair coin based on theoretical probabilities:

Coin Toss Outcome	Head	Tail
Probability	½= 0.5	½ = 0.5

Here is the probability distribution for a weighted coin based on empirical probabilities (relative frequencies):

Coin Toss Outcome	Head	Tail
Probability	201/1000=0.201	799/1000=0.799

Here is a probability distribution for blood type in the U.S. based on empirical probabilities (relative frequencies) from a large study by Stanford University's Blood Center (bloodcenter.stanford.edu):

Blood Type	O	A	B	AB
Probability	0.45	0.41	0.10	0.04

Check your understanding:

1) People with blood type O can donate blood to people with any other blood type. For this reason, people with blood type O are called universal donors.

   a) What is the probability that a randomly selected person from the United States is a universal donor?

   b) If there were 500,000 people in the Stanford study, how many had blood type O?

2) Based on these probability distributions, fill in the properties of a probability distribution below:

   - *All outcomes are assigned a probability. The probabilities are numbers between (and including)* _____ *and* _____. Why does this make sense?

   - *In a probability distribution, the sum of all of the probabilities is* _____. Why does this make sense?

<u>Probability distributions for a quantitative variable</u>

Here is the probability distribution for the roll of a fair die based on theoretical probabilities:

Die	1	2	3	4	5	6
Probability	1/6 ~ 0.167	1/6 ~ 0.167	1/6 ~ 0.167	1/6 ~ 0.167	1/6 ~ 0.167	1/6 ~ 0.167

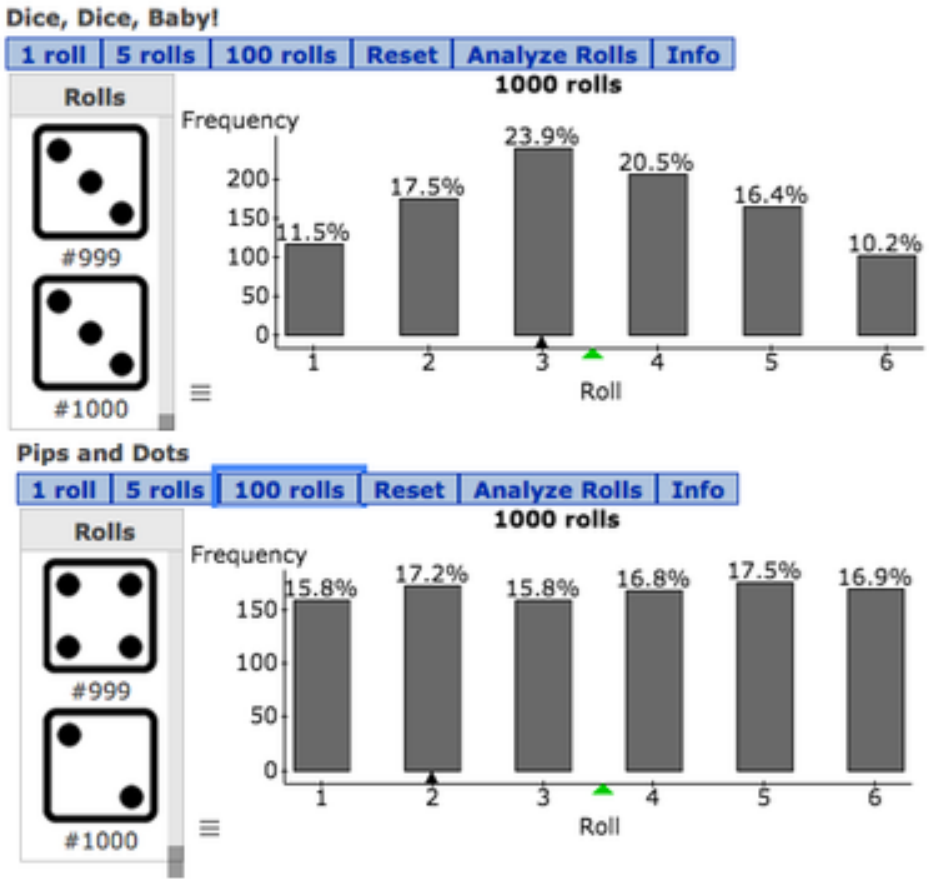
Here is the probability distribution for the number of Boreal owl eggs in a nest based on relative frequencies from large field studies in Canada.

Number of eggs	0	1	2	3	4	5	6
Probability	0.20	0.10	0.10	0.25	0.25	0.05	0.05

Check your understanding:

3) Which is more likely: To find a boreal owl nest with 2 eggs or to find an empty boreal owl nest?

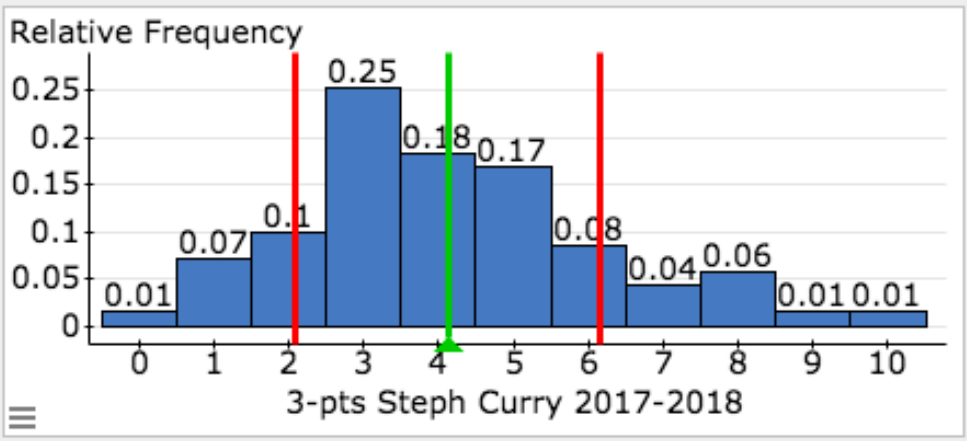
4) *Dice, Dice, Baby* and *Pips and Dots* are two companies that make dice. The graphs below show the probability distributions based on 1000 rolls of a die selected from each company. Is either company producing a fair die? Why do you think so?



Describing the probability distribution for a quantitative variable

When the variable is quantitative, we can use shape, center and spread to describe the probability distribution. We can calculate a mean to describe the center and a standard deviation to describe the spread.

Here is an example. The variable is number of 3-point shots Steph Curry made during a game in 2017-2018 (including pre-season games.) The histogram of the probability distribution shows the mean and one standard deviation above and below the mean.



Summary statistics:

Column	n	Mean	Std. dev.
3-pts Steph Curry 2017-2018	71	4.1267606	2.0488992

Check your understanding:

5) Is it unusual for Steph Curry to not score a 3-point shot during a game? How do you know?

6) Typical values fall within one standard deviation of the mean.

   a) Give an interval that describes Steph Curry's typical number of 3-pointers during a game.

   b) What is the probability that Steph Curry hits within this interval during a game?

Expected value

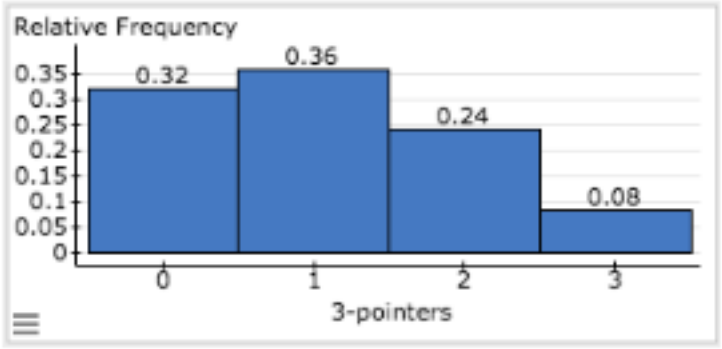
The mean of a probability distribution is called the *expected value.* It is a weighted mean, so values with a higher probability are more influential and values with a lower probability do not affect the expected value as much.

We will use technology to find the expected value, but, just this once, your instructor will demonstrate how to calculate the expected value by hand. The goal of this demonstration is to show how the expected value is a weighted mean.

7) Here is the number of 3-point shots made by a high school basketball player in 25 games, along with the associated frequency table and probability distribution.

0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 3 3

Number of 3-pointers scored per game	Frequency	Relative Frequency (Probability)
0	8	
1	9	
2	6	
3	2	



a) Estimate the expected value for 3-point shots for this player. Briefly explain what you are thinking.

b) Calculate the expected value (the weighted mean).