M5-03: Probability {Mass/Density} Function

Part of the "Polling, Confidence Intervals, and the Normal Distribution" Learning Badge
Video Walkthrough: https://discovery.cs.illinois.edu/m5-03/

.pmf() and .pdf(): Probability {Mass/Density} Functions

The `.pmf()` and `.pdf()` functions find the probability of an event at a specific point in the distribution.

- The Probability Mass Function (PMF) -- or `.pmf()` -- is only defined on discrete distributions where each event has a fixed probability of occurring.
- The Probability Density Function (PDF) -- or `.pdf()` -- is only defined on continuous distributions where it finds the probability of an event occurring within a window around a specific point.

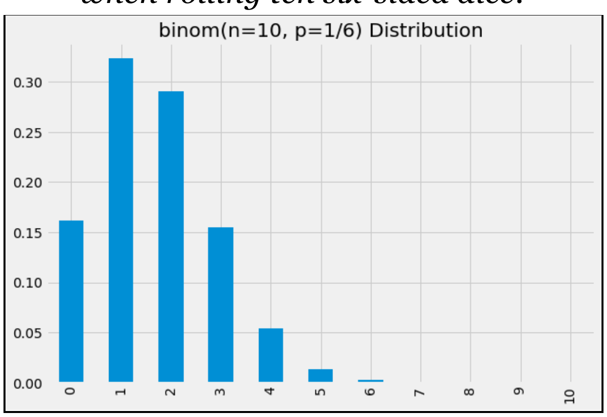
Discrete Distributions: Probability Mass Functions (PMF)

Both the Bernoulli and Binomial distributions are discrete distributions, and we would use the probability mass function (PMF). Define a distribution for the number of times you see a six when you roll ten six-sided dice:

Python:	
Description:	A distribution of the number of sixes that appear when you roll ten six-sided dice.

Visually, this distribution is the following, and we can find our expected values for the probability mass function at various points in the distribution:

Distribution of the number of sixes appearing when rolling ten six-sided dice:



binom(n=10, p=1/6)

P(no sixes):	
P(exactly one six):	
P(sixes == 2):	
P(half are sixes):	
P(sixes == 10):	
P(at least 3 sixes):	

M5-03: Probability {Mass/Density} Function

Part of the "Polling, Confidence Intervals, and the Normal Distribution" Learning Badge

Video Walkthrough: https://discovery.cs.illinois.edu/m5-03/

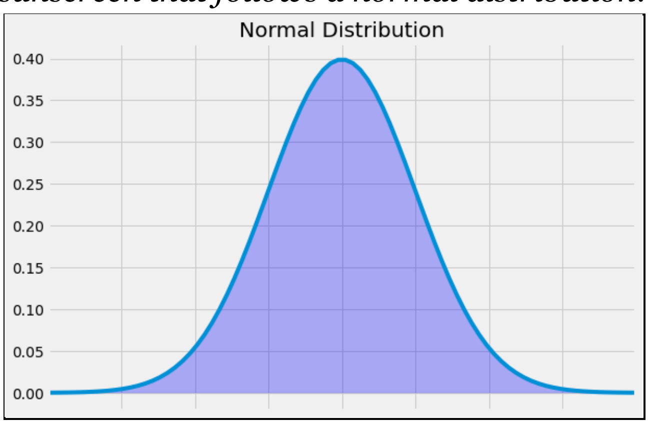
Continuous Distributions: Probability Density Function (PDF)

For continuous distributions, the probability density function tells us the probability of an outcome **near** the specific point. This is necessary since the probability is near-zero for any specific point on the distribution.

For example, the amount of time that sunscreen provides protection from ultraviolet rays at or above a specific level is normally distributed. Specifically, one screen offers SPF 30 production for an average of six hours, with a standard deviation of 20 minutes. This means that:

- Z=0 is **exactly** 6 hours and 0 seconds
- Z=0.002 is **exactly** 6 hours and 1.2 seconds
- Z=0.0022 is **exactly** 6 hours and 1.32 seconds
- *...etc...*

The probability of the sunscreen lasting **exactly** 6 hours is nearly zero -- since the protection might drop below SPF 30 at 5 hours and 59 minutes or 6 hours and 1 minutes. However, it is also most likely that the SPF protection lasts **<u>around</u>** 6 hours. Let's explore the probability density function for this sunscreen:

<table>
<tr>
<td rowspan="3">Distribution of SPF protection of sample sunscreen that follows a normal distribution:


norm()</td>
<td>P(protection expiring at around 6 hours):</td>
</tr>
<tr>
<td>P(protection expiring at around 5 hours):</td>
</tr>
<tr>
<td>P(protection lasting at least 5 hours):</td>
</tr>
</table>

Analysis: If you were a science advisor, how often would you recommend reapplying sunscreen based on the data provided above to ensure continuous SPF 30 protection?