5.1: Probability Distributions of Continuous Random Variables

In unit 4, we organized outcomes and probabilities for a *discrete random variable*. The values of a discrete random variable can be listed in order (can be counted).

Continuous Random Variables

Sometimes the values of a variable cannot be listed in order. This is because, for any given value, it is impossible to list a *next* value. These variables are considered *continuous variables*.

For example, let's try to list the heights of all randomly selected men in order. We are not talking about specific heights, but rather, all possible heights. Consider a man who is 6 feet tall. If we could put these values in order, we should be able to tell what height comes after 6 feet or 6.0000… feet. So what is the next height on the list? Maybe we think 6.1 feet is the next height. But that isn't correct because 6.01 feet is closer to 6 feet. So is the next height 6.01 feet? No, because 6.001 is again closer to 6 feet, and 6.0001 feet is closer to that. We could continue this forever and we would never be able to find the "next height" on the list. This is why the heights of randomly selected men are values of a **continuous random variable**.

A random variable is continuous if it cannot be listed in order, or counted. Usually, continuous random variables are measured (like height, time, or weight).

1. Classify each of the random variables described as either *discrete* or *continuous*.
    a. The number of high school students in a dual enrollment program.

    b. The time it takes to sprint 500 meters.

    c. The number of assignments in an online class.

    d. The time it takes to complete all assignments in an online class.

    e. The distance a planet is from the sun.

    f. The sum of the roll of three dice.

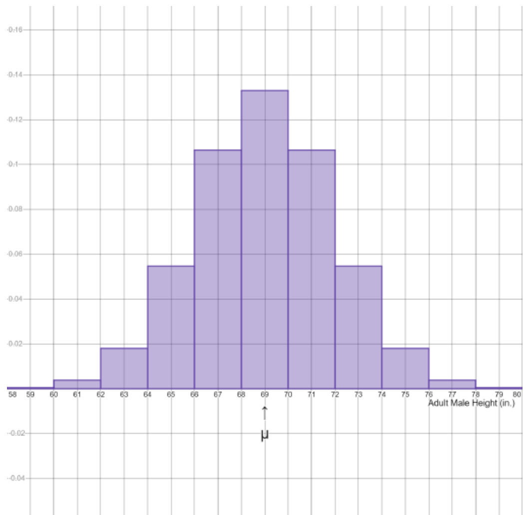
    g. The life expectancy of a person living in a blue zone.

    h. The amount of money in someone's pocket.

Statistical Inference

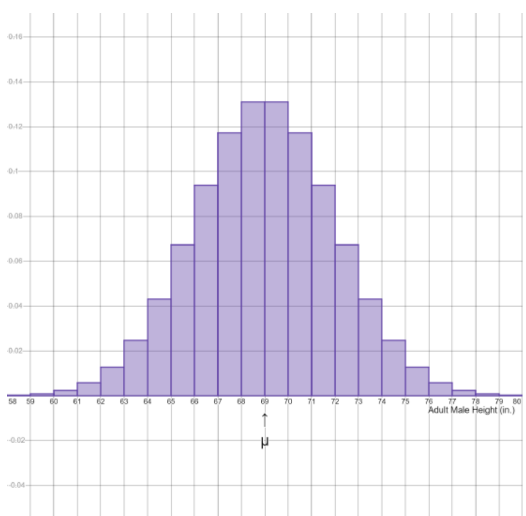
In statistical inference (where we use sample data to make judgements about a population), we often work with continuous random variables, and statistical inference depends on probability. In previous lessons, we visualized probability distributions for discrete random variables using histograms. Each value of a discrete random variable has a probability associated with it, and the probability is the area of the corresponding bar on the probability histogram.

Consider our original example, the distribution of heights of men. Below is a histogram that represents the probabilities that are estimated by relative frequencies from a large random sample of men's heights. The area of each bar represents the probability that a random man's height is within the corresponding range. The mean for all men's heights is 69 inches. Each bar in the histogram has a width of 2 inches.



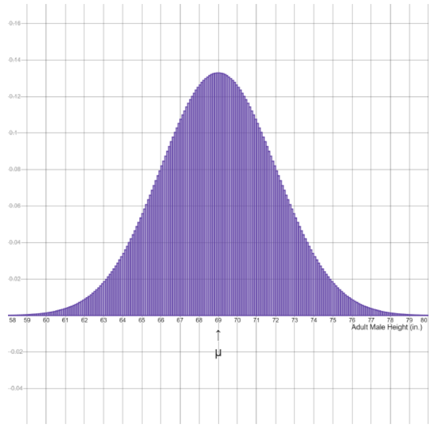
Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

The problem with using this histogram is that it only allows us to estimate probabilities for a predetermined range of values. It does not allow us to find probabilities for other ranges (like the probability that a randomly selected man is between 70 and 71 inches tall). A bar with width 1 would allow us to compute this probability. See the histogram below.



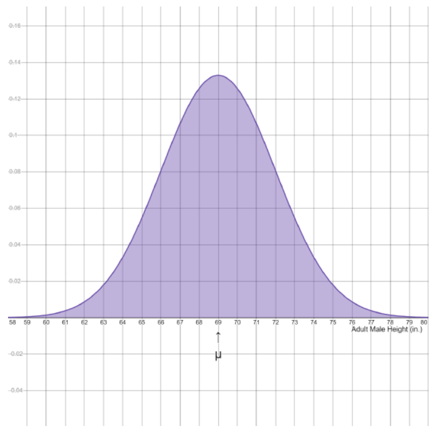
Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

With thinner bars, the graph begins to look closer to a smooth curve. In particular, this histogram is bell-shaped. Unfortunately, we still can't find the probability that a randomly selected man is between 70 and 70.1 inches tall. We can again decrease the width of each bar to be 0.1 inches. See the histogram below.



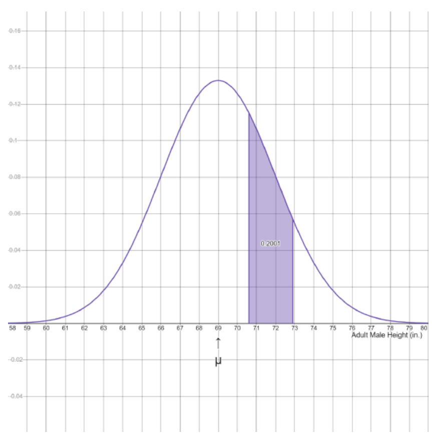
Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

We continue in this way forever until we have achieved a smooth curve. This curve is a **continuous probability distribution.** Mathematicians use Calculus to find areas under this curve which corresponds to probability. The total area under the curve is 1.



Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

Below, the area under the curve from 70.6 to 72.9 inches is shaded.



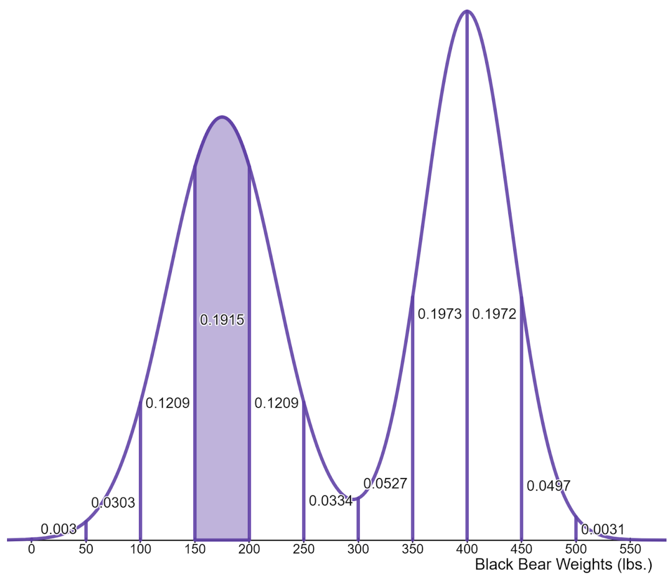
Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

The area of this shaded region is 0.2001. It represents the probability that a randomly selected man is between 70.6 and 72.9 inches tall. If x represents adult male height in inches, we can say this using probability notation: $P(70.6 < x < 72.9) = 0.2001$. The proportion of all adult males that are between 70.6 and 72.9 inches tall is around 20%.

Probabilities from a Bimodal Continuous Probability Distribution

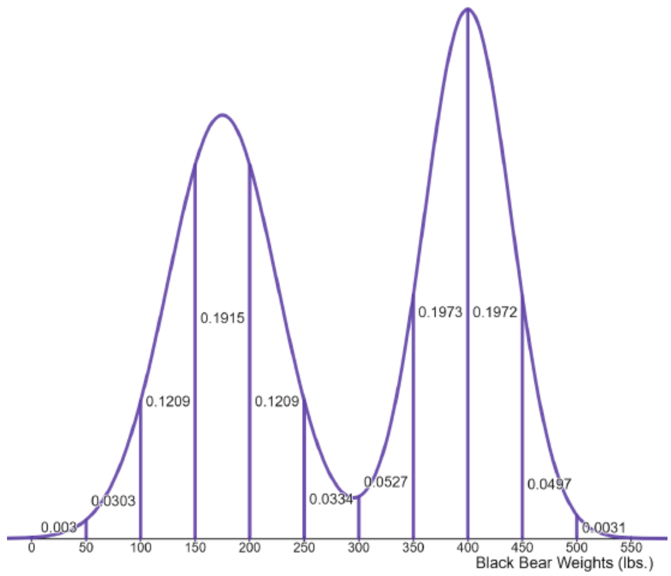
There are many continuous probability distributions. We will use bell-shaped probability curves often in this course. We may see distributions that have other shapes. Sometimes when two populations are merged into one, they form a probability distribution with two peaks. This type of distribution is called **bimodal**.

Black bears weights tend to differ by sex. Female black bears weigh 175 pounds on average, whereas, male black bears weigh 400 pounds on average. Below is the continuous probability distribution of adult black bear weights. Let the continuous random variable take on values of adult black bear weight. The curve has been divided into intervals of equal length. The area of each range is written in each region above the corresponding interval. For example, the area 0.1915 corresponds to the range of values between 150 and 200. In context, this says the probability that a randomly selected black bear weighs between 150 and 200 pounds is 0.1915 or 19.15%.



Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

Using probability notation, we say $P(150 \leq x \leq 200) = 0.1915 = 19.15\%$.



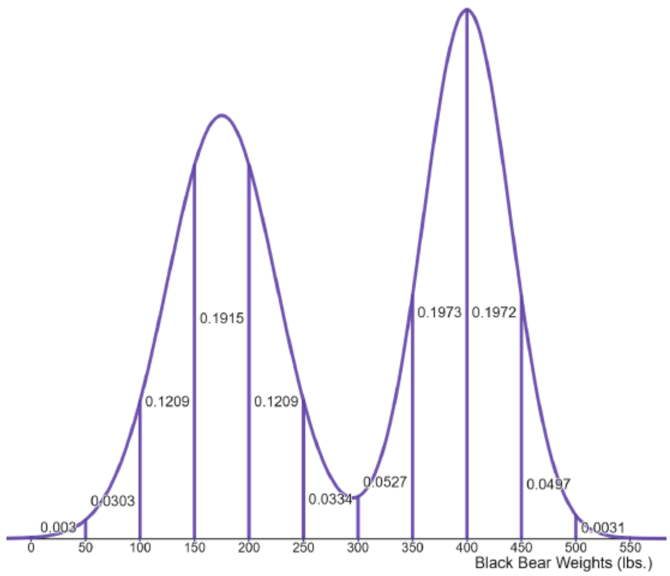
Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

2. Use the continuous probability distribution of randomly selected black bear weights above to answer the following questions.
   a. Compute the total area under the curve above.

   b. Shade the region that represents the probability that a randomly selected black bear weighs between 350 and 500 lbs.

   c. The area you shaded can be written using probability notation as $P(350 < x < 500)$. What is this probability?
   $P(350 < x < 500) =$

   d. Compute the proportion of black bears that weigh between 50 and 250 lbs. Use probability notation in your answer. Shade the area that represents the probability to show your thinking.



Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

3. Use the continuous probability distribution of randomly selected black bear weights above to answer the following questions.
   a. Shade the region that represents the probability that a randomly selected black bear weighs at most 100 lbs.

   b. The area you shaded can be written using probability notation as $P(x \leq 100)$ which is equal to P(x<100) in a continuous probability distribution. What is this probability?

      $P(x \leq 100) =$

   c. Compute the proportion of black bears that weigh at least 100 lbs. Use probability notation in your answer. *Hint: apply the complement rule*.