Densities and Random Variables

As was mentioned before in class, a **random variable** is the measured outcome of some random process. When the random variable is quantitative, we can either have a **discrete** or a **continuous** random variable.

When we have a discrete distribution of a random variable, we can list the probability associated with each possible outcome. As an example, consider a certain population of the freshwater sculpin, *Cottus rotheus*. The distribution of the random variable \( Y = \) number of tail vertebrae is shown in the following table:

<table>
<thead>
<tr>
<th>( y_i )</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(Y = y_i) )</td>
<td>.03</td>
<td>.51</td>
<td>.40</td>
<td>.06</td>
</tr>
</tbody>
</table>

We’ve listed out the entire probability distribution – all the probabilities add up to one. We could graphically represent a discrete distribution with a frequency histogram. Construct a frequency histogram for the number of tail vertebrae in this population of sculpin.
In the case where we have a continuous random variable, we want a different tool to represent this type of distribution. We call this representation of a continuous random variable’s distribution a **density curve**. Consider the distribution of blood glucose levels measured one hour after a subject (from a certain population of women) drinks 50mg of glucose dissolved in water. Example 3.27 depicts this distribution with binwidth set 10, 5 and “0” respectively.

Notice the probability density curve is like having a probability histogram where we’re squeezing the binwidth down to 0 (an infinite number of bins). Then, the way we get probabilities associated with continuous random variables is still an area, just like in the frequency histogram. But, we need area under a curve, so we need calculus.

Integration from $a$ to $b$ of the function that results in the probability density curve will give us the probability of being between those two values under the specified distribution.

Fortunately, your TI calculator will be doing the integration for you in this class! More on that later...

Now, think back to your math classes. 
Questions: What is the length of a single point? What is the area of a line? 
Answers: A single point does not have any length. A line has no area.

The following facts pertaining to probabilities associated with a continuous random variable are consequences of the fact that a line doesn’t have area:

- $P(Y = a) = 0 = P(Y = b)$ (area of a line is zero)
- $P(Y \leq a) = P(Y < a) + P(Y = a) = P(Y < a)$
- And, $P(a \leq Y \leq b) = P(a < Y \leq b) = P(a \leq Y < b) = P(a < Y < b)$
Mean and Variance of a Discrete Random Variable

The population mean of a discrete random variable, $Y$, is given by $\mu_Y = \sum y_i P\{Y = y_i\}$

The mean of a random variable, $Y$, is also known as the expected value of $Y$, denoted $E[Y]$.

The population variance of a discrete random variable, $Y$, is given by $\sigma_Y^2 = \sum (y_i - \mu_Y)^2 P\{Y = y_i\}$

One can show that $\sigma_Y^2 = E[Y^2] - (E[Y])^2 = E[Y^2] - (\mu_Y)^2$

Then, the population standard deviation is $\sigma_Y = \sqrt{\sigma_Y^2}$

Example 3.35 For the number of tail vertebrae in the population of freshwater sculpin, *Cottus rotheus*, find $\mu_Y$ and $\sigma_Y$.

<table>
<thead>
<tr>
<th>$y_i$</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P{Y = y_i}$</td>
<td>.03</td>
<td>.51</td>
<td>.40</td>
<td>.06</td>
</tr>
</tbody>
</table>