General Education Statistics
Class Notes
Discrete Random Variables (Section 6.1)

Roll two six-sided dice. We have looked at the probabilities of getting a sum of 7 or two even numbers or both. Now we ask what would we *expect* from a long string of trials. What would the mean and standard deviation of these many trials look like?

Definition: A **random variable** is a numerical measure of the outcome from a probability experiment, so its value is determined by chance. Random variables are denoted using letters such as X.

Once again, we see the difference between discrete and continuous variables. In this chapter, we will deal with discrete random variables. The next chapter covers continuous ones.

Definition: A **discrete random variable** has either a finite or countable number of values. The values of a discrete random variable can be plotted on a number line with space between each point.

Definition: A **continuous random variable** has infinitely many values. The values of a continuous random variable can be plotted on a line in an uninterrupted fashion.
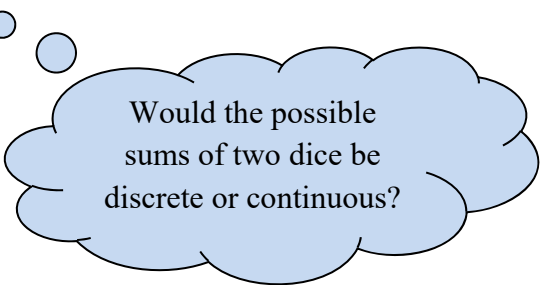
expl 1: Which are discrete and which are continuous random variables?

a.) the number of cars that travel through a Burger King drive-through in one hour

b.) the speed of cars on a highway

c.) the weights of boxes of oranges

d.) the number of free-throws made in a string of ten attempts

1

We will work closely with the probability models we have seen before. Now, they will be called probability distributions.

Definition: The **probability distribution** of a discrete random variable X provides the possible values of the random variable and their corresponding probabilities. A probability distribution can be in the form of a table, graph or mathematical formula.

expl 2: Why are the following two tables *not* proper probability distributions?

a.)

x	$P(x)$
2	.35
4	-.15
6	.45
8	.35

b.)

x	$P(x)$
1	.15
2	.25
3	.35
4	.15

We have done this before. Check that all probabilities are between 0 and 1 and they all add to 1.

We can draw histograms or similar graphs to display probability distributions. The vertical axis will represent the probabilities; the horizontal axis will be the values of X. Instead of drawing a histogram with touching bars, often, we will draw just vertical lines above each value to denote their discrete nature.

expl 3: Draw a graph to display this probability distribution. It is the number of movies streamed from Netflix in a month.

Remember, we use x to denote each value the variable takes on and X to mean the variable itself.

Number of Movies	
x	$P(x)$
0	0.06
1	0.58
2	0.22
3	0.10
4	0.03
5	0.01

Is this data skewed?

Recall, $P(2) = 0.22$ means that, if we select a random customer, the probability that he/she has streamed two movies in the month is 22%.

2

How do we get a clear picture of the traffic for streamed movies on Netflix? Can we find the mean number of movies streamed? What about our old friend, standard deviation? What would that tell us?

Definition: The **mean of a discrete random variable** is given by the formula
$\mu_x = \sum[x \cdot P(x)]$. Here, x is the value of the random variable and $P(x)$ is the probability of the value x.

Take every value that X could be, multiply it by its own probability, and then add those up.

This is also called a **weighted average** or **expected value**.

expl 4: Find the mean number of movies streamed from Netflix in a month. The table is reproduced below.

Number of Movies	
x	**$P(x)$**
0	0.06
1	0.58
2	0.22
3	0.10
4	0.03
5	0.01

This means each customer streams, on average, _____ movies per month.

What does this expected value tell us?
This mean can be thought of as the mean outcome if the experiment was repeated many times. In the case of Netflix, each customer is a "repetition of the experiment". So we get the mean number of streamed movies per month.

**Optional Worksheet: Expected Value of Lottery Ticket:**
We find the expected value of a lottery ticket. Is it worth it to buy a ticket?

Do you remember standard deviation? This number will tell us how spread out the data is. In other words, we will measure how close the data (which is to say, the outcomes of the experiment over many trials) is to that mean value.

Definition: The **standard deviation of a discrete random variable** is given by the formula $\sigma_X = \sqrt{\sum \left[x^2 \cdot P(x) \right] - \mu_X^2}$. Here, x is the value of the random variable , $P(x)$ is the probability of the value x, and μ_X is the mean found earlier.

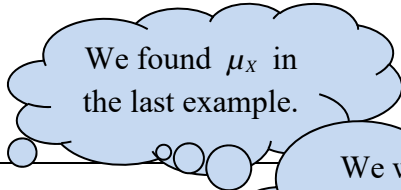The **variance** is the quantity before you square root.

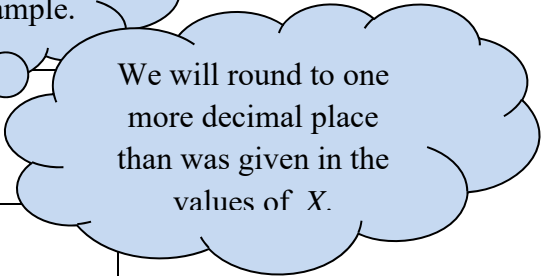Standard deviation is a measure of how far from the mean the data values are.
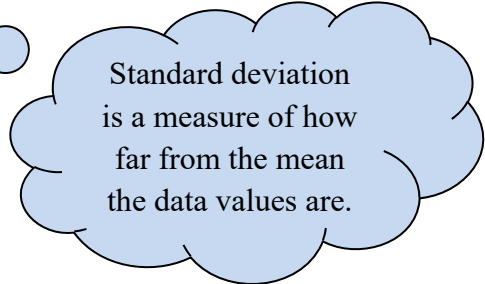
expl 5: Once again, here is the Netflix data on streamed movies. Complete the table to find the variance and standard deviation of this random variable.

Number of Movies		
x	$P(x)$	$x^2 \cdot P(x)$
0	0.06	
1	0.58	
2	0.22	
3	0.10	
4	0.03	
5	0.01	
total $= \sum \left[x^2 \cdot P(x) \right] =$		
variance $= \sum \left[x^2 \cdot P(x) \right] - \mu_X^2 =$		
standard deviation $= \sqrt{\sum \left[x^2 \cdot P(x) \right] - \mu_X^2} =$		

We found μ_X in the last example.

We will round to one more decimal place than was given in the values of X.

This means each customer streams, on average, _____ movies less than *or* greater than the mean.

4
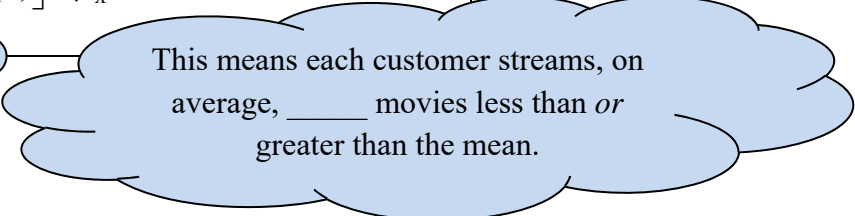
Worksheet: Expected Value and Standard Deviation of a Dart Game:
We will explore (by hand) the formulas for mean and standard deviation.

With few exceptions, you are *not* expected to do these calculations by hand. Here is how we do them using technology.

Finding Mean and Standard Deviation on the Calculator:
Enter the data in the **STAT** editor under columns **L1** and **L2**. Use **L1** for the values the variable takes on (or the values of x) and **L2** for their probabilities. Exit that screen.

Press the **STAT** button again, but now arrow over to **CALC**. Select **1: 1-Var Stats**.

If your calculator is a newer TI-84, you may get a place to tell it to use **L1** as your "List" and **L2** as your "FreqList". If not, you'll have to enter **(L1, L2)** after the "1-Var Stats" on screen. It should now read "1-Var Stats (L₁, L₂)" on screen. [You will find that **L1** and **L2** are second functions of the 1 and 2 buttons.]
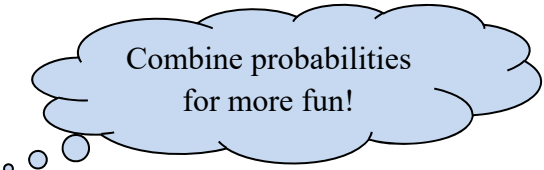
Finding Mean and Standard Deviation on StatCrunch:
Enter the values of the random variable in column **var1** and their probabilities in column **var2**.

Select **Stat > Calculators > Custom**. Tell it that your **Values** are in **var1** and your **Weights** are in **var2**. Then click **Compute!**

StatCrunch throws in the skinny histogram too. Not only that, but check out the dialog box within the display window. Fill in a value of x and step back, baby!

expl 6: Let's do one more thing with this data. Answer the following questions.

Combine probabilities for more fun!

Number of Movies	
x	$P(x)$
0	0.06
1	0.58
2	0.22
3	0.10
4	0.03
5	0.01

a.) What is the probability that a randomly selected customer streams at least three movies a month?

b.) What is the probability that a randomly selected customer streams 1 or 2 movies a month?

c.) What is the probability that a randomly selected customer streams four or more movies a month?