Chapter 5: Discrete Probability Distributions

Section 5.1: Basics of Probability Distributions

As a reminder, a variable or what will be called the random variable from now on, is represented by the letter *x* and it represents a quantitative (numerical) variable that is measured or observed in an experiment.

Also remember there are different types of quantitative variables, called discrete or continuous. What is the difference between discrete and continuous data? **Discrete** data can only take on particular values in a range. **Continuous** data can take on any value in a range. Discrete data usually arises from counting while continuous data usually arises from measuring.

Examples of each:
How tall is a plant given a new fertilizer? Continuous. This is something you measure. How many fleas are on prairie dogs in a colony? Discrete. This is something you count.

If you have a variable, and can find a probability associated with that variable, it is called a **random variable**. In many cases the random variable is what you are measuring, but when it comes to discrete random variables, it is usually what you are counting. So for the example of how tall is a plant given a new fertilizer, the random variable is the height of the plant given a new fertilizer. For the example of how many fleas are on prairie dogs in a colony, the random variable is the number of fleas on a prairie dog in a colony.

Now suppose you put all the values of the random variable together with the probability that that random variable would occur. You could then have a distribution like before, but now it is called a probability distribution since it involves probabilities. A **probability distribution** is an assignment of probabilities to the values of the random variable. The abbreviation of pdf is used for a probability distribution function.

For probability distributions, $0 \le P(x) \le 1$ and $\sum P(x) = 1$

Example #5.1.1: Probability Distribution
The 2010 U.S. Census found the chance of a household being a certain size. The data is in table #5.1.1 ("Households by age," 2013).

Table #5.1.1: Household Size from U.S. Census of 2010

Size of household	1	2	3	4	5	6	7 or more
Probability	26.7%	33.6%	15.8%	13.7%	6.3%	2.4%	1.5%

Solution:
In this case, the random variable is *x* = number of people in a household. This is a discrete random variable, since you are counting the number of people in a household.

This is a probability distribution since you have the *x* value and the probabilities that go with it, all of the probabilities are between zero and one, and the sum of all of the probabilities is one.

You can give a probability distribution in table form (as in table #5.1.1) or as a graph. The graph looks like a histogram. A probability distribution is basically a relative frequency distribution based on a very large sample.

Example #5.1.2: Graphing a Probability Distribution
The 2010 U.S. Census found the chance of a household being a certain size. The data is in the table ("Households by age," 2013). Draw a histogram of the probability distribution.

Table #5.1.2: Household Size from U.S. Census of 2010

| Size of household	1	2	3	4	5	6	7 or more
Probability	26.7%	33.6%	15.8%	13.7%	6.3%	2.4%	1.5%

Solution:
State random variable:
    x = number of people in a household

You draw a histogram, where the *x* values are on the horizontal axis and are the *x* values of the classes (for the 7 or more category, just call it 7). The probabilities are on the vertical axis.

Graph #5.1.1: Histogram of Household Size from U.S. Census of 2010
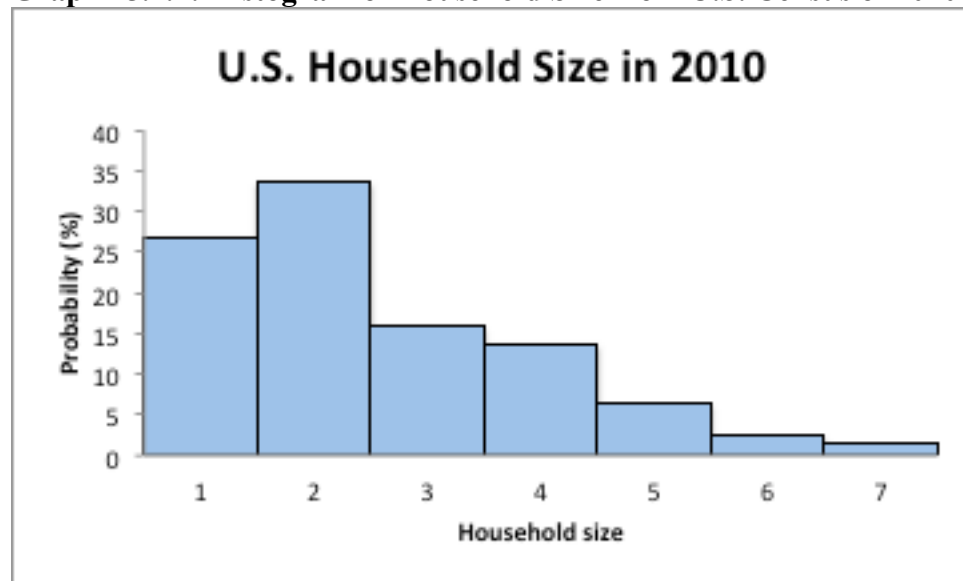
Notice this graph is skewed right.

Just as with any data set, you can calculate the mean and standard deviation. In problems involving a probability distribution function (pdf), you consider the probability distribution the population even though the pdf in most cases come from repeating an experiment many times. This is because you are using the data from repeated experiments to estimate the true probability. Since a pdf is basically a population, the mean and standard deviation that are calculated are actually the population parameters and not the sample statistics. The notation used is the same as the notation for population mean and population standard deviation that was used in chapter 3. Note: the mean can be thought of as the **expected value**. It is the value you expect to get if the trials were repeated infinite number of times. The mean or expected value does not need to be a whole number, even if the possible values of x are whole numbers.

For a discrete probability distribution function,

The mean or expected value is $\mu = \sum xP(x)$

The variance is $\sigma^2 = \sum (x-\mu)^2 P(x)$

The standard deviation is $\sigma = \sqrt{\sum (x-\mu)^2 P(x)}$

where x = the value of the random variable and $P(x)$ = the probability corresponding to a particular x value.

Example #5.1.3: Calculating Mean, Variance, and Standard Deviation for a Discrete Probability Distribution
The 2010 U.S. Census found the chance of a household being a certain size. The data is in the table ("Households by age," 2013).

Table #5.1.3: Household Size from U.S. Census of 2010

Size of household	1	2	3	4	5	6	7 or more
Probability	26.7%	33.6%	15.8%	13.7%	6.3%	2.4%	1.5%

Solution:
State random variable:
    x = number of people in a household

a.) Find the mean

Solution:
    To find the mean it is easier to just use a table as shown below. Consider the category 7 or more to just be 7. The formula for the mean says to multiply the x value by the $P(x)$ value, so add a row into the table for this calculation. Also convert all $P(x)$ to decimal form.

Table #5.1.4: Calculating the Mean for a Discrete PDF

x	1	2	3	4	5	6	7
P(x)	0.267	0.336	0.158	0.137	0.063	0.024	0.015
$xP(x)$	0.267	0.672	0.474	0.548	0.315	0.144	0.098

Now add up the new row and you get the answer 2.525. This is the mean or the expected value, $\mu = 2.525$ people. This means that you expect a household in the U.S. to have 2.525 people in it. Now of course you can't have half a person, but what this tells you is that you expect a household to have either 2 or 3 people, with a little more 3-person households than 2-person households.

b.) Find the variance

Solution:
To find the variance, again it is easier to use a table version than try to just the formula in a line. Looking at the formula, you will notice that the first operation that you should do is to subtract the mean from each x value. Then you square each of these values. Then you multiply each of these answers by the probability of each x value. Finally you add up all of these values.

Table #5.1.5: Calculating the Variance for a Discrete PDF

x	1	2	3	4	5	6	7
P(x)	0.267	0.336	0.158	0.137	0.063	0.024	0.015
$x - \mu$	-1.525	-0.525	0.475	1.475	2.475	3.475	4.475
$(x-\mu)^2$	2.3256	0.2756	0.2256	2.1756	6.1256	12.0756	20.0256
$(x-\mu)^2 P(x)$	0.6209	0.0926	0.0356	0.2981	0.3859	0.2898	0.3004

Now add up the last row to find the variance, $\sigma^2 = 2.023375$ people^2. (Note: try not to round your numbers too much so you aren't creating rounding error in your answer. The numbers in the table above were rounded off because of space limitations, but the answer was calculated using many decimal places.)
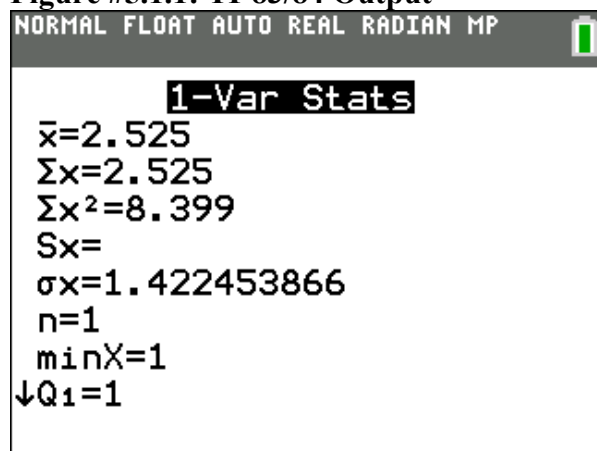
c.) Find the standard deviation

Solution:
To find the standard deviation, just take the square root of the variance, $\sigma = \sqrt{2.023375} \approx 1.422454$ people. This means that you can expect a U.S. household to have 2.525 people in it, with a standard deviation of 1.42 people.

d.) Use a TI-83/84 to calculate the mean and standard deviation.

Solution:
Go into the STAT menu, then the Edit menu. Type the *x* values into L1 and the *P(x)* values into L2. Then go into the STAT menu, then the CALC menu. Choose 1:1-Var Stats. This will put 1-Var Stats on the home screen. Now type in L1,L2 (there is a comma between L1 and L2) and then press ENTER. If you have the newer operating system on the TI-84, then your input will be slightly different. You will see the output in figure #5.1.1.

Figure #5.1.1: TI-83/84 Output

The mean is 2.525 people and the standard deviation is 1.422 people.

e.) Using R to calculate the mean.

Solution:
The command would be weighted.mean(x, p). So for this example, the process would look like:
x<-c(1, 2, 3, 4, 5, 6, 7)
p<-c(0.267, 0.336, 0.158, 0.137, 0.063, 0.024, 0.015)
weighted.mean(x, p)

Output:
[1] 2.525

So the mean is 2.525.

To find the standard deviation, you would need to program the process into R. So it is easier to just do it using the formula.

Example #5.1.4: Calculating the Expected Value
In the Arizona lottery called Pick 3, a player pays $1 and then picks a three-digit number. If those three numbers are picked in that specific order the person wins $500. What is the expected value in this game?

Solution:
To find the expected value, you need to first create the probability distribution. In this case, the random variable x = winnings. If you pick the right numbers in the right order, then you win $500, but you paid $1 to play, so you actually win $499. If you didn't pick the right numbers, you lose the $1, the x value is $-\$1$. You also need the probability of winning and losing. Since you are picking a three-digit number, and for each digit there are 10 numbers you can pick with each independent of the others, you can use the multiplication rule. To win, you have to pick the right numbers in the right order. The first digit, you pick 1 number out of 10, the second digit you pick 1 number out of 10, and the third digit you pick 1 number out of 10. The probability of picking the right number in the right order is $\frac{1}{10} * \frac{1}{10} * \frac{1}{10} = \frac{1}{1000} = 0.001$. The probability of losing (not winning) would be $1 - \frac{1}{1000} = \frac{999}{1000} = 0.999$. Putting this information into a table will help to calculate the expected value.

Table #5.1.6: Finding Expected Value

Win or lose	x	$P(x)$	$xP(x)$
Win	$499	0.001	$0.499
Lose	$-\$1$	0.999	$-\$0.999$

Now add the two values together and you have the expected value. It is $\$0.499 + (-\$0.999) = -\$0.50$. In the long run, you will expect to lose $0.50.

Since the expected value is not 0, then this game is not fair. Since you lose money, Arizona makes money, which is why they have the lottery.

The reason probability is studied in statistics is to help in making decisions in inferential statistics. To understand how that is done the concept of a rare event is needed.

Rare Event Rule for Inferential Statistics
If, under a given assumption, the probability of a particular observed event is extremely small, then you can conclude that the assumption is probably not correct.

An example of this is suppose you roll an assumed fair die 1000 times and get a six 600 times, when you should have only rolled a six around 160 times, then you should believe that your assumption about it being a fair die is untrue.

Determining if an event is unusual
If you are looking at a value of x for a discrete variable, and the P(the variable has a value of x or more) < 0.05, then you can consider the x an unusually high value. Another way to think of this is if the probability of getting such a high value is less than 0.05, then the event of getting the value x is unusual.

> Similarly, if the *P*(the variable has a value of x or less) < 0.05, then you can consider this an unusually low value. Another way to think of this is if the probability of getting a value as small as x is less than 0.05, then the event x is considered unusual.

Why is it "x or more" or "x or less" instead of just "x" when you are determining if an event is unusual? Consider this example: you and your friend go out to lunch every day. Instead of Going Dutch (each paying for their own lunch), you decide to flip a coin, and the loser pays for both. Your friend seems to be winning more often than you'd expect, so you want to determine if this is unusual before you decide to change how you pay for lunch (or accuse your friend of cheating).   The process for how to calculate these probabilities will be presented in the next section on the binomial distribution. If your friend won 6 out of 10 lunches, the probability of that happening turns out to be about 20.5%, not unusual. The probability of winning 6 or more is about 37.7%. But what happens if your friend won 501 out of 1,000 lunches? That doesn't seem so unlikely! The probability of winning 501 or more lunches is about 47.8%, and that is consistent with your hunch that this isn't so unusual. But the probability of winning exactly 501 lunches is much less, only about 2.5%. That is why the probability of getting exactly that value is not the right question to ask: you should ask the probability of getting that value or more (or that value or less on the other side).

The value 0.05 will be explained later, and it is not the only value you can use.

Example #5.1.5: Is the Event Unusual
   The 2010 U.S. Census found the chance of a household being a certain size. The data is in the table ("Households by age," 2013).

Table #5.1.7: Household Size from U.S. Census of 2010

Size of household	1	2	3	4	5	6	7 or more
Probability	26.7%	33.6%	15.8%	13.7%	6.3%	2.4%	1.5%

Solution:
State random variable:
   x = number of people in a household

a.) Is it unusual for a household to have six people in the family?

   Solution:
   To determine this, you need to look at probabilities. However, you cannot just look at the probability of six people. You need to look at the probability of x being six or more people or the probability of x being six or less people. The
   $$P(x \leq 6) = P(x=1) + P(x=2) + P(x=3) + P(x=4) + P(x=5) + P(x=6)$$
   $$= 26.7\% + 33.6\% + 15.8\% + 13.7\% + 6.3\% + 2.4\%$$
   $$= 98.5\%$$
   Since this probability is more than 5%, then six is not an unusually low value. The

$$P(x \geq 6) = P(x = 6) + P(x \geq 7)$$
$$= 2.4\% + 1.5\%$$
$$= 3.9\%$$

Since this probability is less than 5%, then six is an unusually high value. It is unusual for a household to have six people in the family.

b.) If you did come upon many families that had six people in the family, what would you think?

Solution:
Since it is unusual for a family to have six people in it, then you may think that either the size of families is increasing from what it was or that you are in a location where families are larger than in other locations.

c.) Is it unusual for a household to have four people in the family?

Solution:
To determine this, you need to look at probabilities. Again, look at the probability of x being four or more or the probability of x being four or less. The

$$P(x \geq 4) = P(x = 4) + P(x = 5) + P(x = 6) + P(x = 7)$$
$$= 13.7\% + 6.3\% + 2.4\% + 1.5\%$$
$$= 23.9\%$$

Since this probability is more than 5%, four is not an unusually high value. The

$$P(x \leq 4) = P(x = 1) + P(x = 2) + P(x = 3) + P(x = 4)$$
$$= 26.7\% + 33.6\% + 15.8\% + 13.7\%$$
$$= 89.8\%$$

Since this probability is more than 5%, four is not an unusually low value. Thus, four is not an unusual size of a family.

d.) If you did come upon a family that has four people in it, what would you think?

Solution:
Since it is not unusual for a family to have four members, then you would not think anything is amiss.

Section 5.1: Homework

1.) Eyeglassomatic manufactures eyeglasses for different retailers. The number of days it takes to fix defects in an eyeglass and the probability that it will take that number of days are in the table.

Table #5.1.8: Number of Days to Fix Defects

Number of days	Probabilities
1	24.9%
2	10.8%
3	9.1%
4	12.3%
5	13.3%
6	11.4%
7	7.0%
8	4.6%
9	1.9%
10	1.3%
11	1.0%
12	0.8%
13	0.6%
14	0.4%
15	0.2%
16	0.2%
17	0.1%
18	0.1%

a.) State the random variable.
b.) Draw a histogram of the number of days to fix defects
c.) Find the mean number of days to fix defects.
d.) Find the variance for the number of days to fix defects.
e.) Find the standard deviation for the number of days to fix defects.
f.) Find probability that a lens will take at least 16 days to make a fix the defect.
g.) Is it unusual for a lens to take 16 days to fix a defect?
h.) If it does take 16 days for eyeglasses to be repaired, what would you think?

2.)   Suppose you have an experiment where you flip a coin three times. You then
        count the number of heads.
        a.) State the random variable.
        b.) Write the probability distribution for the number of heads.
        c.) Draw a histogram for the number of heads.
        d.) Find the mean number of heads.
        e.) Find the variance for the number of heads.
        f.) Find the standard deviation for the number of heads.
        g.) Find the probability of having two or more number of heads.
        h.) Is it unusual for to flip two heads?

3.)   The Ohio lottery has a game called Pick 4 where a player pays $1 and picks a
        four-digit number. If the four numbers come up in the order you picked, then you
        win $2,500. What is your expected value?

4.)   An LG Dishwasher, which costs $800, has a 20% chance of needing to be
        replaced in the first 2 years of purchase. A two-year extended warrantee costs
        $112.10 on a dishwasher. What is the expected value of the extended warranty
        assuming it is replaced in the first 2 years?

Section 5.2: Binomial Probability Distribution

Section 5.1 introduced the concept of a probability distribution. The focus of the section was on discrete probability distributions (pdf). To find the pdf for a situation, you usually needed to actually conduct the experiment and collect data. Then you can calculate the experimental probabilities. Normally you cannot calculate the theoretical probabilities instead. However, there are certain types of experiment that allow you to calculate the theoretical probability. One of those types is called a **Binomial Experiment**.

Properties of a **binomial experiment** (or Bernoulli trial):
1) Fixed number of trials, *n*, which means that the experiment is repeated a specific number of times.
2) The *n* trials are independent, which means that what happens on one trial does not influence the outcomes of other trials.
3) There are only two outcomes, which are called a success and a failure.
4) The probability of a success doesn't change from trial to trial, where *p* = probability of success and *q* = probability of failure, $q = 1 - p$.

If you know you have a binomial experiment, then you can calculate binomial probabilities. This is important because binomial probabilities come up often in real life. Examples of binomial experiments are:

Toss a fair coin ten times, and find the probability of getting two heads.
Question twenty people in class, and look for the probability of more than half being women?
Shoot five arrows at a target, and find the probability of hitting it five times?

To develop the process for calculating the probabilities in a binomial experiment, consider example #5.2.1.

Example #5.2.1: Deriving the Binomial Probability Formula

Suppose you are given a 3 question multiple-choice test. Each question has 4 responses and only one is correct. Suppose you want to find the probability that you can just guess at the answers and get 2 questions right. (Teachers do this all the time when they make up a multiple-choice test to see if students can still pass without studying. In most cases the students can't.) To help with the idea that you are going to guess, suppose the test is in Martian.

a.) What is the random variable?

Solution:
x = number of correct answers

b.) Is this a binomial experiment?

Solution:
1.) There are 3 questions, and each question is a trial, so there are a fixed number of trials. In this case, $n = 3$.
2.) Getting the first question right has no affect on getting the second or third question right, thus the trials are independent.
3.) Either you get the question right or you get it wrong, so there are only two outcomes. In this case, the success is getting the question right.
4.) The probability of getting a question right is one out of four. This is the same for every trial since each question has 4 responses. In this case,

$$p = \frac{1}{4} \text{ and } q = 1 - \frac{1}{4} = \frac{3}{4}$$

This is a binomial experiment, since all of the properties are met.

c.) What is the probability of getting 2 questions right?

Solution:
To answer this question, start with the sample space.
SS = {RRR, RRW, RWR, WRR, WWR, WRW, RWW, WWW}, where RRW means you get the first question right, the second question right, and the third question wrong. The same is similar for the other outcomes.

Now the event space for getting 2 right is {RRW, RWR, WRR}. What you did in chapter four was just to find three divided by eight. However, this would not be right in this case. That is because the probability of getting a question right is different from getting a question wrong. What else can you do?

Look at just P(RRW) for the moment. Again, that means
P(RRW) = P(R on 1st, R on 2nd, and W on 3rd)
Since the trials are independent, then
P(RRW) = P(R on 1st, R on 2nd, and W on 3rd)
= P(R on 1st) * P(R on 2nd) * P(W on 3rd)
Just multiply p * p * q

$$P(\text{RRW}) = \frac{1}{4} * \frac{1}{4} * \frac{3}{4} = \left(\frac{1}{4}\right)^2 \left(\frac{3}{4}\right)^1$$

The same is true for P(RWR) and P(WRR). To find the probability of 2 correct answers, just add these three probabilities together. You get

$$P(2 \text{ correct answers}) = P(\text{RRW}) + P(\text{RWR}) + P(\text{WRR})$$

$$= \left(\frac{1}{4}\right)^2 \left(\frac{3}{4}\right)^1 + \left(\frac{1}{4}\right)^2 \left(\frac{3}{4}\right)^1 + \left(\frac{1}{4}\right)^2 \left(\frac{3}{4}\right)^1$$

$$= 3\left(\frac{1}{4}\right)^2 \left(\frac{3}{4}\right)^1$$

d.) What is the probability of getting zero right, one right, and all three right?

Solution:
You could go through the same argument that you did above and come up with the following:

r right	P(r right)
0 right	$1 * \left(\frac{1}{4}\right)^0 \left(\frac{3}{4}\right)^3$
1 right	$3 * \left(\frac{1}{4}\right)^1 \left(\frac{3}{4}\right)^2$
2 right	$3 * \left(\frac{1}{4}\right)^2 \left(\frac{3}{4}\right)^1$
3 right	$1 * \left(\frac{1}{4}\right)^3 \left(\frac{3}{4}\right)^0$

Hopefully you see the pattern that results. You can now write the general formula for the probabilities for a Binomial experiment

First, the random variable in a binomial experiment is x = number of successes. Be careful, a success is not always a good thing. Sometimes a success is something that is bad, like finding a defect. A success just means you observed the outcome you wanted to see happen.

Binomial Formula for the probability of r successes in n trials is

$$P(x = r) = {}_n C_r \, p^r q^{n-r} \text{ where } {}_n C_r = \frac{n!}{r!(n-r)!}$$

The ${}_n C_r$ is the number of combinations of n things taking r at a time. It is read "n choose r". Some other common notations for n choose r are $C_{n,r}$, and $\binom{n}{r}$. $n!$ means you are multiplying $n * (n-1) * (n-2) * \cdots * 2 * 1$. As an example, $5! = 5 * 4 * 3 * 2 * 1 = 120$.

When solving problems, make sure you define your random variable and state what n, p, q, and r are. Without doing this, the problems are a great deal harder.

Example #5.2.2: Calculating Binomial Probabilities
    When looking at a person's eye color, it turns out that 1% of people in the world has green eyes ("What percentage of," 2013). Consider a group of 20 people.
    a.) State the random variable.

        Solution:
        x = number of people with green eyes

    b.) Argue that this is a binomial experiment

        Solution:
        1.) There are 20 people, and each person is a trial, so there are a fixed number of trials. In this case, $n = 20$.
        2.) If you assume that each person in the group is chosen at random the eye color of one person doesn't affect the eye color of the next person, thus the trials are independent.
        3.) Either a person has green eyes or they do not have green eyes, so there are only two outcomes. In this case, the success is a person has green eyes.
        4.) The probability of a person having green eyes is 0.01. This is the same for every trial since each person has the same chance of having green eyes.
            $p = 0.01$ and $q = 1 - 0.01 = 0.99$

    Find the probability that
    c.) None have green eyes.

        Solution:
        $$P(x = 0) = {}_{20}C_0 (0.01)^0 (0.99)^{20-0} \approx 0.818$$

    d.) Nine have green eyes.

        Solution:
        $$P(x = 9) = {}_{20}C_9 (0.01)^9 (0.99)^{20-9} \approx 1.50 \times 10^{-13} \approx 0.000$$

    e.) At most three have green eyes.

        Solution:
        At most three means that three is the highest value you will have. Find the probability of x is less than or equal to three.

$$P(x \le 3) = P(x=0) + P(x=1) + P(x=2) + P(x=3)$$
$$= _{20}C_0 (0.01)^0 (0.99)^{20} + _{20}C_1 (0.01)^1 (0.99)^{19}$$
$$+ _{20}C_2 (0.01)^2 (0.99)^{18} + _{20}C_3 (0.01)^3 (0.99)^{17}$$
$$\approx 0.818 + 0.165 + 0.016 + 0.001 > 0.999$$

The reason the answer is written as being greater than 0.999 is because the answer is actually 0.9999573791, and when that is rounded to three decimal places you get 1. But 1 means that the event will happen, when in reality there is a slight chance that it won't happen. It is best to write the answer as greater than 0.999 to represent that the number is very close to 1, but isn't 1.

f.) At most two have green eyes.

Solution:
$$P(x \le 2) = P(x=0) + P(x=1) + P(x=2)$$
$$= _{20}C_0 (0.01)^0 (0.99)^{20} + _{20}C_1 (0.01)^1 (0.99)^{19} + _{20}C_2 (0.01)^2 (0.99)^{18}$$
$$\approx 0.818 + 0.165 + 0.016 \approx 0.999$$

g.) At least four have green eyes.

Solution:
At least four means four or more. Find the probability of x being greater than or equal to four. That would mean adding up all the probabilities from four to twenty. This would take a long time, so it is better to use the idea of complement. The complement of being greater than or equal to four is being less than four. That would mean being less than or equal to three. Part (e) has the answer for the probability of being less than or equal to three. Just subtract that number from 1.
$$P(x \ge 4) = 1 - P(x \le 3) = 1 - 0.999 = 0.001 .$$
Actually the answer is less than 0.001, but it is fine to write it this way.

h.) In Europe, four people out of twenty have green eyes. Is this unusual? What does that tell you?

Solution:
Since the probability of finding four or more people with green eyes is much less than 0.05, it is unusual to find four people out of twenty with green eyes. That should make you wonder if the proportion of people in Europe with green eyes is more than the 1% for the general population. If this is true, then you may want to ask why Europeans have a higher proportion of green-eyed people. That of course could lead to more questions.

The binomial formula is cumbersome to use, so you can find the probabilities by using technology. On the TI-83/84 calculator, the commands on the TI-83/84 calculators when the number of trials is equal to n and the probability of a success is equal to p are binompdf(n,p,r) when you want to find $P(x=r)$ and binomcdf(n,p,r) when you want to find $P(x \leq r)$. If you want to find $P(x \geq r)$, then you use the property that $P(x \geq r) = 1 - P(x \leq r-1)$, since $x \geq r$ and $x < r$ or $x \leq r-1$ are complementary events. Both binompdf and binomcdf commands are found in the DISTR menu. Using R, the commands are $P(x=r) = $ dbinom(r,n,p) and $P(x \leq r) = $ pbinom(r,n,p).

Example #5.2.3: Using the Binomial Command on the TI-83/84
When looking at a person's eye color, it turns out that 1% of people in the world has green eyes ("What percentage of," 2013). Consider a group of 20 people.
a.) State the random variable.
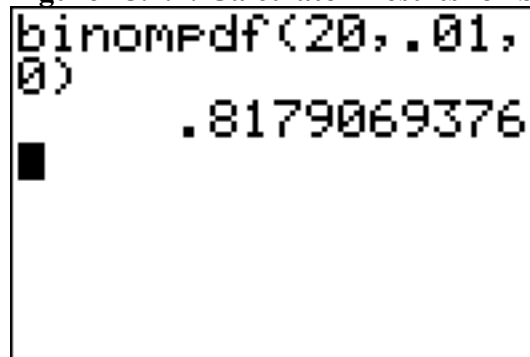
Solution:
$x = $ number of people with green eyes

Find the probability that
b.) None have green eyes.

Solution:
You are looking for $P(x=0)$. Since this problem is $x=0$, you use the binompdf command on the TI-83/84 or dbinom command on R. On the TI-83/84, you go to the DISTR menu, select the binompdf, and then type into the parenthesis your n, p, and r values into your calculator, making sure you use the comma to separate the values. The command will look like binompdf$(20,.01,0)$ and when you press ENTER you will be given the answer. (If you have the new software on the TI-84, the screen looks a bit different.)

Figure #5.2.1: Calculator Results for binompdf

```
binompdf(20,.01,
0)
        .8179069376
■
```

On R, the command would look like dbinom(0, 20, 0.01)

$P(x=0)=0.8179$. Thus there is an 81.8% chance that in a group of 20 people none of them will have green eyes.
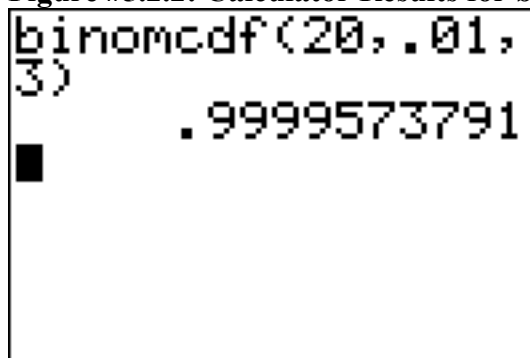
c.) Nine have green eyes.

Solution:
In this case you want to find the $P(x=9)$. Again, you will use the binompdf command or the dbinom command. Following the procedure above, you will have $\text{binompdf}(20,.01,9)$ on the TI-83/84 or dbinom(9,20,0.01) on R. Your answer is $P(x=9)=1.50\times10^{-13}$. (Remember when the calculator gives you $1.50E-13$ and R give you $1.50e-13$, this is how they display scientific notation.) The probability that out of twenty people, nine of them have green eyes is a very small chance.

d.) At most three have green eyes.

Solution:
At most three means that three is the highest value you will have. Find the probability of x being less than or equal to three, which is $P(x\le 3)$. This uses the binomcdf command on the TI-83/84 and pbinom command in R. You use the command on the TI-83/84 of $\text{binomcdf}(20,.01,3)$ and the command on R of pbinom(3,20,0.01)

Figure #5.2.2: Calculator Results for binomcdf

```
binomcdf(20,.01,
3)
         .9999573791
■
```

Your answer is 0.99996. Thus there is a really good chance that in a group of 20 people at most three will have green eyes. (Note: don't round this to one, since one means that the event will happen, when in reality there is a slight chance that it won't happen. It is best to write the answer out to enough decimal points so it doesn't round off to one.

e.) At most two have green eyes.

Solution:
You are looking for $P(x \leq 2)$. Again use binomcdf or pbinom. Following the procedure above you will have binomcdf$(20,.01,2)$ on the TI-83/84 and pbinom(2,20,0.01), with $P(x \leq 2) = 0.998996$. Again there is a really good chance that at most two people in the room will have green eyes.

f.) At least four have green eyes.

Solution:
At least four means four or more. Find the probability of x being greater than or equal to four. That would mean adding up all the probabilities from four to twenty. This would take a long time, so it is better to use the idea of complement. The complement of being greater than or equal to four is being less than four. That would mean being less than or equal to three. Part (e) has the answer for the probability of being less than or equal to three. Just subtract that number from 1.
$P(x \geq 4) = 1 - P(x \leq 3) = 1 - 0.99996 = 0.00004$. You can also find this answer by doing the following on TI-83/84:
$P(x \geq 4) = 1 - P(x \leq 3) = 1 - \text{binomcdf}(20,.01,3) = 1 - 0.99996 = 0.00004$
on R:
$P(x \geq 4) = 1 - P(x \leq 3) = 1 - \text{pbinom}(3,20,.01) = 1 - 0.99996 = 0.00004$
Again, this is very unlikely to happen.

There are other technologies that will compute binomial probabilities.

Example #5.2.4: Calculating Binomial Probabilities
According to the Center for Disease Control (CDC), about 1 in 88 children in the U.S. have been diagnosed with autism ("CDC-data and statistics,," 2013). Suppose you consider a group of 10 children.

a.) State the random variable.

Solution:
x = number of children with autism.

b.) Argue that this is a binomial experiment

Solution:
1.) There are 10 children, and each child is a trial, so there are a fixed number of trials. In this case, $n = 10$.

2.) If you assume that each child in the group is chosen at random, then whether a child has autism does not affect the chance that the next child has autism. Thus the trials are independent.

3.) Either a child has autism or they do not have autism, so there are two outcomes. In this case, the success is a child has autism.

4.) The probability of a child having autism is 1/88. This is the same for every trial since each child has the same chance of having autism. $p = \dfrac{1}{88}$ and $q = 1 - \dfrac{1}{88} = \dfrac{87}{88}$.

Find the probability that

c.) None have autism.

Solution:
Using the formula:

$$P(x = 0) = {}_{10}C_0 \left(\frac{1}{88}\right)^0 \left(\frac{87}{88}\right)^{10-0} \approx 0.892$$

Using the TI-83/84 Calculator:
$$P(x = 0) = \text{binompdf}(10, 1 \div 88, 0) \approx 0.892$$

Using R:
$$P(x = 0) = \text{pbinom}(0, 10, 1/88) \approx 0.892$$

d.) Seven have autism.

Solution:
Using the formula:

$$P(x = 7) = {}_{10}C_7 \left(\frac{1}{88}\right)^7 \left(\frac{87}{88}\right)^{10-7} \approx 0.000$$

Using the TI-83/84 Calculator:
$$P(x = 7) = \text{binompdf}(10, 1 \div 88, 7) \approx 2.84 \times 10^{-12}$$

Using R:
$$P(x = 7) = \text{dbinom}(7, 10, 1/88) \approx 2.84 \times 10^{-12}$$

e.) At least five have autism.

Solution:
Using the formula:
$$P(x \geq 5) = P(x=5) + P(x=6) + P(x=7)$$
$$+ P(x=8) + P(x=9) + P(x=10)$$

$$=_{10}C_5\left(\frac{1}{88}\right)^5\left(\frac{78}{88}\right)^{10-5} +_{10}C_6\left(\frac{1}{88}\right)^6\left(\frac{78}{88}\right)^{10-6}$$

$$+_{10}C_7\left(\frac{1}{88}\right)^7\left(\frac{78}{88}\right)^{10-7} +_{10}C_8\left(\frac{1}{88}\right)^8\left(\frac{78}{88}\right)^{10-8}$$

$$+_{10}C_9\left(\frac{1}{88}\right)^9\left(\frac{78}{88}\right)^{10-9} +_{10}C_{10}\left(\frac{1}{88}\right)^{10}\left(\frac{78}{88}\right)^{10-10}$$

$$= 0.000 + 0.000 + 0.000 + 0.000 + 0.000 + 0.000$$
$$= 0.000$$

Using the TI-83/84 Calculator:
To use the calculator you need to use the complement.
$$P(x \geq 5) = 1 - P(x < 5)$$
$$= 1 - P(x \leq 4)$$
$$= 1 - \text{binomcdf}(10, 1 \div 88, 4)$$
$$\approx 1 - 0.9999999 = 0.000$$

Using R:
To use R you need to use the complement.
$$P(x \geq 5) = 1 - P(x < 5)$$
$$= 1 - P(x \leq 4)$$
$$= 1 - \text{pbinom}(4, 10, 1/88)$$
$$\approx 1 - 0.9999999 = 0.000$$

Notice, the answer is given as 0.000, since the answer is less than 0.000. Don't write 0, since 0 means that the event is impossible to happen. The event of five or more is improbable, but not impossible.

f.) At most two have autism.

Solution:
Using the formula:
$$P(x \le 2) = P(x = 0) + P(x = 1) + P(x = 2)$$

$$=_{10} C_0 \left(\frac{1}{88}\right)^0 \left(\frac{78}{88}\right)^{10-0} +_{10} C_1 \left(\frac{1}{88}\right)^1 \left(\frac{78}{88}\right)^{10-1}$$

$$+_{10} C_2 \left(\frac{1}{88}\right)^2 \left(\frac{78}{88}\right)^{10-2}$$

$$= 0.892 + 0.103 + 0.005 > 0.999$$

Using the TI-83/84 Calculator:
$$P(x \le 2) = \text{binomcdf}(10, 1 \div 88, 2) \approx 0.9998$$

Using R:
$$P(x \le 2) = \text{pbinom}(2, 10, 1/88) \approx 0.9998$$

g.) Suppose five children out of ten have autism. Is this unusual? What does that tell you?

Solution:
Since the probability of five or more children in a group of ten having autism is much less than 5%, it is unusual to happen. If this does happen, then one may think that the proportion of children diagnosed with autism is actually more than 1/88.

Section 5.2: Homework
1.) Suppose a random variable, x, arises from a binomial experiment. If $n = 14$, and $p = 0.13$, find the following probabilities using the binomial formula.

a.) $P(x = 5)$

b.) $P(x = 8)$

c.) $P(x = 12)$

d.) $P(x \le 4)$

e.) $P(x \ge 8)$

f.) $P(x \le 12)$

2.)   Suppose a random variable, x, arises from a binomial experiment. If $n = 22$, and p = 0.85, find the following probabilities using the binomial formula.

a.) $P(x = 18)$

b.) $P(x = 5)$

c.) $P(x = 20)$

d.) $P(x \leq 3)$

e.) $P(x \geq 18)$

f.) $P(x \geq 20)$

3.)   Suppose a random variable, x, arises from a binomial experiment. If $n = 10$, and p = 0.70, find the following probabilities using technology.

a.) $P(x = 2)$

b.) $P(x = 8)$

c.) $P(x = 7)$

d.) $P(x \leq 3)$

e.) $P(x \geq 7)$

f.) $P(x \leq 4)$

4.)   Suppose a random variable, x, arises from a binomial experiment. If $n = 6$, and p = 0.30, find the following probabilities using technology.

a.) $P(x = 1)$

b.) $P(x = 5)$

c.) $P(x = 3)$

d.) $P(x \leq 3)$

e.) $P(x \geq 5)$

f.) $P(x \leq 4)$

5.)   Suppose a random variable, x, arises from a binomial experiment. If $n = 17$, and p = 0.63, find the following probabilities using technology.

a.) $P(x = 8)$

b.) $P(x = 15)$

c.) $P(x = 14)$

d.) $P(x \leq 12)$

e.) $P(x \geq 10)$

f.) $P(x \leq 7)$

6.)   Suppose a random variable, x, arises from a binomial experiment. If $n = 23$, and p = 0.22, find the following probabilities using technology.

a.) $P(x = 21)$

b.) $P(x = 6)$

c.) $P(x = 12)$

d.) $P(x \leq 14)$

e.) $P(x \geq 17)$

f.) $P(x \leq 9)$

7.)   Approximately 10% of all people are left-handed ("11 little-known facts," 2013). Consider a grouping of fifteen people.

a.) State the random variable.

b.) Argue that this is a binomial experiment

Find the probability that

c.) None are left-handed.

d.) Seven are left-handed.

e.) At least two are left-handed.

f.) At most three are left-handed.

g.) At least seven are left-handed.

h.) Seven of the last 15 U.S. Presidents were left-handed. Is this unusual? What does that tell you?

8.)   According to an article in the American Heart Association's publication *Circulation*, 24% of patients who had been hospitalized for an acute myocardial infarction did not fill their cardiac medication by the seventh day of being discharged (Ho, Bryson & Rumsfeld, 2009). Suppose there are twelve people who have been hospitalized for an acute myocardial infarction.

a.) State the random variable.

b.) Argue that this is a binomial experiment

Find the probability that

c.) All filled their cardiac medication.

d.) Seven did not fill their cardiac medication.

e.) None filled their cardiac medication.

f.) At most two did not fill their cardiac medication.

g.) At least three did not fill their cardiac medication.

h.) At least ten did not fill their cardiac medication.

i.) Suppose of the next twelve patients discharged, ten did not fill their cardiac medication, would this be unusual? What does this tell you?

9.)   Eyeglassomatic manufactures eyeglasses for different retailers.   In March 2010, they tested to see how many defective lenses they made, and there were 16.9% defective lenses due to scratches. Suppose Eyeglassomatic examined twenty eyeglasses.
a.) State the random variable.
b.) Argue that this is a binomial experiment
Find the probability that
c.) None are scratched.
d.) All are scratched.
e.) At least three are scratched.
f.) At most five are scratched.
g.) At least ten are scratched.
h.) Is it unusual for ten lenses to be scratched? If it turns out that ten lenses out of twenty are scratched, what might that tell you about the manufacturing process?

10.)   The proportion of brown M&M's in a milk chocolate packet is approximately 14% (Madison, 2013). Suppose a package of M&M's typically contains 52 M&M's.
a.) State the random variable.
b.) Argue that this is a binomial experiment
     Find the probability that
c.) Six M&M's are brown.
d.) Twenty-five M&M's are brown.
e.) All of the M&M's are brown.
f.) Would it be unusual for a package to have only brown M&M's? If this were to happen, what would you think is the reason?

Section 5.3 Mean and Standard Deviation of Binomial Distribution

If you list all possible values of *x* in a Binomial distribution, you get the **Binomial Probability Distribution** (pdf). You can draw a histogram of the pdf and find the mean, variance, and standard deviation of it.

For a general discrete probability distribution, you can find the mean, the variance, and the standard deviation for a pdf using the general formulas

$$\mu = \sum xP(x), \ \sigma^2 = \sum (x-\mu)^2 P(x), \text{ and } \sigma = \sqrt{\sum (x-\mu)^2 P(x)}.$$

These formulas are useful, but if you know the type of distribution, like Binomial, then you can find the mean and standard deviation using easier formulas. They are derived from the general formulas.

For a Binomial distribution, μ, the expected number of successes, σ^2, the variance, and σ, the standard deviation for the number of success are given by the formulas:

$$\mu = np \qquad \sigma^2 = npq \qquad \sigma = \sqrt{npq}$$

Where *p* is the probability of success and $q = 1 - p$.

Example #5.3.1: Finding the Probability Distribution, Mean, Variance, and Standard Deviation of a Binomial Distribution

　　　When looking at a person's eye color, it turns out that 1% of people in the world has green eyes ("What percentage of," 2013). Consider a group of 20 people.
　　　a.) State the random variable.

　　　Solution:
　　　x = number of people who have green eyes

　　　b.) Write the probability distribution.

　　　Solution:
　　　In this case you need to write each value of *x* and its corresponding probability. It is easiest to do this by using the binompdf command, but don't put in the *r* value. You may want to set your calculator to only three decimal places, so it is easier to see the values and you don't need much more precision than that. The command would look like $\text{binompdf}(20, .01)$.
　　　This produces the information in table #5.3.1

Table #5.3.1: Probability Distribution for Number of People with Green Eyes
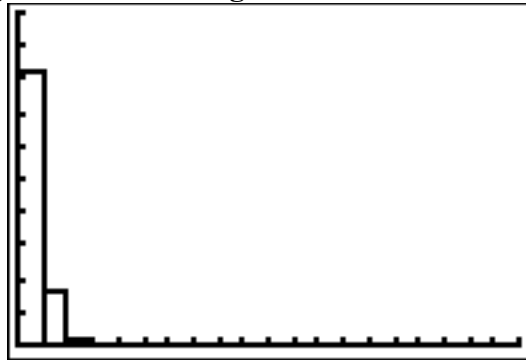
x	$P(x=r)$
0	0.818
1	0.165
2	0.016
3	0.001
4	0.000
5	0.000
6	0.000
7	0.000
8	0.000
9	0.000
10	0.000
⋮	⋮
20	0.000

Notice that after x = 4, the probability values are all 0.000. This just means they are really small numbers.

c.) Draw a histogram.

Solution:
You can draw the histogram on the TI-83/84 or other technology. The graph would look like in figure #5.3.1.

Figure #5.3.1: Histogram Created on TI-83/84

This graph is very skewed to the right.

d.) Find the mean.

Solution:
Since this is a binomial, then you can use the formula $\mu = np$. So
$\mu = 20(0.01) = 0.2$ people
You expect on average that out of 20 people, less than 1 would have green eyes.

e.) Find the variance.

> **Solution:**
> Since this is a binomial, then you can use the formula $\sigma^2 = npq$.
> $q = 1 - 0.01 = 0.99$
> $\sigma^2 = 20(0.01)(0.99) = 0.198$ people^2

f.) Find the standard deviation.

> **Solution:**
> Once you have the variance, you just take the square root of the variance to find the standard deviation.
> $\sigma = \sqrt{0.198} \approx 0.445$ people

Section 5.3: Homework

1.) Suppose a random variable, x, arises from a binomial experiment. Suppose $n = 6$, and $p = 0.13$.
   a.) Write the probability distribution.
   b.) Draw a histogram.
   c.) Describe the shape of the histogram.
   d.) Find the mean.
   e.) Find the variance.
   f.) Find the standard deviation.

2.) Suppose a random variable, x, arises from a binomial experiment. Suppose $n = 10$, and $p = 0.81$.
   a.) Write the probability distribution.
   b.) Draw a histogram.
   c.) Describe the shape of the histogram.
   d.) Find the mean.
   e.) Find the variance.
   f.) Find the standard deviation.

3.) Suppose a random variable, x, arises from a binomial experiment. Suppose $n = 7$, and $p = 0.50$.
   a.) Write the probability distribution.
   b.) Draw a histogram.
   c.) Describe the shape of the histogram.
   d.) Find the mean.
   e.) Find the variance.
   f.) Find the standard deviation.

4.)   Approximately 10% of all people are left-handed. Consider a grouping of fifteen people.
      a.) State the random variable.
      b.) Write the probability distribution.
      c.) Draw a histogram.
      d.) Describe the shape of the histogram.
      e.) Find the mean.
      f.) Find the variance.
      g.) Find the standard deviation.

5.)   According to an article in the American Heart Association's publication *Circulation*, 24% of patients who had been hospitalized for an acute myocardial infarction did not fill their cardiac medication by the seventh day of being discharged (Ho, Bryson & Rumsfeld, 2009). Suppose there are twelve people who have been hospitalized for an acute myocardial infarction.
      a.) State the random variable.
      b.) Write the probability distribution.
      c.) Draw a histogram.
      d.) Describe the shape of the histogram.
      e.) Find the mean.
      f.) Find the variance.
      g.) Find the standard deviation.

6.)   Eyeglassomatic manufactures eyeglasses for different retailers.   In March 2010, they tested to see how many defective lenses they made, and there were 16.9% defective lenses due to scratches. Suppose Eyeglassomatic examined twenty eyeglasses.
      a.) State the random variable.
      b.) Write the probability distribution.
      c.) Draw a histogram.
      d.) Describe the shape of the histogram.
      e.) Find the mean.
      f.) Find the variance.
      g.) Find the standard deviation.

7.)   The proportion of brown M&M's in a milk chocolate packet is approximately 14% (Madison, 2013). Suppose a package of M&M's typically contains 52 M&M's.
      a.) State the random variable.
      b.) Find the mean.
      c.) Find the variance.
      d.) Find the standard deviation.

Data Sources:

11 little-known facts about left-handers. (2013, October 21). Retrieved from
http://www.huffingtonpost.com/2012/10/29/left-handed-facts-lefties_n_2005864.html

CDC-data and statistics, autism spectrum disorders - ncbdd. (2013, October 21).
Retrieved from http://www.cdc.gov/ncbddd/autism/data.html

Ho, P. M., Bryson, C. L., & Rumsfeld, J. S. (2009). Medication adherence. *Circulation*,
119(23), 3028-3035. Retrieved from http://circ.ahajournals.org/content/119/23/3028

Households by age of householder and size of household: 1990 to 2010. (2013, October
19). Retrieved from http://www.census.gov/compendia/statab/2012/tables/12s0062.pdf

Madison, J. (2013, October 15). *M&M's color distribution analysis*. Retrieved from
http://joshmadison.com/2007/12/02/mms-color-distribution-analysis/

What percentage of people have green eyes?. (2013, October 21). Retrieved from
http://www.ask.com/question/what-percentage-of-people-have-green-eyes