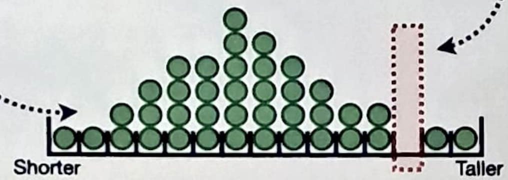
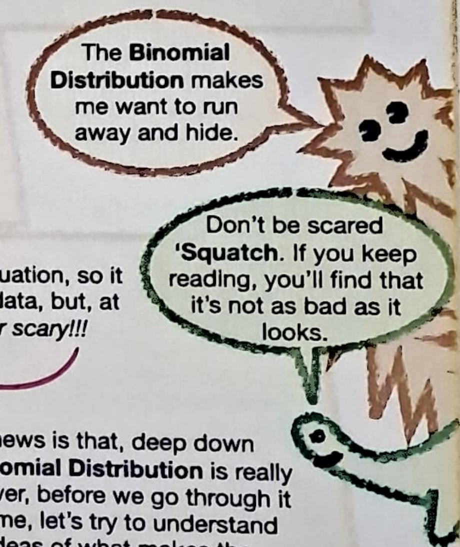
Discrete Probability Distributions: Main Ideas

1 **The Problem:** Although, technically speaking, histograms are **Discrete Distributions**, meaning data can be put into discrete bins and we can use those to estimate probabilities...

...they require that we collect a lot of data, and it's not always clear what we should do with blank spaces in the histograms.



2 **A Solution:** When we have discrete data, instead of collecting a ton of data to make a histogram and then worrying about blank spaces when calculating probabilities, we can let ***mathematical equations*** do all of the hard work for us.



3 One of the most commonly used **Discrete Probability Distributions** is the **Binomial Distribution**.

As you can see, it's a mathematical equation, so it doesn't depend on collecting tons of data, but, at least to **StatSquatch**, it looks *super scary!!!*

$$p(x \mid n, p) = \left(\frac{n!}{x!(n-x)!} \right) p^x (1-p)^{n-x}$$

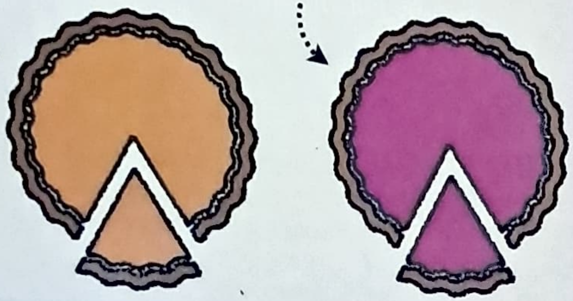
The good news is that, deep down inside, the **Binomial Distribution** is really simple. However, before we go through it one step a time, let's try to understand the main ideas of what makes the equation so useful.

The Binomial Distribution: Main Ideas Part 1

1 First, let's imagine we're walking down the street in **StatLand** and we ask the first **3** people we meet if they prefer pumpkin pie or blueberry pie...

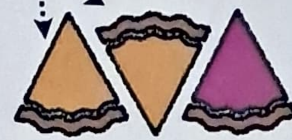
Pumpkin Pie

Blueberry Pie



2 ...and the first **2** people say they prefer pumpkin pie...

...and the last person says they prefer blueberry pie.



Based on our extensive experience judging pie contests in **StatLand**, we know that **70%** of people prefer pumpkin pie, while **30%** prefer blueberry pie. So now let's calculate the probability of observing that the first two people prefer pumpkin pie and the third person prefers blueberry.

3 The probability that the first person will prefer pumpkin pie is 0.7...

0.7

...and the probability that the first two people will prefer pumpkin pie is **0.49**...

(Psst! If this math is blowing your mind, check out **Appendix A.**)

$0.7 \times 0.7 = \mathbf{0.49}$

...and the probability that the first two people will prefer pumpkin pie and the third person prefers blueberry is **0.147**.

(Again, if this math is blowing your mind, check out **Appendix A.**)

$0.7 \times 0.7 \times 0.3 = \mathbf{0.147}$

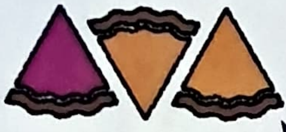
NOTE: 0.147 is the probability of observing that the first two people prefer pumpkin pie and the third person prefers blueberry...

...it is *not* the probability that **2** out of **3** people prefer pumpkin pie.

Let's find out why on the next page!!!

The Binomial Distribution: Main Ideas Part 2

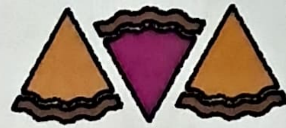
4 It could have just as easily been the case that the first person said they prefer blueberry and the last two said they prefer pumpkin.



In this case, we would multiply the numbers together in a different order, but the probability would still be **0.147** (see **Appendix A** for details).

$0.3 \times 0.7 \times 0.7 = \mathbf{0.147}$

5 Likewise, if only the second person said they prefer blueberry, we would multiply the numbers together in a different order and still get **0.147**.



$0.7 \times 0.3 \times 0.7 = \mathbf{0.147}$

6 So, we see that all three combinations are equally probable...


$0.3 \times 0.7 \times 0.7 = \mathbf{0.147}$


$0.7 \times 0.3 \times 0.7 = \mathbf{0.147}$


$0.7 \times 0.7 \times 0.5 = \mathbf{0.147}$

7 ...and that means that the probability of observing that **2** out of **3** people prefer pumpkin pie is the ***sum*** of the **3** possible arrangements of people's pie preferences, **0.441**.


$0.3 \times 0.7 \times 0.7 = \mathbf{0.147}$

+


$0.7 \times 0.3 \times 0.7 = \mathbf{0.147}$

+


$0.7 \times 0.7 \times 0.3 = \mathbf{0.147}$

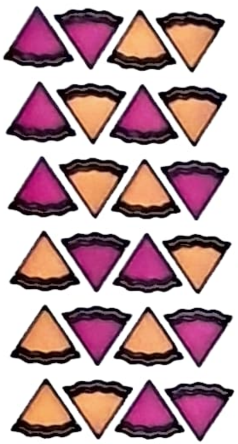
$= \mathbf{0.441}$

NOTE: Calculating by hand the probability of observing that **2** out of **3** people prefer pumpkin pie was not that bad. All we did was draw the **3** different ways **2** out of **3** people might prefer pumpkin pie, calculate the probability of each way, and add up the probabilities.

Bam.

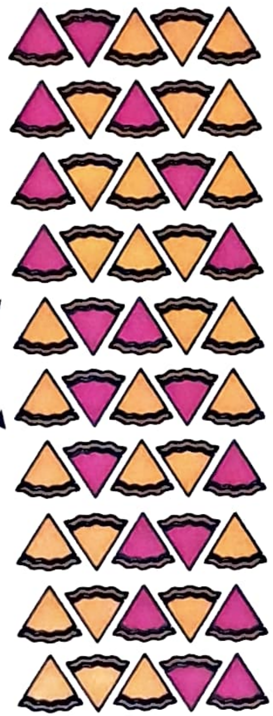
The Binomial Distribution: Main Ideas Part 3

8 However, things quickly get tedious when we start asking more people which pie they prefer.



For example, if we wanted to calculate the probability of observing that **2** out of **4** people prefer pumpkin pie, we have to calculate and sum the individual probabilities from **6** different arrangements...

...and there are **10** ways to arrange **3** out of **5** people who prefer pumpkin pie.



UGH!!! Drawing all of these slices of delicious pie is super tedious.

:(

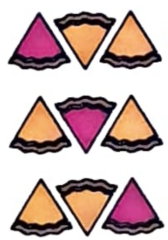
9 So, instead of drawing out different arrangements of pie slices, we can use the equation for the **Binomial Distribution** to calculate the probabilities directly.

$$p(x \mid n, p) = \left(\frac{n!}{x!(n-x)!} \right) p^x (1-p)^{n-x}$$

BAM!!!

In the next pages, we'll use the **Binomial Distribution** to calculate the probabilities of pie preference among **3** people, but it works in any situation that has binary outcomes, like wins and losses, yeses and noes, or successes and failures.

Now that we understand why the equation for the **Binomial Distribution** is so useful, let's walk through, one step at a time, how the equation calculates the probability of observing **2** out of **3** people who prefer pumpkin pie.



$0.3 \times 0.7 \times 0.7 = \mathbf{0.147}$

+

$0.7 \times 0.3 \times 0.7 = \mathbf{0.147}$

+

$0.7 \times 0.7 \times 0.3 = \mathbf{0.147}$

$= \mathbf{0.441}$

The Binomial Distribution: Details Part 1

1 First, let's focus on just the left-hand side of the equation.

In our pie example, x is the number of people who prefer pumpkin pie, so in this case, $x = 2$...

...n is the number of people we ask. In this case, $n = 3$...

...and p is the probability that someone prefers pumpkin pie. In this case, $p = 0.7$...

$$p(x \mid n, p) = \left(\frac{n!}{x!(n-x)!} \right) p^x (1-p)^{n-x}$$

2 ...p means **probability**...

...the vertical bar or pipe symbol means **given** or **given that**...

...and the comma between n and p means **and**...

So, the left-hand side reads: "The probability we meet $x = 2$ people who prefer pumpkin pie, given that we ask $n = 3$ people *and* the probability of someone preferring pumpkin pie is $p = 0.7$."

BAM!

Gentle Reminder: We're using the equation for the **Binomial Distribution** to calculate the probability that **2** out of **3** people prefer pumpkin pie...

0.3 x 0.7 x 0.7 = **0.147**

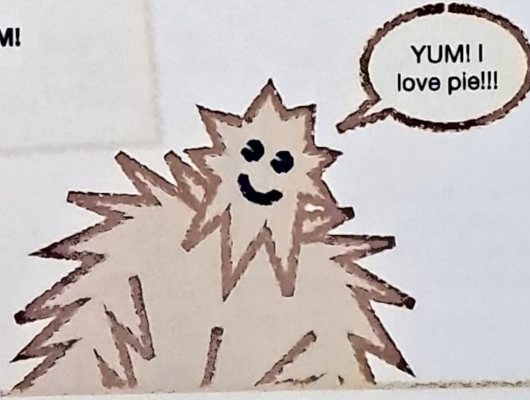
+

0.7 x 0.3 x 0.7 = **0.147**

+

0.7 x 0.7 x 0.3 = **0.147**

= 0.441



The Binomial Distribution: Details Part 2

3 Now, let's look at the first part on the right-hand side of the equation. **StatSquatch** says it looks scary because it has factorials (the exclamation points; see below for details), but it's not that bad.

Despite the factorials, the first term simply represents the number of different ways we can meet **3** people, **2** of whom prefer pumpkin pie...

...and, as we saw earlier, there are **3** different ways that **2** out of **3** people we meet can prefer pumpkin pie.

$$p(x \mid n, p) = \left(\frac{n!}{x!(n-x)!}\right) p^x (1-p)^{n-x}$$

Gentle Reminder: We're using the equation for the **Binomial Distribution** to calculate the probability that **2** out of **3** people prefer pumpkin pie...

0.3 x 0.7 x 0.7 = **0.147**

\+

0.7 x 0.3 x 0.7 = **0.147**

\+

0.7 x 0.7 x 0.3 = **0.147**

= 0.441

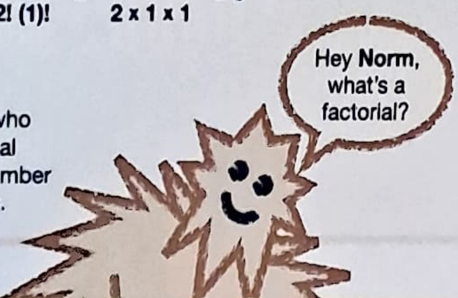
4 When we plug in $x = 2$, the number of people who prefer pumpkin pie...

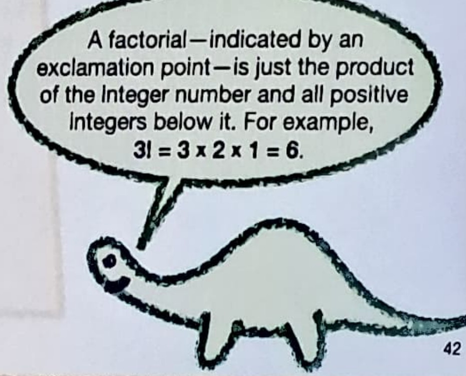
...and $n = 3$, the number of people we asked, and then do the math...

...we get **3**, the same number we got when we did everything by hand.

$$\frac{n!}{x!\,(n-x)!} = \frac{3!}{2!\,(3-2)!} = \frac{3!}{2!\,(1)!} = \frac{3 \times 2 \times 1}{2 \times 1 \times 1} = 3$$

NOTE: If x is the number of people who prefer pumpkin pie, and n is the total number of people, then $(n - x)$ = the number of people who prefer blueberry pie.





The Binomial Distribution: Details Part 3

5 Now let's look at the second term on the right hand side.

The second term is just the probability that **2** out of the **3** people prefer pumpkin pie.

In other words, since ***p***, the probability that someone prefers pumpkin pie, is **0.7**...

...and there are ***x* = 2** people who prefer pumpkin pie, the second term = $0.7^2 = 0.7 \times 0.7$.

$$p(x \mid n, p) = \left(\frac{n!}{x!(n-x)!} \right) p^x (1-p)^{n-x}$$

Gentle Reminder: We're using the equation for the **Binomial Distribution** to calculate the probability that **2** out of **3** people prefer pumpkin pie...

$0.3 \times 0.7 \times 0.7 = 0.147$

+

$0.7 \times 0.3 \times 0.7 = 0.147$

+

$0.7 \times 0.7 \times 0.3 = 0.147$

= 0.441

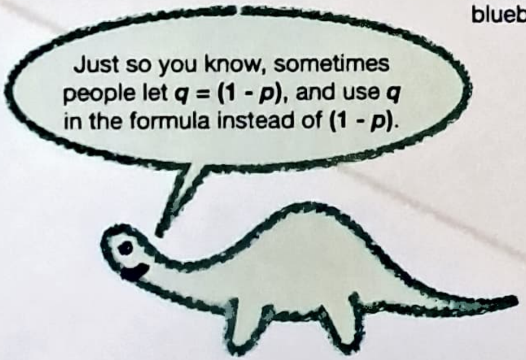
6 The third and final term is the probability that **1** out of **3** people prefers blueberry pie...

...because if ***p*** is the probability that someone prefers pumpkin pie, **(1 - *p*)** is the probability that someone prefers blueberry pie...

...and if ***x*** is the number of people who prefer pumpkin pie and ***n*** is the total number of people we asked, then ***n* - *x*** is the number of people who prefer blueberry pie.

So, in this example, if we plug in ***p* = 0.7**, ***n* = 3**, and ***x* = 2**, we get **0.3**.

$$(1 - p)^{n-x} = (1 - 0.7)^{3-2} = 0.3^1 = 0.3$$



The Binomial Distribution: Details Part 4

7 Now that we've looked at each part of the equation for the **Binomial Distribution**, let's put everything together and solve for the probability that **2** out of **3** people we meet prefer pumpkin pie.

We start by plugging in the number of people who prefer pumpkin pie, x = **2**, the number of people we asked, n = **3**, and the probability that someone prefers pumpkin pie, p = **0.7**...

$$p(x = 2 \mid n = 3, p = 0.7) = \left(\frac{n!}{x!(n-x)!}\right) p^x (1-p)^{n-x}$$

...then we just do the math...

$$= \left(\frac{3!}{2!(3-2)!}\right) 0.7^2 (1-0.7)^{3-2}$$

(Psst! Remember: the first term is the number of ways we can arrange the pie preferences, the second term is the probability that 2 people prefer pumpkin pie, and the last term is the probability that 1 person prefers blueberry pie.)

$$= 3 \times 0.7^2 \times (0.3)^1$$

$$= 3 \times 0.7 \times 0.7 \times 0.3$$

...and the result is **0.441**, which is the same value we got when we drew pictures of the slices of pie.

$$\boxed{= 0.441}$$

Gentle Reminder: We're using the equation for the **Binomial Distribution** to calculate the probability that **2** out of **3** people prefer pumpkin pie...

0.3 x 0.7 x 0.7 = **0.147**

\+

0.7 x 0.3 x 0.7 = **0.147**

\+

0.7 x 0.7 x 0.3 = **0.147**

= 0.441

TRIPLE BAM!!!

classmate

Date ______

Page ______

* Discrete Probability Distribution :-

→ when we have discrete data instead of collecting a ton of data to make a histogram & then worrying about blank spaces when calculating probabilities. we can let mathematical equations do all of the hard work for us.

(1) Binomial Distribution :-

$$\boxed{P(x \mid n, p) = \left(\frac{n!}{x!\,(n-x)!}\right) p^{x} (1-p)^{n-x}}$$

P → Probability

| → given or given that

Left hand side is read as : " The Probability we meet $x = 2$ people who prefer pumkin pie, given that we ask $n = 3$ people & the probability of someone preferring pumkin pie is $\underbrace{0.7}$."

took as an example!

classmate

Date

Page

* Example:-

$$P(x=2 \mid n=3, p=0.7) =$$

$$\therefore \boxed{P(x \mid n,p) = \left(\frac{n!}{x!\,(n-x)!}\right) p^x (1-p)^{n-x}}$$

$$P(x \mid n,p) = \left(\frac{3!}{2!\,(1!)}\right) 0.7^{(2)} (1-0.7)^{3-1}$$

$$= \left(\frac{6}{2}\right) (0.7)^2 (0.3)^1$$

$$= 3 \times 0.7 \times 0.7 \times 0.3$$

$$= \underline{\underline{0.441}}$$