We're Back!!! Only today though

Unit 5 – Discrete Probability Distributions, Day 2

Your Bakery Professor Colton

Unit 5, Day 2 - Outline

Unit 5 – Discrete Probability Distributions

Binomial Experiments

- Binomial Distribution
- Examples
- Unusual Data

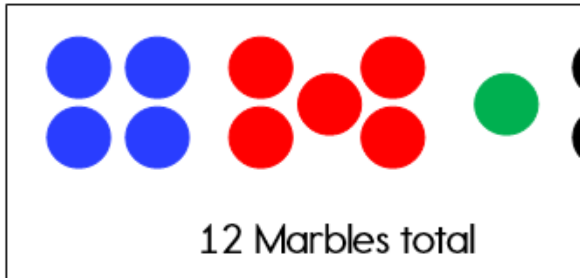
Review

- We had just looked at how to study a general discrete probability distribution, the context could be about anything..

Example: Let *X* represent the sum of two dice. Then the probability distribution of X is as follows:

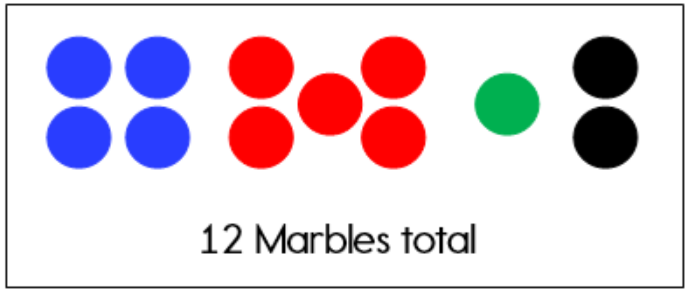
x	2	3	4	5	6	7	8	9	10	11	12
P(x)	1/36	2/36	3/36	4/26	5/36	6/36	5/36	4/36	3/36	2/36	1/36

- We have also looked at more complicated events, where we were selecting multiple marbles by <u>sampling with replacement (i.e. independent events)</u>.
- Could answer questions like:
  - P(2 Blues) = P(Blue AND Blue) = 4/12 x 4/12 = 16/144 = 0.111
    - Can think of this as Blue and then another Blue (kinda sequentially)
    - Because these events are independent, we can just multiply the marginal (original) probabilities
  - P(Black, Red, Green) = P(Black AND Red AND Green)
    = P(Black) x P(Red) x P(Green) = (2/12) x (5/12) x (1/12) = 0.005

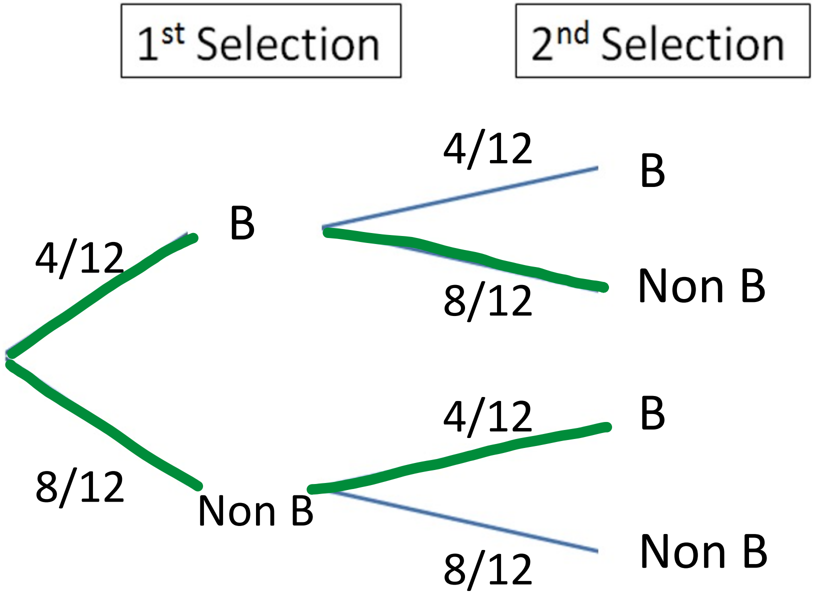


New

- Here is a slightly new twist!



- Lets think of Blue as our **success**, and everything else as **failure**; we are going to select two marbles
- Now for some probabilities:
  - P(2 Blues) = (4/12)^2
  - P(2 Non-Blues) = P(Non-Blue) x P(Non-Blue) = (8/12) x (8/12) = 0.444
  - P(1 Blue and 1 Non-Blue) = ?? = ends up equaling 0.444
    - Would need to use a tree diagram to get all the possible sequences (branches) of interest



X = # of Blues	0	1	2
P(X)	0.444	P(X = 1)	0.111

*** Even before calculating P(X = 1) via the tree, we already know what it must equal based on our Probability rules:*
1 = P(X = 0) + P(X = 1) + P(X = 2)
= 0.444 + P(X = 1) + 0.111
Then just algebra…

- Now let's say we extend this scenario to selecting 5 marbles (so 5 selections on our tree), or 10 marbles, and so on…
- This actually follows a special situation that allows us to use rules to more quickly analyze and understand this scenario!

Binomial Distribution

Binomial Distribution

- **Binomial random variable** is specific type of discrete random variable that counts how often a particular event occurs in a fixed number of trials

Conditions

- For a variable to follow a binomial distribution, ALL the following conditions must be met:
    - Fixed number of trials *n*
    - Binary outcome ("success" and "failure")
    - Probability of "success" *p* does not change
    - Independent, the outcome of one trial doesn't affect others

- X ~ Binomial(n, p), n and p are the parameters of this distribution!

- Can find probabilities using our calculator!!
    - There is a fancy formula behind these probabilities, but we aren't going to look at it

Mean and Standard Deviation

- Has special calculations for mean (expected value) and standard deviation:
    - Mean = μ = E(X) = np
    - Standard Deviation = **σ** = SD(X) = $\sqrt{np(1-p)}$
        - Might see this written as $\sqrt{npq}$, where q = 1 – p and represents the probability of failure.

Examples

- Number of correct guesses at 30 true-false questions when you randomly guess all answers
- X = Number of tails when flipping a coin 10 times
- Number of left-handers in a randomly selected sample of 100 unrelated people

Binomial Distribution Long LCQ

Setup: The 1.69 oz bag of m&m’s contains 57 pieces of m&m’s. Red is my favorite color, which Mars claims that 20% of all m&m’s produced in the traditional bags are red.

Can we describe the number of red m&m’s in a standard 1.69 oz bag with a Binomial distribution?

Binomial Distribution Long LCQ

Setup: The 1.69 oz bag of m&m's contains 57 pieces of m&m's. Red is my favorite color, which Mars claims that 20% of all m&m's produced in the traditional bags are red.

Can we describe the number of red m&m's in a standard 1.69 oz bag with a Binomial distribution?

To answer this, we have to check the **assumptions** of the Binomial Distribution!!

- Fixed number of trials? *Yes, we have 57 pieces of candy in a bag* → n = 57
- Binary outcome? *Yes, either we have a red m&m or we do not*
- Probability does not change? *Yes, the 20% probability does not change from m&m to m&m* → *p* = 0.2
- Independent trials? *Yes, one m&m doesn't affect the color of another*

All conditions were met, we can use a Binomial Distribution!!

Binomial Distribution Long LCQ

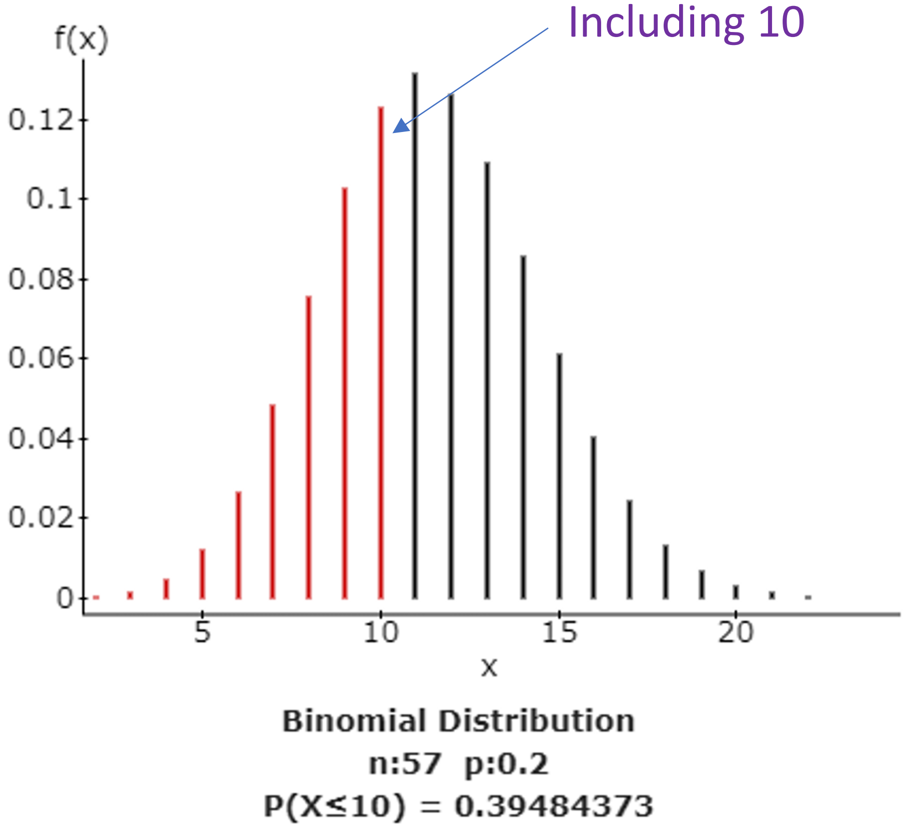
Since we just saw that this situation can be described with a binomial distribution, let use it to find some probabilities and the expected number of red m&m's in a bag.

But first!!! Let's write what we know fancily!

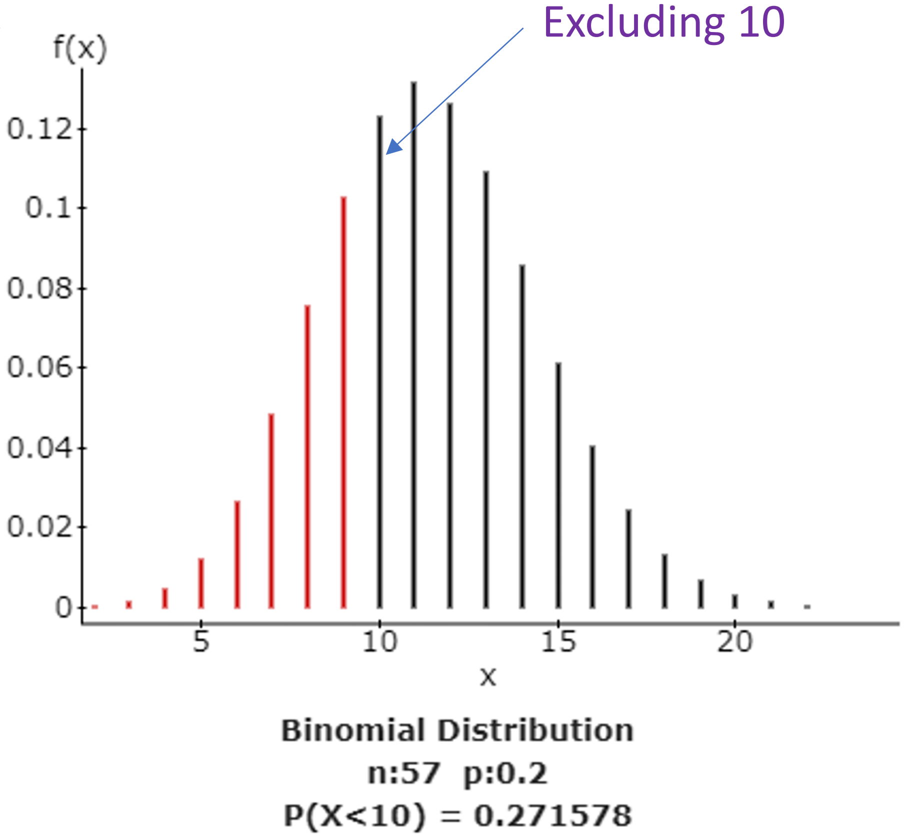
- Lets say X is a random variable that represents the number of red m&ms.
  - *We know X follows a Binomial Distribution, with probability of success = 0.2 and there is 57 m&ms (trials).*
  - **$X \sim \text{Binomial}(n = 57, p = 0.2)$**
- What is the probability that at most 10 m&m's are red?
- Is this different than asking for the probability that less than 10 m&m's are red?
- Think about this in terms of a probability statement: P(X ?? 10), which symbol do we use: <, >, ≤, or ≥
  - So we have to translate the words description to a probability notation!!

Binomial Distribution Long LCQ

Probability of at most 10? $P(X \leq 10)$



Probability of less than 10? $P(X < 10) = P(X \leq 9)$

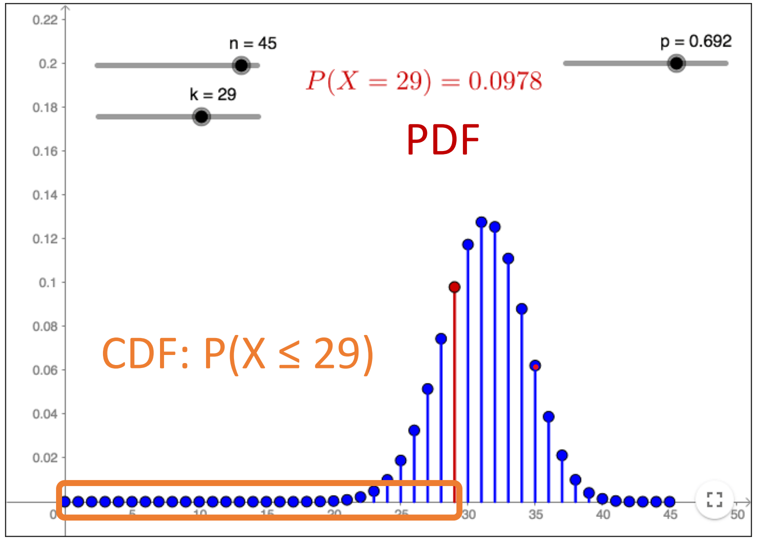


Binomial Distribution Long LCQ (Calc fun!!)

GOAL: Use the Binomial distribution to calculate probabilities!

1. Choose correct Dist: 2ND → VARS
   - binompdf() or binomcdf()
2. Enter in information (parameters and X)
   - trials = n
   - p = probability of success
   - x value = number of successes we want
   - If you have TI-83, you would type binompdf(n,p,x) or binomcdf(n,p,x)

Pdf = Probability distribution function, a **single point**!
Cdf = Cumulative distribution function, **including and everything to the left**!



http://www.malinc.se/math/statistics/binomialen.php

X ~ Binomial(n = 57, p = 0.2)

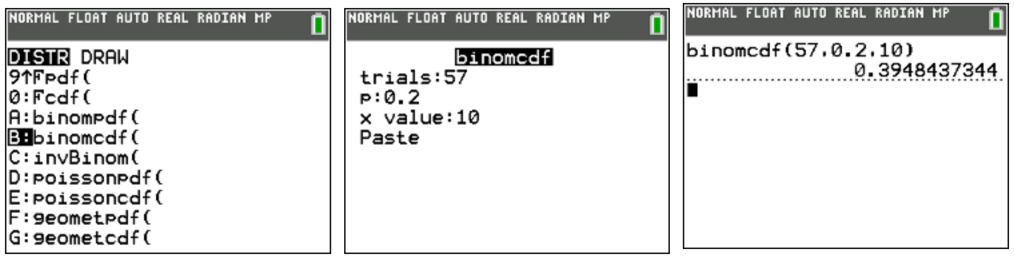
a) P(X ≤ 10) = ??
b) P(X < 10) = ??
c) P(X = 3) = ??
d) P(X > 5) = ??
e) P(X ≥ 6) = ??

Binomial Distribution Long LCQ (Calc fun!!)

a) $P(X \leq 10)$ = ??

$P(X \leq 10)$ = binomcdf(n = 57, p = 0.2, x = 10) = 0.394

This confirms the visual we saw earlier

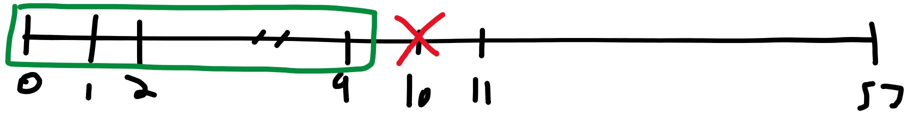


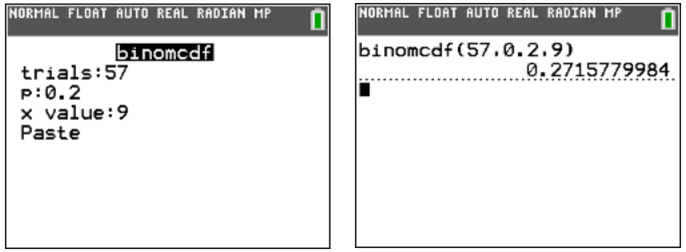
b) $P(X < 10)$ = ??

CDF on calc AWLAYS gives ≤, so need to adjust our statement to correctly take this into account

Drawing a picture helps!

$P(X < 10) = P(X = 0) + P(X = 1) + \ldots + P(X = 9)$ → this is everything we want to include





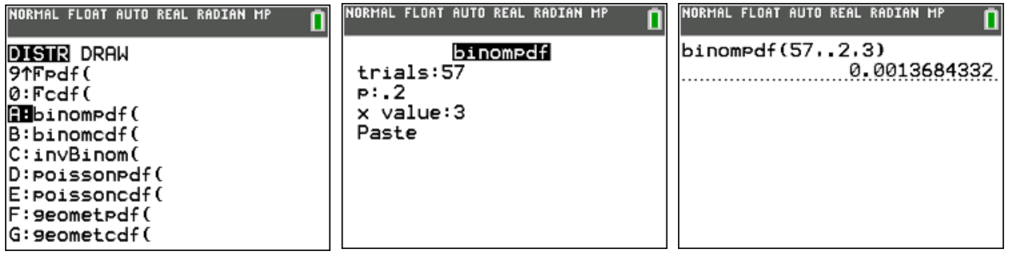
$P(X < 10) = P(X \leq 9)$ = binomcdf(n = 57, p = 0.2, x = $\cancel{10}$ 9) = 0.2715

Had to decrease the X value by one to make our calculator give us what we want

c) $P(X = 3)$ = ??

This is just a single point now, so need to use PDF!

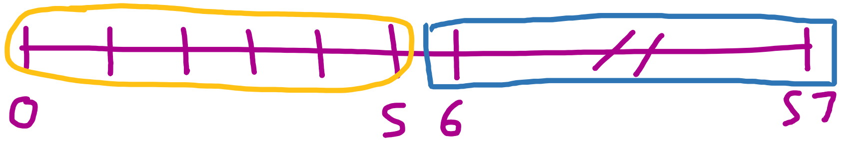
$P(X = 3)$ = binompdf(n = 57, p = 0.2, x = 3) = 0.001



Binomial Distribution Long LCQ (Calc fun!!)

d) $P(X > 5)$ = ??

- *This probability is to the right, but Calc always gives us to the left: $P(X \leq x)$*
- *So to get to the right, we have to use the complement → 1 – CDF (left)*



- *First rewrite use CDF (meaning: 1 minus and with a less than or equal to sign):*
  - $P(X > 5) = 1 - P(X \leq 5)$
- *Then show work! 1 – binomcdf(n = 57, p = 0.2, x = 5) = 0.981*

e) $P(X \geq 6)$ = ??

- *Same strategy: we want a greater than or equal to probability*
  - *Need to rewrite as 1 – CDF (with the correct X value)*
- *We know we want to include 6 in our final answer, so we don't want to take it away in our subtraction (again, a sketch helps*
  - *Thus we have to decrease our X value in the CDF from 6 to 5*

$P(X \geq 6) = 1 - P(X \leq 5) = P(X > 5)$ = < same as previous >

Binomial Distribution Long LCQ

What is the expected number of red found in a bag of m&m's? (Recall X~ Binomial(n = 57, p = 0.2)

What is the standard deviation of number of red m&m's found in a bag of m&m's?

Binomial Distribution Long LCQ

What is the expected number of red found in a bag of m&m's? (Recall X~ Binomial(n = 57, p = 0.2)

- Because we know it follows a binomial distribution, it makes our life easier! Can just use the easy formulas!
- *Expected number E(X) = mean = np = 57 x 0.2 = 11.4 red m&m's*

What is the standard deviation of number of red m&m's found in a bag of m&m's?

- Because we know it follows a binomial distribution, it makes our life easier! Can just use the easy formulas!
- *Standard Deviation SD(X) =* $\sqrt{np(1-p)} = \sqrt{57(0.2)(1-0.2)} = 3.019$

Don't have to use these formulas!

$$E(X) = \sum x \bullet P(X = x)$$

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

Which is what we usually have to do to find the mean of a discrete probability distribution!

LCQ: Binomial Distribution

Determine if a Binomial distribution is appropriate for each of the following scenarios. If not, **explain** why not.

NOTE!! For our class, it will be stated in the problem that we have a Binomial distribution!

1) Flip a fair coin until you get heads.

2) The number of diamonds when selecting 5 cards from a standard 52 card deck without replacement.

3) Selecting 10 cards from a standard 52 card deck with replacement. We are interested whether the card is a diamond, a heart or other.

4) The number of makes out of 7 free throws for an NBA player that shoots 67% from the free throw line.

5) Find the mean and standard deviation for the scenario in 4? Interpret both in context.

LCQ: Binomial Distribution

Determine if a Binomial distribution is appropriate for each of the following scenarios. If not, **explain** why not.

If not ALL of the 4 binomial distribution conditions are met, then we CAN NOT use it!
So just think if there is at least that isn't true for each of these scenarios!

NOTE!! For our class, it will be stated in the problem that we have a Binomial distribution!

1) Flip a fair coin until you get heads.

NO, because there is NOT a fixed number of trials

2) The number of diamonds when selecting 5 cards from a standard 52 card deck without replacement.

NO, probability of success is not constant , not independent trials

3) Selecting 10 cards from a standard 52 card deck with replacement. We are interested whether the card is a diamond, a heart or other.

NO, because more than two outcomes (not binary)

4) The number of makes out of 7 free throws for an NBA player that shoots 67% from the free throw line.

YES, meets all conditions
So we can use the Binomial distribution formulas and calc functions!

5) Find the mean and standard deviation for the scenario in 4? Interpret both in context.

E(X) = np = 7(0.76) = 4.69, this means we expect this player to make 4.69 out of 7 shots in the long run
SD(X) = $\sqrt{np(1-p)} = \sqrt{7(0.67)(1-0.67)} = 1.22$*, on average his makes will be 1.22 away from the expected value of 4.69*

Unusual Data

Unusual Events

- Same as we have seen before….
- An unusual event is one whose probability is very small.
- A rule of thumb is that any event whose probability is less than 0.05 is considered to be unusual.

Example

- If X ~ Binomial(n = 24, p = 0.6), what is the interval of events that would be considered not unusual?????

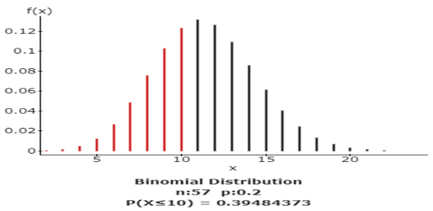
- First, let's answer these:
  - Is X = 11 unusual??
  - Is X = 10 unusual??

Unusual Data

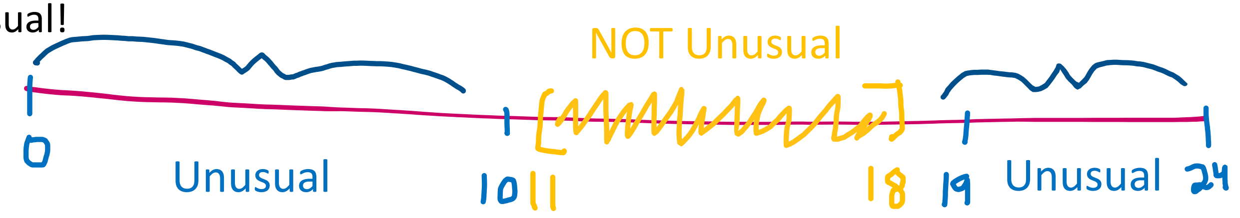
Unusual Events

- Same as we have seen before….
- An unusual event is one whose probability is very small.
- A rule of thumb is that any event whose probability is less than 0.05 is considered to be unusual.

Example



- If X ~ Binomial(n = 24, p = 0.6), what is the interval of events that would be considered not unusual?????
- First, let's answer these:
  - Is X = 11 unusual?? P(X = 11) = binompdf(n = 24, p = 0.6, x = 11) = 0.06 → No because probability is > 0.05
  - Is X = 10 unusual?? P(X = 10) = binompdf(n = 24, p = 0.6, x = 11) = 0.032 → Yes because probability is < 0.05
- Probabilities are decreasing towards the edges (kinda like in the plot on right)
  - So all events (Xs) less than or equal to 10 would also be unusual
- We can do this same process to find the upper X cutoff for unusual!
  - And we end up with the interval shown here:



P(X = 11) = 0.06

P(X = 18) = 0.056

P(X = 10) = 0.032

P(X = 19) = 0.027

PROBLEM SESSION!!!!!!!!!!!!

B-dubs

Buffalo Wild Wings carefully monitors customer orders and has found that 20% of all customers ordering food ask for wings (W), while the remainder order something else (E).

a) Suppose three customers are selected at random. Let the random variable X be the number of customers who order wings. Can we use a Binomial distribution to describe this situation?
b) Find the probability distribution for X.
c) Find the probability that two customers order wings.
d) Find that probability that fewer than three customers order wings.
e) What is the probability that at least one customer orders wings?

Parts (a) and (b) Answers

Buffalo Wild Wings carefully monitors customer orders and has found that 20% of all customers ordering food ask for wings (W), while the remainder order something else (E). Suppose three customers are selected at random. Let the random variable X be the number of customers who order wings.

a) To answer this, we have to check the **assumptions** of the Binomial Distribution!!

- Fixed number of trials? Yes, we have 3 customers
- Binary outcome? Yes, either wings or else
- Probability does not change? Yes, the 20% probability does not change customer to customer
- Independent trials? Yes because we are selecting customers at random, we can assume they are independent

So yes!! Meets all the conditions. Thus X ~ Binomial(n=3, p = 0.20

b)

- To find the probability distribution, we can set up a table just like we have done with general discrete prc distributions
- Then we can find the probabilities using the PDF of the binomial distribution!

x	0	1	2	3
P(x)	0.5120	0.3840	0.0960	0.0080

P(X = x) = binomial(n = 3, p = 0.2, x), for each X = 0,1,2,3

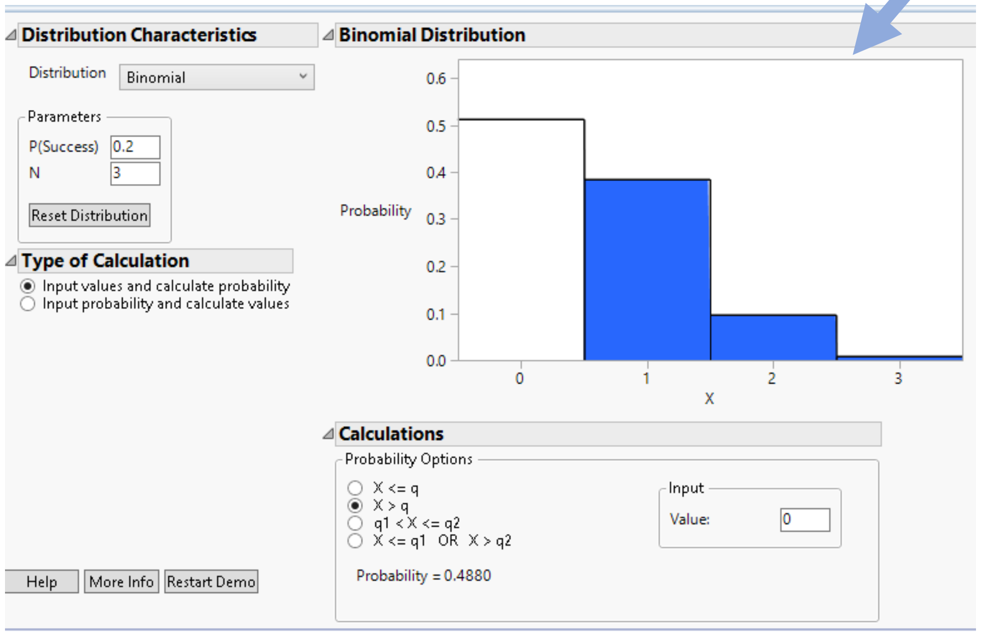
Parts (c), (d) and (e) Answers

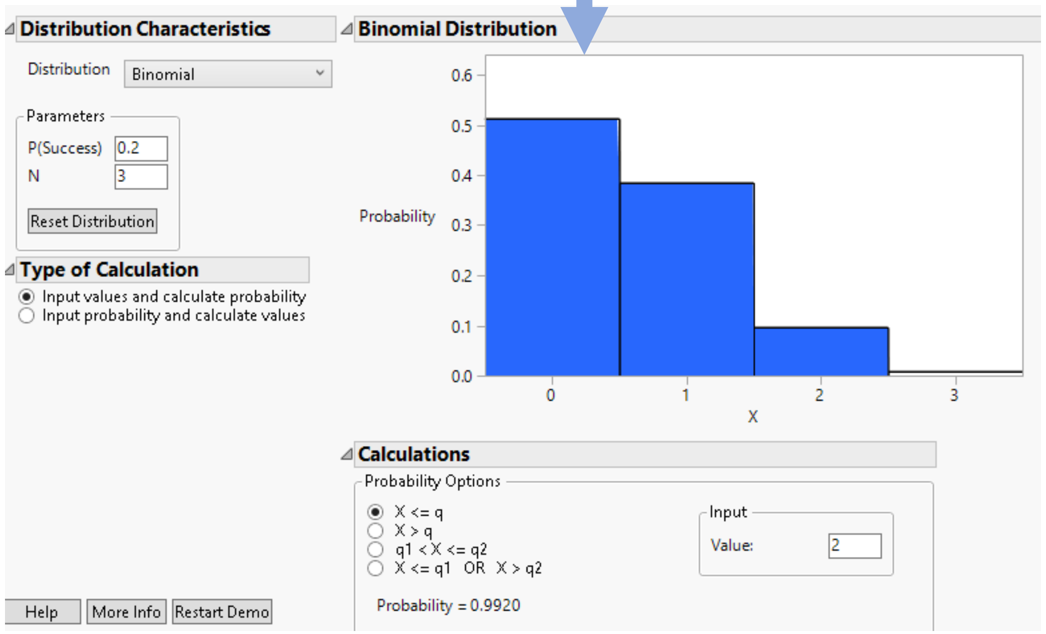
c) $P(X = 2) = 0.096$

d) $P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2) = P(X \leq 2) = 0.9920$

e) $P(X \geq 1) = 1 - P(X = 0) = 0.4880$

Here's what it looks like visually from another software. It would be a more accurate visual if it was just lines above the numbers to represent the probabilities instead of a histogram look, but you get the idea





Problem #51

A manufacturer of game controllers is concerned that their controller may be difficult for left-handed users. They set out to find lefties to test. About 13% of the population is left-handed. If they select a sample of five customers at random in their stores, what is the probability of each of these outcomes?

a) There are some lefties among the 5 people.
b) There are exactly 3 lefties in the group.
c) There are at least 3 lefties in the group.
d) There are no more than 3 lefties in the group.

Problem #51 Solution

- $P = 0.13$, $q = 0.87$, $n = 5$

a) $P(X \geq 0) = 1 - P(X = 0) = 0.5016$

b) $P(X = 3) = 0.0166$

c) $P(X \geq 3) = P(X = 3) + P(X = 4) + P(X = 5) = 1 - P(X \leq 2) = 0.0179$

d) $P(X \leq 3) = 0.9987$

Problem #57

A manufacturer of game controllers is concerned that their controller may be difficult for left-handed users. They set out to find lefties to test. About 13% of the population is left-handed. If they select a sample of five customers at random in their stores:

a) How many lefties do you expect?
b) With what standard deviation?

Problem #57 Solution

a) E(X) = np = 5(.13) = 0.65

b) SD(X) = $\sqrt{npq} = \sqrt{5(0.13)(0.87)} \approx 0.752$

Problem #13

Which of these situations fit the conditions for using a Binomial experiment? Explain.

a) You are rolling 5 dice and need to get at least two 6s to win the game.
b) We record the distribution of home states of customers visiting our website.
c) A committee consisting of 11 men and 8 women selects a delegation of 4 to attend a professional meeting at random. What is the probability they choose all women?
d) A study found that 56% of M.B.A. students admit to cheating. A business school dean surveys all the students in the graduating class and gets responses that admit to cheating from 250 of 481 students.

Problem #13 Solution

a) Yes, 2 mutually exclusive outcomes; probability of success is the same for each trial, 1/6, and we have independent trials, the roll of one dice does not affect the others.

b) No, more than 2 outcomes are possible.

c) No, the probability of choosing a man or woman changes depending on who has already been chosen.

d) Yes, assuming responses (and cheating) are independent among the students.

Textbook Problem #14

At the airport entry sites, a computer is used to randomly decide whether a traveler’s baggage should be opened for inspection. If the chance of being selected is 12%, can you model your chance of having your baggage opened with a Binomial model? Check each condition specifically.

Textbook Problem #14 Solution

- Each trial consists of a computer deciding whether a traveler’s bag should be opened for inspection.
- Two possible outcomes: either the bag is opened for inspection or it is not.
- The trials are independent; if one traveler’s bag is selected to be opened for inspection that does not affect the probability of other travelers’ bags being selected to be opened for inspection.
- The probability of selecting a traveler’s to be opened for inspection is the same, 0.12 or 12%, on every trial.