Sustaining Canada's Forest Resources...

Background Information

British Columbia's rainforest is home to some of the largest and most beautiful trees in the world. It is no coincidence that it is also home to a successful logging industry that relies on clear-cutting to keep profit levels high.

Lately, timber companies have been targeting old growth red cedar stands of forest as this wood attracts high prices in the market. These old growth stands are home to red cedars that can grow to 75 metres high with a base diameter close to two metres!

B.C.'s red cedars have a history, longer than the loggers that have been cutting them down. Canada's First People used red cedars to make clothing, shelters, and canoes. The trees that they used are still standing and are reminders of B.C.'s history.

Red cedars also play an important role in the health of B.C.'s forest ecosystem and provide habitat for many wildlife species such as bears, seabirds, and owls.

In this activity, you will gain an appreciation for what it takes to keep B.C.'s red cedar stands from disappearing. You will design a sustainable management approach for saving the old growth stands.

You will use a hypothetical but realistic data set that appears in table 1 on the next page. Assume the data was collected by an ecologist who sampled a large number of red cedar forest stands in the province of B.C.

Table 1: Grouped data set showing the number of trees with a diameter within the range indicated.

Tree diameter range (m)	Number of trees
0.00 – 0.25	100
0.25 – 0.50	150
0.50 – 0.75	215
0.75 – 1.00	300
1.00 – 1.25	218
1.25 – 1.50	149
1.50 – 1.75	98
1.75 – 2.00	25

Activity

1. Why do you think the ecologist was not able to sample the entire population of red cedar stands in the province of B.C.?

2. Create a histogram showing the ecologist's data from table 1. Be sure to include a title and label the axis' accordingly including units of measurement.

3. Does the histogram that you created show strong traits of being normally distributed? Explain your reasoning.

4. What does the shape of the histogram tell you about the red cedar stands in the province of B.C.

5. The ecologist concludes that the true diameter of red cedar trees in the province of B.C. is normally distributed with a mean of 0.9 m and a standard deviation of 0.4 m. Sketch a continuous probability distribution that illustrates the ecologist's conclusion. Label accordingly.

6. Calculate the probability of finding a red cedar tree in the province of B.C. that has a diameter less than 0.6 m. Use z-scores to do this and show your steps.

7. Now calculate the probability of finding an “old growth” red cedar tree in the province of B.C. that has a diameter of more than 1.7 m. Use z-scores and show your steps.

8. Considering the fact that red cedars have been around in B.C. for close to 5,000 years and your answer to Q7, what can you say about what has happened to the “old growth” red cedar stands?

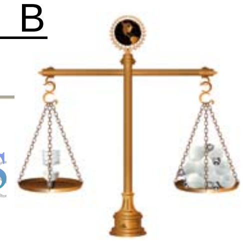
9. Using the table below, list all of the arguments for using the red cedars and all the arguments for not using them.

<u>Using them…</u>	<u>Not using them…</u>
▪	▪
▪	▪
▪	▪
▪	▪
▪	▪

10. What section of your probability distribution should be focused on in terms of clearing red cedars if loggers want to practice a sustainable yield management approach for B.C.'s red cedar stands such that both sides of the argument from the table in question 9 are balanced. Why is this the case? Explain.

11. What are some everyday things that you can do to help save B.C.'s red grove cedar stands and the rest of the trees in B.C.'s rainforest?

Confidence intervals...weighing the options



You have been hired by the government of a developing country to evaluate the safety of a large hydro dam project. The large mega-dam is located on a steep section of a river that carves through the mountains on its way to lower land. Another country located downstream has financed the construction of the dam as they have a vested interest in purchasing the electricity that will be generated.

The construction of the dam began almost 10 years ago and has taken a long time to complete. The river is a mighty one and doesn't like to cooperate. Finally, the final blocks on concrete have set and the dam is set to begin producing electricity in a couple of weeks. The government has called upon you to do a final check on the structural safety of the dam.

Upon arriving in the country, you quickly learn that not everyone has supported the construction of the dam. Native residents along the river were forced to relocate elsewhere giving up the land that their ancestors once lived on and fought to protect. This was a huge loss for them as the land was their livelihood. Environmentalists still protest today citing environmental damage in exchange for business and money.

Flying over the area, you can see the magnitude of the project. Huge areas of forest had to be cleared to make way for transport and access roads while the immediate area upstream is all water as an artificial lake was created when construction began flooding millions of hectares of habitat. No wonder everyone was upset.

Arriving at the hydro plant located at the site of the dam, the atmosphere is very different. Everyone is looking forward to the grand opening and delegates from both countries are signing the final agreements concerning how they will share the water and electricity. You are ready to begin your project

The chief engineer of the project has informed you that that the dam has been designed to support an artificial lake with a true mean maximum depth up to 210 ft. Anything above this and the dam would not pass safety regulations. The chief engineer has been concerned lately though because the region has been receiving more rainfall over the previous years.

You inform the engineer that you will measure the maximum depth of the artificial lake once a day on a daily basis over 50 days. Then you will use the data to construct a confidence interval concerning the true mean maximum depth of the lake. The data appears below and is measured in ft.

200	210	205	218	220	230	204	199	180	195
201	207	175	192	200	200	202	191	180	196
170	200	207	216	212	204	201	198	197	230
201	204	200	200	206	198	190	188	187	200
200	280	201	201	204	200	200	200	198	199

<u>Follow-up</u>

1. Calculate the mean maximum depth of the lake based on the 50 days of data you sampled

Mean ____ ft

2. Calculate the standard deviation based on the sample above.

Standard deviation _____ ft

3. What information does the standard deviation provide? Explain.

Now it is time to use your knowledge of probability distributions to put the data to the test. Although you don't know if the underlying population of the daily maximum depth of the lake is normally distributed, you do know that the sample means are normally distributed.

4. What will be the standard deviation of the sample means based on the data that you collected. Why is it significantly lower than the standard deviation you calculated in Q3?

5. Sketch a properly labeled probability distribution of the sample means of the maximum depth of the lake based on the sample mean that you calculated in Q1 and the standard deviation you calculated in Q4.

6. Why is it not feasible to know the true mean maximum depth of the lake from a statistical point of view? Explain.

7. Construct a 90% confidence interval for the true mean maximum depth of the lake based on the probability distribution that you created in Q5.

8. Why is this not good enough in terms of making a decision as to whether the dam is safe from a structural point of view. Explain.

9. In light of your answer to Q8, construct a confidence interval so that you can be absolutely sure that the dam is safe / unsafe from a statistical and structural point of view.

10. Based on the your new confidence interval, should the dam pass safety inspection? Explain.

11. What confidence level(s) do you think the following parties would use if they were creating the confidence interval? Use the chart below to organize your results. Then calculate the confidence interval for each of the parties.

Party involved	Confidence level	Confidence interval
Government officials	____	______________
Environmentalists	____	_______________
Local residents	____	_______________
Engineers	____	_______________

12. Which of the parties above would accept the dam project in terms of passing safety regulations? Why do you think this is the case? Explain.

Where does 80% of the population live...

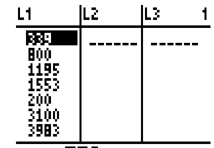
A long time ago, a famous economist by the name of Pareto described the phenomenon that a larger portion of the wealth of a society is owned by a smaller percentage of the people living in that society. This idea is sometimes expressed more simply as the “80-20 rule”, that is, 20% of the population owns 80% of the wealth. Today, this concept is widely applicable to a variety of contexts including human settlement patterns. To what extent does the famous Pareto distribution apply to human settlements and their respective populations across the Earth?

Humans have made decision about where to live since the beginning of time. Traditionally, people settled near water for transportation and trade. These settlements have now become many of the large metropolis cities. At the same time, many other smaller settlements have also developed near and far from these large cities. The data below provides an estimate of the population of a sample of cities/towns from the province of Ontario.

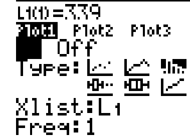
Goulais River	339	Napanee	15,132	Stoney Creek	57,327
Kearney	800	Midland	16,700	Kanata	59,700
Coldwater	1,195	Huntsville	17,338	Sarnia	70,876
Nipissing	1,553	Uxbridge	17,377	Sault Ste. Mar	74,566
Bronte	2,000	Trenton	19,374	Peterborough	74,600
Stayner	3,100	Dundas	20,000	Waterloo	86,543
Picton	3,983	Lincoln	20,612	Ajax	90,000
Deep River	4,135	Grimbsby	21,297	Thunder Bay	109,016
Elora	4,546	Owen Sound	21,431	Unionville	110,000
Angus	6,000	Collingwood	21,500	Whitby	110,000
Perth	6,003	Orangeville	25,248	Kingston	114,195
Parry Sound	6,500	Fort Erie	28,143	Burlington	150,836
Aylmer	7,018	Bolton	30,000	Sudbury	155,219
Acton	7,767	Bowmanville	30,000	Oakville	155,700
Dryden	8,198	Innisfil	31,000	Richmond Hill	163,000
Kirkland Lake	8,616	Orillia	32,692	Vaughan	182,022
Wellesley	9,365	Georgetown	36,500	Kitchener	190,399
Alliston	9,679	Timmins	43,686	Windsor	208,402
Fergus	10,017	Cornwall	45,640	Markham	208,615
Meaford	10,500	Halton Hills	47,600	Brampton	325,428
Ingersoll	11,000	Aurora	48,000	London	336,539
Dorchester	12,000	Welland	48,402	Hamilton	499,268
Streetsville	12,040	Belleville	49,060	Mississauga	612,925

Part 1: Creating a histogram

1. Enter the population data from table 1 in a list on your graphing calculator

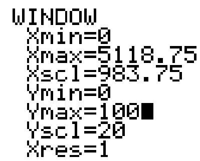


2. Turn the histogram option on by pressing the 2nd key followed the Y= key. Then set up the screen so that it looks like the picture to the right.



3. Press the Zoom key and choose option 9 to create the histogram

4. Press the Window key and adjust the x-scale so that the frequency is calculated for a population interval of every 100,000 people



Part II: Creating the probability distribution

1. Press the Trace key on your calculator and record the frequencies you see on the table below. Use the data to calculate the relative frequencies.

Population Interval	Frequency	Relative Frequency
0 - 100		
100 - 200		
200 - 300		
300 - 400		
400 - 500		
500 - 600		
600 - 700		
700 - 800		
800 - 900		
900 - 1,000		
> 1,000		

2. Plot the relative frequencies from the previous table on the graph below. Turn it into a continuous probability distribution. State any assumptions that you are making.

Relative Frequency

Population Interval

3. Does your distribution show characteristics that Pareto originally noticed when looking at the wealth of individuals across a society? Explain your reasoning clearly.

4. Discuss the advantages that the continuous probability distribution has over the bar graph distribution that you created? Be specific.

Part III – Making inferences and extrapolating

1. What percentage of the cities/towns that were sampled in Ontario had a population less than 50,000 according to table 1? Tip: You can press the 2^{nd} key followed by the Stat key and scroll over to OPS and sort L1 in ascending order.

2. Describe how could you use the probability distribution that you created to determine the probability that a city/town in Ontario has a population less that 50,000 people.

3. Using calculus, it can be shown that the probability that the population of a city/town in Ontario is greater than some number "c" for the distribution that you created is approximately...

$$P(x \geq c) = \left(\frac{c}{339}\right)^{-0.22373478}$$

Use this to quantify your answer to Q2 from this section and compare it to your answer to Q1. What are the advantages of using the probability distribution rather than the sample from table 1. What are the disadvantages.

4. To what extent does the “80-20” rule apply to the distribution of human settlements in Ontario based on the data from table 1? Based on the information that you have gathered regarding the continuous probability distribution? Tip: To find the total population of the cities sampled press the 2nd key followed by the Stat key and scroll over to Math and the Sum function.

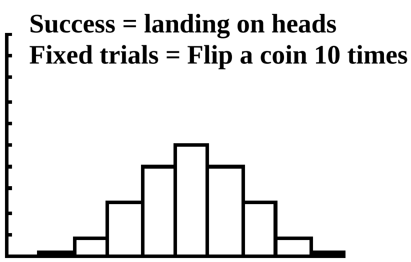
5. To what extent do you think the Pareto distribution applies to human settlement patterns in other provinces of Canada? In other words, could your probability distribution be used as a model for other provinces or even countries in the world? Explain your reasoning clearly.

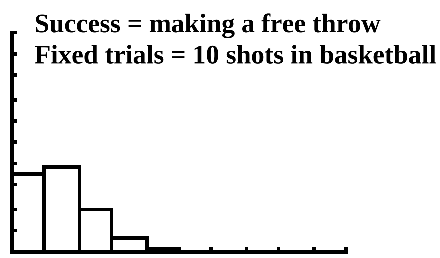
Normal approximation to discrete probability

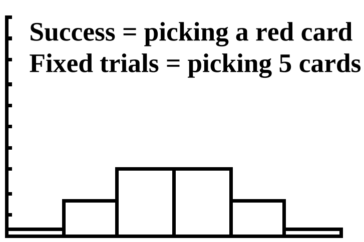
Background Information

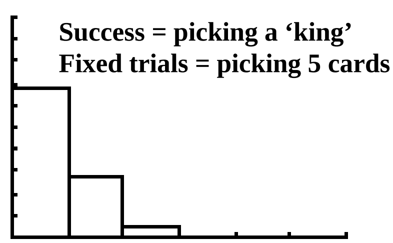
You have seen that the binomial distribution is useful for calculating the probability that a certain number of successes will occur out of a fixed number of trials. A classic example is flipping a coin 4 times and calculating the probability of getting exactly 2 heads out of the 4 tosses. The probability works out to 37.5%, not 50% that comes to mind.

You have also seen that the binomial distribution looks like a normal distribution at times – and sometimes it doesn't. For example, here are some examples of binomial distributions from different binomial experiments









Is there any framework for determining which discrete probability experiments will resemble a normal distribution? What are the implications if it does resemble a normal distribution?

Part A

Suppose that someone asked you to determine the probability of getting at least 20 heads if you toss a coin 100 times? Well if you created the binomial distribution for this, you would see that it would look exactly like a normal distribution. So why not use a normal distribution to calculate the answer!

1. If you were to prepare the binomial distribution for this experiment by hand, what major obstacle would you face?

2. What information would you need if you were to try to use the normal distribution to approximate this probability?

Part B

By now you hopefully realize that the normal distribution can be used to approximate the binomial distribution when calculating probability ranges for success. To do this, you will need to know the mean and standard deviation of the binomial experiment. The mean is the simple part. It is just the number of trials multiplied by the probability of success. The standard deviation is the hard part and this is what this part of the activity focuses on.

To derive the standard deviation of a binomial experiment follow the steps listed on the next page.

Step 1

Since the standard deviation for a continuous variable is simply the square root of the 'variance', why not start with the formula for the variance of a continuous random variable and see what we can go from here.

Variance = $\dfrac{\sum\left(x_i - \overline{x}\right)^2}{n}$

Step 2

Since binomial experiments deal with discrete variables, not continuous ones, the formula for variance won't help us much in this format. However, since standard deviation is an 'average' measure of deviation, think about what part of the formula creates the 'average' part of the measure and think about why the above formula for variance can be rewritten as follows:

Variance = $E\left[\sum\left(x_i - \overline{x}\right)^2\right]$

Step 3

Consider a very simple binomial experiment with 1 single trial. This experiment will have just two possible outcomes "1" and "0" where the probability of getting "1" is p .

1. Create a probability distribution below to illustrate the outcomes of this simple experiment.

2. Calculate the expected value of this experiment using the distribution you just created.

Step 4

Now you will focus on the ‘fancy’ way of writing the variance of a discrete variable.

$$E\left[\sum (x_i - \bar{x})^2\right]$$

3. Forget about the expected value for a second – just write out the sum of the square deviations (what’s inside the brackets) for this simple experiment.

4. Now find the expected value of your answer Q3 by preparing a probability distribution of the sum of the square deviations you wrote in Q3. What is the probability of these deviations happening?

5. Congratulations! You have found the variance for this simple binomial experiment for a single trial. What do you think the variance is for n number of trials?

6. Okay, now use the variance from Q5 to calculate the standard deviation of a binomial experiment with n number of trials.

Part C

Now you are ready to use the normal distribution to approximate a binomial distribution. Please acknowledge that the main practical reason for using a normal distribution to approximate a binomial distribution is because it is sometimes very time consuming to compute binomial probabilities.

For example, recall the example of calculating the probability of getting at least 20 heads if you toss a coin 100 times. To do this without a graphing calculator would very time consuming as you would have to calculate 80 binomial probabilities and add them up!

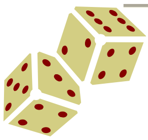
1. Determine the probability of getting at least 20 heads if you toss a coin 100 times using your graphing calculator

2. Sketch the normal distribution that approximates this binomial distribution. You will need to determine the mean and standard deviation using the formulas that you derived from the last part.

3. Use a z-score to determine the probability of getting at least 20 heads out of 100 tosses of a coin. Show the section of the probability distribution above that corresponds with your answer.

4. Compare your answer from Q3 to your answer from Q1. How well does the normal distribution in terms of approximating the real answer?

5. What do you think a “continuity correction” means and explain how this would change your answer to Q4.



Create your own game of chance...

Your task is to work as part of a team to create an **original and creative** game involving dice, spinners, cards, or any other reasonable items that introduce an element of chance.

Your game must be based on the binomial distribution and involve payouts for success and penalties for failures. The **expected value** of your game must be in favour of you, the owners of the game. However, they must be such that the game does not appear to be unfair to a player. You will use the **normal approximation** to approximate the probabilities for your game.

A **Games Fair Day** will be held where participants will have the opportunity to play your game. Each participant will receive 20 free chips to use to play games. Every time a player visits to play your game, you will record their **net payout** when they are finished. You should create a sign with the name of your game, a brief caption to catch players' attention, and the expected payouts. The rules of the game should also be available for players to read so that you don't need to explain the game to every player that visits.

You are responsible for supplying the materials and running your game on the day of the fair.

Your game will be marked on the **Games Fair Day** based on the following...

- was the game original, creative, and simple to play?

- did the game create enthusiasm and attract players?

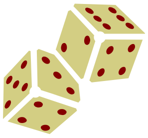
- did the payouts appear to be fair and enticing?

You will also submit a **written report** including the following...

- description of the rules of your game and materials required
- list of the payouts of your game and the rules for payout
- payouts of your game communicated using "odds"
- theoretical probability of each payout calculated
- net expected payout per game calculation
- a chart showing the actual net payouts per game based on the participants that played your game
- comparison of empirical net payout per game to theoretical net payout per game. Account for similarities/differences
- probability distribution function created for your game and approximated with the normal distribution
- player feedback survey on whether the game was based on skill or luck?
- a discussion of the role that casinos play in society and how your game relates to some of these social issues
- include your physical game along with your report

Probability Distributions in the Real World Unit

Create your own game of chance rubric



Group members _______________ **Game** __________________

	Level 1	Level 2	Level 3	Level 4
Knowledge (25%)				
Payouts communicated using odds				
Theoretical probability calculated accurately				
Accuracy of probability distribution for your game				
Net expected payout per game calculated accurately				
Application (25%)				
Game based on binomial distribution				
Approximation using normal distribution				
Teamwork and administration of your game				
Empirical vs. theoretical payouts chart				
Thinking (25%]				
Original and creative game designed				
Thoughtful design of net payouts				
Role of casinos in society				
Games fair administration				
Communication (25%)				
Overall presentation of your game				
Player feedback survey				
Clear, well organized report				