Socially-based Curriculum Unit: Probability Distributions in the Real World

Unit Title: Probability Distributions in the Real World

Time Frame: Approximately 14 full class periods (3 weeks) + 3 full class period for Unit Project

Unit Developer(s): T. Nicitopoulos

Developed for Course Name and Course Code: Mathematics of Data Management, MDM4U

Strand(s) and Curriculum Learning Expectations Addressed:

Probability Distributions Strand

PD2.01 - recognize and identify a continuous random variable (i.e., a variable that assumes values from the infinite number of possible outcomes in a continuous sample space), and distinguish between situations that give rise to discrete frequency distributions (e.g., counting the number of outcomes for drawing a card or tossing three coins) and situations that give rise to continuous frequency distributions (e.g., measuring the time taken to complete a task or the maximum distance a ball can be thrown).
PD2.02 -recognize standard deviation as a measure of the spread of a distribution, and determine, with and without technology, the mean and standard deviation of a sample of values of a continuous random variable.
PD2.03 - describe challenges associated with determining a continuous frequency distribution (e.g., the inability to capture all values of the variable, resulting in a need to sample; uncertainties in measured values of the variable), and recognize the need for mathematical models to represent continuous frequency distributions.
PD2.04 - represent, using intervals, a sample of values of a continuous random variable numerically using a frequency table and graphically using a frequency histogram and a frequency polygon, recognize that the frequency polygon approximates the frequency distribution, and determine, through investigation using technology (e.g., dynamic statistical software, graphing calculator), and compare the effectiveness of the frequency polygon as an approximation of the frequency distribution for different sizes of the intervals.
PD2.05 2.5 recognize that theoretical probability for a continuous random variable is determined over a range of values (e.g., the probability that the life of a lightbulb is between 90 hours and 115 hours), that the probability that a continuous random variable takes any single value is zero, and that the probabilities of ranges of values form the probability distribution associated with the random variable
PD2.06 - recognize that the normal distribution is commonly used to model the frequency and probability distributions of continuous random variables, describe some properties of the normal distribution (e.g., the curve has a central peak; the curve is symmetric about the mean; the mean and median are equal; approximately 68% of the data values are within one standard deviation of the mean and approximately 95% of the data values are within two standard deviations of the mean), and recognize and describe situations that can be modeled using the normal distribution (e.g., birth weights of males or of females, household incomes in a neighbourhood, baseball batting averages).

PD2.07 - make connections, through investigation using dynamic statistical software, between the normal distribution and the binomial and hypergeometric distributions for increasing numbers of trials of the discrete distributions (e.g., recognizing that the shape of the hypergeometric distribution of the number of males on a 4-person committee selected from a group of people more closely resembles the shape of a normal distribution as the size of the group from which the committee was drawn increases) Sample problem: Explain how the total area of a probability histogram for a binomial distribution allows you to predict the area under a normal probability distribution curve.
PD2.08 - recognize a z-score as the positive or negative number of standard deviations from the mean to a value of the continuous random variable, and solve probability problems involving normal distributions using a variety of tools and strategies (e.g., calculating a z-score and reading a probability from a table; using technology to determine a probability), including problems arising from real-world applications Sample problem: The heights of 16-month-old maple seedlings are normally distributed with a mean of 32 cm and a standard deviation of 10.2 cm. What is the probability that the height of a randomly selected seedling will be between 24.0 cm and 38.0 cm?

Desired Results

Unit Description:

Continuous probability distributions are extremely useful when modeling real world situations for random variables. Students will gain experience using the normal and exponential (Pareto) probability models applied to contemporary social issues in the real world. Students will gain an appreciation of how these models are derived and will use a variety of methods to calculate probability of various events occurring based on these models.

Enduring Understandings / Learning:

Students will be able to…

- understand how continuous probability distributions are created from histograms
- understand the advantages of using a continuous probability distribution over a histogram
- understand that the area under a continuous probability distribution can be used to determine the probability of various events occurring
- recognize situations that can be modeled using the normal distribution and why so many real life situations follow this particular model
- sketch normal probability models given a population mean and standard deviation

- calculate the mean and standard deviation of continuous random variables
- calculate the probability of events from a population that is normally distributed using z-scores and technology
- identify skewness as it relates to probability distributions such as the exponential (Pareto) distribution and real life examples that can be modeled by such
- understand the challenges faced when trying to determine the population mean for a continuous random variable
- calculate a range of values for which a true population mean lies (confidence interval) using a sample of data and z-scores and technology
- approximate the binomial distribution using a normal approximation for a probability model for a game of chance that they create

Assessment Tasks

Performance Tasks and Other Evidence That Will Demonstrate the Knowledge and Skills Acquired:

Quizzes, Unit project (Appendices E and F), Unit Test

Assessment Criteria:

Rubric attached for Unit Project (Appendix F)

Unit Planning Notes

Prior Learning Necessary (if any):

Students need to know how to…
1. Create a frequency distribution (histogram) and relative frequency distribution
2. Calculate probabilities of events based on a binomial and hyper-geometric distributions
3. Have some prior experience using the graphing calculator and its statistical features

Preparation Notes (if any):

1. Teachers will need access to a class set of graphing calculators
2. Teachers will need to ensure that "Probability Simulator" program is loaded on all calculators
3. Teachers will need to coordinate school participation in the Unit Project (see Appendix E) including: organizing dates, room locations, and overall administration of the project.

Learning Plan

Lesson 1

1. Present an example of a discrete variable and a continuous variable on the blackboard and see if students can discover the difference between the two
2. Create a histogram and relative frequency distributions for each of the variables and calculate the mean and standard deviation for each using the grouped data via the distributions
3. Determine probability of various events occurring using the relative frequency distributions
4. Have students practice similar problems for the rest of the period and for homework

Lesson 2

1. Start with a data set for student heights (cm) in the class and create a relative frequency distribution
2. Provoke students to think about whether this would be a good model for another class of the same grade
3. Turn the relative frequency distribution into a continuous probability model and ensure students understand that the sample space represents 100% of the area under the distribution
4. Use a simple uniform distribution for student height to section off and calculate different areas to determine the probability of students having heights between a specific range
5. Now have students work with a continuous probability distribution that isn't uniform to gain an appreciation of the challenges in determining the area under the curve
6. Have students practice creating continuous probability distributions from relative frequency distributions

Lesson 3

1. Introduce the normal distribution using a variety of examples from the real world (weight, height, etc.)
2. Have students discover that the mean, median, and mode are all the same and the normal distribution is symmetrical relative to them
3. Show students that sectioning off different areas of the normal distribution based on the number of standard deviations away from the mean generates in well defined probabilities and areas under the curve. (for example, 68% of the area is contained within 1 standard deviation plus/minus from the mean)
4. Introduce the concept of a z-score as the number of standard deviations that a value for the continuous random variable is away from the mean
5. Practice calculating z-scores and introduce the students to the standard normal distribution which the area under the curve is known to a high degree of accuracy.
6. Use standard normal distributions and z-score tables to calculate the probabilities of various sections under the normal distribution to solve a variety of real life applications
7. Have students practice sketching various normal distributions with specified means and standard deviations plus calculating probabilities using z-scores and tables for the standard normal distribution

Lesson 4

1. Distribute the activity titled "Sustaining Canada's Forest Resources" (see Appendix A)
2. All the entire period to work on this activity

Lesson 5

1. Take up the activity "Sustaining Canada's Forest Resources"
2. Allow students to continue to practice working with normal distributions and calculating probabilities using z-scores

Lesson 6

1. Quiz students on material covered to date
2. Start lesson afterwards outlining the difference between a sample mean and a population mean (the latter being the true mean and very often difficult to measure)

3. Pose the question as to what the true mean snowfall is at a ski resort over the past 200 years and entice students to think about why this would not be known
4. Provoke students to ask whether the mean over a sample of 50 years would be a good place to start in terms of making an inference as to a range where the true mean could lie
5. Introduce the "central limit theorem" which basically says that regardless of the underlying population, the sample means will become normally distributed through repeated sampling and a large enough sample size
6. Students create a probability distribution of the sample mean of the ski resort's annual snowfall and sketch a normal distribution with a much tighter standard deviation (s/sqrt(n))
7. Students calculate probabilities that the sample mean lies within a range of values using z-scores

Lesson 7

1. Start lesson by introducing the concept of a "confidence intervals" for the true mean of a population based on the sample means (eg. 98% chance that the true mean lies within this interval)
2. Construct a confidence interval for the true mean snowfall received by the ski resort in the previous activity
3. Distribute the activity called "Confidence intervals – weighing the options…" and allow students the remainder of class time to work on this activity.
4. Students are to finish the rest for homework.

Lesson 8

1. Take up activity from previous day
2. Assign more questions regarding confidence intervals

Lesson 9

1. At this point, students are ready to look at another continuous probability distribution
2. Handout the activity "Where does 80% of the population live?" (see Appendix C)
3. Students use the remainder of the time to work on this activity

Lesson 10

1. Quiz confidence intervals
2. Allow the remainder of the period to finish yesterday's activity

Lesson 11

1. Take up yesterday's activity
2. Distribute activity called "Normal Approximation to Discrete Probability" (see Appendix D)
3. Allow students the remainder of the period to work on this activity

Lesson 12

1. Take up yesterday's activity
2. Allow students the remainder of the period to practice using the normal approximation to solve binomial and hyper-geometric probability problems

Lesson 13

1. Distribute the Unit Project (see Appendices E and F)
2. Ensure students understand what needs to be done and the timelines.
3. Allow students the remainder of the period to review for unit test

Lesson 14

1. Students write Unit Test

Attachments

Appendix A: Sustaining Canada's forest resources
Appendix B: Confidence intervals…weighing the options…

Appendix C: Where does 80% of the population live?
Appendix D: Normal approximation to discrete probability
Appendix E: Create your own game of chance
Appendix F: Create your own game of chance rubric

Other Possible Course Applications

Senior level Geography and Economics teachers may find the activities useful in terms of topics covered in their respective courses. The Mathematics component of each activity should be removed or made accessible to students in these courses