Why np min?

Math Objectives

- Students will understand that normal distributions can be used to approximate binomial distributions whenever both *np* and *n*(1 – *p*) are sufficiently large.
- Students will understand that when either *np* or *n*(1 – *p*) is small, the normal distribution probabilities for impossible numbers of successes (less than 0 or greater than *n*) are unreasonable.
- Students will formulate guidelines to determine what they mean by sufficiently large for good approximations.

Vocabulary

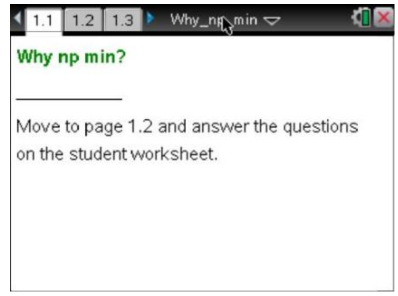
- binomial random variable
- normal random variable
- probability distribution function
- standard deviation
- mean
- number of trials, *n*
- probability of success, *p*

About the Lesson

- This lesson involves examining the general shape of binomial distributions for a variety of values of *n* and *p*.
- As a result, students will:
    - Compare the shapes of binomial distributions to those of related normal distributions, recognizing the distinction between discrete and continuous random variables.
    - Recognize that normal approximations of binomial probabilities become less and less accurate as either *np* or *n*(1–*p*) falls below 5 (or 10 or 15) by examining the probabilities calculated from the normal distribution for having the "number of successes" be less than 0 or greater than *n*.

Prerequisites Knowledge

- Familiarity with binomially distributed random variables, including their fundamental definition and formulas for determining exact probabilities, mean, and standard deviation.
- Familiarity with normally distributed random variables, including the general shape of the normal *pdf* and how the graph of the *pdf* depends on the mean and standard deviation.



Tech Tips:

- This activity includes class captures taken from the TI-Nspire CX handheld. It is also appropriate for use with the TI-Nspire family of products including TI-Nspire software and TI-Nspire App. Slight variations to these directions may be required if using other technologies besides the handheld.
- Watch for additional Tech Tips throughout the activity for the specific technology you are using.
- Access free tutorials at http://education.ti.com/calculators/pd/US/Online-Learning/Tutorials

Lesson Files:

Student Activity

- Why_np_min_Student.*pdf*
- Why_np_min_Student.doc

TI-Nspire document

- Why_np_min.tns

TI-Nspire™ Navigator™ System

- Send out the *Why_np_min.tns* file.
- Monitor student progress using Class Capture.
- Use Live Presenter to spotlight student answers.

Activity Materials

- Compatible TI Technologies: TI-Nspire™ CX Handhelds, TI-Nspire™ Apps for iPad®, TI-Nspire™ Software

Discussion Points and Possible Answers

Teacher Tip: This lesson uses the binomial random variable "count of successes" in all calculations and graphs in order that it can be used during or immediately after the study of random variables, before beginning formal inference methods. If you want to wait for the study of inference, you might also want to replace the count of successes, $\boldsymbol{x}$, with the more familiar proportion of successes, $\hat{\boldsymbol{p}}$, making appropriate adjustments in the corresponding means and standard deviations. This lesson refers to (1 – $\boldsymbol{p}$) as the probability of a "failure" outcome. Some texts use the notation $\boldsymbol{q}$ for that quantity.

Tech Tip: Be sure students understand how to use the *n* and *p* "slider" controls. For *n*, select the up or down indicator to step through values of *n* from 0 to 30 in increments of 5. For *p*, grab and drag the slider left or right to move through values of *p* in the interval [0, 1].

1. You have already studied binomial random variables. This question reviews important facts about the distributions of such random variables.
   a. Describe exactly what a binomial random variable measures (counts).

   <u>Answer:</u> A binomial random variable counts the number of successes in exactly $\boldsymbol{n}$ identical and independent random trials, each having probability exactly $\boldsymbol{p}$ of producing a success.

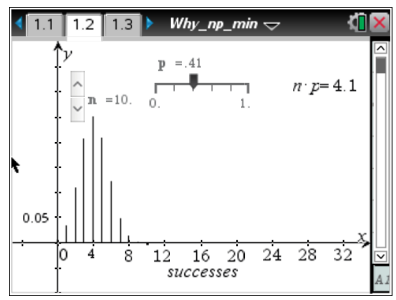
   b. Write the formula for the binomial probability of exactly $\boldsymbol{x}$ successes. Be sure to state clearly what each letter in your formula represents.

Answer: $B(x)=\binom{n}{x}p^x(1-p)^{n-x}$, where ***n*** is the number of trials, ***p*** is the probability of success on each trial, and ***x*** is the number of successes for which the probability is being calculated, which is ***B*(*x*)**.

c. In the context of a binomial scenario, what do the values of ***np*** and ***n***(1 – ***p***) mean?

Answer: *np* is the average number of successes and *n*(1 – *p*) is the average number of failures in *n* trials.

Move to page 1.2.



2. The image on Page 1.2 is the probability distribution function (*pdf*) for a binomially-distributed random variable defined by ***n*** trials and probability ***p*** of success on each trial.

   a. Explain why the graph consists of a set of separated vertical bars.

   Answer: Every binomial random variable counts the number of successes in some fixed number of trials. A count can only be a positive integer, so only positive integers can have a non-zero probability of occurring and appear as a location for a "bar."

   b. What does the height of each bar represent, and what formula is used to calculate that height?

   Answer: The height of the bar gives the probability of having "*x*" number of successes using the given number *n* of trials and probability *p* of success per trial, $\binom{n}{x}p^x(1-p)^{n-x}$

3. A description of any distribution usually includes information about its center and shape, among other things. The questions below ask you to use the sliders to investigate how values of ***n*** and ***p*** affect the center and shape of binomial distributions.

   a. Set ***n*** to 30. Then vary ***p*** from 0 to 1. How are ***np*** and ***n***(1 – ***p***) related to the "center" of the distribution?

   <u>Sample Answers:</u> *np* is an *x*-coordinate near the visual center of the graph and represents the expected number of successes. *n*(1 – *p*) is the expected number of failures, so it is approximately the distance from that center to the right end of the domain *n*.

   b. Shape usually includes both modality (unimodal, bimodal, multimodal) and symmetry (symmetric, skewed left, skewed right). Summarize your observations regarding shape as you changed values of ***p*** in question 3a.

   <u>Sample Answers:</u> For most values of *p,* the shape seems to be about the same, staying unimodal and mound-shaped. As *p* increases, the location of the mode moves gradually from left to right. For most values of p, the distribution is approximately symmetric; however, as values of *p* approach 0 or 1, the distribution becomes more and more skewed—to the right for *p* near 0 and to the left for *p* near 1. Also, when *p* =0 (or 1), the distribution degenerates to a single bar at 0 or *n*, respectively.

   c. Repeat questions 3a and b with other values for ***n***. (You might need to adjust ***p*** more slowly as its value nears 0 or 1 in order to observe the changes more clearly.)

   <u>Sample Answers:</u> The general behavior remains the same. The center seems to be around *np;* and for most values of *p,* the shape seems to be unimodal, mound-shaped, and generally symmetric. However, the larger the value of *n*, the wider the interval of values for *p* for which the graph stays reasonably symmetric.

   d. Explain why some choices of ***n*** and ***p*** lead to graphs that are very non-symmetric.

   <u>Sample Answers:</u> When *p* is close to 0 or 1, the average number of successes is very near the edge of the domain. Thus one "side" of the graph is essentially cut off.

TI-Nspire Navigator Opportunity: *Class Capture*
See Note 1 at the end of this lesson.

Move to page 1.3.

4. The plot on Page 1.3 contains exactly the same information as that on Page 1.2. However, the appearance of the plot is different and probably looks more familiar. Describe the difference between these two plots, and then comment on their relative advantages.

   Answer: The new plot shows unit-wide bars instead of line segments at each "success" value. The new plot is more familiar and makes the overall pattern a little easier to see by "smoothing out" the graph. The earlier plot emphasizes the discrete nature of the variable. Both graphs are correct.

5. Select a value of your own choosing for ***n***. Then choose ***p*** as far from 0.5 as you can while still displaying a probability histogram that appears reasonably symmetric. Record your values for ***n*** and ***p***.

   Sample Answers: I used $n = 10$. Using $p = 0.3$ seemed to be about the smallest value of p for which the graph appeared to remain symmetric.

> **Teacher Tip:** For question 5, you might want to assign different values of n to various groups in the class, and have some groups look at values of p above 0.5 while others examine values of p below 0.5. That way the full range of possible graphs can be observed among the groups.

TI-Nspire Navigator Opportunity: *Class Capture*
See Note 2 at the end of this lesson.

> **Teacher Tip:** Discuss student answers to questions 4-6 so that all students see that one effect of replacing the individual "spikes" with adjacent bars is to "smooth out" the overall shape of the discrete histograms. This might suggest to some students that a continuous curve might be a reasonable approximation. Knowing the values of the mean and standard deviation help determine exactly which curve to use.

6. You found values for ***n*** and ***p*** in question 5 that produce an approximately symmetric binomial probability histogram. Thus it might seem reasonable to try to overlay a normal curve over your graph. But which one?
   a. What two numbers (parameters) are needed in order to graph any specific normal curve?

      Answer: Its mean and standard deviation.

b. What are the mean and standard deviation of the binomial *pdf* you made in question 5? Show the formulas you use to calculate these values.

Sample Answers: For n = 10 and p = 0.3 the mean is 3, and the standard deviation is $\sqrt{2.1}$ or about 1.45. mean = np and standard deviation = $\sqrt{np(1-p)}$.

c. Explain why knowing just ***n*** and ***p*** is sufficient to define both the binomial *pdf* and its related normal *pdf*.

Answer: The binomial is defined by n and p. And since the normal curve is determined by its mean and standard deviation, and the mean and standard deviation of the binomial depend only on n and p, those two values completely determine both *pdf*s.

Move to page 2.1.

7. The plot on Page 2.1 graphs a normal *pdf* on top of the corresponding binomial probability histogram. Set the controls of Page 2.1 to match your graph in question 5.

a. Describe how well the normal *pdf* fits the binomial probability histogram.

Sample Answers: It seems to be a good fit. The graphs are very close to each other.

b. Look back at your answer to question 3d. Based on what you saw earlier, predict exactly how the graph of a normal *pdf* might differ from its corresponding binomial *pdf* if the average number of successes or failures is very small. Be as specific as you can in describing the graphs. Remember, normal *pdf*s are *always* symmetric.

Sample Answers: I expect that as the average number of successes gets small the left end of the normal will extend to the left of x = 0 and include areas that are not part of the binomial. A similar problem will occur on the right end for small values of average number of failures (large values of average number of successes).

c. Use the sliders to check your predictions. Describe your observations.

Sample Answers: My predictions were pretty much correct. As p got near 0 or 1, the normal curve seemed to "hang over" the end of the graph too much.

Teacher Tip: Note that decisions in question 7 will be subjective. Steer discussion of student answers toward methods that can be replicated and used by others. One such method is defined in question 8 ("restrict the approximating probabilities of impossible events"), but students should be encouraged to come up with ideas of their own. For example, they might suggest that approximations are poor whenever two or three standard deviations away from the mean falls outside the domain of [0, n].

Tech Tip: The unlabeled slider to the left of the vertical axis on Page 3.1 permits the binomial histogram to be displayed or removed for easier visualization of the shaded normal probabilities for $P(x < 0)$ and $P(x > n)$.

Move to page 3.1.

8. a. What are the possible values for a binomial random variable (count of successes)?

   Answer: Only integers from 0 through *n*, inclusive.

   b. Explain why your answer to part a is problematic when using a normal distribution to approximate a binomial distribution.

   Sample Answers: A normal distribution always allows values from the entire number line, so numbers less than 0 and greater than *n* will cause conflict between the two models.

   c. Page 3.1 displays the normal probabilities for $x < 0$ and $x > n$. For the actual binomial distribution, what are the exact values of these two probabilities? In general, how do the normal and binomial probabilities compare?

   Sample Answers: The binomial probabilities are always 0. In general, the two models are pretty close unless the mean of the distributions gets close to 0 or *n*.

d. Observe these two probabilities as you vary ***n*** and ***p***. Record the combinations of ***n*** and ***p*** for which the normal approximation gives a value of at least 0.01 for either of these two "extreme" probabilities. Be sure to check for values of ***p*** near 1 as well as for values near 0. (Note: Use the unlabeled slider to the left of the vertical axis to hide/show the binomial histogram bars.)

Sample Answers: For $n = 30$, $P(x < 0)$ first exceeds 0.01 at about $p = 0.15$. For $n = 20$, it happens around $p = 0.21$. For $n = 10$, it's around $p = 0.36$. For $n = 5$, at least one of these probabilities is always greater than 0.01. For $P(x > n)$, replace each p with $1 - p$ and the results remain valid.

TI-Nspire Navigator Opportunity: *Quick Poll*
See Note 3 at the end of this lesson.

9. Many textbooks include guidelines for using normal random variables to approximate binomial random variables. Those guidelines usually are of the form, "Check that each of the values of np and $n(1 - p)$ is at least ____ ." Different books disagree on what they prefer to put in the blank as their "magic number."

a. What number would you put there, and why?

Sample Answers: I would use 5 since it seems to be big enough for the rule we used in question 8.

b. How does the quality of the approximation of a binomial distribution by a normal distribution vary as the "magic number" increases? Explain.

Sample Answers: As the products increase the approximations get better, at least in the sense that less and less of the normal distribution corresponds to impossible events.

Wrap Up

Upon completion of the discussion, the teacher should ensure students are able to understand:

- Binomial distributions are actually discrete and finite.
- The mean and standard deviation of a binomial random variable are determined by ***n*** and ***p***.
- Since a normal distribution is determined by its mean and standard deviation, ***n*** and ***p*** completely determine the appropriate normal distribution to approximate a binomial distribution.
- Normal distributions are reasonable approximations to binomial distributions whenever both ***np*** and ***n***(1 – ***p***) are sufficiently large (usually taken to be 5, 10 or 15).

Assessment

Use Quick Poll or an exit quiz to assess students' understanding of the key concepts in the lesson.

T or F. (Students should be asked for reasons for their choices.)

1. If ***n*** = 20 and ***p*** =.15, the binomial distribution can be fairly well approximated using a normal distribution.

   Answer: F. According to usual guidelines, *np* should be at least 5 (or 10 or 15). Here *np* is only 3.

2. The quantity ***n***(1 – ***p***) represents the average number of failures in ***n*** trials in the context of a binomial setting.

   Answer: T.

3. When ***p*** is close to 1, the average number of successes is small.

   Answer: F. *np* is the average number of successes, so it would be near *n*, its maximum possible value.

4. When ***p*** is close to 1, the binomial distribution is skewed right and has no or a very small left tail.

   Answer: F. The exact opposite is true. The description in the stem applies to *p* near 0.

5. If ***np*** >10, you do not have to worry about the size of ***n***(1 - ***p***) in order to approximate the binomial with a normal distribution.

   Answer: F. If the average number of successes is large then the average number of failures can be too small, so it has to be checked as well.

6. An essential difference between a normal distribution and a binomial distribution is that the binomial distribution can be used only for discrete outcomes while the normal distribution can be used for continuous outcomes.

   Answer: T.

TI-Nspire Navigator

Note 1

Question 3, *Class Capture*

Use Class Capture for the entire class during their exploration of question 2. You might either ask students to produce graphs that they find interesting or assign various combinations of *n* and *p* for a more systematic look at the various possible graphs.

Note 2

Question 5, *Class Capture*

Use Class Capture for the entire class during their exploration of question 5, especially if different values for *n* have been assigned. Have students begin to notice how symmetry is related to both *n* and *p*.

Note 3

Question 9a, *Quick Poll*

Use a quick poll to collect and display students' suggested cut-off values. Discuss how various values were selected.