



Distinguishing Between Binomial, Hypergeometric and Negative Binomial Distributions

[Jacqueline Wroughton](#)

Northern Kentucky University

[Tarah Cole](#)

Northern Kentucky University

Journal of Statistics Education Volume 21, Number 1 (2013),
www.amstat.org/publications/jse/v21n1/wroughton.pdf

Copyright © 2013 by Jacqueline Wroughton and Tarah Cole all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of the editor.

Key Words: Active Learning; Binomial; Hypergeometric; Negative Binomial; Project.

Abstract

Recognizing the differences between three discrete distributions (Binomial, Hypergeometric and Negative Binomial) can be challenging for students. We present an activity designed to help students differentiate among these distributions. In addition, we present assessment results in the form of pre- and post-tests that were designed to assess the effectiveness of the activity. Pilot study results show promise that the activity may help students recognize the differences in these three distributions.

1. Introduction

Recently, a great deal of research has focused on active learning and hands-on activities in undergraduate statistics courses (e.g., [Ledolter 1995](#); [Chance 1997](#); [delMas, Garfield and Chance 1999](#); [Pfaff and Weinberg 2009](#)). Most of the activities presented have been designed for the algebra-based introductory statistics class. However, this does not mean that students in the calculus-based introductory statistics class could not also benefit from the introduction of activities designed to improve understanding of certain statistical concepts. In addition to the complexity and level of the topics in the calculus-based course, there are issues in this course which are not completely attributed to the content, but also to the make-up of the student population. Many students who take a calculus-based course are confident in their mathematical ability, unlike many in the algebra-based course. This may pose a different kind of issue as students who are confident that they know what they are doing don't show much of their work

when solving problems or don't feel they need to explain their method of reasoning. However, this can also be true for the confused student. Without the work and explanation it becomes hard to distinguish between these two groups of students. It is believed that through well designed activities, teachers are able to better understand a student's thought process, and thus, should be more able to distinguish between these groups of students. This assumes that the activity is designed in a way that encourages students to explain their thought processes in more detail.

The recommendation for hands-on activities is not new. The GAISE report encourages statistics educators to increase student involvement in classroom activities. In particular it is recommended that activities do not lead students step by step but rather allow students to discuss and think critically about the problem at hand ([GAISE College Report 2005](#)). [Lunsford, Rowell, and Goodson-Espy \(2006\)](#) discuss how lectures, or just merely demonstrations by the professor, are not as effective as hands-on activities for developing student understanding. [Mills \(2002\)](#) found that students who were actively collecting and analyzing data developed a better understanding of the concepts.

Although many of the activities that have been designed for an algebra-based course could be adapted for use in a calculus-based course, there are new topics and issues that arise. Students in the calculus-based course are introduced to numerous distributions. Students often have difficulty recognizing the difference(s) between the Binomial, Hypergeometric and Negative Binomial distributions. For example, students may have trouble identifying the appropriate distribution in the following scenario: *When taking the written driver's license test, they say that about 7 out of 8 people pass the test. An examiner is interested in the number of test takers that will have to take the test in order to find 6 people who pass.*

Here we present and assess a hands-on activity whose primary goal is to help students better distinguish among these three distributions. This paper will discuss where the confusion arises, the motivation for the activity, the activity itself, and assessment of the activity.

2. Confusing the Distributions

In a calculus-based introductory statistics course, there is generally a lesson on some of the named discrete distributions. This lesson often addresses the following distributions: Discrete Uniform, Poisson, Binomial, Hypergeometric and Negative Binomial. Generally, all of these distributions, except the Poisson, have already been utilized in a probability unit in the course without the students even realizing it. That is, they have calculated probabilities "from scratch" using the ideas of combinations, permutations, conditional probability, etc. For example, a student may be asked to find the probability when a fair coin is tossed five times, that exactly two are heads. Although this example follows a Binomial Distribution, students learn how to construct this probability prior to ever hearing its name. Thus, the calculations of these probabilities are not new, but the names and specific properties are. We may believe that the application of these distributions should not be troublesome if students have already had exposure to the calculations. However, students often struggle in the probability unit distinguishing between permutations and combinations as well as "with replacement" and "without replacement." From our experience, adding the names and properties to these distributions simply confuses students more. So, one might ask why add them? Three of the motivational factors for students to learn the specific named distributions are:

1. To simplify the implementation of probability calculations by utilizing their probability mass functions (pmfs) and to understand the general form of a pmf.
2. To determine and utilize their specific expected value and variance.
3. To learn about the concept of modeling outcomes of a given situation through demonstration.

In order to understand how students may confuse three of these distributions, we will go through each of the three distributions' characteristics in more detail.

2.1 Binomial Distribution

When the Binomial Distribution is introduced, it is often done so by a list of conditions that must be satisfied. These five conditions (adapted from [Wackerly, Mendenhall and Scheaffer 2008](#)) are:

1. There is a fixed number, n , of identical trials.
2. For each trial, there are only two possible outcomes (success/failure).
3. The probability of success, p , remains the same for each trial.
4. The trials are independent of each other.
5. The random variable Y = the number of successes observed for the n trials.

To determine if an experiment is Binomial, one just needs to examine whether or not each of the characteristics listed above is met.

If the conditions of the Binomial Distribution are satisfied, then the following pmf can be used:

$$f(y; n, p) = \binom{n}{y} p^y (1-p)^{n-y} \quad y = 0, 1, 2, \dots, n.$$

In addition, the expected value and variance can be utilized:

$$E(Y) = np \qquad \text{Var}(Y) = np(1-p).$$

2.2 Hypergeometric Distribution

The Hypergeometric Distribution arises when sampling is performed from a finite population without replacement thus making trials dependent on each other. However, when the Hypergeometric Distribution is introduced, there is often a comparison made to the Binomial Distribution. More specifically, it is said that if n is small relative to the population size, N , then (assuming all other conditions are met) Y could be approximated by a Binomial Distribution. This case is made due to the fact that not replacing the item has a negligible effect on the conditional p . However, when this is not the case, the independence condition is no longer met and the Binomial Distribution will no longer do an efficient job at an approximation since not replacing the item will have an effect on the conditional p . Consequently, the Hypergeometric Distribution should be used instead. So, when one compares the Hypergeometric to the Binomial conditions, one will see that conditions 1, 2, 3, and 5 still hold whereas condition 4 (given in Section 2.1), independence, no longer holds.

When the Hypergeometric Distribution is of interest, the following pmf can be used:

$$f(y; N, n, r) = \frac{\binom{r}{y} \binom{N-r}{n-y}}{\binom{N}{n}} \quad y \in [\max(0, n - (N - r)), \min(r, n)].$$

Where r represents the number of “success” items out of the N total items. In addition, the expected value and variance can be utilized:

$$E(Y) = \frac{nr}{N} \quad \text{Var}(Y) = \frac{(N-n)nr}{(N-1)N} \left(1 - \frac{r}{N}\right).$$

Comparing the pmf of the Binomial Distribution to that of the Hypergeometric Distribution, one can see that they are different due to the “with replacement” aspect of the Binomial Distribution compared to the “without replacement” aspect of the Hypergeometric Distribution. In addition, the support of y looks quite different between the two. This again is due to “with replacement” vs. “without replacement” aspect between the two distributions. For example, when drawing cards from a standard 52-card deck, suppose we are interested in $Y = \#$ of clubs drawn out of n . If $n = 15$, the most clubs we could draw is 13. If $n = 45$, the fewest clubs we could draw is 6. When one compares the expected value and variance of the two distributions, they appear to be very different. However, this is not really the case. If one were to see that r/N (of the Hypergeometric Distribution) is similar to p (of the Binomial Distribution), the expected values are the same and the variances are only different by the factor of $(N-n)/(N-1)$, where the variances are identical in $n=1$; the variance of the Hypergeometric is smaller for $n > 1$. This relates back to the idea that the Hypergeometric Distribution is used when the sample size, n , is no longer small in relation to the population size, N . However, when n is small in relation to N , this factor is negligible (and thus a Binomial Distribution may be appropriate as an approximation).

2.3 Negative Binomial Distribution

When the Negative Binomial Distribution is introduced, it is often compared (and contrasted) to the Binomial Distribution. It has some of the same characteristics (conditions) as the Binomial Distribution, but has two distinct differences: The value of n (the number of trials) is no longer a fixed value and the random variable Y is defined differently; here Y is the number of trials needed to obtain r successes. That is, when looking at the conditions needed for a Binomial Distribution, conditions 2, 3, and 4 still hold whereas conditions 1 and 5 (given in Section 2.1) no longer hold due to the fact that the random variable and n basically change places. The number of successes, r , now becomes fixed and the number of trials, n , becomes the random variable.

When the Negative Binomial Distribution is of interest, the following pmf can be used:

$$f(y, r, p) = \binom{y-1}{r-1} p^r (1-p)^{y-r} \quad y = r, r+1, \dots$$

In addition, the expected value and variance can be utilized:

$$E(Y) = \frac{r}{p} \qquad \text{Var}(Y) = \frac{r(1-p)}{p^2}.$$

Comparing the pmf of the Negative Binomial Distribution to that of the Binomial Distribution, one can see that they look pretty similar in construction, but the placement of the y is different (and of course, the number of combinations is slightly smaller). The support of y is quite different due to how the random variable, Y , is defined. In order to achieve r successes, one must have at least r trials. The maximum number of trials needed is not known.

2.4 Activity Motivation

Looking at sections 2.1-2.3 as the instructor, it appears that the differences in these three distributions are easy to discern; it seems as though one should just go through these five conditions that are needed for the Binomial Distribution to be met and figure out which ones hold and which ones don't hold. However, after having taught this class for multiple semesters, it appeared that the students had a tough time determining which of these distributions a situation followed, and why it would follow that distribution. More specifically, it appeared that the students had a tendency to believe the distribution would be Binomial more than any other of these three distributions. Thus, we felt that there was a need for the students to discover what it meant for independence to be violated, as well as seeing when n is no longer a fixed value. A hands-on activity was believed to be a good method because the impact of student discovery seems to trump the best teacher explanations.

3. The Activity

In order to make scenarios similar and the situation straightforward, the design of this activity only requires a standard 52-card deck. Although the set-up is very straightforward, one could adapt what the cards represent to make the activity mimic a real-life application. For example, one could design a deck of cards where each red card represented a "republican vote" and each black card represented a "democratic vote" for a given election scenario. Another example might be to have red cards represent "defective" items and black cards represent "satisfactory" items where the students could be part of the quality control component of the company. (See [Appendix A](#) for the Activity Handout)

For the activity, the students encounter three distinct scenarios (one for each distribution: Negative Binomial, Binomial, Hypergeometric respectively) and for each scenario the student is told how to shuffle and replace, as well as how the random variable, Y , is defined. This process and definition is what make the distinction among the three scenarios. Below we describe each scenario in more detail.

In scenario one, the student is asked to draw a single card, then replace the card, shuffle, and repeat this process until two hearts have been obtained. The random variable, Y , is defined as the number of draws to achieve two hearts. Due to the fact that the card is replaced (and cards shuffled), independence is achieved. However, since the student does not know how many times that they will have to draw, they only know when to stop by achieving the second heart, the

number of draws is the random variable, thus making scenario one the Negative Binomial Distribution.

In scenario two, the student is asked to draw a single card, then replace the card, shuffle, and repeat this process five times. The random variable, Y , is defined as the number of hearts drawn in the five draws. Due to the fact that the card is replaced (and cards shuffled), independence is achieved. The student knows that they will stop after drawing a total of five cards. Thus, this scenario meets all five conditions to follow the Binomial Distribution.

In scenario three, the student is asked to draw five cards from the deck all at once. The random variable, Y , is defined as the number of hearts drawn in the five cards. Due to the fact that all five cards are drawn at once, independence no longer holds. The student knows how many cards they will draw. Thus, this scenario follows the Hypergeometric Distribution.

In order for the activity to achieve enough data for demonstration of other values of interest, students are asked to go through each scenario multiple times (5-10).

While students wait for the rest of the class to obtain their data, they can work on calculating the expected value and standard deviation of Y for each scenario and compute a theoretical probability using the appropriate pmf. After all students have obtained their data, they are also asked to compare the theoretical probabilities that they calculated to the empirical probabilities obtained by the class. Time and applicability permitting, one may be able to incorporate some extra topics to do as a class (or by group). For example, the class may decide to graph the theoretical and empirical pmf's for the given scenarios. This would give students a nice visual demonstration of how "close" the observed data is to the expected. In addition, if one would have designed an "unfair" deck, students would get the opportunity to notice the discrepancies between the theoretical and observed values.

At the end of the activity, students are given a follow-up assignment for this activity which involves an activity write-up. These write-ups are completed for each activity that is performed in the class and are similar in nature. The group is asked to include "purpose", "design", and "analysis/results" with discussion. Students are asked to complete a write-up to give them experience on how to communicate statistical results in a meaningful way as well as to give them ownership of the results. Questions used for the write-up for this activity can be found at the end of [Appendix A](#).

As students completed the activity, some did not take it all that seriously (which happens during many in-class activities), while some experienced the "light bulb" moment. However, having the students work together has been anecdotally observed as beneficial as the students get the chance to explain, or question, a distribution to another student in the course. We believe that there is a difference in the level of understanding when students can *explain* a topic to other students versus merely being able to *perform* it themselves. In addition, students get confirmation when they reach the same conclusion as another student.

4. Assessment Design

After implementing this activity a few times, it became of interest to know the extent to which this activity was helping students distinguish among these three distributions. In order to assess

the activity, we used a pre- and post-test design to investigate if students were really having a problem recognizing the distributions before the activity compared to after the activity. The pilot study was conducted at our institution in a class taught by one of the authors.

The pre- and post-tests each consisted of multiple-choice questions where each set of answers were Binomial, Hypergeometric, or Negative Binomial. Students were asked to:

Read each question carefully and circle which of the three discrete distributions (Binomial, Hypergeometric, Negative Binomial) best fits the description given. If you are unsure, please leave the question blank (i.e. – please don't guess).

There were a total of 16 questions created (See [Appendix B](#)) and two tests (of eight questions each) were constructed from this pool of questions. All questions were examined by three statistics professors (one author included) as well as two former students of the course to ensure that the question wording would be clear enough to students and that the correct answer choice was recognizable. In addition, as there may be concern about one of the two tests being more challenging than the other, half of the students were given version A as their pre-test and half were given version B as their pre-test. The students were then given the other version for their post-test. (Thus, half the students did A-B and the other half did B-A). Students were randomly assigned to which test they took first.

It is important to note that these tests were given during class time and were not counted toward the student's grade. That is, a low score wouldn't hurt a grade nor would a high score help a grade. This part is crucial because of the option for students to leave a question blank. Generally, if students are being graded on a test, they will not leave a question blank but instead will take a "wild guess." We wanted to be able to distinguish between a student who got the answer wrong (and thus, thought that they understood things correctly) and a student who wasn't sure (and thus, hadn't necessarily gotten "incorrect" knowledge, but instead was lacking knowledge). However, very few students left any questions blank.

When considering the timing of the pre- and post-tests, we needed to determine what would be best for determining if the activity helped students to better recognize these distributions. This class was taught on Monday and Wednesday afternoons. The lectures on these three distributions were given on a Wednesday (Binomial and Hypergeometric), and Monday (Negative Binomial with practice problems worked out together). The students were then given a homework assignment that was due on Wednesday. This assignment was checked off by the instructor as being completed or not. Note that all students included in the results presented had completed the assignment. The pre-test was given that same Wednesday before the activity. The post-test was given the following Monday at the beginning of class (at which time their activity write-up was due). [Table 1](#) summarizes the timeline.

Table 1: Timeline

Day	Class Material	Assessments Given/Collected	
Wednesday	Lecture on Binomial and Hypergeometric Distributions		
Monday	Lecture on Negative Binomial; Practice Problems	Homework assignment given at the end of class	
Wednesday	Activity	Checked-off homework assignment	Pre-test given at the beginning of class
Monday	Review for Exam	Post-test given at the beginning of class	Activity Write-up due

The option to leave a question blank complicated the scoring method as a blank answer (lack of knowledge) was believed to be different than an incorrect answer (wrong knowledge). The scoring method that was utilized was adapted from the SAT scoring method ([College Board 2012](#)):

Correct answer: 1 point
 Incorrect answer: -0.5 point
 Blank answer: 0 point

When utilizing this method, the highest score would be 8 points and the lowest score would be -4 points.

5. Results and Discussion

At the time of this activity in the class, there were 25 students who were still registered for the course, 22 of whom gave permission for their data to be used and completed the pre-test. Two of these students were not in class the day of the post-test so their results are not included. Thus, the data are reported for 20 students with results for both the pre- and post-tests (see [Table 2](#) below).

Table 2: Individual Student Results

Version for Pre- Test	PRE-TEST				POST-TEST			
	correct	wrong	blank	Score	correct	wrong	blank	Score
A	8	0	0	8	7	1	0	6.5
A	8	0	0	8	8	0	0	8
A	8	0	0	8	6	2	0	5
A	6	2	0	5	5	3	0	3.5
A	6	2	0	5	4	3	1	2.5
A	5	3	0	3.5	6	2	0	5
A	4	4	0	2	6	2	0	5
A	3	4	1	1	4	4	0	2
A	3	5	0	0.5	5	3	0	3.5
A	1	7	0	-2.5	7	1	0	6.5
B	7	1	0	6.5	8	0	0	8
B	7	1	0	6.5	7	1	0	6.5
B	5	3	0	3.5	8	0	0	8
B	4	4	0	2	6	2	0	5
B	4	4	0	2	5	3	0	3.5
B	3	4	1	1	7	1	0	6.5
B	2	6	0	-1	4	4	0	2
B	2	6	0	-1	2	6	0	-1
B	1	7	0	-2.5	8	0	0	8
B	1	7	0	-2.5	2	4	2	0

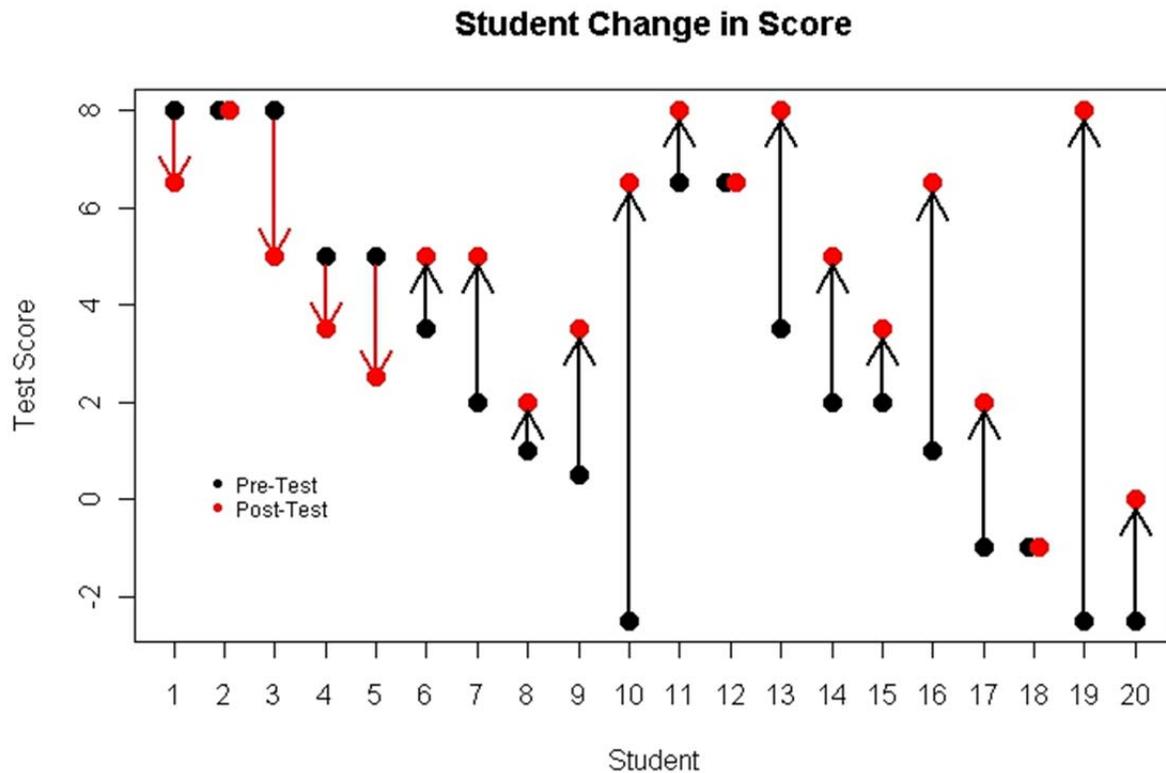


Figure 1: Pre-Test/Post-Test Change

[Figure 1](#) is a graphical representation of how students' scores changed from pre-test to post-test. A black arrow indicates a student whose score increased whereas a red arrow indicates a student whose score decreased. When looking at the [Table 2](#) and [Figure 1](#), one can see that there were quite a few students whose pre-test score is perfect (or near perfect). That is, there were five students who got at least seven of the questions correct on the pre-test. Thus seeing that some of their post-tests scores were lower is not as shocking since they didn't have as far to go up. However, there were also many students (11) who got at least 4 of the questions incorrect on the pre-test. Overall, the majority of the incorrect answers on both versions of the tests were "Binomial" leading authors to believe that this is the "default" answer to a student in doubt. We examined the scores to determine if post-test scores were significantly higher than pre-test scores. In order to assess this, the Wilcoxon Signed-Rank Sum Test was conducted and was found to be statistically significant ($p\text{-value} = 0.0092$). Although the ordering of the versions was randomized between students, it was still of interest to see if there was a significant difference in the change in scores from the pre-test to the post-test depending on which version the student was given as the pre-test. This was found to not be statistically significant ($p\text{-value} = 0.1161$) again using the Wilcoxon Signed-Rank Sum Test. As the semester progressed I discovered that my "least successful" students in the class all had version B first which can partly explain the observed difference in version ordering.

6. Conclusion

Although we are pleased with the promising results that were found in this pilot study, we cannot disregard that this was just that, a pilot study. We have only conducted this assessment on one set of very particular students – ones that had one of the authors as the professor of a certain course designed by a specific university. Another limitation was the lack of an available comparison group. It would have proved beneficial to have had another section of the course spend the same amount of time on the topic and have those students complete the pre-test and post-test to see how the scores would have compared. However, a comparison group was not possible at this time due to the limited number of sections of this course offered at our university, and the inability to include a comparison group without adding another confounding factor (such as instructor).

In addition, the increase in student performance may be contributed to other factors as well. The following questions address some other potential factors. How much did the students self-study between the completion of the pre-test and the administration of the post-test? Would any activity or sample problems completed have similar results? Did the activity increase their understanding or their attitudes about the subject? (And then lead to more time outside of class studying?)

One possible addition that may be beneficial would be for the pre- and post-tests to include an “explain why” part to each question. This would aid in determining whether or not the students are arriving at their answer using the correct information (as opposed to guessing) thus indicating they understood the characteristics of the given distribution. Thus, the results presented here are promising, but more investigation is warranted.

Appendix A: Activity Worksheet

Activity – Working with Discrete Distributions

Scenario 1: With your 52-card deck you will shuffle, randomly draw a card, replace it, shuffle again, randomly draw a card, replace it,...

We will be interested in Y = the number of draws to obtain 2 hearts.

1. Repeat this process 10 times. (Thus, you should have 10 numbers that represent the number of draws it took to get 2 hearts). Record (tally) your values on the board with the rest of the class.
2. Based on the appropriate pmf, calculate the probability of getting the 2nd heart on the 7th draw.
3. Based upon our sample (on the board), what proportion of the time did the 2nd heart come on the 7th draw? Is this close to the answer in 2? Explain!
4. For this scenario, what is the expected number of draws to get the 2nd heart?
5. For this scenario, what is the standard deviation for the number of draws to get the 2nd heart? Use this as well as your answer in #4 above to interpret the standard deviation.

Scenario 2: With your 52-card deck you will shuffle, randomly draw a card, replace it, shuffle again, randomly draw a card, replace it...

We will be interested in Y = the number of hearts drawn in 5 draws.

1. Repeat this process 10 times. (Thus, you should have 10 numbers that represent the number of hearts drawn in 5 draws). Record (tally) your values on the board with the rest of the class.
2. Based on the appropriate pmf, calculate the probability of getting exactly 1 heart in 5 draws.
3. Based upon our sample (on the board), what proportion of time did we get exactly 1 heart in 5 draws? Is this close to the answer in 2? Explain!
4. For this scenario, what is the expected number of hearts in 5 draws?
5. For this scenario, what is the standard deviation of the number of hearts in 5 draws? Use this as well as your answer in #4 above to interpret the standard deviation.

Scenario 3: With your 52-card deck you will shuffle and draw 5 cards at once, replace all of them, shuffle, draw 5 cards at once, replace it...

We will be interested in $Y =$ the number of hearts drawn in 5 draws.

1. Repeat this process 10 times. (Thus, you should have 10 numbers that represent the number of hearts drawn in 5 draws). Record (tally) your values on the board with the rest of the class.
2. Based on the appropriate pmf, calculate the probability of getting exactly 1 heart in the 5 cards.
3. Based upon our sample (on the board), what proportion of time did we get exactly 1 heart out of the 5 cards? Is this close to the answer in 2? Explain!
4. For this scenario, what is the expected number of hearts in the 5 cards?
5. For this scenario, what is the standard deviation for the number of hearts in the 5 cards? Use this as well as your answer in #4 above to interpret the standard deviation.

Take-home questions/exercises for write-up:

1. For each scenario, find the theoretical cdf and the empirical cdf and compare. (For scenario 1, only go up to $y = 20$).
2. Describe what you would expect to have happen to empirical probabilities if instead each individual did each scenario 20 times.
3. While completing the activity, you had to decide which of the discrete distributions that we have talked about fit each scenario. Explain why you chose the distribution you did for each scenario.
4. What are the differences between #2 and #3 for each scenario?
5. What if you were to have found that in Scenario 2, the answer to #3 was 0.1436? What would you wonder about the deck of cards (had you not already verified its components)?
6. Why is it important to not only have an expected value, but to also calculate the standard deviation?
7. Make sure you summarize your findings in the write-up as well as identify which distribution each scenario follows.

Appendix B: Question Pool

Note: Questions 1 – 8 were Version A and questions 9-16 were Version B.

Each question had the following options:

Binomial

Hypergeometric

Negative Binomial

1. At a certain manufacturing company, approximately 5% of the products are defective. We are interested in calculating the probability that the third defective is the 20th one sampled.
2. An assembly line produces products that they put into boxes of 50. The inspector then randomly picks 3 items inside a box to test to see if they are defective. In a box containing 4 defectives, they are interested in the probability that at least one of the three items sampled is defective.
3. When rolling a pair of dice, we generally care about the sum of the two dice. We are interested in the number of rolls that we perform before we get our first sum of 7.
4. The ACT is a standardized test that many students take in order to enter college. It is said that 4 out of every 5 students at NKU take the ACT. We are interested in the number of students in a given class of 30 that took the ACT.
5. A husband has 7 tasks on his to-do list and a wife has 10 things on her to-do list. Five tasks are randomly picked out of these 17 tasks. We are interest in the expected number of tasks the wife will have to do.
6. A certain stoplight, when coming from the North, is green approximately 31% of the time. Over the next few days, someone comes to this light 8 times from the North. We are interested in finding the expected number of green lights the person will come to.
7. A certain radio station's phone lines are all busy approximately 98% of the time when trying to call during a contest. We are interested in finding the probability that the 5th time that you call is the 1st time you get through during a contest.
8. Type O blood is one of the best to be donated since it can be used for many people. Approximately 42% of people have type O blood. In a given day at a blood bank, about 120 people come in to donate. The blood bank is interested in the number of donors who are type O.
9. A police officer has found that approximately 15% of the vehicles he pulls over are from out of state. We are interested in the number of vehicles that are out of state from the next 50 vehicles that he pulls over.

10. In a capture-recapture experiment, 20 animals were captured, tagged and released. A few weeks later, a sample of 40 of these animals is captured and we are interested in the number of animals in our sample that are tagged.
11. When taking the written drivers license test, they say that about 7 out of 8 people pass the test. A test-taker is interested in the number of times they will have to take the test in order to pass.
12. There is an urn containing 15 balls, 40% of which are green. A person gets to pull three balls out at the same time. For each ball that is green, he/she wins \$200. For each ball that is not green, he/she must pay \$50. We are interested in the number of green balls pulled out of the 3.
13. A student is taking a true/false test that consists of 15 questions. The student has approximately a 72% chance of getting any individual question correct. We are interested in the probability that the student gets at least 9 of the 15 questions correct.
14. A company has a pool of 15 applicants (10 male, 5 female) for a particular position that has 3 current openings. They are interested in the probability that none of the positions are filled by females.
15. When rolling a pair of dice, we generally care about the sum of the two dice. We are interested in how many times we get a sum of 7 out of 30 rolls.
16. A couple likes to play darts together. However, the female is not as good at the game and only has about a 8% chance of winning any individual game. Being stubborn, the couple will play until the female wins two games. We are interested in the expected number of games the couple will need to play.

Acknowledgements

This material is based upon work supported by the National Science Foundation's STEM Talent Expansion Program (DUE-STEP) program under Grant No. 0969280. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Aliaga, M., Cuff, C., Garfield, J., Lock, R., Utts, J. and Witmer, J. (2005), *Guidelines for Assessment and Instruction in Statistics Education (GAISE): College Report*. American Statistical Association. Available at: <http://www.amstat.org/education/gaise/>

Chance, B. L. (1997), "Experiences with Authentic Assessment Techniques in an Introductory Statistics Course," *Journal of Statistics Education*, 5(3). Available at:
www.amstat.org/publications/jse/v5n3/chance.html

College Board. (2012), *How the SAT is Scored*. Last accessed on March 11, 2012 from:
<http://sat.collegeboard.org/scores/how-sat-is-scored>

delMas, R. C., Garfield, J., & Chance, B. (1999), "A Model of Classroom Research in Action: Developing Simulation Activities to Improve Students' Statistical Reasoning," *Journal of Statistics Education*, 7(3). Available at:
www.amstat.org/publications/jse/secure/v7n3/delmas.cfm

Ledolter, J. (1995), "Projects in Introductory Statistics Courses," *The American Statistician*, 49, 364 - 367.

Lunsford, M. L., Rowell, G. H., & Goodson-Espy, T. (2006), "Classroom Research: Assessment of Student Understanding of Sampling Distributions of Means and the Central Limit Theorem in Post-Calculus Probability and Statistics Classes," *Journal of Statistics Education*, 14(3). Available at: www.amstat.org/publications/jse/v14n3/lunsford.html

Mills, J. D. (2002), "Using Computer Simulation Methods to Teach Statistics: A Review of the Literature," *Journal of Statistics Education*, 10(1). Available at:
www.amstat.org/publications/jse/v10n1/mills.html

Pfaff, T. J. & Weinberg, A. (2009), "Do Hands-On Activities Increase Student Understanding?: A Case Study," *Journal of Statistics Education*, 17(3). Available at:
www.amstat.org/publications/jse/v17n3/pfaff.html

Wackerly, D.D., Mendenhall, W., & Scheaffer, R.L. (2008), *Mathematical Statistics with Applications*, Belmont, CA: Thomson Brooks/Cole.

Jacqueline Wroughton
Nunn Drive
MP 422
Highland Heights, KY 41099
wroughtonj1@nku.edu
Phone: (859)572-5331

Tarah Cole
311 Upham Hall
Oxford, Ohio 45056
coletl2@muohio.edu