

PROBABILITY ACTIVITIES -- INSTRUCTOR'S GUIDE

LIST OF ACTIVITIES:

- *P1: ACTIVITY: Some questions on probability
- *P1: ACTIVITY: Tossing and spinning a poker chip
- *P3: ACTIVITY: Mothers and babies
- P3: ACTIVITY: Sampling from a bag
- P4: ACTIVITY: Mixed-up letters
- P4: ACTIVITY: The longest run
- P4: ACTIVITY: Sampling people from a room
- *P4: ACTIVITY: The birthday problem
- P5: TECHNOLOGY ACTIVITY: Rolling two dice
- P5: TECHNOLOGY ACTIVITY: How many defectives?
- P6: TECHNOLOGY ACTIVITY: Constructing a probability distribution by simulation
- *P6: TECHNOLOGY ACTIVITY: Playing roulette
- P6: ACTIVITY: How many keys?
- *P6: ACTIVITY: Investing money: comparing safe and risky investments
- *P7: ACTIVITY: Coin flipping: Is it real or fake?
- P7: TECHNOLOGY ACTIVITY: Simulated coin flipping
- P7: ACTIVITY: Is a professional athlete streaky?
- P7: ACTIVITY: Graphing binomial and negative binomial experiments
- P8: TECHNOLOGY ACTIVITY: Spinning away
- P8: TECHNOLOGY ACTIVITY: Waiting for the shuttle
- P8: TECHNOLOGY ACTIVITY: A test with a bimodal distribution
- P9: TECHNOLOGY ACTIVITY: Sampling heights
- P9: TECHNOLOGY ACTIVITY: Rolling biased dice

TOPIC P1: PROBABILITY – A MEASURE OF UNCERTAINTY

P1: ACTIVITY: SOME QUESTIONS ON PROBABILITY

MATERIALS NEEDED: None

This is a good activity for introducing the three interpretations of probability. Students generally have been exposed to the concepts of "equally likely" outcomes, the frequency viewpoint and the subjective viewpoint in their mathematics classes in the schools, but they are unsure when to apply a particular viewpoint for a given situation. By use of the answers to these questions, the instructor can understand some of the popular misconceptions about probability that hopefully can be cleared up in this class.

One can collect the answers to the probability questions and analyze the data. One can graph the probability values for a given question. Also one can look at the reasons for the probability specification for a given question and categorize the answers by the three viewpoints. As a follow-up to this activity, you can discuss some of the general answer patterns with the students.

Here are the probability values for one statistics class:

First set of questions:

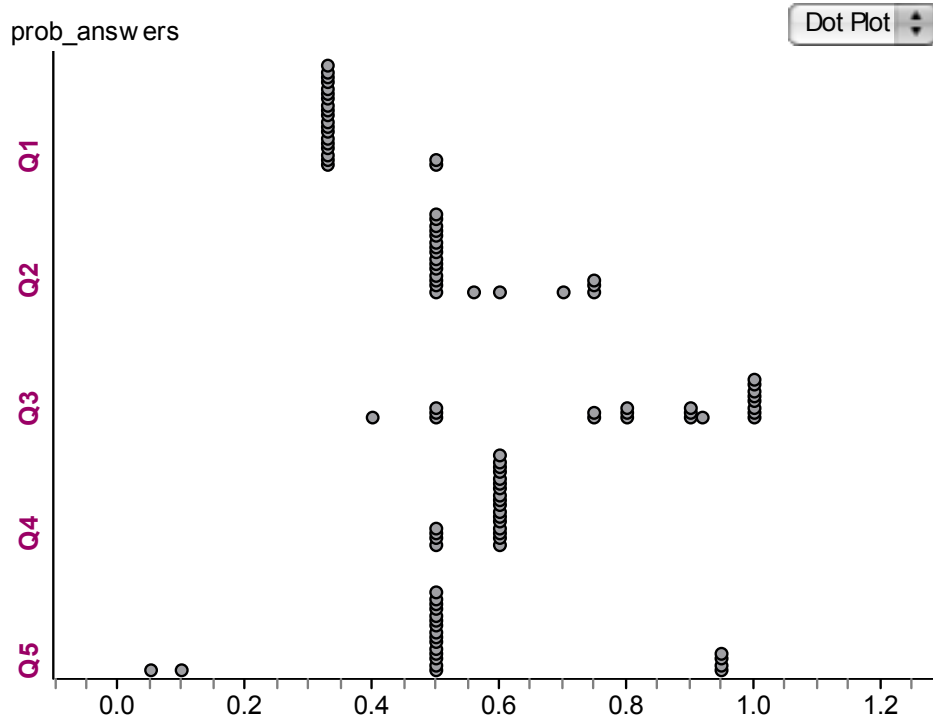
Q1: Probability of choosing white from bag of 4 white and 8 black balls

Q2: Probability of bumping in a female student at the bookstore

Q3: Probability that our team defeats another team in football game this Saturday

Q4: Probability thumbtack lands point-side up with some data given.

Q5: Probability coin lands heads after observing 19 heads and one tail



Comments about the students' answers.

Q1: This question was easy for the students to answer since it was a well-defined problem where each ball was equally likely to be selected.

Q2: Our school has a sizeable majority of female students, but many students said it was equally likely to bump into a female or a male.

Q3: Our college team is better than the opponent, so the probability our team wins is larger than .5. Some students used the equally likely viewpoint and assigned a probability of .5. Other students took the attitude that the more favorable outcome was certain to occur and assigned a probability of 1.

Q4: Since the students had little experience with this thumbtack experiment, most used the data provided to obtain their probability estimate.

Q5: Here the students could have even taken the classical (equally-likely) viewpoint or the frequency viewpoint based on the data provided. Most students had a strong opinion about the fairness of a coin and assigned a probability of .5.

Second set of questions:

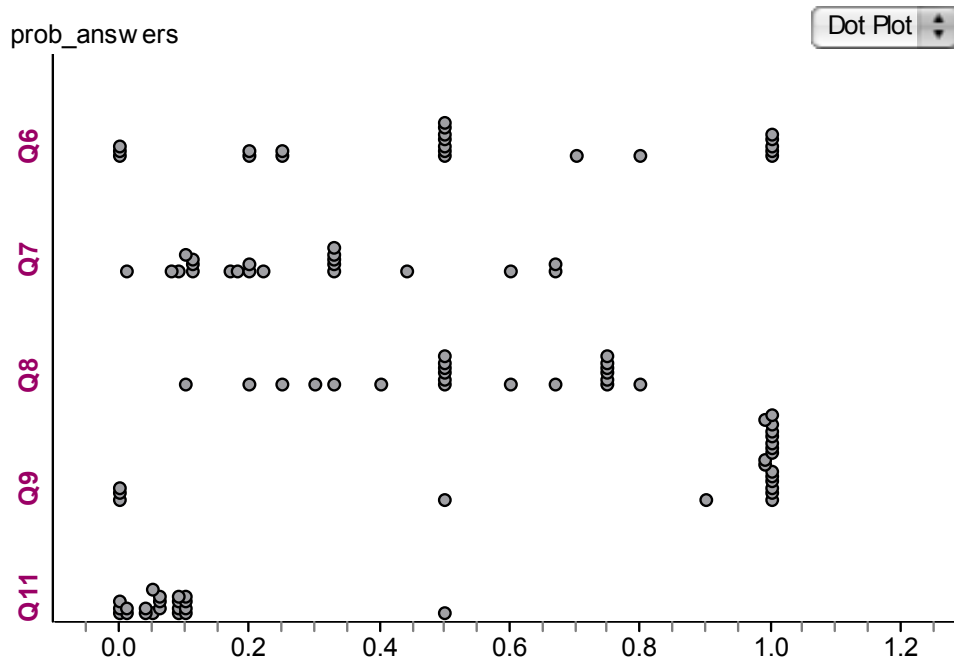
Q6: Probability of being married when you are 25 years old.

Q7: Probability the sum of two dice is equal to 5.

Q8: Probability a high school math teacher is male.

Q9: Probability you graduate in five or fewer years.

Q11: Probability two people in the class have the same birthday.



Comments about the students' answers.

Q6: Students have a strong opinion about their likelihood of getting married, so they were more apt to specify a subjective probability. This explains the high variation in the probability responses.

Q7: Although rolling dice is a familiar experiment, students generally aren't able to compute the probability the sum is equal to 5.

Q8: Many students use the equally likely viewpoint and said that a math teacher is equally likely to be male or female. Only a few students were aware of the fact that most math teachers are male.

Q9: Students are very knowledgeable about their time to complete their degrees – this explains the large number of 0 and 1 responses.

Q11: Students are unaware of the birthday problem and so their guess at the probability of a match is too low. Actually, for a class of 23 students, the chance of a matching probability is close to .5.

Here are some general misconceptions that are made:

1. Students will generally assume outcomes are equally likely even when this assumption is inappropriate. For example, if team A is playing team B in a given sport, some students will say that the probability of team A winning is .5 since there are two possible winners. This comment ignores the differences in the talents of the two teams.
2. Students will be reluctant to use the subjective notion of probability. They think that their opinion doesn't count and so they might use some alternative viewpoint such as the classical one in assigning their probability. They are more likely to use the subjective viewpoint on issues that are important, say the chance he or she graduates from college in five years and the chance that they will be married by the age of 25.
3. Students have the tendency to use a calculation, even a silly calculation, to obtain a probability.

P1: ACTIVITY: TOSSING AND SPINNING A POKER CHIP

MATERIALS NEEDED: A poker chip for each student in the class. This activity can be done using a heavy coin such as a nickel or a quarter. Also a small container of putty is needed for weighting the chips.

The frequency notion of probability rests on the important assumption that one is able to repeat an experiment, say tossing a coin, in exactly the same manner. If one changes the manner in which one does the experiment, then the relative frequency of an outcome will not approach the associated probability.

For example, consider the simple experiment of tossing a chip. How exactly is this done? How high do you toss the chip? Or do you toss the chip for a particular amount of seconds? Do you require the chip to turn over a few times before it lands? What happens if you drop the chip and it lands on the floor?

The instructor can lead a class discussion on how a poker chip is to be tossed. If the students have a clear idea about how the chip is to be tossed, then it is more likely that the students will toss the chips in a consistent manner.

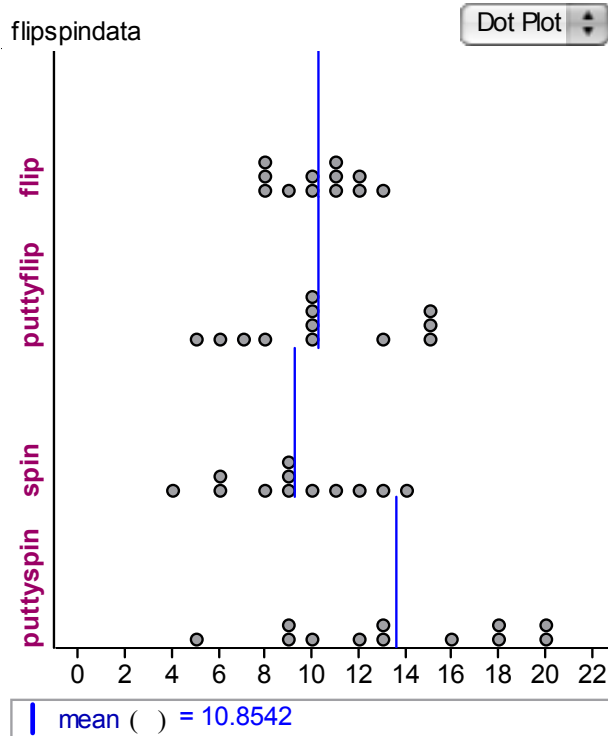
Spinning a chip is less familiar to students than tossing. So the manner in which the chip is to be spun needs to be discussed. How many spins of the chip are required before it is considered a legal spin? Does it count if the chip falls on the floor?

Each student will toss a chip twenty times and record the number of heads; likewise he or she will spin the chip twenty times and record the number of heads. After some putty is applied to the heads side, the student will toss the modified chip twenty times and spin the chip twenty times.

When all students have completed tossing and spinning, then you collect the number of heads for four situations

- tossed, no putty
- spun, no putty
- tossed, putty
- spun, putty

Here are the number of heads in the four situations for one recent class.



What do we find? The addition of the putty should have little impact on the probability of a head when tossed. But the putty will increase the probability of heads when spun.

TOPIC P3: LET ME COUNT THE WAYS ...

P3: ACTIVITY: MOTHERS AND BABIES

MATERIALS NEEDED: For each pair of students, you need four red and four black cards from a bridge deck where the numbers match. For example, an 8-card set could consist of the 2, 3, 6, 7 of spades, and the 2, 3, 6, 7 of hearts.

This activity is a different flavor of the famous hat-checking problem. In the hat-check problem, n men go to a restaurant and check their hats with an attendant. When the men leave the restaurant, the attendant returns the hats to the men in a random order. We are interested in the probability that a particular number of the men get their own hats back.

In this activity, the students find the probability of no matches, one match, etc. by two methods. In the first method, the students simulate this experiment by use of the cards. After 20 trials, they get approximate probabilities. In the second method, all possible orders of four babies are listed. By assuming these different orders are equally likely, one finds the exact probabilities of different number of matches.

P3: ACTIVITY: SAMPLING FROM A BAG

MATERIALS NEEDED: Each pair of students gets a bag containing five blocks - 3 black and 2 white.

Suppose we select two blocks from the bag without replacement, how many blacks will we select?

This activity illustrates three methods for obtaining the probabilities of choosing 0, 1, and 2 blacks in the sample. The first method simulates the process 20 times and the probabilities are approximated by the relative frequencies of the outcomes. In the second method, the student lists all ten possible ways of choosing two from five blocks. By assuming the outcomes are equally likely, one obtains exact probabilities of the different number of blacks selected. The third method computes probabilities based on counting arguments from Topic P3.

TOPIC P4: COMPUTING PROBABILITIES BY SIMULATION

These activities illustrate the use of simulation experiments to compute probabilities for some interesting problems. I believe it is important for students to get experience in simulation experiments both of the hands-on and technology varieties. It is easy for students simply to follow the recipe for performing a simulation experiment without really understanding the process. By simulating the experiment different ways and comparing the empirical probability estimates with exact calculations, hopefully the students will better understand the process. One can assess the students understanding of

experiments by having them design and implement different experiments such as those described in the exercises.

P4: ACTIVITY: MIXED-UP LETTERS

This activity illustrates finding probabilities of particular arrangements of four letters. This example differs from the usual problem of arranging distinct objects in that there are two t's, one s, and one a. Given the indistinguishable t's, there are not $4! = 24$ possible arrangements, but rather $4!/(2! 1! 1!) = 12$ arrangements.

There might be some confusion about the meaning of “word” in this activity. We let a word denote an arrangement of the letters that may or may not actually correspond to a real English word.

P4: ACTIVITY: THE LONGEST RUN

This activity illustrates an arrangement problem where the outcome of interest is the length of the longest run of a single color. There is no simple formula for computing the probability that the longest run will be a specific value and so this problem is well-suited for simulation.

In Part I, the student should first note there are $6!/(3! 3!) = 20$ possible ways of mixing up three red and three blue balls. The longest run will be 1 if you observe the sequences RBRBRB, BRBRBR and so the probability the longest run is 1 is $2/20$. The longest run will be three if you observe BBRRR, RRRBBB, RBBRR, RRBBB, BRRRBB, BBRRRB and so the associated probability is $6/20$.

Parts II and III of the activity illustrate simulating this experiment using the TI-83 plus (TI-84 plus) calculator and Fathom. Depending on the available technology, you will likely have the class work on either part II or part III.

P4: ACTIVITY: SAMPLING PEOPLE FROM A ROOM

This activity demonstrates the basic “sampling without replacement” situation. In Part I, the use of Fathom to simulate this experiment is carefully described. There are five steps in designing this simulation.

Step 1. One first sets up the population – the group of men and women from which you will sample.

Steps 2 and 3. Next one has to indicate how the sample will be taken. How many people will be sampled? Is it taken with or without replacement?

Step 4. When you take a sample, you will record a measure that is the number of women in the sample.

Step 5. One repeats the sampling many times, and the probabilities of interest will be found from the collection of measures from the different samples.

In Part III, the student works through the exact probability calculations using counting arguments. The number of ways of choosing four people from a room of 10 is given by the combination formula (10 choose 4). To find the probability of selecting, say one woman – we first find the number of ways of choosing one woman and three men. There are (6 choose 1) ways of choosing the one woman and (4 choose 3) ways of choosing the men. By the multiplication rule, there are (4 choose 3) (6 choose 1) ways of choosing the woman and the men. Dividing this product by (10 choose 4) gives you the probability.

P4: ACTIVITY: THE BIRTHDAY PROBLEM

This is a nice activity since the students initially think it is very unlikely for two students to share the same birthday and so they are surprised to learn that this event of matching birthdays is pretty likely. You might have the students initially summarize the answers to the question in the “probability questions” activity of topic P1 – they will see that most, if not all students believe that the probability of a match is very small.

At the beginning of this activity, it is helpful to actually collect the birthdays from the students in the class and check for matches. The students will get a good understanding of the birthday problem. Also it is better than even chance that you will get a match (if the number of students in the class exceeds 23), and an observed match will motivate the problem of finding the probability of a match.

This is a very easy experiment to simulate. Simply, one has the computer (or calculator) generate n integers between 1 and 365, where n is the size of the class, and one checks if you observe any duplicate numbers. After one performs the experiment

enough times to approximate the probability, then they can investigate how the probability of a match depends on the class size. At size $n = 23$, the probability of a match is approximately .5 and the matching probability increases quickly for class sizes larger than 23. The last part of the activity guides the student through the exact calculation of the probability. The trick here is that it is easy to compute the probability of the complementary event that there are no matches, and then the probability of interest is given by one minus the probability of no matches.

TOPIC P5: CONDITIONAL PROBABILITY

P5: TECHNOLOGY ACTIVITY: ROLLING TWO DICE

This is a simulation activity using Fathom although it could be done using the calculator. One rolls two dice and records two measures: the sum of the rolls and the maximum of the two rolls. By repeating this experiment many times and summarizing the outcomes of the two measures by a two-way frequency table, one can compute many probabilities of interest. In particular, one computes conditional probabilities by computing conditional row or conditional column proportions from the table.

TOPIC P6: PROBABILITY DISTRIBUTIONS

P6: TECHNOLOGY ACTIVITY: CONSTRUCTING A PROBABILITY DISTRIBUTION BY SIMULATION

In this topic, we introduce probability distributions by constructing a list of equally likely outcomes, labeling each outcome by the value of the random variable, and then using this information to find the probability distribution for X . An alternative way of finding a probability distribution is by simulation. The computer provides an attractive way of obtaining an approximate probability distribution by simulation for complicated experiments where the listing of all outcomes is tedious.

In Fathom, a random variable is represented by a measure computed on a sample. Here we consider a bag with equal numbers of balls labeled 1, 2, and 3, and we have

Fathom sample from this bag until one chooses a ball numbered 3. We define a random variable or Fathom measure that is the number of balls selected. By taking repeated samples and tallying the values of the measure, we obtain an approximate probability distribution.

The activity gives instructions on constructing the Fathom simulation to construct the probability distribution for the number of balls selected to find one numbered 1. Then the activity asks the student to write a simulation for the random variable Y equal to the number of selections to obtain all three types. The only change in this Fathom simulation is the definition of the measure and the special function `uniquevalues()` is useful here.

P6: TECHNOLOGY ACTIVITY: PLAYING ROULETTE

This activity simulates the game of roulette. The student decides to bet \$5 on a particular bet such as the Corner Bet that pays at odds 8 to 1. By use of Fathom or a calculator, the student plays this game 20 times in one day, keeping track of his payoff each time. After 20 plays, the student can compare his total payoff with the \$100 that he or she paid to pay.

Suppose the student plays the game 20 times for each day in the year. On Fathom, one is able to simulate the total payoffs for all days. If this activity is done on a calculator, the instructor can combine the total payoff for all students.

There are several important things to be learned here.

1. For every single bet, the average payoff for twenty plays will be smaller than the \$100 that is bet. For all bets (with the exception of the five-number bet), you will lose about 5 cents to the dollar (or \$5 for a \$100 bet).
2. Although the average payoffs are the same, the bets differ with respect to the standard deviation or variation of the total payoff for 20 plays. If you bet on a single number, for example, the standard deviation of your total payoff will be larger than the total payoff for betting on red.

To reinforce some of these ideas, it is helpful to consider a single bet and let X denote the payoff for this bet. One can construct the probability distribution for X and

compute the mean and standard deviation. One will find that the mean of different bets will be the same, but the standard deviations for the bets will differ.

P6: ACTIVITY: HOW MANY KEYS?

Here one has 6 keys, 2 good ones (labeled G, G) and 4 bad keys (labeled B, B, B, B) and you sequentially try keys to open the door until you find a good one. This activity illustrates two methods for constructing a probability distribution. In the exact method (parts a, b, and c), one lists the five possible outcomes

G, BG, BBG, BBBG, BBBBG

(corresponding to the events $X = 1, 2, 3, 4,$ and 5), and find the probability of each outcome by the multiplication rule. Then the experiment is simulated many times on Fathom and the relative frequencies of the events $X = 1, 2, 3, 4,$ and 5 should be approximately equal to the actual probabilities computed in part (c).

When writing the Fathom simulation, it is important to remember that sampling is done without replacement and you indicate in the sample box that you are sampling until you find a key labeled G. The measure is the number of keys sampled.

P6: ACTIVITY: INVESTING MONEY: COMPARING SAFE AND RISKY INVESTMENTS

MATERIALS NEEDED: Groups of dice of three colors so that each pair of students has a set of three dice. The activity describes yellow, red and white dice, but the activity can be done using any three die colors available.

This activity is very helpful for illustrating the importance of variation of the probability distribution of a random variable.

There are three possible investments called YELLOW, RED, and WHITE. The percentage return for a particular investment depends on the roll of a single die. Which

investment is best? If one compares probability distributions, it is natural to focus on comparing the means. From this criterion, the RED investment is best since the average return is an impressive 71%. But is it really best? This activity should communicate that the variation in the investment, as measured by the standard deviation, is also an important criterion in choosing an investment strategy.

This activity is best done in two stages. Students will roll the three dice 20 times - the value of the rolls will determine the value multipliers for each year of the investment. After the rolls are made, the students can go back to compute the value of the initial \$1000 investment after each year.

After the groups have computed the value of the YELLOW, RED, and WHITE investments after 20 years, then collect the data from all groups. The students are asked to graph the collections of final investment values for YELLOW, and RED, WHITE. They should discover that the final investment values for the safe YELLOW investment show little variation, and in contrast, the final values for RED show wide variation. Some groups will go bankrupt with RED, while other groups will make a lot of money with this same investment. After the students see these graphs, they are asked to apply what they have learned and give advice to a young couple who is starting to invest.

After the activity is done, the students should have some understanding about the risk in an investment and how this risk can be measured in terms of the standard deviation of the yearly average returns.

TOPIC P7: COIN-TOSSING DISTRIBUTIONS

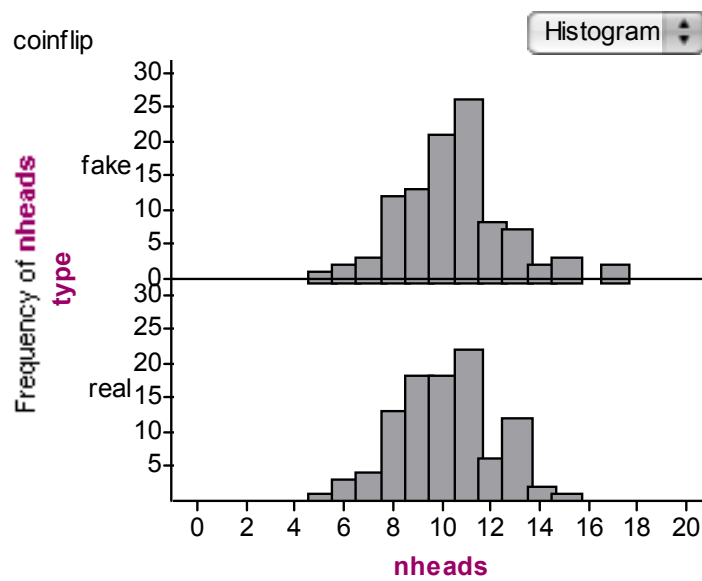
P7: ACTIVITY: COIN FLIPPING: IS IT REAL OR FAKE?

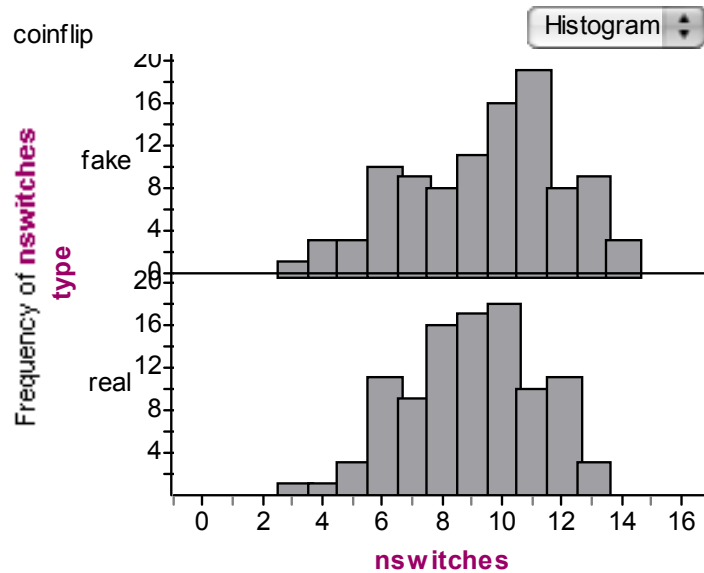
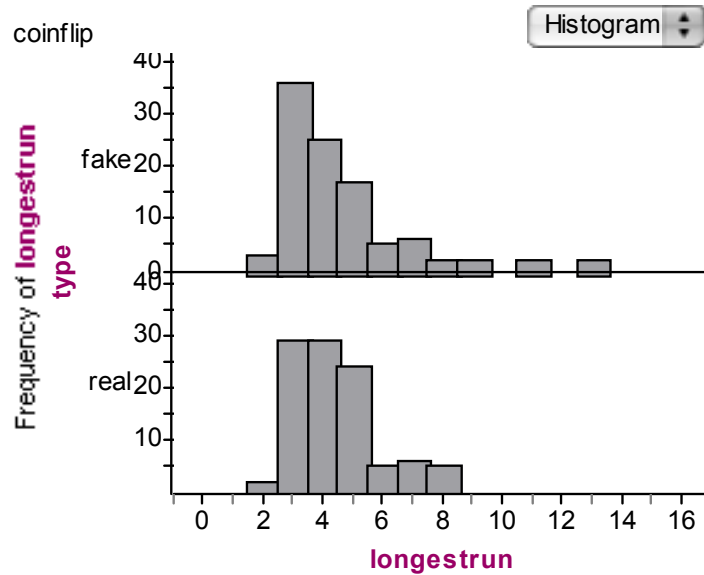
MATERIALS NEEDED: A set of coins of the same denomination (pennies, quarters, etc.)

This activity illustrates that people generally don't understand the patterns of coin tossing. We demonstrate this here by having the students first pretend to flip coins and then actually flip coins. By comparing the fake data from the real data, we can understand our misconceptions about coin flipping.

The students flip coins in groups of twenty. For each group of flips, they compute the values of three random variables: the number of heads, the length of the longest run of either heads or tails, and the number of switches.

Here we compare the distributions of fake and real flips for each random variable. One can compare the distributions by back-to-back stemplots or parallel dotplots. (We use Fathom to construct parallel histograms below.) Although it may not be obvious in the graphs, there are some discoveries from this comparison of fake and real coin-flipping.





1. The numbers of heads of the real and fake coin flips tend to have the same average value. But the spread of the number of heads for the real flips tends to be larger. Real coins tend to have more extreme values of number of heads.
2. Real coin flips tend to have longer streaks than fake flips. It is hard to write down long streaks if you are making up the results.
3. Also real coin flips tend to have a smaller number of switches than fake flips. The number of switches is actually closely related to the number of runs in the sequence.

Do these misconceptions about coin flipping matter? Yes, since they impact our perception of streakiness in our world. For example, people are often surprised by the presence of long streaks of success or failure among athletes, but these long streaks can be very common for a coin-flipping experiment. We explore the streakiness of performances of athletes in another activity in this topic.

P7: TECHNOLOGY ACTIVITY: SIMULATED COIN FLIPPING

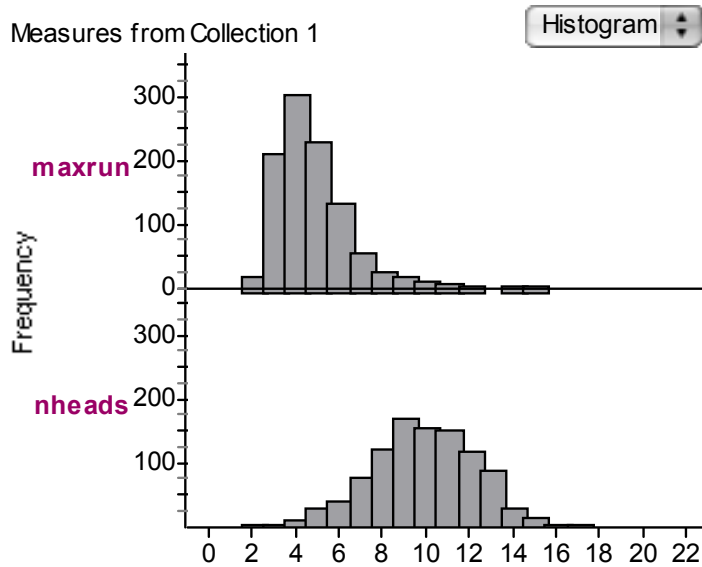
[ALTERNATIVE – CHOOSE COIN A RANDOM FROM ($P=.3$, $P=.5$) – FLIP 20 TIMES AND COMPUTE X AND Y – CONSTRUCT 2-WAY TABLE OF P AND STATISTIC]

In the earlier hands-on coin-tossing activity, students get experience flipping coins and they see that the patterns of real coin flipping are different from the patterns observed if they try to make up the results. This activity illustrates that coin flipping is easy to simulate on a computer. Also the computer can generalize the usual coin-slipping scenario by allowing the chance of heads on a single flip to be a number different from $p = .5$. By exploring the probability distribution of the number of heads in 20 flips and the length of the longest run of heads or tails, the students should understand that the probability distributions do depend on the underlying assumptions – that the probability of heads on a single flip is equal to $.5$ and that results on different trials are independent.

The probability distributions of the number of heads and the length of the longest run found in this simulation activity should resemble the probability distributions found of these two random variables in the earlier hands-on activity. Also, the students should notice changes in the patterns of these probability distributions when one changes the probability of heads from $p = .5$ to $p = .3$. Specifically, the number of heads in 20 flips should decrease -- from an average of 10 to an average of 6. Also one should observe more streakiness in the sequence of 20 flips when $p = .3$ resulting in larger values of the longest run.

Here is some sample output from Fathom. First we show histograms and summary statistics for the number of heads and the length of the longest run when $p = .5$ and then the displays when $p = .3$.

Fathom output for a simulation with $p = .5$



Measures from Collection 1

nheads	9.865
	2.3418495

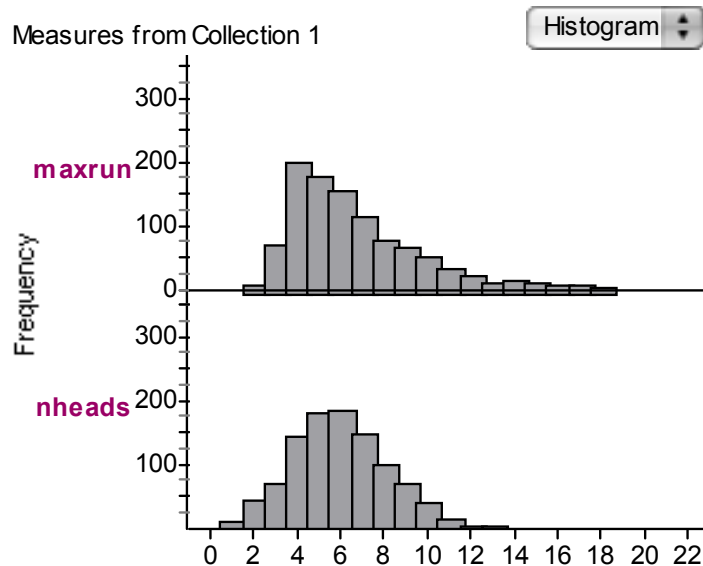
S1 = mean ()
S2 = s ()

Measures from Collection 1

maxrun	4.715
	1.6530298

S1 = mean ()
S2 = s ()

Fathom output for a simulation with $p = .3$



Measures from Collection 1	
nheads	5.897 2.1332937

S1 = mean ()
S2 = s ()

Measures from Collection 1	
maxrun	6.487 2.832641

S1 = mean ()
S2 = s ()

P7: ACTIVITY: IS A PROFESSIONAL ATHLETE STREAKY?

In the previous activities, the students explored the patterns in coin flipping. One thing that was learned was that real coin flipping shows more streaky patterns than one would expect based on the students prior opinions. This misconception has implications when people look at the streakiness patterns in performances of athletes.

Here we focus on the shooting pattern of the great professional player Kobe Bryant during a recent NBA game. We see Kobe make (or miss) a number of consecutive shots – does this mean that he is unusually hot (or cold)?

In this activity, the student first finds the length of the longest run of makes or misses for Kobe's data. To see if this value is unusually small or large, the student compares this value with the longest run of heads or tails from flips of 40 coins of fair

coins. We consider 40 flips since Kobe took 40 shots during this particular game. We consider a fair coin with $p = .5$ since it is reasonable to assume that Kobe's chance of making a particular shot is .5.

For each of the sequences of 40 flips, the student computes the longest run Y . One graphs the 20 values of Y and then draws a vertical line on the graph corresponding to the observed longest run for Kobe. If Kobe's value is in the extreme tails of the distribution, then one would say that Kobe is indeed unusually hot or cold for this game. But it is typical that the value is in the middle of this distribution. This means that Kobe's streakiness is similar to the streakiness in 40 flips of a fair coin.

P7: ACTIVITY: GRAPHING BINOMIAL AND NEGATIVE BINOMIAL EXPERIMENTS

This activity demonstrates a graphical way of recording the outcomes of binomial and geometric experiments. A Bernoulli trial is simulated using a ten-sided die or a random digit and then the outcome is recorded on a graph. One starts at the origin of a x-y grid and then moves to one unit to the right (or one unit up) if the trial is a success (or failure). For binomial experiments, one continues simulating trials until you hit the diagonal line corresponding to 10 trials. For negative binomial experiments, one continues simulating until one hits the horizontal line corresponding to 3 failures.

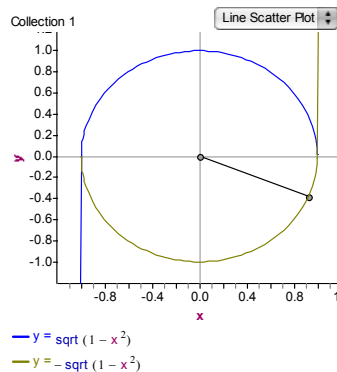
This graph is helpful for seeing the fluctuation in the numbers of successes and failures as the experiment is performed. Essentially one is graphing a random walk where one is either moving north (up) or east (right) on each step. For a binomial experiment, one is taking a total of 10 steps. For a negative binomial experiment, one continues walking until he or she has walked a particular distance north.

Although the questions focus on the binomial or negative binomial outcomes, one could also ask questions regarding the random paths. For the basketball player, what is the probability that the number of makes will always exceed the number of misses during the 10 shots? You could ask the student to find the probability distribution of the maximum difference between made shots and missed shots during the sequence. [ADD THESE TYPE OF QUESTIONS TO THE ACTIVITY?]

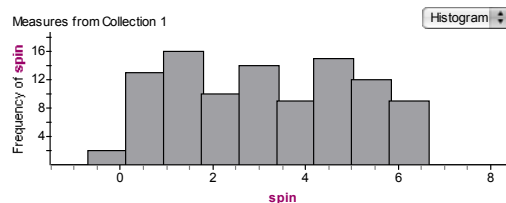
TOPIC P8: CONTINUOUS DISTRIBUTIONS

P8: TECHNOLOGY ACTIVITY: SPINNING AWAY

This activity constructs a random spinner on Fathom. One simulates an angle that is uniformly distributed on the interval from 0 to 2π and the corresponding location of the spinner is displayed on a Fathom graph.



The student is asked to spin this spinner 100 times. A histogram of the spinner results approximates the density function of a uniform random variable X .



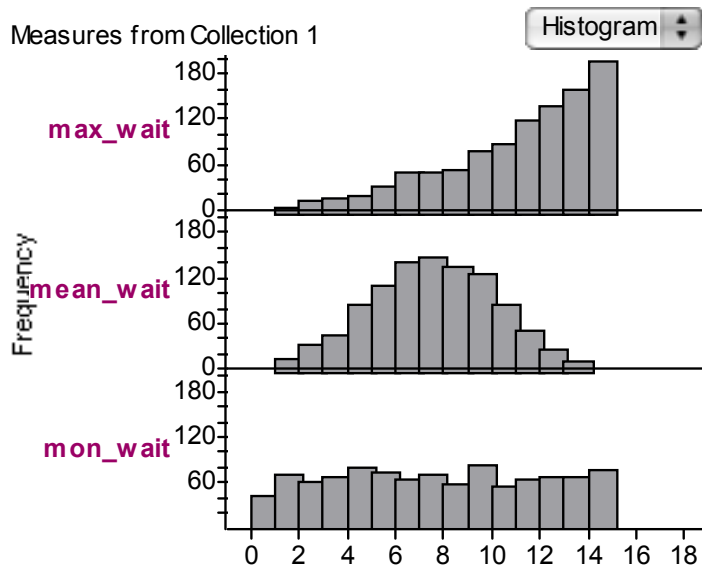
One can approximate the cumulative density function (cdf) of X by computing the probability that X is less than x for several values of x .

This activity is helpful for reinforcing two ideas. The students will see that the shape of the histogram of the random draws is flat or uniform that relates to the uniform shape of the density function of X . By computing the cdf of X at several values, the student will note that the cdf has a linear shape between 0 and 2π .

P8: TECHNOLOGY ACTIVITY: WAITING FOR THE SHUTTLE

In this activity, the student gets some experience with different shapes of density functions. We start with a simple problem – a student waits for a bus for three days and the waiting time X for one day is uniformly distributed on the interval from 0 to 15 minutes. After three days, the student records three random variables: the time that she waited on the first day, the mean waiting time for all three days, and the longest time she had to wait for the three days. Each of these three random variables has a distinctive shape:

1. The time she waits on Monday will be uniformly distributed.
2. The mean time will have a symmetric bell-shape about the average value 7.5 minutes.
3. The longest time will be left-skewed with a maximum value at the right endpoint 15 minutes.



At the end, the student is asked to compute the mean and standard deviation of the simulated Monday waiting times. He and she should be able to compute the exact values of the mean and standard deviation from the definition of the uniform density function and these exact values should be close to the values computed in the simulation.

P8: TECHNOLOGY ACTIVITY: A TEST WITH A BIMODAL DISTRIBUTION

This activity illustrates the use of a “complicated” probability density where the normalizing constant is not obtainable by analytical integration. Although one can’t analytically integrate this density function, it is straightforward to compute the normalizing constant and probabilities by the use of numerical integration.

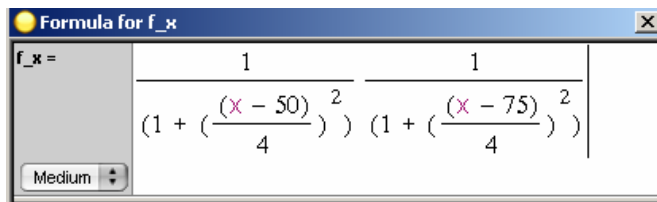
We illustrate the calculations on Fathom, although they can also be done on a calculator. The Fathom document `integrate.ftm` illustrates the work for the particular density function in this activity, although it can be easily modified for a different density function.

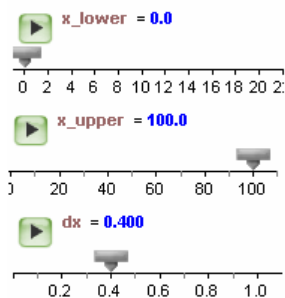
We approximate an integral $\int_{-\infty}^{\infty} f(x)dx$ by the sum

$$\sum_{i=1}^n f(x_i)(\Delta x)$$

where x_1, \dots, x_n is a grid of equally spaced values and Δx is the distance between consecutive values. To obtain an accurate estimate at the integrand, we need to choose the endpoints x_1 and x_n that contain almost all of the probability of the density f , and choose a small value of the distance Δx so that f is evaluated at a large number of values.

In the Fathom document `integrate.ftm`, there is a Collection with two attributes – x and f_x . The attribute x contains the values of the grid and f_x will contain values of the function on the grid. We choose the grid by specifying the endpoints, called x_{lower} and x_{upper} , and the distance dx – all three quantities are chosen by changing the slider values. We specify the function by defining a formula for f_x – the formula for the bimodal density in this activity is shown below. The document also graphs the function f on the grid that is selected.



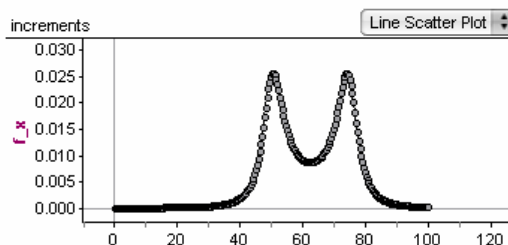


increments

	x	f_x	<new>
1	0.4	1.85241e...	
2	0.8	1.90274e...	
3	1.2	1.95481e...	
4	1.6	2.00869e...	
5	2.0	2.06445e...	
6	2.4	2.12217e...	
7	2.8	2.18194e...	
8	3.2	2.24384e...	
9	3.6	2.30797e...	

increments	
f_x	0.58146678

S1 = sum () * dx



Here we have chosen $x_{lower} = 0$, $x_{upper} = 100$ and $dx = .4$. We are told in this problem that the test grade X falls in the interval $(0, 100)$ and $dx = .4$ is sufficiently small so that the graph is an accurate representation of the function. The summary box contains the value of the integrand – so $\int_{-\infty}^{\infty} f(x)dx = .5815$. We make f a legitimate density function by dividing f by $.5815$:

$$f(x) = \frac{1}{(.5815) \left(1 + \left(\frac{x-50}{4} \right)^2 \right) \left(1 + \left(\frac{x-75}{4} \right)^2 \right)}, \quad 0 < x < 100$$

Probabilities that X falls in particular regions are also integrals computed by numerical integration. The probability a student scores below 60 is given by the integral

$$P(X < 60) = \int_0^{60} \frac{1}{(.5815) \left(1 + \left(\frac{x-50}{4} \right)^2 \right) \left(1 + \left(\frac{x-75}{4} \right)^2 \right)} dx$$

We make several adjustments in the Fathom program to compute this integral:

1. We adjust the definition of the density f by dividing by the normalizing constant .5815.

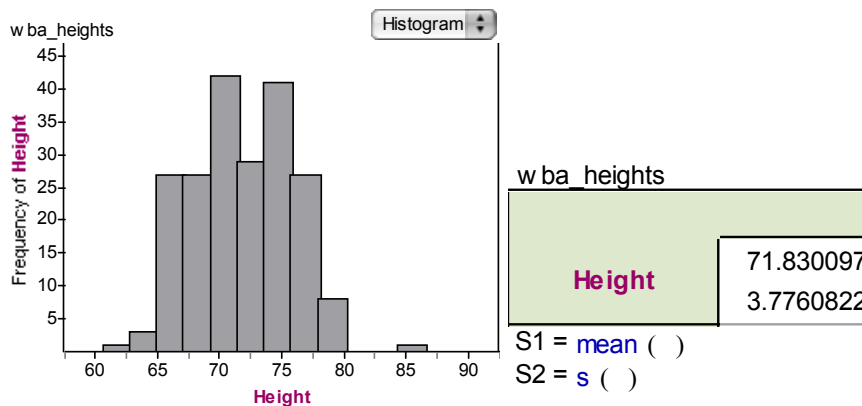
2. We change the endpoints of the interval to 0 and 60.

We find the probability $P(X < 60)$ is given by .473. Other probabilities about X by changing the values of the endpoints.

TOPIC P9: THE NORMAL DISTRIBUTION

P9: TECHNOLOGY ACTIVITY: SAMPLING HEIGHTS

This activity is a good illustration of the Central Limit Theorem. One begins with a population of heights of professional woman basketball players. The student loads this population of heights into Fathom and explores the population, making comments about shape, location, and spread. Note below that the shape of the population is bell-shaped, although there is one non-normal looking outlying height at 85 inches. The mean μ and the standard deviation σ of this population are computed.

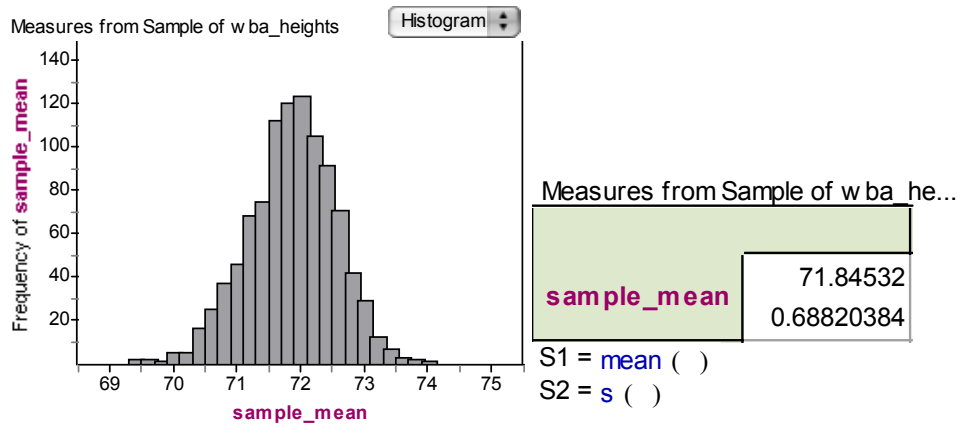


Then one uses Fathom to take random samples of size $n = 25$ without replacement from this population of heights. This is done in two steps:

1. First one takes a sample from the collection of heights. By changing properties of the inspector, one takes a sample without replacement, changes the sample size to 25, and defines a measure equal to the mean of the sampled heights.

2. Next, one collects measures to take repeated samples, collecting the sample means. By changing properties of this collection, one indicates that a total of 1000 samples are to be taken.

Here is a histogram of 1000 sample means from this population. The summary box gives the mean and standard deviation of this sampling distribution.



Here are some things to observe here:

1. The sampling distribution of means is approximately normal shaped.
2. The mean of the sampling distribution (71.84) is approximately equal to the mean of the population (71.83).
3. The standard deviation of the sampling distribution (.688) is much smaller than the standard deviation of the population (3.77).

At the end of this activity, the student is asked to check the Fathom work by use of the Central Limit Theorem. One can compute the exact mean and standard deviation of the distribution of the sample mean. Using these values and the normal-shape assumption, one can find, say the probability the sample mean is larger than 73, and this computed probability should agree with the Fathom calculation.

P9: TECHNOLOGY ACTIVITY: ROLLING BIASED DICE

This activity is a variation of the usual rolling dice where one is interested in the distribution of the sum of the rolls. The Central Limit Theorem states that the

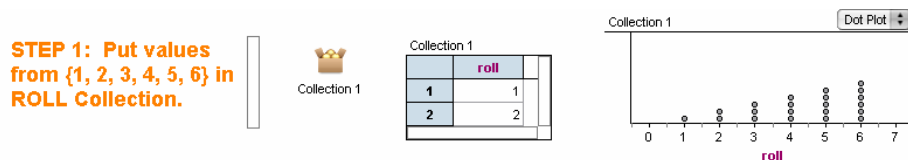
distribution of the sum will be approximately normally distributed for any population with mean μ and standard deviation σ . A student is given the opportunity to construct an arbitrary probability distribution for the roll of a single die. It is quite remarkable that the distribution of the sum of rolls of five of these special dice will look approximately bell-shaped for any choice of the probability distribution.

The first step is for the student to construct his biased dice. This is done by assigning whole-number weights to each number. In the example below, I assign a weight of 1 to the roll 1, a weight of 2 to the roll 2, and so on. My die is loaded to roll high numbers. Once I assign the weights, the probabilities are found by dividing each weight by the sum of the weights.

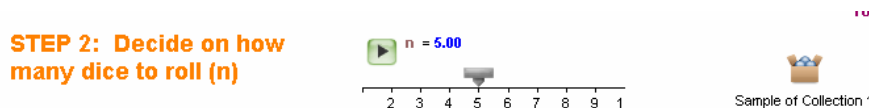
Roll	1	2	3	4	5	6
Weight	1	2	3	4	5	6
Probability	1/21	2/21	3/21	4/21	5/21	6/21

From this distribution, one finds by hand the mean μ and standard deviation σ . Here I find $\mu = 4.33$ and $\sigma = 1.49$.

Then the student uses the Fathom program `rolling_biased_dice.ftm` to roll his/her biased dice. The first step is to put the weights (one 1, two 2's, three 3's, ...) into the Collection.



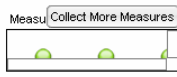
The next step is to decide on how many biased dice to throw. You change the number of dice n by a slider. We start throwing $n = 5$ dice.



Now by pressing the “Collect More Measures” button, we roll five biased dice 500 times, collecting each time S , the sum of the rolls. The Measures collection contains the values of the 500 sums and a histogram of the sums is shown below. Note that even though the

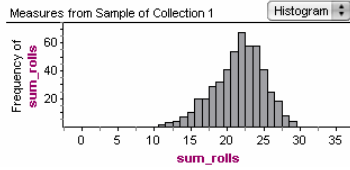
population was very skewed and not normal, the histogram of sums of these dice are approximately bell-shaped.

STEP 3: Roll dice 500 times by Collecting More Measures



Measures from Sample of Colle...

	sum_ro...	<new>
1	23	
2	22	
3	18	
4	22	



Histogram of sum of dice rolls

From the Fathom output, the student is asked to compute various probabilities about the sum of rolls S . Then the student checks the answers by computing the probabilities assuming S has a normal distribution with mean $n\mu$ and standard deviation $n\sigma$. This is good practice in applying the Central Limit Theorem and normal tables and uses the mean and variance calculations that the student did in the first part of the activity.