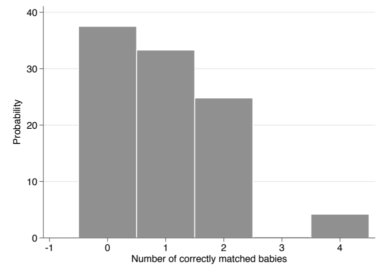
Activity #6: Discrete Random Variables; Expected Value; Variance

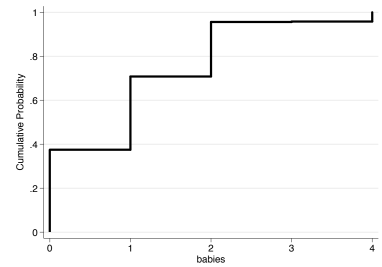
So far, all the probability models we've constructed in class have dealt with **discrete random variables**.
One example was the probability model we constructed for randomly returning babies to their correct mothers

Matched Babies	Probability
0	0.375
1	0.333
2	0.25
3	0
4	0.042

Table



Probability Mass Function



Cumulative Function

When we can begin to count (and list) all the possible outcomes (or values a variable can take), we are dealing with a **discrete random variable**.

Discrete Random Variable: A variable that can take a finite (or countably infinite) number of possible values

Probability models for discrete variables can be displayed via probability mass functions or cumulative distributions.

Probability Mass Function: Displays the probability of each possible outcome, $P(X = x)$
Properties: a) For each outcome, $0 \leq P(X=x) \leq 1$
b) The sum of all probabilities in the sample space is 1.00

Cumulative function: Displays the probability of each outcome and all preceding outcomes, $P(X \leq x)$
Properties: a) It is monotonically increasing
b) It always starts at 0 and ends at 1.0

For other variables, it's impossible to count and list all possible outcomes. These variables are continuous.

Continuous Random Variable: A variable that can take on an infinite number of possible values within a region

We'll display probability models for continuous variables (density functions or cumulative distributions) in the next unit.

Classify each of the following variables as either discrete or continuous:

________ Number of children in a randomly chosen family

________ Age (in years) of a randomly selected SAU student

________ A random person's height (rounded to the nearest inch)

________ Number of grains of sand on the beach

Over the next few weeks, we'll work with discrete (and then continuous) variables in an attempt to:
- Display a probability model (via a graph, table, or formula)
- Calculate the expected value and variance (which we'll learn in this activity)
- Calculate probabilities of interest

Expected value and variance

Plinko debuted in 1983 on *The Price is Right* gameshow. To play Plinko, a contestant climbs up to the top of the Plinko pegboard and drops a chip. The chip then falls down the board and into one of nine spaces at the bottom. Depending on where the chip lands, the contestant can win between $0 and $10,000 with each chip.

From 2000-2011, Plinko was played 308 times with 1,227 chips for winnings totaling $2,214,600 (an average of $1805 per chip).

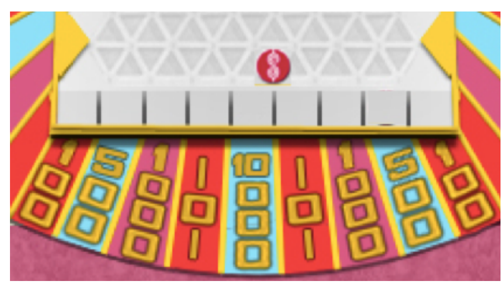
To watch a video clip of Plinko in action, go to:
http://www.youtube.com/watch?v=vuMir11YFPs&t=3m50s

1) It's your lucky day! You've been invited to play Plinko on *The Price is Right*. As the host hands you some Plinko chips, you think to yourself, "Finally, this is my chance to pay off all my student loans!"

   If you want to maximize your chances of winning the most money, where should you initially place each chip? Should you drop it from the middle, the left, the right? Does it matter?

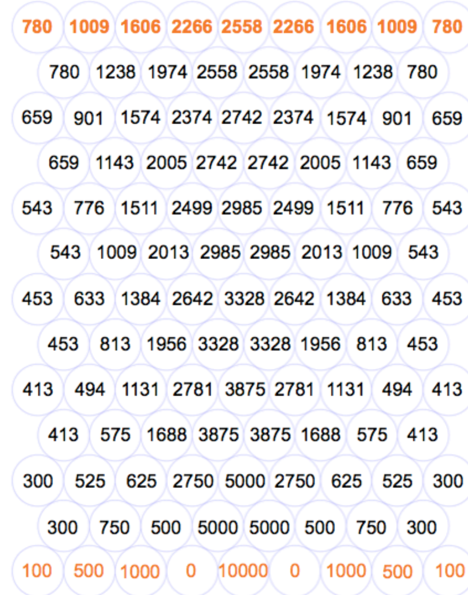
2) We could run a computer simulation to see what happens if we were to drop a Plinko chip from each initial position. We can also try to calculate the best initial position. To do this, let's start at the bottom of the board.

   Suppose I offer to sell you the Plinko chip pictured to the right: Based on the picture, what would be a fair price to pay for this chip? Explain.



3) Using this logic, how could we decide which initial position would give us the greatest pay-out?

4) Extending this process to the entire Plinko board gives us the figure to the right. What do the numbers in the cells represent? Where should you drop your Plinko chip?



Of the 1,227 Plinko chips that were dropped from 2000-2011, the average amount won was \$1,805. By not dropping the chip from the center slot, contestants have (on average) lost \$753 per chip.

Sources: Gameshow Central: http://gscentral.net/pricing.htm
Patterns in Practice: http://patternsinpractice.wordpress.com

5) When we went through the *Lazy Nurse* activity, we found the following probability model:

Number of correctly matched babies:	0	1	2	3	4
Probability of getting this number:	0.375	0.333	0.250	0.000	0.042

If we were going to do this (randomly return 4 babies to 4 mothers) only one time, how many correct matches would you expect?

6) If we were going to do this 10,000 times, how many correct matches do you think we would get on average?

Expected value: The long-run average value that would result from repeatedly running an experiment.

Expected value of a discrete random variable: $E(X) = \sum_{i=1}^{N} x_i P(x_i)$, where x_i represents each possible outcome.

7) What is the expected value of the outcome you would get from rolling a single six-sided die?

8) What is the expected value of X in the lazy nurse activity?

X = # of correctly matched babies:	0	1	2	3	4
Probability of getting this number:	0.375	0.333	0.250	0.000	0.042

9) For 50¢, you can play a lottery where you pick a 3-digit number between 000 and 999. If you pick the winning number, you collect \$275. What is the expected amount you will win or lose in this game?

10) If you're in your early 20s, the probability you live another year is 0.9985 (estimated from the U.S. National Center for Health Statistics). Suppose I offer to sell you a \$100,000 life insurance policy for \$250. If you buy this policy, what's the expected value? How much should I charge for the policy if I want to break even (in the long run)?

11) Bernie Madoff offers you a choice of three investments, each with an expected return of \$50.

Investment Choice A	
Possible Return	Probability
\$45	0.50
\$55	0.50

Investment Choice B	
Possible Return	Probability
\$0	0.50
\$100	0.50

Investment Choice C	
Possible Return	Probability
\$0	0.30
\$50	0.40
\$100	0.30

Which investment would you choose? Why?

The expected value of a random variable tells us the long-run average value that we should expect, but it does not tell us anything about how spread out the values of a variable may be.

Variance: The expected value of the squared deviation of a variable from its expected value. (Huh?)
A measure of how spread out a distribution is.

Variance of a discrete random variable: $\mathrm{Var}(X) = E\left(\left[X - E(X)\right]^2\right) = E\left(X^2\right) - \left[E(X)\right]^2$

12) Explain what the formula for variance represents? What is the variance of a distribution?

13) Calculate the variance of the outcome of rolling one 6-sided die. Recall the expected value was 3.5.

14) Calculate the variance of each investment option (from question #11).

15) Suppose I tell you the expected value of your score on the first exam is 45 and the variance is 16. Since I want to appear to be an effective teacher, I've decided that I'm going to go back and double everyone's test score. What would happen to the expected value and variance if I calculated it for these new scores?

E(X) = 45 E(2x) = ____________________

Var(X) = 16 Var(2x) = ____________________

16) Let's see what happens to the expected value if we transform the values of X. In this example, let's suppose we multiply every score by **a** and add **b**.

Original Data

Values of X	Probability
X_1	$P(X_1)$
X_2	$P(X_2)$

Transformed Data

Values of X	Probability
$aX_1 + b$	$P(X_1)$
$X_2 + b$	$P(X_2)$

Let's first substitute (ax+b) for each value of X in the formula:

$$E(X) = x_1 P(x_1) + x_2 P(x_2)$$

The probabilities don't change when we transform a variable:

$$E(aX+b) = (ax_1+b)P(ax_1+b) + (ax_2+b)P(ax_2+b)$$

We can then distribute, collect like terms, and factor out terms:

$$E(aX+b) = (ax_1+b)P(x_1) + (ax_2+b)P(x_2)$$
$$E(aX+b) = ax_1P(x_1) + bP(x_1) + ax_2P(x_2) + bP(x_2)$$
$$E(aX+b) = a\left[\underline{x_1P(x_1) + x_2P(x_2)}\right] + b\left[\underline{P(x_1) + P(x_2)}\right]$$

Look at the two underlined parts of the formula. What can we substitute? Why?

17) What happens to the variance if we transform the values of X (by multiplying every value by **a** and add **b**)?

We start with the formula for variance and substitute ax+b for each value of x:

$$Var(X) = E(X^2) - [E(X)]^2$$
$$Var(aX+b) = E((aX+b)^2) - [E(aX+b)]^2$$
$$= E(a^2X^2 + 2abX + b^2) - [aE(X)+b]^2$$
$$= E(a^2X^2) + E(2abX) + E(b^2) - \left[a^2[E(X)]^2 + 2abE(X) + b^2\right]$$
$$= a^2E(X^2) + 2abE(X) + b^2 - a^2[E(X)]^2 - 2abE(X) + b^2$$
$$= a^2E(X^2) - a^2[E(X)]^2$$
$$= a^2\left[E(X^2) - [E(X)]^2\right]$$
$$=$$

We just derived some important properties of expected values and variances: $E(aX+b) = aE(X)+b$

$$\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$$

18) Recall that Bernie Madoff offered us an investment (Choice C) with an expected value of $50 and a variance of $25

Investment Choice A	
Possible Return	Probability
$45	0.50
$55	0.50

Bernie Madoff really wants your money, so he offers to triple your return and give you an extra $10. Calculate the expected value and variance for this new deal.

19) Suppose we have two independent variables, X and Y. We also know the following:

E(X) = 25 Var(X) = 4
E(Y) = 13 Var(Y) = 3

Calculate the following:

E(X + Y) =

E(X – Y) =

Var(X + Y) =

Var(X – Y) =