MTH245   Unit 4 Module 3   Binomial Distribution

Let's look at a specific type of probability distribution called a binomial distribution. "Bi" means two, so this distribution occurs when we have two possible outcomes: success and failure; heads and tails; male and female, etc.

If we flip a fair coin once we have two outcomes: heads or tails. If we are interested in the number of heads our sample space would be 0, 1. The probability distribution for the number of heads would be:

x	P(x)
0	.5
1	.5

If we flip a coin twice and count the number of heads our sample space would be 0, 1, and 2 (for the events TT, TH, HT, and HH) and the probability distribution would be:
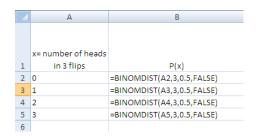
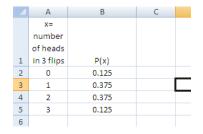| x	P(x)
0	.25
1	.5
2	.25

If we were to flip TEN coins it starts to get really complicated. We could set up an Excel page and model tossing ten coins, but there is a way to do this theoretically in Excel also.

If this were a statistics class we would be using a formula: $P(x) = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$. We don't need to go there. Instead we will use Excel, as this formula is already programmed in. There are several parameters we will need to enter. Remember, this function is used when we have an experiment with two possible outcomes, but also with the probability of success being "p", and what we want to find is how likely it is to have "x" successes when the experiment is repeated "n" times.

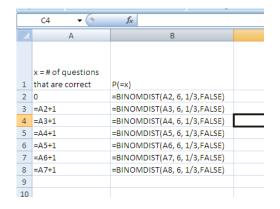Format of the Binomial Distribution in Excel is =BINOMDIST(#of successes, # of trials, Prob of success in one trial, False). # of success is how many successes in a particular trial (for example, 0 heads for our 3 coins), # of trials is the number of trials or times the experiment is repeated…the number of coins, the number of questions, the number of passengers, etc.. (We will come back to what False is all about later.)

Let's start with modeling flipping three fair coins and counting the number of heads. The probability of heads is .5 on any one flip. We are interested in the number of successes being all of the possible outcomes so we will first find the probability of 0 heads in 3 flips, then 1 head in three flips, then 2 out of 3 and last 3 out of 3. We can arrange this in a table and create a probability distribution for the sample space:

	A	B
1	x= number of heads in 3 flips	P(x)
2	0	=BINOMDIST(A2,3,0.5,FALSE)
3	1	=BINOMDIST(A3,3,0.5,FALSE)
4	2	=BINOMDIST(A4,3,0.5,FALSE)
5	3	=BINOMDIST(A5,3,0.5,FALSE)
6		

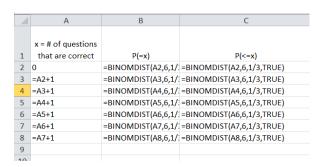|	A	B	C
1	x= number of heads in 3 flips	P(x)	
2	0	0.125	
3	1	0.375	
4	2	0.375	
5	3	0.125	
6			

Let's do another one and see what the false business was about. We are going to take a test with 6 questions. We have not studied at all but fortunately the test is multiple choice, with only 3 options for each question. How likely are we to guess our way to a passing grade? Our possible outcomes for this experiment are to get 0, 1, 2, 3, 4, 5, or 6 answers correct. The probability of getting any one question correct is 1/3. There are a total of 6 questions we are going to guess at.
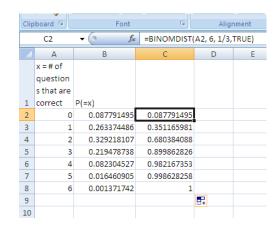
C4		fx	
	A	B	
1	x = # of questions that are correct	P(=x)	
2	0	=BINOMDIST(A2, 6, 1/3,FALSE)	
3	=A2+1	=BINOMDIST(A3, 6, 1/3,FALSE)	
4	=A3+1	=BINOMDIST(A4, 6, 1/3,FALSE)	
5	=A4+1	=BINOMDIST(A5, 6, 1/3,FALSE)	
6	=A5+1	=BINOMDIST(A6, 6, 1/3,FALSE)	
7	=A6+1	=BINOMDIST(A7, 6, 1/3,FALSE)	
8	=A7+1	=BINOMDIST(A8, 6, 1/3,FALSE)	
9			
10			

	A	B	C	D	E
1	x = # of questions that are correct	P(=x)			
2	0	0.087791			
3	1	0.263374			
4	2	0.329218			
5	3	0.219479			
6	4	0.082305			
7	5	0.016461			
8	6	0.001372			
9					
10					
11					

So what does this table tell us? The first row tells us that the probability of getting exactly 0 of the 6 questions correct is 0.0878, or about 8.78%. A couple rows down and we see that the probability of getting exactly 2 of the 6 questions correct is about 32.9%. What would be the probability of getting less than 5 questions correct? This would be the P(0 correct)+P(1 correct)+P(2 correct)+P(3 correct)+P(4 correct). We can add those all up, but a useful addition to our table would be a third column with the **cumulative** probabilities. We can add that column by copying over our binomdist column and changing the "false" to "true".

Now take a look at the x = 4 row. The probability of getting exactly 4 answers correct is 8.23%, The probability of getting 4 or fewer answers correct is 98.22%. So the probability of getting less than 5 questions correct is also 98.22. How about the probability of getting 5 or more answers correct? This would be a complementary event, P(x>=5) = 1 - .9822 = 0.0178, or 1.78% chance of doing really well on this test.

	A	B	C
1	x = # of questions that are correct	P(=x)	P(<=x)
2	0	=BINOMDIST(A2,6,1/:	=BINOMDIST(A2,6,1/3,TRUE)
3	=A2+1	=BINOMDIST(A3,6,1/:	=BINOMDIST(A3,6,1/3,TRUE)
4	=A3+1	=BINOMDIST(A4,6,1/:	=BINOMDIST(A4,6,1/3,TRUE)
5	=A4+1	=BINOMDIST(A5,6,1/:	=BINOMDIST(A5,6,1/3,TRUE)
6	=A5+1	=BINOMDIST(A6,6,1/:	=BINOMDIST(A6,6,1/3,TRUE)
7	=A6+1	=BINOMDIST(A7,6,1/:	=BINOMDIST(A7,6,1/3,TRUE)
8	=A7+1	=BINOMDIST(A8,6,1/:	=BINOMDIST(A8,6,1/3,TRUE)
9			
10			

Clipboard   Font   Alignment

C2   fx =BINOMDIST(A2, 6, 1/3,TRUE)

	A	B	C	D	E
1	x = # of questions that are correct	P(=x)			
2	0	0.087791495	0.087791495		
3	1	0.263374486	0.351165981		
4	2	0.329218107	0.680384088		
5	3	0.219478738	0.899862826		
6	4	0.082304527	0.982167353		
7	5	0.016460905	0.998628258		
8	6	0.001371742	1		
9					
10					

One more thought. When we set up a simulation to, say, toss ten coins and count the number of heads observed a large number of times and use those results to estimate the probability distribution of observing "x" heads when 10 coins are tossed, that would be an empirical distribution table, based on a large sample. And each of us running the simulation will get slightly different results. When the binomial function is used to build a probability distribution for observing "x" heads in ten tosses of a coin, that is a theoretical distribution table, and we should all be getting exactly the same result.

x = # of heads in 10 coins	P(=x)	P(<=x)
0	0.000977	0.000977
1	0.009766	0.009766
2	0.043945	0.043945
3	0.117188	0.117188
4	0.205078	0.205078
5	0.246094	0.246094
6	0.205078	0.205078
7	0.117188	0.117188
8	0.043945	0.043945
9	0.009766	0.009766
10	0.000977	0.000977