http://wiki.stat.ucla.edu/socr/index.php/SOCR_Courses_2008_Thomson_ECON261

Discrete Probability Distributions and application in Business

By Grace Thomson

DISCRETE PROBALITY DISTRIBUTIONS

In this section we will learn about some types of probability distributions studied in Statistics: Discrete Probability Distributions: Binomial, Poisson, Hypergeometric.

Discrete Probability Distributions:

There are 3 important discrete probability distributions: Binomial, Poisson and Hypergeometric. The formulas might look difficult but its computation is simple if we use Excel.

Table 1

Summary of Discrete Probability Distributions

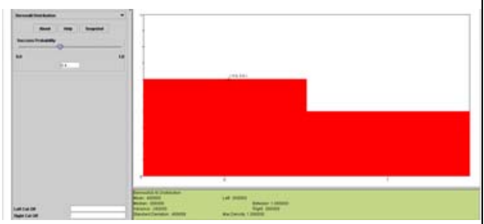
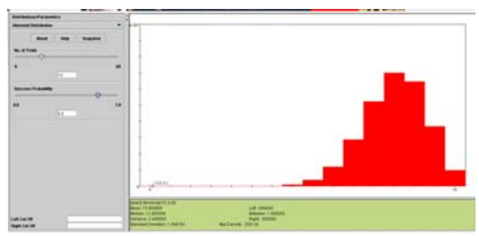
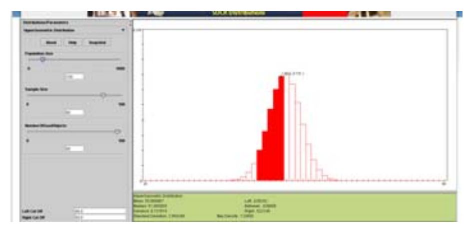
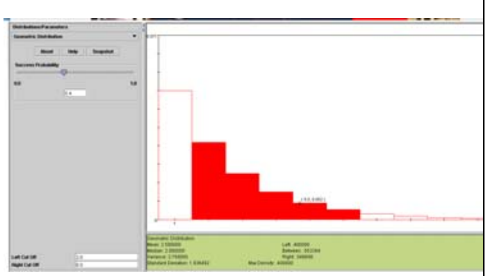
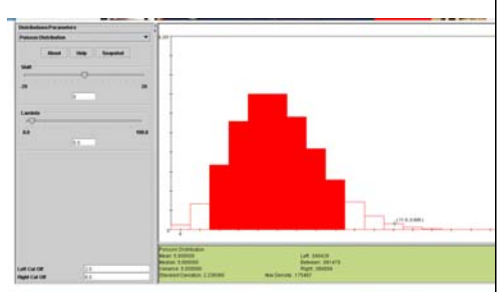
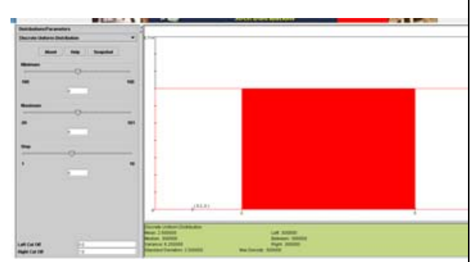
Binomial	Poisson	Hypergeometric
$P_x^n = C_x^n p^x q^{(n-x)}$ *Combination of **x** objects selected (successes) from a sample of size **n**, times probability of success **(p)** times probability of failure **(q)**.*	$P(x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}$ λt = average # of successes in a period of time t= period of time x = # of successes on a segment e= 2.71828 base of ln	$P(x) = \frac{C_{n-x}^{N-X} * C_x^X}{C_n^N}$ Measures the # of ways to select an event of interest x out of subgroup X, multiplied by the # of ways to select (n-x) failures out of the remaining (N-X) population.

Binomial	Poisson	Hypergeometric
Only 2 possible outcomes: success (p) /failure (q) *Defective / acceptable* *Get / not get a contract* *Yes, I will buy/ No, I won't buy* *Accept /reject an applicant*	**When success and failures are not easily counted** How many customers won't come How many cars won't be purchased today How many calls won't enter a PBX **Describes a process that extends over time or space**	**When there are more than 2 not independent events to be counted.** **Used when there are small populations and large samples** 10 articles: 3 brand A, 4 brand B, 3 brand C. Notice 3, 4 and 3 account for 30% and 40% respectively. **Compares failure vs. success.**

Simple rule: if n= 5%N, use Binomial Distribution to ensure independence and constant **p**	If **lambda** exists and no probability of success is available, Poisson must be used.	It's used any time that you have sets and subsets of events.

All of these distributions are available via SOCR Distributions (http://socr.ucla.edu/htmls/SOCR_Distributions.html). Notice that:

1. **You can control with the mouse the areas to compute the probabilities of interest (below, between or above).**
2. **You can also see the mean and Std Deviation for all of these distributions.**

Bernoulli 	**Binomial** 	**Hyper-Geometric** 
Geometric 	**Poisson** 	**Discrete Uniform** 

Steps to compute binomial, poisson and hypergeometric with excel

Binomial	Poisson	Hypergeometric
Steps to compute Binomial distributions using Excel =binomdist(x,n,p,cumulative)	**Steps to compute Poisson distributions using Excel** =poisson(x,lambda,cumulative)	**Steps to compute Hypergeometric distributions using Excel** =hypgeomdist(x,n,X,N)

x: Number_s (# of successes x) n: Trials (sample size n) p: Probability (p given in the exercise) Cumulative: type "false" Click oK	x: variable of interest lambda: Mean (lambda λ) cumulative: type "False"	=hypergeom(success in sample x, number in sample n, success in population X, number in population N)
Mean of a Binomial Distribution Σx = E(x) = np **Standard Deviation of a Binomial Distribution** $\sigma_x = \sqrt{npq}$	**Mean of a Poisson Distribution** Ex = λt **Standard Deviation of a Poisson Distribution** $\sigma_x = \sqrt{\lambda t}$	

Applications of Binomial Distribution to Business problems

We have 30 applicants for a loan. We know that 25% of the businesses will file for bankruptcy in 5 years. What is the probability that 4 of our applicants file for bankruptcy in this period?

1. We identify the problem: Probability of 4 applicants filing for bankruptcy

2. We identify the elements of the problem:

Probability of bankruptcy: p= 0.25

Probability of no bankruptcy: p=0..75

Number of observations: 30

Event of interest: Probability that 4 applicants file for bankruptcy.

3. We identify the type of probability distribution that the variables would follow:

This is clearly a binomial distribution as there are only 2 possible outcomes and there is a known probability.

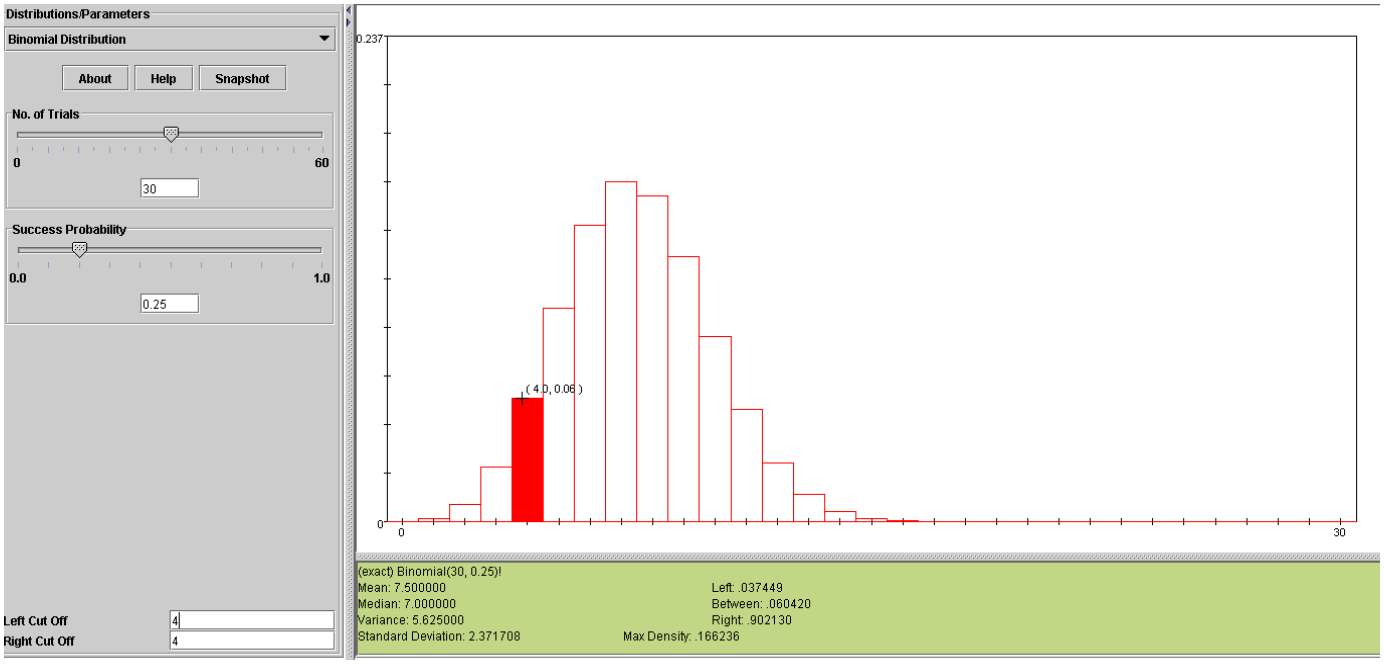
4. We compute the probability distribution:

Compute P(x= r) r = 4, where r is # of applicants.

Use formula for binomial distributions: $p(x=r)=\frac{n!}{(n-r)!r!}p^r(1-p)^{n-r}$

Use **SOCR Distributions** (http://socr.ucla.edu/htmls/SOCR_Distributions.html) to compute the answer:

Bring the Mouse over X=4 to see the answer P (X=4) = 0.06. Or use the "Left" and "Right" Cut-Off values to get the result in the bottom text area:



Or use EXCEL

=BINOMDIST(r,n,p,cumulative)

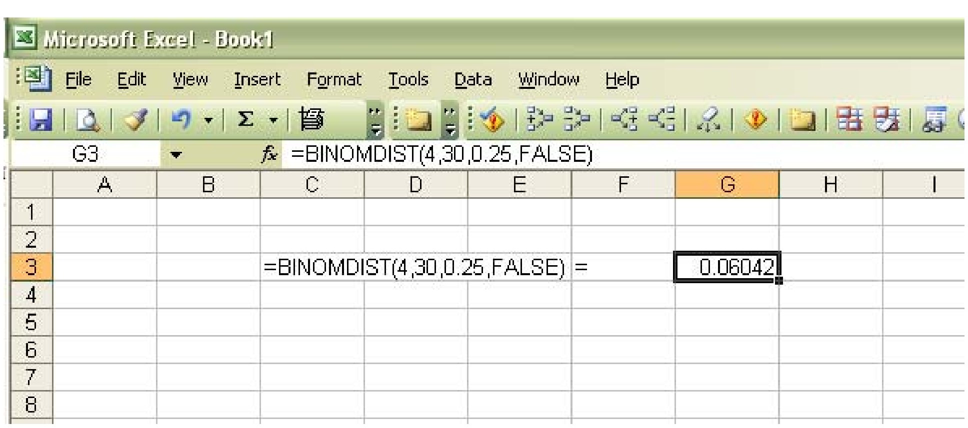
Where:

r = number of successes

n= # of trials or # of observations or events

p= known probability of success

cumulative= “false” means individual distribution, not cumulative.



Now, if you want to build a distribution table, with all the possible occurrences of bankruptcy for your 30 applicants, you can do it easily. From 0 applicants in bankruptcy to 30 applicants in bankruptcy.

Format Tools Data Window Help

=BINOMDIST(D7,30,0.25,FALSE)

Probability distribution table for bankruptcy in a group of 30 applicants

x	Formula	P(x)
0	=BINOMDIST(d7,30,0.25,FALSE)	0.0002
1	=BINOMDIST(d8,30,0.25,FALSE)	0.0018
2		0.0086
3		0.0269
4		0.0604
5		0.1047
6		0.1455
7		0.1662
8		0.1593
9		0.1298
10		0.0909
11		0.0551
12		0.0291
13		0.0134
14		0.0054
15		0.0019
16		0.0006
17		0.0002
18		0.0000
19		0.0000
20		0.0000
21		0.0000
22		0.0000
23		0.0000
24		0.0000
25		0.0000
26		0.0000
27		0.0000
28		0.0000
29		0.0000
30		0.0000

Notice how the sum of the probabilities will be exactly 1.00, since we have included all the possible number of applicants. Notice also, that in the formula I have left the cell for “x” as a formula, so you can easily copy it by “pulling and dragging down”

You can answer all the questions regarding these applicants, such as:

What is the probability that at least 5 applicants filed for bankruptcy?

$P(x \geq 5) = 1 - P(x<5) = 1 - [P(x=0) + P(x=1) + P(x=2) + P(x=3) + P(x=4)$

$P(x \geq 5) = 1 - 0.0002 – 0.0018 - 0.0086 - 0.0269 - 0.0604 = 0.9021$

LEARNING TEAM ACTIVITY

What is the probability that 10 or more applicants but less than 20 applicants, filed for bankruptcy?

R: $P(20 > x \geq 10) = 0.1966$

Expected value in a binomial distribution

In binomial distributions, the expected value is $E(x) = np$ and the Standard deviation is

$$\sigma = n\,p(1-p).$$

What is the expected number of borrowers that could file for bankruptcy in our example? $E(x) = np$

$$E(x) = 30\,(0.25) = 7.50 \rightarrow 8 \text{ applicants}$$

And what is the standard deviation of our applicants?

$$\sigma = n\,p(1-p).$$

R: The expected value times the probability of failure.

$$\sigma = (30)\,(0.25)\,(0.75) = 5.625$$

Application of Poisson distribution to Business problems

Verizon found out that during peak hours the number of calls per minute in each one of their towers was 10 calls. They know that once a call comes in the follow a Poisson pattern. Using this average, Verizon wants to compute the probability of having more than 15 calls in a given minute. If they have more than 15 calls they will expand their towers.

1. What is the problem?

Probability of having more than 15 calls in a minute given that they have a Poisson pattern

2. Elements of the problem:

Number of successes: 15/minute

Average (lambda): 10/minute

Probability: not given (not needed)

Number of total calls: unknown, they could be endless, we don't know them.

3. We can compute Poisson using the formula

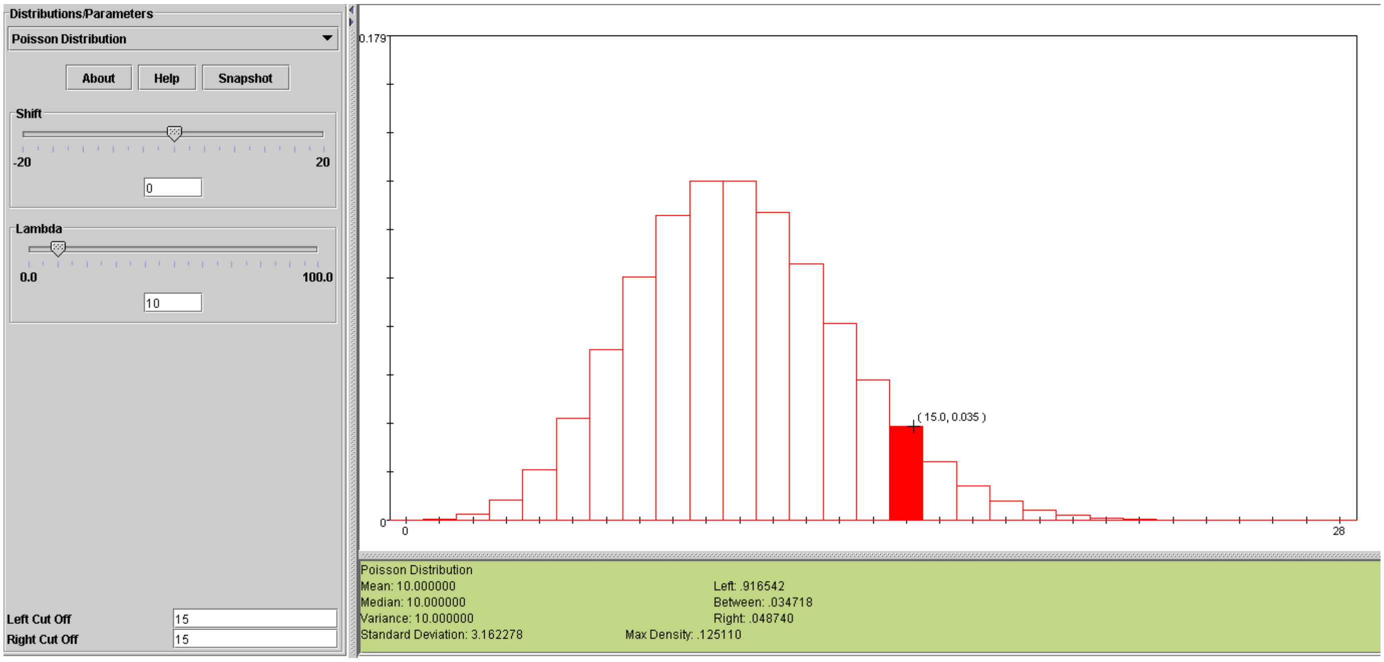
Poisson
$P(x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}$ λt = average # of successes in a period of time t= period of time x = # of successes on a segment e= 2.71828 base of ln

Use SOCR Distributions (http://socr.ucla.edu/htmls/SOCR_Distributions.html) to solve the problem:

Use the "Left Cut-off=15" and "Right Cut-Off=15": to get the answer:

P(X=15) = 0.0347 and

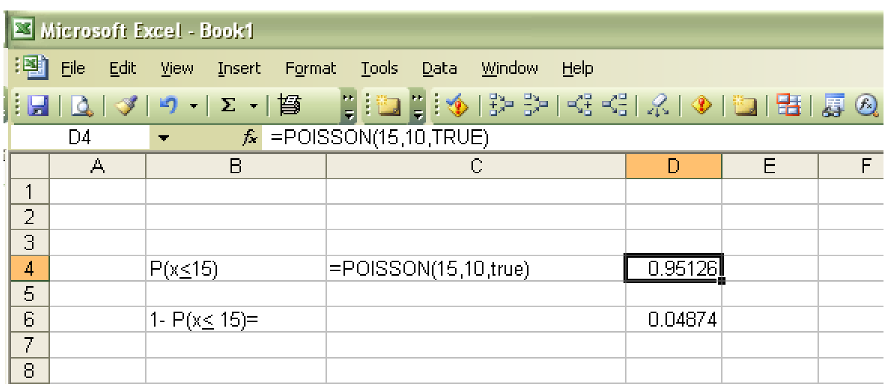
P(X>15) = 0.04874.



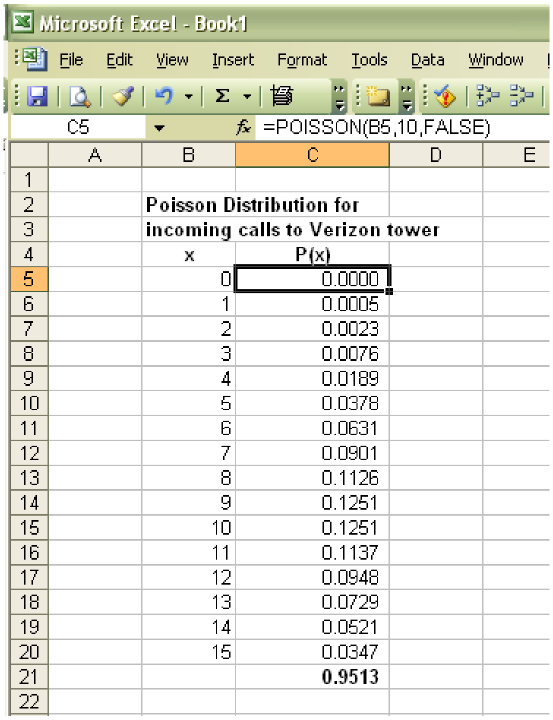
Or using excel =poisson(x,lambda,cumulative)

Let's operate, The probability that more than 15 calls occur in a given minute is:

$$P(x > 15) = 1 - P(x \leq 15) = 0.04874$$



If you want to build a distribution table for all the possible results, you can do it easily by repeating the formula in each cell. Let's prepare the Probability distribution table from 0 to 15 calls.



	A	B	C	D	E
1					
2		Poisson Distribution for			
3		incoming calls to Verizon tower			
4		x	P(x)		
5		0	0.0000		
6		1	0.0005		
7		2	0.0023		
8		3	0.0076		
9		4	0.0189		
10		5	0.0378		
11		6	0.0631		
12		7	0.0901		
13		8	0.1126		
14		9	0.1251		
15		10	0.1251		
16		11	0.1137		
17		12	0.0948		
18		13	0.0729		
19		14	0.0521		
20		15	0.0347		
21			**0.9513**		
22					

Notice how the sum of these probabilities doesn't reach 1.00 yet, as there might be endless number of calls that the tower registers per minute, not only 15. Notice also, that in the formula I have left the cell for "x" as a formula, so you can easily copy it by "pulling and dragging down"

Using this table, you can answer all the questions regarding the distribution of the incoming calls to the Verizon tower, let's see:

LEARNING TEAM ACTIVITY

What is the probability to receive no more than 8 calls in a given minute?

R: $P(x \leq 8) = 0.3328$

Application of Hypergeometric distribution to Business

Company A has 15 executive positions, 10 are male and 5 are female. The company is having a problem with the employees and they need a high-power committee to deal with employee grievance. The committee requires 6 members selected at random (with the same chance to be selected).

What is the probability that 4 members in the committee are females?

1. What is the problem

Probability to select 4 females for the committee

2. Elements of the problem

Total number of executives= 15

Number of members to serve in the committee= 6

Probability identified for this problem= none

Number or male executives= 10

Number of female executives= 5

3. What type of probability distribution is this?

When there are samples and sub-samples for which to compute combinations, we are in front of a hypergeometric distribution.

4. Compute the probability using the formula:

$$P(x) = \frac{C_{n-x}^{N-X} * C_x^X}{C_n^N}$$

Measures the # of ways to select an event of interest x out of subgroup X, multiplied by the # of ways to select (n-x) failures out of the remaining (N-X) population. In our case:

N = 15 number in population N

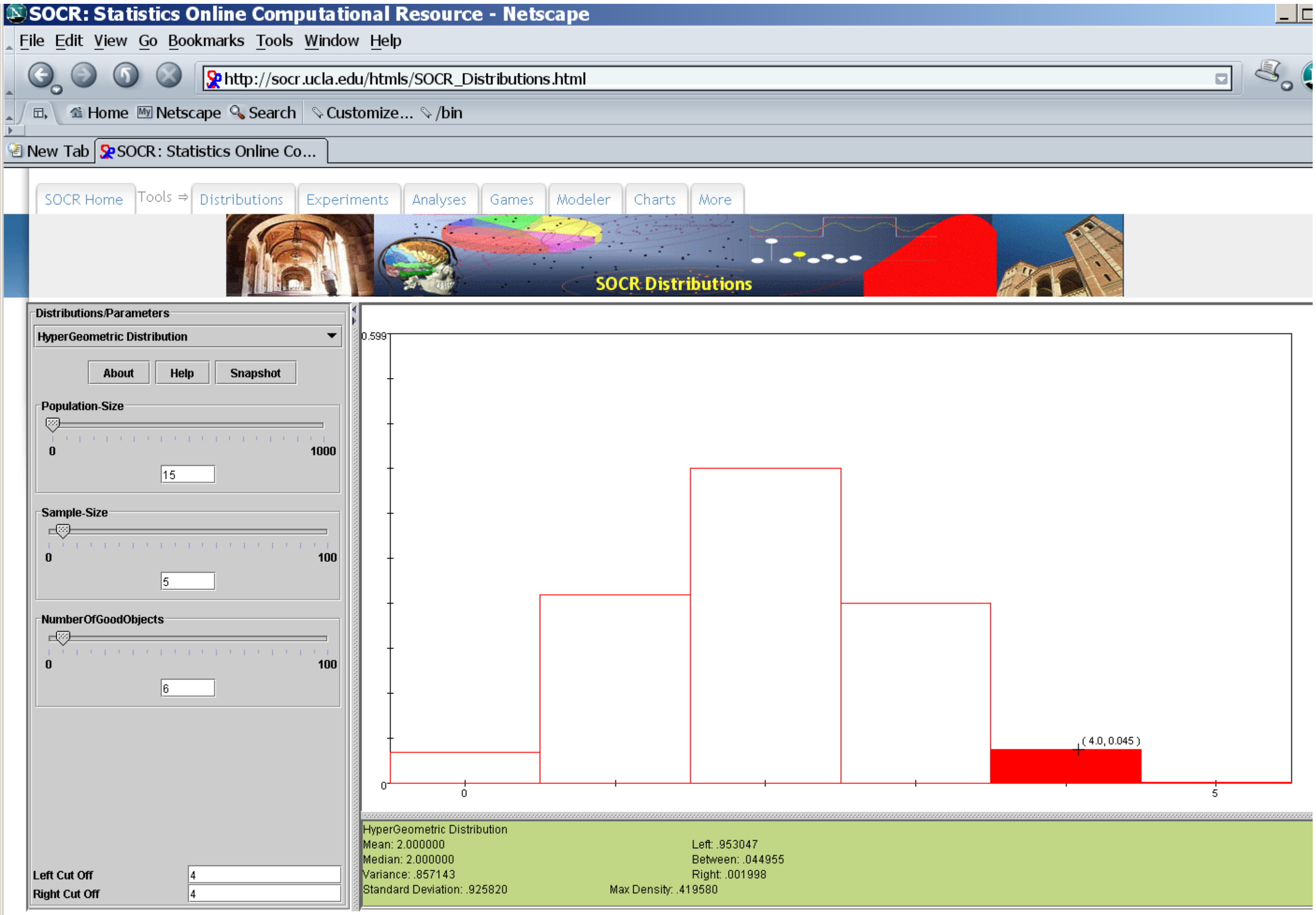
X = 6 success in population X

n = 5 number in sample n

x = 4 success in sample x

Using SOCR for Hypergeometric:

P(X=4) = 0.045



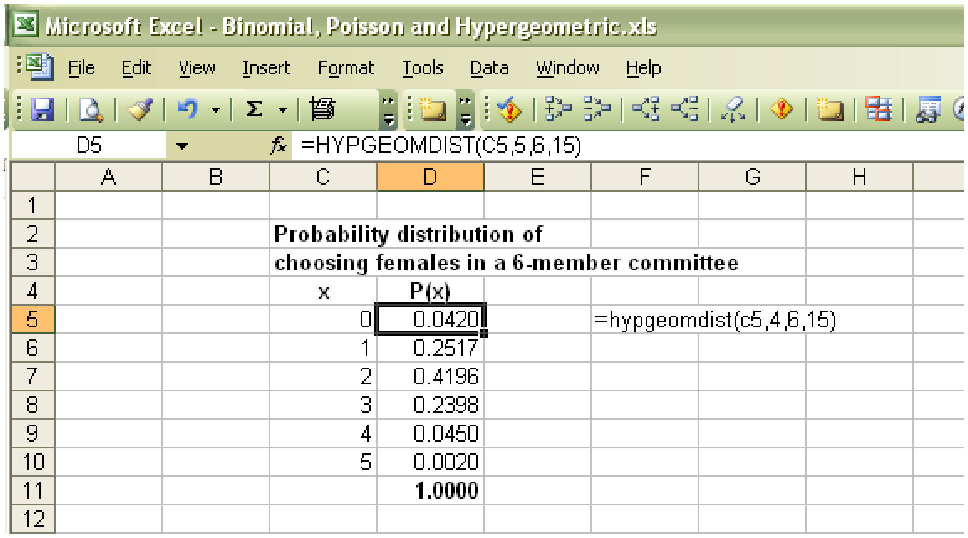
We can use excel formula: =hypergeom(x,n,X,N)

=hypergeom(success in sample x, number in sample n, success in population X, number in population N)

=HYPGEOMDIST(4,5,6,15) = 0.0450

And naturally, we can build a probability distribution table for all the possible number of females in the 6-member committee. We build the table from 0 female to 5 females and then compute the hypergeometric probability for each row, take a look.

Notice how the sum of the probabilities is exactly 1.00, since we have included all the possible number of females. Notice also, that in the formula I have left the cell for “x” as a formula, so you can easily copy it by



“pulling and dragging down”

Another important note to this table is that this only portrays the information for females. Should you need the distribution table for males, you need build it with a slight change in the formula, as the number of males in the company is 10. Let’s say that you want to know about the probability of having 4 males in the committee?

=hypgeomdist(4,10,6,15)= 0.41958

If you are a good observer, you will notice that the probability of having 4 males in the committee MUST BE exactly the same as the probability of having 2 females in the committee, since the committee only has 6 members.

LEARNING TEAM ACTIVITY

What is the probability to have 3 females and 3 males in the committee?

R: P(x_1= 3) = 0.2398

MORE SOCR Exercises and help is available here:

http://wiki.stat.ucla.edu/socr/index.php/SOCR_EduMaterials_Activities_Binomial_Distributions