Statistical Analysis

Johns Hopkins
Carey Business School

Discrete Random Variables

1. Two Types of Random Variables
2. Discrete Probability Distributions
3. The Binomial Distribution
4. The Poisson Distribution
5. The Hypergeometric Distribution

Two Types of Random Variables

Random variable
A variable whose outcome occurs by chance is called random variable. It is the result of a random event

- Discrete
- Continuous

Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

Discrete Random Variable

Discrete random variable

- Possible values can be counted or listed
- Whole numbers, not fractions
- *Examples*
  - The number of defective units in a batch of 20
  - A listener rating (on a scale of 1 to 5) in an AccuRating music survey
  - The number of students attending a lecture

Continuous Random Variable

Continuous random variable
May assume any numerical value in one or more intervals

- Variable can take on an infinite number of values
- The values do not have to be whole number, but include all fractions.
- *Example*
  - the precise weight of a box of candy can be 15.845 oz or 4 oz or .99999 oz

Discrete Probability Distributions

The **probability distribution** of a discrete random variable is a table, graph or formula that gives the probability associated with each possible value that the variable can assume

Notation: Denote the values of the random variable by x and the value's associated probability by p(x)

Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

Discrete Probability Distribution Properties

1. For any value x of the random variable, $p(x) \geq 0$
2. The probabilities of all the events in the sample space must sum to 1, that is…

$$\sum_{\text{all } x} p(x) = 1$$

Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

Expected Value of a Discrete Random Variable

The mean or expected value of a discrete random variable X is:

$$EV = \mu_X = \sum_{All\,x} x\,p(x)$$

μ is the value expected to occur in the long run and on average

Adapted from Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

Variance

The variance is the average of the squared deviations of the different values of the random variable from the expected value

The variance of a discrete random variable is:

$$\sigma_X^2 = \sum_{All\,x} (x - \mu_X)^2 p(x)$$

Standard Deviation

The standard deviation is the square root of the variance

$$\sigma_X = \sqrt{\sigma_X^2}$$

The variance and standard deviation measure the spread of the values of the random variable from their expected value

Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

Ferrari Car Dealership

Data are collected for 50 weeks

# of cars sold	# of weeks	%
0	5	.10 (5/50)
1	6	.12 (6/50)
2	18	.36 (18/50)
3	13	.26 (13/50)
4	7	.14 (7/50)
5	1	.02 (1/50)
Total	**50**	**1.00**

Expected Value

Expected value is the mean for a probability distribution

$$EV = (.10)(.0) + (.12)(1) + (.36)(2) + (.26)(3) + (.14)(4) + (.02)(5)$$

$$EV = 2.28 \text{ cars / week}$$

Interpretation: on average the dealership sells 2.28 cars per week

# of cars sold	# of weeks	%
0	5	.10
1	6	.12
2	18	.36
3	13	.26
4	7	.14
5	1	.02
Total	**50**	**1.00**

Standard Deviation

To calculate the dispersion

$$\sigma = \sqrt{(0-2.28)^2(.10)+(1-2.28)^2(.12)+\ldots+(5-2.28)^2(.02)}$$

$$\sigma = 1.6$$

Interpretation

68.4% of the time the car dealership sells 2.28 ± 1.6 cars or between .68 and 3.88 cars.

Binomial Distribution Characteristics

The **binomial experiment** characteristics:

1. There must be only two possible outcomes, such as “heads or tails”, “success” or “failure”
2. The probability remains the constant, such as with tossing coin, always 50%
3. Trials are independent
4. The experiment can be repeated many times

If x is the total number of successes in n trials of a binomial experiment, then x is a **binomial random variable**

Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

Binomial Formula

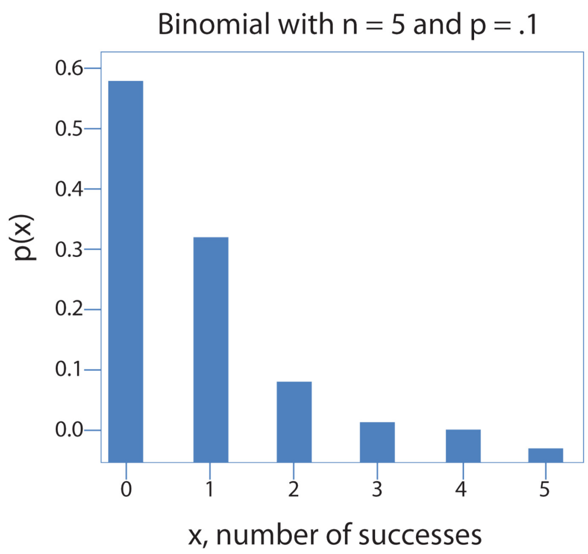
For a binomial random variable x, the probability of x successes in n trials is given by the binomial distribution:

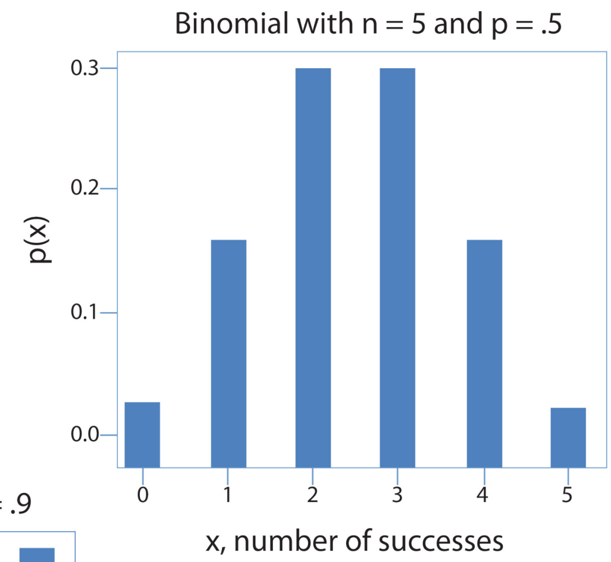
$$p(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

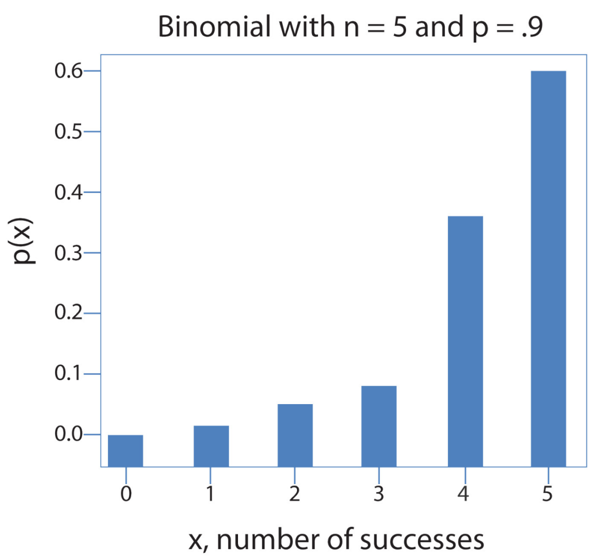
- $n!$ is read as "n factorial" and $n! = n \times (n-1) \times (\mathrm{n}-2) \times \ldots \times 1$
- $0! = 1$
- Not defined for negative numbers or fractions

Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

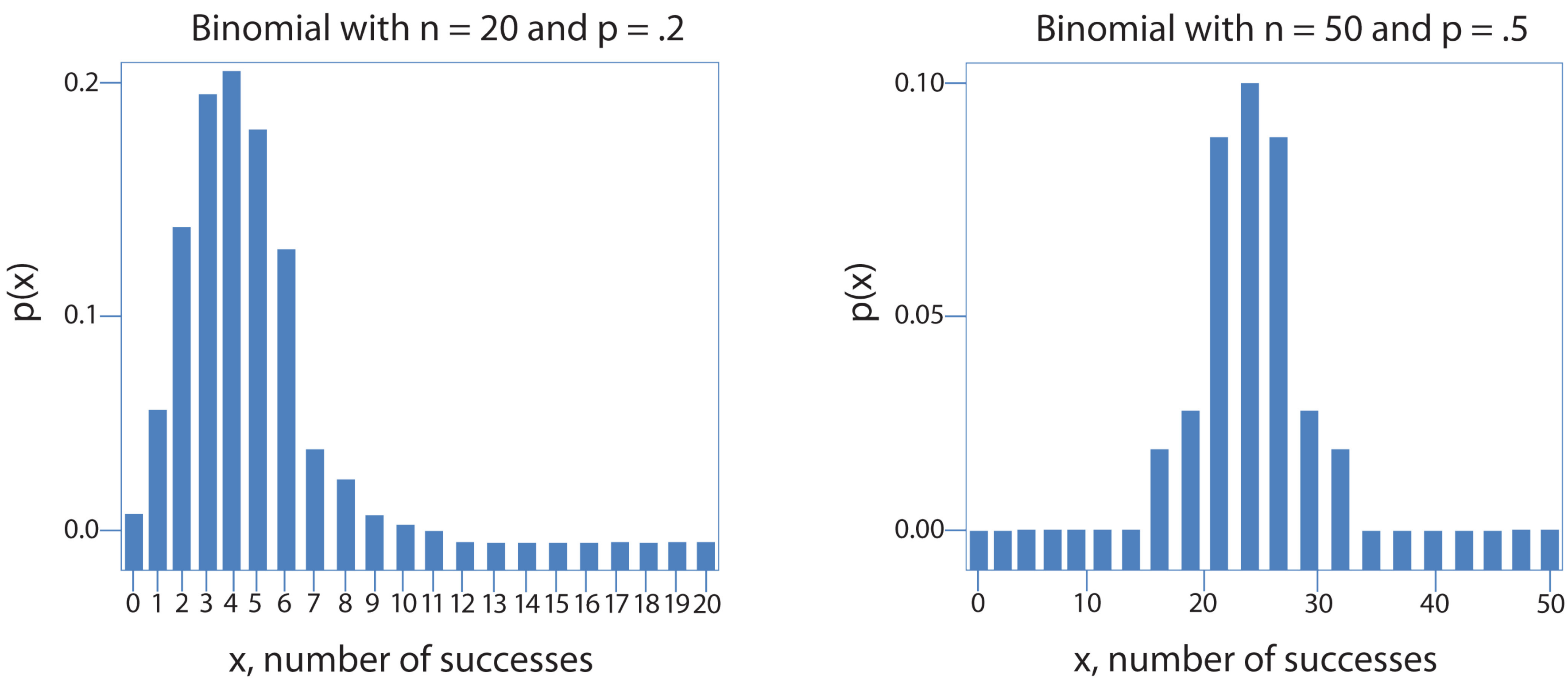
Several Binomial Distributions







Several Binomial Distributions *continued*



Mean and Variance of a Binomial Random Variable

If x is a binomial random variable with parameters n and p (so $q = 1 - p$), then

- Mean $= n \times p$
- Variance $s^2{}_x = n \times p \times q$
- Standard deviation $s_x =$ square root $n \times p \times q$

$$\sigma_X = \sqrt{npq}$$

Binomial Table

Table 4 Binomial Probability Distribution $C_{n,r}\ p^r\ q^{n-r}$

This table shows the probability of r successes in n independent trials, each with probability of success p.

		p																			
n	*r*	.01	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65	.70	.75	.80	.85	.90	.95
2	0	.980	.902	.810	.723	.640	.563	.490	.423	.360	.303	.250	.203	.160	.123	.090	.063	.040	.023	.010	.002
	1	.020	.095	.180	.255	.320	.375	.420	.455	.480	.495	.500	.495	.480	.455	.420	.375	.320	.255	.180	.095
	2	.000	.002	.010	.023	.040	.063	.090	.123	.160	.203	.250	.303	.360	.423	.490	.563	.640	.723	.810	.902
3	0	.970	.857	.729	.614	.512	.422	.343	.275	.216	.166	.125	.091	.064	.043	.027	.016	.008	.003	.001	.000
	1	.029	.135	.243	.325	.384	.422	.441	.444	.432	.408	.375	.334	.288	.239	.189	.141	.096	.057	.027	.007
	2	.000	.007	.027	.057	.096	.141	.189	.239	.288	.334	.375	.408	.432	.444	.441	.422	.384	.325	.243	.135
	3	.000	.000	.001	.003	.008	.016	.027	.043	.064	.091	.125	.166	.216	.275	.343	.422	.512	.614	.729	.857
4	0	.961	.815	.656	.522	.410	.316	.240	.179	.130	.092	.062	.041	.026	.015	.008	.004	.002	.001	.000	.000
	1	.039	.171	.292	.368	.410	.422	.412	.384	.346	.300	.250	.200	.154	.112	.076	.047	.026	.011	.004	.000
	2	.001	.014	.049	.098	.154	.211	.265	.311	.346	.368	.375	.368	.346	.311	.265	.211	.154	.098	.049	.014
	3	.000	.000	.004	.011	.026	.047	.076	.112	.154	.200	.250	.300	.346	.384	.412	.422	.410	.368	.292	.171
	4	.000	.000	.000	.001	.002	.004	.008	.015	.026	.041	.062	.092	.130	.179	.240	.316	.410	.522	.656	.815
5	0	.951	.774	.590	.444	.328	.237	.168	.116	.078	.050	.031	.019	.010	.005	.002	.001	.000	.000	.000	.000
	1	.048	.204	.328	.392	.410	.396	.360	.312	.259	.206	.156	.113	.077	.049	.028	.015	.006	.002	.000	.000
	2	.001	.021	.073	.138	.205	.264	.309	.336	.346	.337	.312	.276	.230	.181	.132	.088	.051	.024	.008	.001
	3	.000	.001	.008	.024	.051	.088	.132	.181	.230	.276	.312	.337	.346	.336	.309	.264	.205	.138	.073	.021
	4	.000	.000	.000	.002	.006	.015	.028	.049	.077	.113	.156	.206	.259	.312	.360	.396	.410	.392	.328	.204
	5	.000	.000	.000	.000	.000	.001	.002	.005	.010	.019	.031	.050	.078	.116	.168	.237	.328	.444	.590	.774
6	0	.941	.735	.531	.377	.262	.178	.118	.075	.047	.028	.016	.008	.004	.002	.001	.000	.000	.000	.000	.000
	1	.057	.232	.354	.399	.393	.356	.303	.244	.187	.136	.094	.061	.037	.020	.010	.004	.002	.000	.000	.000
	2	.001	.031	.098	.176	.246	.297	.324	.328	.311	.278	.234	.186	.138	.095	.060	.033	.015	.006	.001	.000
	3	.000	.002	.015	.042	.082	.132	.185	.236	.276	.303	.312	.303	.276	.236	.185	.132	.082	.042	.015	.002
	4	.000	.000	.001	.006	.015	.033	.060	.095	.138	.186	.234	.278	.311	.328	.324	.297	.246	.176	.098	.031
	5	.000	.000	.000	.000	.002	.004	.010	.020	.037	.061	.094	.136	.187	.244	.303	.356	.393	.399	.354	.232
	6	.000	.000	.000	.000	.000	.000	.001	.002	.004	.008	.016	.028	.047	.075	.118	.178	.262	.377	.531	.735
7	0	.932	.698	.478	.321	.210	.133	.082	.049	.028	.015	.008	.004	.002	.001	.000	.000	.000	.000	.000	.000
	1	.066	.257	.372	.396	.367	.311	.247	.185	.131	.087	.055	.032	.017	.008	.004	.001	.000	.000	.000	.000
	2	.002	.041	.124	.210	.275	.311	.318	.299	.261	.214	.164	.117	.077	.047	.025	.012	.004	.001	.000	.000
	3	.000	.004	.023	.062	.115	.173	.227	.268	.290	.292	.273	.239	.194	.144	.097	.058	.029	.011	.003	.000
	4	.000	.000	.003	.011	.029	.058	.097	.144	.194	.239	.273	.292	.290	;268	.227	.173	.115	.062	.023	.004
	5	.000	.000	.000	.001	.004	.012	.025	.047	.077	.117	.164	.214	.261	.299	.318	.311	.275	.210	.124	.041
	6	.000	.000	.000	.000	.000	.001	.004	.008	.017	.032	.055	.087	.131	.185	.247	.311	.367	.396	.372	.257
	7	.000	.000	.000	.000	.000	.000	.000	.001	.002	.004	.008	.015	.028	.049	.082	.133	.210	.321	.478	.698

Example 1

New hires for a biotech company must pass a test on lab safety before they are assigned tasks in the lab. The probability that a new hire will pass the test is 35%.

If 20 test results are selected at random from the new hires, what is the probability that exactly 10 will pass?

Answer: .069 or 6.9%

What is the probability that exactly 5 will pass the test?

Answer: .127 or 12.7%

Example 2

A computer manufacturing company has received the latest shipments a computer board. The source has a defect rate of 15%.

a. If a sample of ten is selected at random, what is the probability that exactly 4 will be defective?

   Answer: .0401

b. What is the probability that more than 5 will be defective?

   Answer: .001

c. What is the probability that less than 3 will be defective?

   Answer: .8202

d. What is the probability that none will be defective?

   Answer: 0.197

Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.),

Poisson Distribution

Developed by the French mathematician Simeon Poisson (1781–1840), the *Poisson distribution* measures the probability of a random event over some interval of time or space.

Two assumptions

1. The probability of the occurrence of the event is constant for any two intervals of time or space
2. The occurrence of the event in any interval is independent of the occurrence in any other interval

If x equals the number of occurrences in a specified interval, then x is a Poisson random variable

Poisson Distribution

Suppose μ is the mean or expected number of occurrences during a specified interval

The probability of x occurrences in the interval when μ are expected is described by the Poisson distribution

$$p(x) = \frac{e^{-\mu}\,\mu^{x}}{x!}$$

- where x can take any of the values $x = 0,1,2,3, \ldots$
- and $e = 2.71828$ (e is the base of the natural logs)

Poisson Probability Table

x, Number of Occurrences	μ, Mean Number of Occurrences									
	.1	.2	.3	$\boxed{.4}$	.5	.6	.7	.8	.9	1.0
0	.9048	.8187	.7408	*.6703*	.6065	.5488	.4966	.4493	.4066	.3679
1	.0905	.1637	.2222	*.2681*	.3033	.3293	.3476	.3595	.3659	.3679
2	.0045	.0164	.0333	*.0536*	.0758	.0988	.1217	.1438	.1647	.1839
$\boxed{3}$	.0002	.0011	.0033	$\boxed{.0072}$	.0126	.0198	.0284	.0383	.0494	.0613
4	.0000	.0001	.0003	*.0007*	.0016	.0030	.0050	.0077	.0111	.0153
5	.0000	.0000	.0000	*.0001*	.0002	.0004	.0007	.0012	.0020	.0031
6	.0000	.0000	.0000	*.0000*	.0000	.0000	.0001	.0002	.0003	.0005

$$P(x=3) = \frac{e^{-0.4}(0.4)^3}{3!} = 0.0072$$

Poisson Probability Calculations

x, *the Number of Errors in a Week*	$p(x) = \frac{e^{-\mu}\mu^x}{x!}$
0	$p(0) = \frac{e^{-.4}(.4)^0}{0!} = .6703$
1	$p(1) = \frac{e^{-.4}(.4)^1}{1!} = .2681$
2	$p(2) = \frac{e^{-.4}(.4)^2}{2!} = .0536$
3	$p(3) = \frac{e^{-.4}(.4)^3}{3!} = .0072$
4	$p(4) = \frac{e^{-.4}(.4)^4}{4!} = .0007$
5	$p(5) = \frac{e^{-.4}(.4)^5}{5!} = .0001$
6	$p(6) = \frac{e^{-.4}(.4)^6}{6!} = .0000$

Example

Assume industrial records show that 10% of employees steal from their employers. The personnel manager of a firm wishes to determine the probability that from a sample of 20 employees, three have illegally taken company property.

Step 1 calculate the average: n(prob.) = 20 (.10) = 2.0

Step 2 use Poisson table = 0.1804

Step 3 Interpretation- there is an probability of 18% that three of the 20 employees have stolen company property.

Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

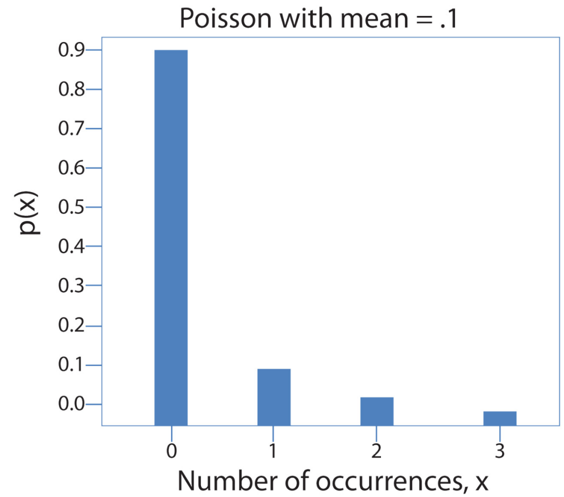
Mean and Variance of a Poisson Random Variable

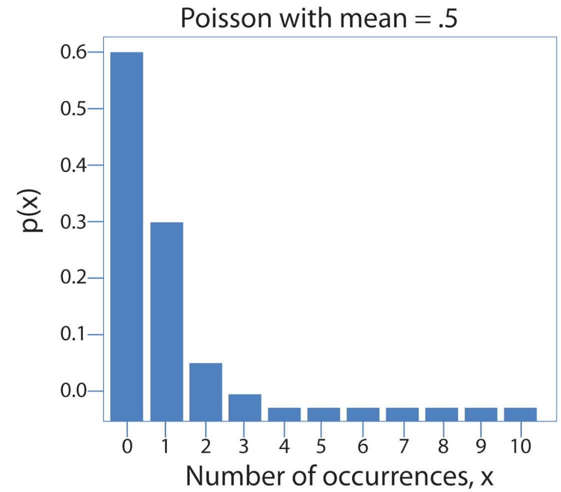
If x is a Poisson random variable with parameter μ, then

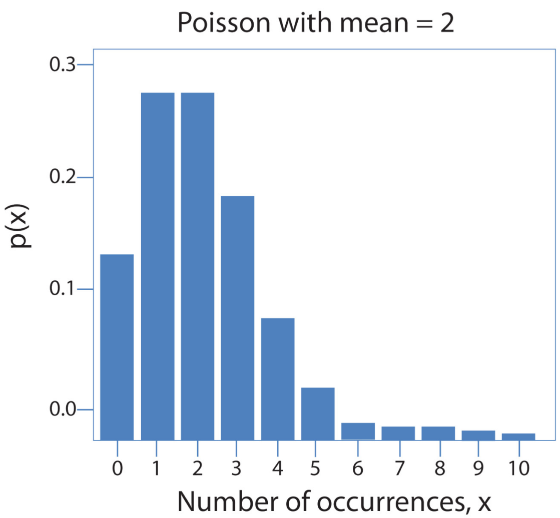
- Mean $\mu_x = \mu$
- Variance $\sigma^2{}_x = \mu$
- Standard deviation σ_x is square root of variance $\sigma^2{}_x$

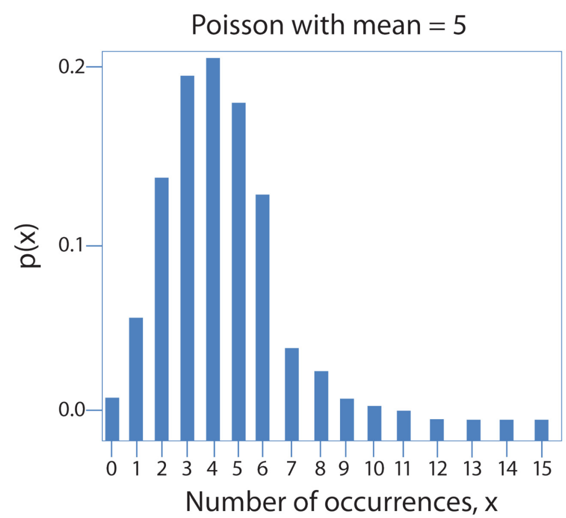
Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

Several Poisson Distributions









Adapted from Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

Hypergeometric Distribution

If a sample is selected without replacement from a known finite population and contains a relatively large proportion of a the population, such that the probability of a success is measurably altered from one selection to the next, the hypergeometric distribution should be used.

Hypergeometric Distribution

Population consists of N items

- r of these are successes
- $(N-r)$ are failures

If we randomly select n items without replacement, the probability that x of the n items will be successes is given by the hypergeometric probability formula

$$P(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

Example

In a recent court case three women brought suit against a local company charging sex discrimination. Of nine people who were eligible for promotion, four were women. Three of the nine were given promotions and only one woman was promoted.

The consultant used the Hypergeometric distribution to show the probability of this outcome.

Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

Example *continued*

$N = 9$; the number of people eligible for promotion.

$r = 4$; the number in the population identified as successes (women).

$n = 3$; the number in the sample (those chosen for promotion).

$x \le 1 \, (x = 0 \text{ and } x = 1)$; the number of successes (women) in the sample.

The probability that no more than one woman was promoted is

$$P(X = 0) + P(X = 1)$$

$$P(X = 1) = \frac{({}_4C_1)({}_5C_2)}{({}_9C_3)} = \frac{4 \times 10}{84} = 0.4762$$

$$P(X = 0) = \frac{({}_4C_0)({}_5C_3)}{({}_9C_3)} = \frac{1 \times 10}{84} = 0.1190$$

Example *continued*

Thus, $P(X \leq 1) = 0.4762 + 0.1190 = 0.5952$

Interpretation

There was almost a 60 percent probability that without any consideration given to gender, no more than one woman would be promoted. On the basis of these findings, as well as other evidence presented in the case, the court ruled that there was not sufficient evidence of discrimination.

Mean and Variance of a Hypergeometric Random Variable

Mean

$$\mu_x = n\left(\frac{r}{N}\right)$$

Variance

$$\sigma^2{}_x = n\left(\frac{r}{N}\right)\left(1-\frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

Hypergeometric Example

Population of six trucks: N=6

Four trucks have mechanical problems: r=4

We randomly select three trucks: n=3

What is the probability that two trucks have problems: x=2

Find P(x=2), mean, and variance

$$P(x=2)=\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}=\frac{\binom{4}{2}\binom{2}{1}}{\binom{6}{3}}=\frac{(6)(2)}{20}=0.6$$

$$\mu_{\mathrm{x}}=n\left(\frac{r}{N}\right)=3\left(\frac{4}{6}\right)=2$$

$$\sigma^{2}{}_{\mathrm{x}}=n\left(\frac{r}{N}\right)\left(1-\frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)=3\left(\frac{4}{6}\right)\left(1-\frac{4}{6}\right)\left(\frac{6-3}{6-1}\right)=0.4$$