Chapter 5 Discrete Probability Distributions

1 Random Variable X

a. A numerical value resulting from a Random Experiment that, by chance, can assme different values.

b. A numerical description of the outcome of a Random Experiment.

Remember:

Random Experiment

1. well defined outcomes
2. only 1 outcome on each Trial
3. outcome occurs by chance

c. Examples:

customers coming into store 0,1,2,3...
weight of cereal Box 10 oz., 10.21 oz.
Defect = 1 Not Defect = 0

2 Discrete Random Variable X (usually counting)

a. Discrete = "Gaps" between numbers 1,2,3 or 1.1, 1.2, 1.3

b. May assume either:

A finite number of values like: 1,2,3
An infinite sequence of numbers like: 1,2,3...

c. Examples:

customers coming into a store like: 0,1,2,3...
scores for a Dancer like: 0, 0.1, 0.2, ... 9.8, 9.9, 10
product Quality like Defect = 1 Not Defect = 0

3 Continuous Random Variable 1 2 {Lots of possible numbers

a. May assume any numerical value in an interval or collection of intervals. Depends on measurement instrument.

b. Examples:

weight of cereal Box 10.1 oz., 10.11 oz., 10.112 oz. $0 \le X \le 12$
Time between customers in line at Disneyland $X \ge 0$ min.
% score on Test $0\% \le X \le 100\%$
Money (even though it seems Discrete)

4 Probability Distribution

A description / presentation of how the probabilities are distributed over the values of the random variable.

5 Discrete Probability Distribution

(a) A description/presentation of how the probabilities are distributed over the values of a discrete random variable

(b) Description / presentation can be:

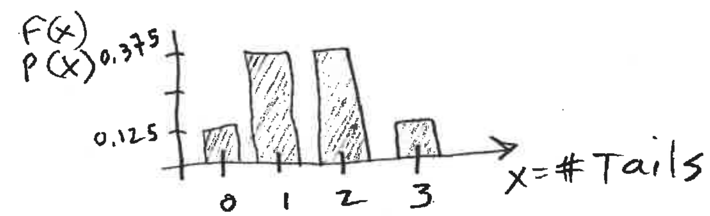
1. Table like:

Experiment: Flip coin 3 times

# of Tails	f(x) or P(x)
0	0.125
1	0.375
2	0.375
3	0.125
$\Sigma =$	1

Each $P(x) \geq 0$

2. Chart like:



3. Discrete Probability Function like: $f(x) = P(x) = \frac{n!}{x!(n-x)!} * p^x * (1-p)^{n-x}$

6 Discrete Probability Function f(x) or P(x)

A probability function, f(x) or P(x), that provides the probability for each value of the discrete random variable.

7 Requirements for f(x) or P(x)

$\boxed{1}$ $f(x) = P(x) \geq 0$ AND $\boxed{2}$ $\Sigma f(x) = \Sigma P(x) = 1$

8 Methods for assigning Probabilities that are useful for creating Discrete Probability Distributions are: classical, Relative & subjective.

8 Examples for creating Discrete Probability Distr.:

a classical

Experiment: Roll 1 die, $X = 1, 2, 3, 4, 5, 6$

Roll Number X	f(x), P(x)
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6
Σ	1 ✓

Each $P(x) \geq 0$ ✓

Discrete Uniform Probability Function

$f(x) = P(x) = 1/n$

n = # of values random variable may assume

b Relative (from past data)

Experiment = count # Restaurant Banquet Rooms Used in 1 day

Relative Frequency method based on large data sets is called "Empirical Discrete Probabity Distribution"

X = # of Rooms Used	# of Days Frequency	Relative Frequency f(x) P(x)
0	2	0.02
1	21	0.21
2	42	0.41
3	27	0.27
4	8	0.08
Total	100	Σ = 1.00

C Subjective (Little past data & Not equally likely outcomes)

Experiment: # of compressor sales in 1 day

# of compressor sales X	f(x), P(x)
0	0.10
1	0.35
2	0.40
3	0.12
4 or more	0.03
Σ	1 ✓

$P(x) \geq 0$ ✓

Sales manager just estimated from memory. Did not directly analyze past data. First year of operation, No past data.

9 Advantage of Probability Distribution?

Easy to calculate probability of a variety of events.

Example:

# of compressor sales X	$f(x)$, $P(x)$
0	0.10
1	0.35
2	0.40
3	0.12
4 or more	0.03
Σ	1 ✓

$P(x) \geq 0$ ✓

Random Variables and Probability Distributions are models that we can use to make predictions about population data!

$$P(X \leq 1) = 0.35 + 0.1 = 0.45$$

$$P(X \geq 2) = 0.40 + 0.12 + 0.03 = 0.55$$

10 steps for building a Discrete Probability Distribution

(a) steps:

1. Define Random Variable
2. Build Frequency Distribution
3. calculate Relative Frequency $f(x) = P(x)$
4. check Requirements: $P(x) \geq 0$ AND $\Sigma P(x) = \underline{1}$
5. create column chart (if desired) to visually portray Distribution (Discrete = columns NOT Touch)
6. make predictions

(b) Examples next two pages:

Example 1:

How to Make a Discrete Probability Distr. (#5)

consider random experiment: A coin tossed 3 times

X = random discrete variable = # of Heads

H = Heads

T = Tails

Experimental outcomes

sample space (list of all possible sample points) = SS

{# of Trials} = TOSS = 3 times = 3

outcomes = 2

$2 * 2 * 2 = 8$ = count of all sample Points = SS

count sample points	coin toss 1st	2nd	3rd	# of Heads
1	H	H	H	3
2	H	H	T	2
3	H	T	H	2
4	T	H	H	2
5	H	T	T	1
6	T	H	T	1
7	T	T	H	1
8	T	T	T	0

possible values of x ⇒ 0, 1, 2, 3

Discrete Probability Distribution

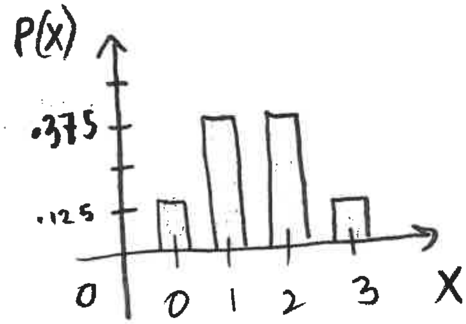
# of Heads	P(x)
0	1/8 = .125
1	3/8 = .375
2	3/8 = .375
3	1/8 = .125
	Σ = 1.00

P(0 Heads in 3 toss) = .125

P(1 Head in 3 toss) = .375

P(2 Heads in 3 toss) = .375

P(3 Heads in 3 toss) = .125



2nd method

# of Heads	P(x)
0	.5 * .5 * .5 * 1 = .125
1	.5 * .5 * .5 * 3 = .375
2	.5 * .5 * .5 * 3 = .375
3	.5 * .5 * .5 * 1 = .125

Example 2:

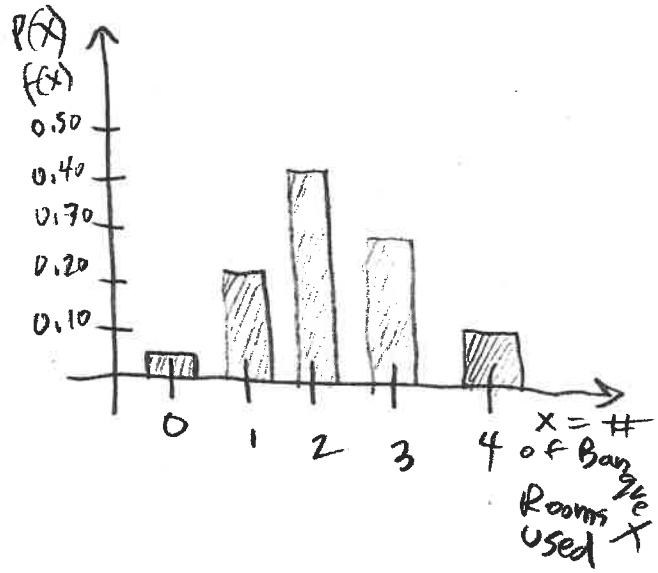
Create Discrete Probability Distribution

Isaac's Italian Restaurant has 4 banquet rooms. Over the past year they collected the following data for weekend room usage: on two days 0 rooms were used, on 21 days 1 room was used, on 42 days 2 rooms were used, on 27 days 3 rooms were used, on 4 days 8 rooms were used.

Create a Discrete Probability Distribution.

Let Discrete Random variable = X = # of banquet rooms used during day

X = # Rooms used	# of Days Frequency	Relative Frequency f(X), P(X)
0	2	2/100 = 0.02
1	21	21/100 = 0.21
2	42	42/100 = 0.42
3	27	27/100 = 0.27
4	8	8/100 = 0.08
	Σ = 100	Σ = 1.00



$P(X \leq 1) = 0.02 + 0.21 = 0.23$

$P(X = 2 \text{ OR } X = 3) = 0.42 + 0.27 = 0.69$

$P(X > 0) = 1 - 0.02 = 0.98$

Conclusion:
- People like rooms
- Maybe we don't need 4 rooms
- Staffing should anticipate 1-3 rooms being used.

11 Expected Value for Discrete Random Variable

$$E(x) = \mu = \sum X * f(x) = \sum X * P(x)$$

1. mean
2. weighted mean
3. Expected Value
4. Long-Run Ave
5. Does not have to be a value that Random Variable ca[n] assume
6. central locatio[n]

* yes, Greek Letters are used for Discrete Random Variables.

(★) Remember from ch. 3?

X # of Rooms	f(x) P(X)	X * P(X)
0	0.02	0*0.02 = 0
1	0.21	1 * 0.21 = 0.21
2	0.42	2 * 0.42 = 0.84
3	0.27	3 * 0.27 = 0.81
4	0.08	4 * 0.08 = 0.32

$\sum P(x)$ 1

Just a weighted mean

$\sum X * P(x) = 2.18 = E(X)$

* In Excel use: SUMPRODUCT

12 Standard Deviation for Discrete Random Variable

1. measures variation
2. How fairly mean represent data points

$$\sigma = \sqrt{\sum (x - E(x))^2 * P(X)}$$

Random variable

Expected value weighted Mean

Probability for Random variable

X # of Rooms	X − E(X)	$(X - E(x))^2$	$(X - E(x))^2 * P(x)$
0	0 - 2.18 = -2.18	4.7524	0.095048
1	1 - 2.18 = -1.18	1.3924	0.292404
2	2 - 2.18 = -0.18	0.0324	0.013608
3	3 - 2.18 = 0.82	0.6724	0.181548
4	4 - 2.18 = 1.82	3.3124	0.264992
	Σ = -0.9		Σ = 0.8476 = Var.

Not add up to Zero BECAUSE we don't have ALL Raw Data (All Numbers)

Standard Deviation $= \sigma = \sqrt{0.8476} = 0.920651943$

Conclusion: Average about 2 rooms used w/ SD of about 1 room.

Back to chapter 4: Adding & Multiplying

An insurance agent has appointments with 4 prospective clients tomorrow. From the past she knows that the probability of making a sale on any one appointment is 1 in 5 what is likelihood that she will sell 3 policies in 4 tries?

Event = sell 3 policies in 4 tries
number of steps or Trials = 4
number of outcomes on any sale attempt = 2
success = sale = S
Failure = Not sale = NS
Each attempt at a sale is an independent event
P (sale) = 0.2 = P
P (Not sale) = 1 − 0.2 = 0.8 = (1−P)
X = Discrete Random variable = # of sales in 4 tries
x = 0, 1, 2, 3, 4

# sample points	sample points	multiply independent events to get P(s.p.)
1	S, S, S, NS	.2 * .2 * .2 * .8 = .0064
2	S, S, NS, S	.2 * .2 * .8 * .2 = .0064
3	S, NS, S, S	.2 * .8 * .2 * .2 = .0064
4	NS, S, S, S	.8 * .2 * .2 * .2 = .0064
		P(3 sales in 4 Tries) = .0256

Add to get Probability of event

Define: Probability of Event: Add Prob. of all sample Points!!

Whole Distribution

(14) Build

probability Distribution (Discrete)

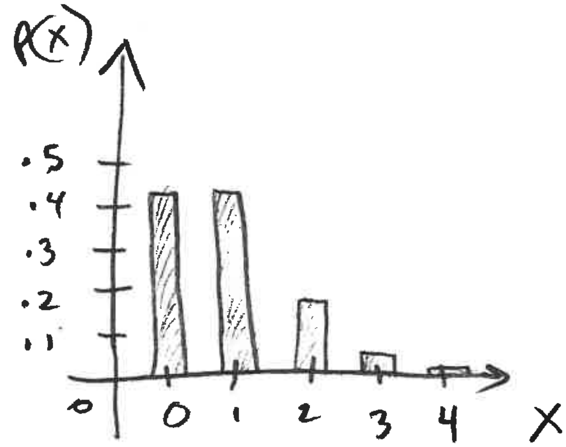
\# ss = 2*2*2*2 = 16

Possible Outcomes sample points	Attempt sale 1st	2nd	3rd	4th	# of sales	Probability of occurrence	
1	S	S	S	S	4	.2*.2*.2*.2 =	.0016
2	S	S	S	NS	3	.2*.2*.2*.8 =	.0064
3	S	S	NS	S	3	.2*.2*.8*.2 =	.0064
4	S	NS	S	S	3	.2*.8*.2*.2 =	.0064
5	NS	S	S	S	3	.8*.2*.2*.2 =	.0064
6	S	S	NS	NS	2	.2*.2*.8*.8 =	.0256
7	S	NS	NS	S	2	.2*.8*.8*.2 =	.0256
8	NS	NS	S	S	2	.8*.8*.2*.2 =	.0256
9	S	NS	S	NS	2	.2*.8*.2*.8 =	.0256
10	NS	S	S	NS	2	.8*.2*.2*.8 =	.0256
11	NS	S	NS	S	2	.8*.2*.8*.2 =	.0256
12	S	NS	NS	NS	1	.2*.8*.8*.8 =	.1024
13	NS	S	NS	NS	1	.8*.2*.8*.8 =	.1024
14	NS	NS	S	NS	1	.8*.8*.2*.8 =	.1024
15	NS	NS	NS	S	1	.8*.8*.8*.2 =	.1024
16	NS	NS	NS	NS	0	.8*.8*.8*.8 =	.4096

$\Sigma = 1$

x # of Sales Random variable	P(x)	P(x)	
0	P(0) =	.4096*1 =	.4096
1	P(1) =	.1024*4 =	.4096
2	P(2) =	.0256*6 =	.1536
3	P(3) =	.0064*4 =	.0256
4	P(4) =	.0016*1 =	.0016

$\Sigma = 1$



Discrete Probability Distributions with n = 4 pi = .2

But there must be an easier way!!

For our sales agent problem we have 16 total possible sample points but we needed only 4 of them:

2	S, S, S, NS
3	S, S, NS, S
4	S, NS, S, S
5	NS, S, S, S

4 total ⇓

How can we calculate this 4?

"sample points"

15 Number of Experimental outcomes that provide exactly X successes in n Trials

$$\left\{\begin{array}{c}\text{\# experimental outcomes} \\ \text{that have x successes} \\ \text{in n Trials}\end{array}\right\} = \frac{n!}{X!(n-x)!}$$

X = # successes of Random Discrete variable

* we use X instead of n, because X = successes in n trials

n = # Fixed Trials

* we use n instead of N because n = # of Fixed Trials

sample points

$n = 4$
$x = 3$

$$\left\{\begin{array}{c}\text{\# of experimental} \\ \text{outcomes that} \\ \text{have x successes} \\ \text{in n trials}\end{array}\right\} = \frac{4!}{3!(4-3)!} = \frac{\cancel{1}*\cancel{2}*\cancel{3}*4}{\cancel{1}*\cancel{2}*\cancel{3}(1)!} = \frac{4}{1} = 4$$

* earlier formula $= \frac{N!}{n!(N-n)!}$ N = count of all objects = pop size
n = size of subset = sample size

16 Also Notice from sales agent problem

# sample point Experimental outcomes	sample point	multiply Independent Events to get P(sample point)	
1	S, S, S, NS	.2 * .2 * .2 * .8	= 0.0064
2	S, S, NS, S	.2 * .2 * .8 * .2	0.0064
3	S, NS, S, S	.2 * .8 * .2 * .2	0.0064
4	NS, S, S, S	.8 * .2 * .2 * .2	0.0064

P(3 in 4 Try) = 0.0256
= 0.0256

multiplication can be done in any order

$p = 0.2$ = probability of success
$(1-p) = 0.8$ = probability of Not success
$x = 3$ = # of successes
$n = 4$ = # of Trials

$$4 * 0.2 * 0.2 * 0.2 * 0.8$$

$$4 * 0.2^3 * 0.8^1$$

$$\underbrace{\frac{n!}{x!(n-x)!} \quad p^x * (1-p)^{n-x}}$$

Binomial Probability Function

$$\frac{4!}{3!(4-3)!} * 0.2^3 * (1-0.2)^{4-3} = \frac{\cancel{1}*\cancel{2}*\cancel{3}*4}{\cancel{1}*\cancel{2}*\cancel{3}*(1)!} * 0.008 * 0.8^1$$

$$= 4 * 0.008 * 0.8 = 4 * 0.0064 = 0.0256$$

(17) Binomial Probability Distribution (Discrete P.D.)

List of all outcomes for a Binomial Experiment (common multi-step experiment that has many useful applications) and the probabilities associated with each experimental outcome (sample point). (P.12)

(18) Requirements for Binomial Experiment

These 3 met = Bernoulli process

1. The experiment consists of a sequence of n identical Trials. (Random variable counts the # of successes in a Fixed # of Trials. Fixed # of Trials = n)
2. 2 outcomes are possible on each Trial. one is defined as a "success" and the other is "Not success" or "failure." S or F.
3. Probability of success, denoted as "p" or the greek letter π ("pi"), remains the same on each Trial. 1 − p does Not change. (Stationary Assumption)
4. The Trials are independent (one does Not affect Next).

Think of sales person loosing enthusiasm... P.221

Examples:

1. Make 4 sales calls. (1) n = 4, (2) S = Sale, (3) P = .2 (4) yes
2. Test with 15 T/F Questions. (1) n = 15, (2) S = get correct, (3) P = .5 (4) yes
3. Flip coin 3 times. (1) n = 3, (2) S = H, (3) P = .5 (4) yes
4. Drive across Bridge 7 times During Rush Hour Traffic (1) n = 7 (2) S = stuck in Traffic (3) P = .15 (4) yes
5. Air Flight from Oak. to Seattle. 6 Flights per day. (1) n = 6 (2) S = late (3) P = .1 (4) yes

19 Binomial Probability Function

$$P(x) = f(x) = \frac{n!}{x!(n-x)!} * p^x * (1-p)^{(n-x)}$$

Equals height of column when you plot on column chart → $P(x) = f(x)$ = Probability of x successes in n Trials

n = # of Fixed Trials

p = Probability of success on any 1 Trial

$1-p$ = Probability of failure on any 1 Trial

20 Excel Binomial Probability Function

→ =BINOM.DIST(number_s, trials, probability_s, cumulative)

number_s = X = Discrete Random variable count # successes

Trials = n = # of Fixed Trials

probability_s = p = Probability of success.

cumulative = 0 for exactly X $\boxed{P(X=2)}$

1 for less than or equal to $\boxed{P(X \le 2)}$

Example:

For our sales Agent Problem, what is probability of making exactly 3 sales in 4 Attempts, p = .2 ?

$n = 4$
$x = 3$
$p = .2$

$$f(3) = P(3) = \frac{4!}{3!(4-3)!} * .2^3 * .8^{(4-3)} =$$

$$= 4 * .008 * .8 = .0256$$ ✓

or

= BINOMDIST(3, 4, .2, 0) = .0256

Binomial?

1. Fixed # Trials? yes.
2. 2 outcomes on each Trial? yes.
3. Probability same each Trial? yes.
4. Events independent? yes.

P 13.5

New Excel Binomial Function:

BINOM.DIST.RANGE(trials, probability_s, number_s, [number_s])

trials	probability_s	number_s	[number_s]
Fixed number Trials = n	probability of success = p	Discrete Random variable that counts successes = X	2nd X value
		→ * if you only use this argument it calculates probability of X value	→ * if you put 2nd X function calculates probability between low X value & high X value
		→ * If you use this argument and next argument, this argument is X value on low end.	this must be X value on high end.

Example:

If $p = .2$, $n = 4$

$\boxed{\text{Find } P(X \le 2) = f(X \le 2)}$

$X = 0$ or $X = 1$ or $X = 2$

(a) $P(0 \text{ or } 1 \text{ or } 2) = f(0 \text{ or } 1 \text{ or } 2) = f(0) + f(1) + f(2) =$

$$= \frac{4!}{0!(4-0)!} * .2^0 * .8^{(4-0)} + \frac{4!}{1!(4-1)!} * .2^1 * .8^{(4-1)} + \frac{4!}{2!(4-2)!} * .2^2 * .8^{(4-2)}$$

$$= .4096 + .4096 + .1536$$

$$= .9728 = P(x \le 2) = f(x \le 2)$$

(b) $= \text{BINOMDIST}(2, 4, .2, 1) = .9728$

Example:

$p = .2$, $n = 4$

Find $f(X \ge 3) = P(X \ge 3)$

$X = 3$ or $X = 4$

(a) $P(X = 3 \text{ or } X = 4) = P(3) + P(4) =$

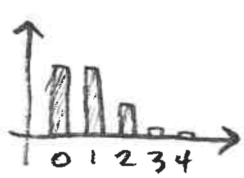
$$= \frac{4!}{3!(4-3)!} * .2^3 * .8^{(4-3)} + \frac{4!}{4!(4-4)!} * .2^4 * .8^{(4-4)}$$

$$= .0256 + .0016$$

$$= .0272 = P(X \ge 3) = f(X \ge 3)$$

TO GO 1 below so 3 is included.

(b) $= 1 - \text{BINOMDIST}(3-1, 4, .2, 1) = .0272$



* Excel function always does cumulative from Low End up $\longrightarrow$ And All area = 1

Expected Value & Standard Deviation for the Binomial Distribution

$$E(X) = \mu = \text{Mean} = n * p$$

$$\sigma = \text{Standard Deviation} = \sqrt{n*p*(1-p)}$$

n = # of Fixed Trials
p = probability of success

Example:

For our sales Agent Problem, what is the mean number of sales she will make in 4 attempts, and what is the standard Deviation?

$n = 4$
$p = .2$

$E(X) = \text{mean} = 4 * .2 = .8$

"For every 4 calls she can expect to sell .8 policies. If she has 40 calls planned, she can expect to sell

$$\frac{40}{4} * .8 = 8 \text{ policies}$$

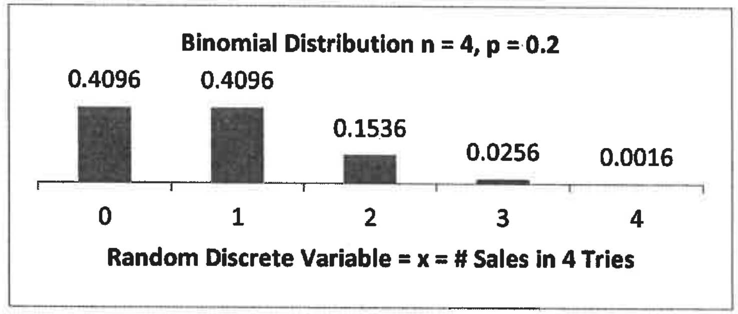
Standard Deviation $= \sqrt{.8 * (1-.2)} = .8$

measures dispersion & could be used to compare to other sales people.

An insurance agent has appointments with 4 clients tomorrow.
From past data, the chance of making a sale is 1 in 5. What is likelihood that she will sell 3 policies in 4 tries?

Binomial Experiment?

1	Fixed # of Identical Trials = n	Yes
2	Each trial only results in S or F	Yes
3	p remains the same for each trial	Yes
4	All events are independent	Yes



n = # of Fixed Trials = Appointments	4
p = Prob of Success =	0.2
Random Discrete Variable = x = # Sales in 4 Tries	
Mean = μ = E(x) = Expected Value =	0.8
Mean = μ = E(x) = Expected Value =	0.8
Standard Deviation =	0.64
Standard Deviation =	0.64

E12: =SUMPRODUCT(A18:A22,B18:B22)
E13: =E9*E10
E14: =SUMPRODUCT((A18:A22-E12)^2,B18:B22)
E15: =E9*E10*(1-E10)

X	P(x) = f(x)
0	0.4096
1	0.4096
2	0.1536
3	0.0256
4	0.0016
Total	1

B18: =BINOM.DIST.RANGE(E9,E10,A18:A22)

X	P(x) = f(x)
0	0.4096
1	0.4096
2	0.1536
3	0.0256
4	0.0016
Total	1

I18:
=COMBIN(E9,H18:H22)*E10^H18:H22*(1-E10)^(E9-H18:H22)

P(x) = f(x)	P(x) = f(x)	SUM	X Lower	X Upper
P(x = 3)	0.0256	0.0256	3	3
P(x <= 0)	0.9984	0.9984	0	3
P(x > 4)	0.0016	0.0016	4	4
P(x >= 3)	0.0272	0.0272	3	4
P(x < 0)	0.9728	0.9728	0	2
P(x <= 3)	0.9984	0.9984	3	

B26: =BINOM.DIST.RANGE(E9,E10,D26:D30,E26:E30)

C26: =SUM(B21)
C27: =SUM(B18:B21)
C28: =SUM(B22)
C29: =SUM(B21:B22)
C30: =SUM(B18:B20)
C31: =SUM(B18:B21)

Math Formula >	X	P(x) = f(x)	P(x) = f(x)
	3	P(x = 3)	0.0256

E34: =COMBIN(E9,C34)*E10^C34*(1-E10)^(E9-C34)

P16

Binomial Experiment Eample 1:

22 ① A flight from oakland to Seattle occurs 6 times per day. The probability that any one flight is late is 0.1. What is the probability that exactly 2 planes are late? What is the probability that less than 2 planes are late? Is this a binomial experiment? Mean? SD?

Binomial Experiment?

① Fixed # of trials (each count S/F)? yes ✓ $n = 6$

② Each Trial Independent? yes ✓ (more or less)

③ S/F each time? yes late or Not late

④ Probability of success same each trial? yes $p = .1$

Variables

$p = .1$ = success = late $x = 2$

$1 - p = 1 - .1 = .9$ = Not late $x < 2$

$n = 6$ = Fixed # of trials

$P(2) = \frac{6!}{(6-2)!2!} * (.1)^2 * (1-.1)^{(6-2)}$

$P(2) = \frac{6 \cdot 5}{2} * (.01) * (.9)^4$

$P(2) = 15 * .01 * .6561$

$P(2) = .098415$

Probability of exactly 2 flights late is .098415

Excel:

=BINOMDIST(x,n,P,0)

$P(x<2) = P(1) + P(0)$

$P(x<2) = .3543 + .5314$

$P(x<2) = .8857$

The probability that less than 2 flight will be late is .8857

Excel:

=BINOMDIST(x,n,P,1)

$\mu = np$

$\mu = 6 * .1$

$\mu = .6$

The mean amount late per day is .6 flights

$\sigma = \sqrt{n\pi * (1-\pi)}$

$\sigma = \sqrt{6 * .1 * (1-.1)}$

$\sigma = \sqrt{.6 * .9}$

$\sigma = \sqrt{.54} = .7348$

The Standard Deviation is .7348

Binomial Experiment Example 2:

The probability of sitting in traffic on the West Seattle Bridge during rush hour is .15. During your next 7 rush hour bridge crossings, what is the probability that you will sit in traffic 3 times? 5 or more times? Mean? SD?

Binomial?

Fixed # of trials? yes $n = 7$

$* \Uparrow = p$

Independent? yes

S/F? stuck in traffic / Not stuck in traffic

P same each time? yes $p = .15$

$(1-p) = .85$

$$P(3) = \frac{7!}{(7-3)!\,3!} * .15^3 (1-.15)^{(7-3)} = \frac{7 \cdot 6 \cdot 5}{3 \cdot 2} * .003375 * .85^4 =$$

$$= 35 * .003375 * .5220625 = .061662$$

$$P(x \geq 5) = P(7) + P(6) + P(5) = .0011522 + .0000678 + .0000017 = .0012217$$

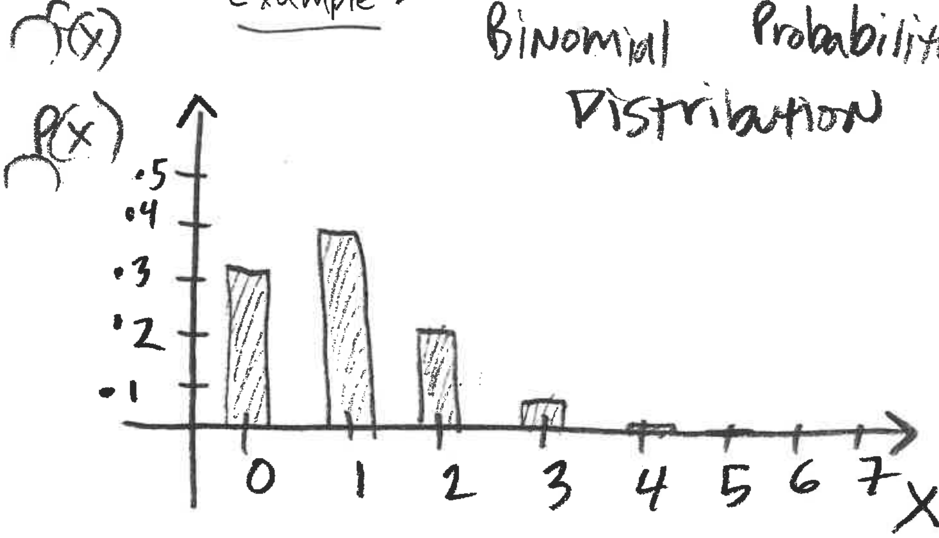
$$\mu = .15 * 7 = 1.05$$

$$\sigma = \sqrt{.15 * 7 * (1-.15)} = .944722181$$

Example 3: Binomial Probability Distribution

$n = 7$

$\pi = .15$



\# of successes
\# of stuck in traffic

② Plot Table with column chart

① Use BINOM.DIST function to create whole table

X # of times stuck in traffic	P(x)
0	.3206
1	.3960
2	.2097
3	.0617
4	.0109
5	.0012
6	.0001
7	.0000

23 For Binomial Distribution :

1. As p (π) approaches $.5$, the Distribution becomes symmetrical

2. As n gets larger, the Distribution becomes symmetrical.

24 Example from diffent Book

Exercise 6.2

Check your answers against those in the ANSWER section.

It is known that 60 percent of all registered voters in the 42^{nd} Congressional District are Republicans. Three registered voters are selected at random from the district. Compute the probability that exactly 2 of the 3 selected are Republicans, using:

a. The rules of probability b. The binomial formula. c. table

Binomial ?

Fixed trials ? yes $n = 3$

Independent ? yes

S or F ? R or Not yes

π same each time ? yes $\pi = .6$

$n = 3$

$\pi = .6$

$x = 2$

$1 - \pi = .4$

a)

location NR	order of occurance	Probability of occurance	" "
3	R, R, NR	$(.6)(.6)(.4) =$	.144
1	NR, R, R	$(.4)(.6)(.6) =$	.144
2	R, NR, R	$(.6)(.4)(.6) =$	.144

$\Sigma = .432$

P(exacty 2 of 3 selected are Republicans) = .432

b) $P(x) = \frac{n!}{x!(n-x)!} \pi^x (1-\pi)^{n-x}$

$P(2) = \frac{3!}{2!(3-2)!} (.6)^2 (1-.6)^{3-2} = 3(.36)(.4) = .432$

c)

$P(2) = .432$

Poisson Probability Distribution

ⓐ Discrete Probability Distribution used to estimate the number of occurrences over a specified interval of time or space.

- Good for number of arrivals in a waiting line situation over a certain time period
  - like how many people arrive to stand in line @ Disneyland or at Dick's Hamburgers in Seattle
- Good for number of repairs needed over a distance of road or pipe.

ⓑ Properties or Requriements for Poisson Experiment

1. Actual Relative Frequency Pattern from past data "fits" Poisson Pattern
2. Variance is about equal to mean
3. Probability of an occurrence is same for any two intervals of equal length.
4. The occurrence or nonoccurrence in any interval is independent of the occurrence / non-occurrence in any other interval.

ⓒ Formula:

$$P(x) = \frac{\mu^x * e^{-\mu}}{x!}$$

x = # occurances
μ = mean (calculated from Past Data)
$e = 2.71828$ ← constant

Excel: =POISSON.DIST(x, mean, cumulative)

cumulative = 0 for exact x
1 less than or equal to x

* x = # of occurrences, No upper limit, 0, 1, 2...
But for occurrences like people arriving to get in line...

d Poisson Example:

The mean number of people arriving at Dick's Drive In Hamburgers in Seattle at Saturday noon lunch period (Noon to 1 PM) during a 1 minute period is 3.88 people.
Looking at past Data:

1. The mean & variance are about equal
2. Probability of a person arriving is the same for any two time periods of equal length.
3. Arrival / Non-arival of a person in any time period is idependent of the arrival / nonarrival of a person in any other time period.
4. M = mean = 3.88 people.

$$P(x) = \frac{M^x * e^{-M}}{x!}$$

$$P(4 \text{ arrivals in } 1 \text{ minute}) = \frac{3.88^4 * 2.71828^{-3.88}}{4!}$$

$= 0.19501$

= POISSON.DIST(4, 3.87969, 0)

$= 0.0206572$

P (3 or 4 or 5 arrivals in 1 minute) POISSON.DIST(5, 3.87969, 1)
− POISSON.DIST(3−1, 3.87969, 1)

= 0.5473729

Subtract 1:
Have to go back to 2 so you can include 3

Power Query to import data:

1) When you import data from a CSV file, the file usually has just one object: a table. You can use the From Text/CSV button in the Get & Transform group in the Data tab in the Excel Ribbon to import CSV data.

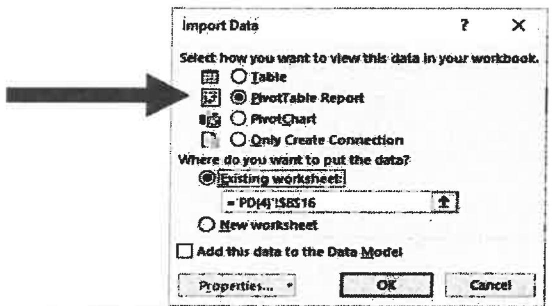
2) When you import data from an Excel file, the file can have more than one object. An Excel file can have objects such as Worksheets, Excel Tables, or Defined Names. Therefore, when you use Power Query to import a table of data from an Excel file, during the import process you must select the object that you want to import. In this video, we will select the Worksheet object named "MonclaireGolf".

To import a table of data from a single Excel file, you can follow these steps:

1) Get & Transform group in the Data tab in the Excel Ribbon
2) Click Get Data dropdown arrow
3) From the dropdown menu, hover over "From File"
4) Click on "From Excel Workbook"
5) Select file from Import Data File dialog box
6) In the Navigator dialog box, on the left side, select the object that has the data table that you want to import
7) Click the "Transform Data" button to bring the object into the Power Query Editor
8) Make transformations you want in Power Query Editor, then Close and Load your table.

Power Query Import Data dialog box to import table directly to PivotTable Cache:

To import a table of data directly to the PivotTable Cache: Select the "PivotTable Report" dialog button in the Import Data dialog box, as shown here:



If you do not need the table of data in the worksheet, loading directly to the PivotTable Cache, avoids loading the table to two locations (worksheet and PivotTable Cache). This can reduce file size and avoid have to refresh both the query and the PivotTable report when source data changes.

Binomial Example in Excel:

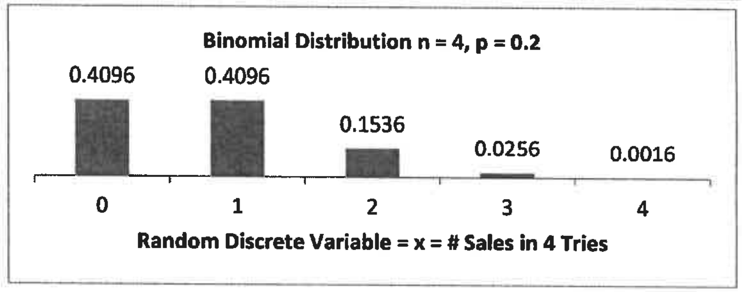
An insurance agent has appointments with 4 clients tomorrow.

From past data, the chance of making a sale is 1 in 5. What is likelihood that she will sell 3 policies in 4 tries?

Binomial Experiment?

1	Fixed # of Identical Trials = n	Yes
2	Each trial only results in S or F	Yes
3	p remains the same for each trial	Yes
4	All events are independent	Yes

n = # of Fixed Trials = Appointments	4
p = Prob of Success =	0.2
Random Discrete Variable = x = # Sales in 4 Tries	
Mean = μ = E(x) = Expected Value =	0.8
Mean = μ = E(x) = Expected Value =	0.8
Standard Deviation =	0.8
Standard Deviation =	0.8



E13: =SUMPRODUCT(A25:A29,B25:B29)
E14: =E10*E11
E15: =SQRT(SUMPRODUCT((A25:A29-E13)^2,B25:B29))
E16: =SQRT(E10*E11*(1-E11))

Binomial Probability Function

$$P(x) = f(x) = \frac{n!}{x!(n-x)!} * p^x * (1-p)^{(n-x)}$$

Math Formula >

X	P(X)	P(X)
3	P(x = 3)	0.0256

E22: =COMBIN(E10,C22)*E11^C22*(1-E11)^(E10-C22)

X	P(x) = f(x)
0	0.4096
1	0.4096
2	0.1536
3	0.0256
4	0.0016
Total	1

B25: =BINOM.DIST.RANGE(E10,E11,A25:A29)

P(X)	SUM	P(X)	X Lower	X Upper
P(x = 3)	0.0256	0.0256	3	3
P(x <= 0)	0.9984	0.9984	0	3
P(x > 4)	0.0016	0.0016	4	4
P(x >= 3)	0.0272	0.0272	3	4
P(x < 0)	0.9728	0.9728	0	2
P(x <= 3)	0.9984	0.9984	3	

B33: =SUM(B28)   C33: =BINOM.DIST.RANGE(E10,E11,D33:D37,E33:E37)
B34: =SUM(B25:B28)
B35: =SUM(B29)
B36: =SUM(B28:B29)
B37: =SUM(B25:B27)
B38: =SUM(B25:B28)   C38: =BINOM.DIST(D38,E10,E11,1)

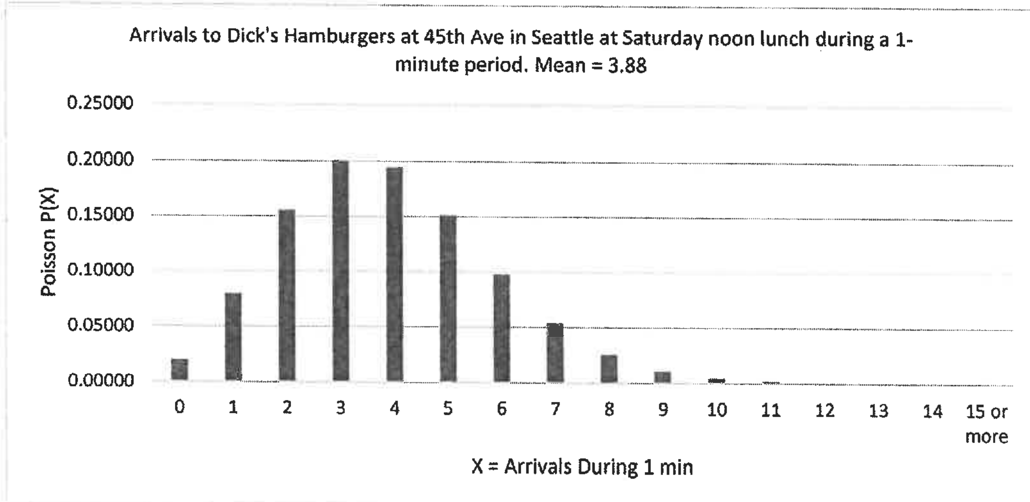
25

Poisson Excel in Excel:

Arrivals to Dick's Hamburgers at 45th Ave in Seattle at Saturday noon lunch during a 1-minute period. Mean = 3.88

Mean	3.88
Variance	3.75

X = Arrivals During 1 min	Poisson P(X)	
0	0.02066	=POISSON.DIST(A8:A23,B4,0)
1	0.08014	
2	0.15547	
3	0.20105	
4	0.19501	
5	0.15131	
6	0.09784	
7	0.05423	
8	0.02630	
9	0.01134	
10	0.00440	
11	0.00155	
12	0.00050	
13	0.00015	
14	0.00004	
15 or more	0.00001	
Total	1.0000	=SUM(B8:B23)



$$P(X) = \frac{\mu^X * e^{-\mu}}{X!}$$

Low X	High X		SUM	POISSON	Formula			
	0	P(0)	0.020657	0.020657	0.020657	=SUM(B8)	=POISSON.DIST(B27,B4,0)	=B4^B27*EXP(1)^-B4/FACT(B27)
3	5	P(3 <= No. Arvls. <= 5)	0.547373	0.547373		=SUM(B11:B13)	=POISSON.DIST(B28,B4,1)-POISSON.DIST(A28-1,B4,1)	
2		P(No. Arvls > 2)	0.743730	0.743733		=SUM(B11:B23)	=1-POISSON.DIST(A29,B4,1)	

26

Hypergeometric Distribution Example in Excel:

Similar to Binomial Distribution except: 1) the trails are not independent AND 2) the probability of success changes from trial to trial.

In Excel use: HYPGEOM.DIST(x,n,population_s,number_pop,cumulative)

x = Number of Successes in Sample = Discrete Random Variable = **sample_s**

n = Trials/Steps in Experiment = Sample Size = **number_sample**

Number Successes in Pop. = **population_s**

Population Size = **number_pop**

cumulative = FALSE = 0 = probability mass function. Use to calculate exactly x

OR

cumulative = TRUE = 1 = probability mass function. Use to calculate <= x (everything from smallest up to and including x)

Probability of pulling 2 Face Cards in 5 tries?

Success = Face card	Face Card (Not Ace)	
Event = pull 2 face cards (not an ace) in 5 tries		
x = Number of Successes in Sample = Discrete Random Variable = **sample_s**	2	
n = Trials/Steps in Experiment = Sample Size = **number_sample**	5	
Number Successes in Pop. = **population_s**	12	
Population Size = **number_pop**	52	
cumulative = FALSE = 0 = probability mass function. Use to calculate exactly x	0	
Number of Sample Points	10	=COMBIN(G16,G15)
Probability of pulling 2 Face Cards in 5 tries	0.2509	=HYPGEOM.DIST(G15,G16,G17,G18,0)
Probability of pulling 2 Face Cards in 5 tries	0.2509	=SUM(G27:G36)

Sample Point	Step 1	Step 2	Step 3	Step 4	Step 5	Prob.	
Sample Point 1	F	F	NF	NF	NF	0.02509	=12/52*11/51*(50-10)/50*(49-10)/49*(48-10)/48
Sample Point 2	F	NF	F	NF	NF	0.02509	=12/52*(51-11)/51*11/50*(49-10)/49*(48-10)/48
Sample Point 3	F	NF	NF	F	NF	0.02509	=12/52*(51-11)/51*(50-11)/50*11/49*(48-10)/48
Sample Point 4	F	NF	NF	NF	F	0.02509	=12/52*(51-11)/51*(50-11)/50*(49-11)/49*11/48
Sample Point 5	NF	F	F	NF	NF	0.02509	=(52-12)/52*12/51*11/50*(49-10)/49*(48-10)/48
Sample Point 6	NF	F	NF	F	NF	0.02509	=(52-12)/52*12/51*(50-11)/50*11/49*(48-10)/48
Sample Point 7	NF	F	NF	NF	F	0.02509	=(52-12)/52*12/51*(50-11)/50*(49-11)/49*11/48
Sample Point 8	NF	NF	F	F	NF	0.02509	=(52-12)/52*(51-12)/51*12/50*11/49*(48-10)/48
Sample Point 9	NF	NF	F	NF	F	0.02509	=(52-12)/52*(51-12)/51*12/50*(49-11)/49*11/48
Sample Point 10	NF	NF	NF	F	F	0.02509	=(52-12)/52*(51-12)/51*(50-12)/50*12/49*11/48

27

Hypergeometric Distribution Example in Excel:

Example 2:

Success = Face card	Face Card (Not Ace)
Event = pull 2 face cards (not an ace) in 5 tries	
x = Number of Successes in Sample = Discrete Random Variable = sample_s	2
n = Trials/Steps in Experiment = Sample Size = number_sample	5
Number Successes in Pop. = population_s	12
Population Size = number_pop	52
cumulative = FALSE = 0 = probability mass function. Use to calculate exactly x	0

X = Pull Face in 5 tries	P(x)
0	0.25318127
1	0.42196879
2	0.25090036
3	0.06602641
4	0.00761843
5	0.00030474
Total	1

Calculate the probability of getting 0 Face cards OR 1 face card OR 2 face cards in 5 tries:

X	P(X)	SUM
2	P(x<=2)	0.92605042
		HYPGEOM.DIST
		0.92605042

Example 3:

During the financial crisis of 2008, of the ten biggest banks, only three increase lending after they were given TARP funds of about 2 Billion dolars were given out.
If you took at random sample of 4 of the ten biggest banks at that time, what is the probability that 1 of them would have increased lending, all 3?

Population size =	10
success in pop	3
Sample size	4
Success =	Did lend $
x = Number	# that did lend in sample

x	P(x)
1	0.5
3	0.03333333

Combinations
4

P(x)
0.5

Sample Point	Step 1	Step 2	Step 3	Step 4	Prob.	
Sample Point 1	Lend	Not Lend	Not Lend	Not Lend	0.125	=3/10*(9-2)/9*(8-2)/8*(7-2)/7
Sample Point 2	Not Lend	Lend	Not Lend	Not Lend	0.125	=7/10*3/9*(8-2)/8*(7-2)/7
Sample Point 3	Not Lend	Not Lend	Lend	Not Lend	0.125	=7/10*(7-1)/9*3/8*(7-2)/7
Sample Point 4	Not Lend	Not Lend	Not Lend	Lend	0.125	=7/10*(7-1)/9*(7-2)/8*3/7

28

Notes from Excel Workbook File "Ch05-ESA.xlsm":

Chapter 4 Terms:
Probability
E = Event, P = Probability, then P(E) = A Number between 0 and 1, inclusive.
Estimate of the unknown future
Random Experiment:
1) Well Defined Outcomes
2) Any single Trial only one outcome can occur
3) Outcome occurs by chance
Multi-Step Experiment:
Random Experiment with more than 1 Step/Trial
Sample Point
Sample Point = Experimental Outcome
Sample Space
List of all Possible Sample Points/Experimental Outcomes
{Total Number Sample Points} for Multi-Step Experiment:
k = Number of Steps
n_k = Outcomes for particular Step
{Total Number Sample Points} = $n_1 + n_2 + \dots n_k$
If number of outcomes is same for each Step, {Total Number Sample Points} = n^k
Assigning Probability
Classical = All Sample Points are equally likely
Relative Frequency = Data from past is Relative Frequency Table that has categories which are:
Mutually Exclusive (no item fits into more than one category) and Collectively Exhaustive (enough categories to count everything).
Subjective = Outcomes not equally likely, Little past data, Use best information you can find but it will be your personal belief
Requirements for Assigning Probabilities:
1) All Sample Points must have a probability between 0 and 1, inclusive
2) Sum of all Sample Point Probabilities = 1
Event
Collection of 1 or more Sample Points
Probability of an Event
Sum of the probabilities of the Sample Points in the Event

Chapter 5:
Random Variables
A numerical description of the outcome of an experiment
Discrete Random Variables
May assume either: a finite number of values like: 1, 2, 3 or an infinite sequence of numbers like 1, 2, 3... (Gaps between numbers)
Continuous Random Variables
May assume any numerical value in an interval or collection of intervals. Depends on measuring instrument
Many possible numbers between 1 & 2. "No Gaps"
Probability Distributions
A description/Presentation of how the probabilities are distributed over the values of the random variable.
Models
Random Variables and Probability Distributions are models for populations of data that we can use to estimate the unknown future.
Discrete Probability Distributions
A description/Presentation of how the probabilities are distributed over the values of a discrete random variable.
Discrete Probability Function
A probability function, f(x) or P(x), that provides the probability for each value that the discrete random variable can assume.
Requirements for f(x) or P(x)
f(x) = P(x) >= 0 AND ∑f(x) = ∑P(x) = 1
Methods for creating Discrete Probability Distributions:
Classical, Relative Frequency (for large data sets it is called Empirical Discrete Distribution) and Subjective
Advantage of Probability Distributions?
Once you have the table, it is easy to calculate probabilities for a variety of Events!!!
Steps for building a Discrete Probability Distribution:
1) Define Random Variable
2) Build Frequency Distribution
3) Calculate Relative Frequency - P(x) = f(x)
4) Check Requirement #1: f(x) >= 0 , Check Requirement #2: Σf(x) = 1
5) Create a Column Chart for a Discrete Variable (Columns do not touch)
6) Make predictions

Expected Value for a Discrete Random Variables
E(x) = μ = Weighted Average = ∑x*f(x) = ∑x*P(x), use: =SUMPRODUCT(XRange,P(x)Range)
Variance and Standard Deviation for a Discrete Random Variables
Var = ∑(x-E(x))^2*P(x), use: =SUMPRODUCT((Xrange-E(x))^2,P(x)Range)
SD = SQRT(∑(x-E(x))^2*P(x)), use: =SQRT(SUMPRODUCT((Xrange-E(x))^2,P(x)Range))
Binomial Experiment
Requirements:
1) Fixed Number of Trials, n = # Trials
2) Each Trial can only result in two outcomes: Success or Failure
3) Probability of success is the same for each trial
4) Trials are independent (one does not affect the next)
Binomial Probability Function
f(x) = P(x) = n!/(x!*(n-x)!)*p^x*(1-p)^(n-x)
In Excel use: BINOM.DIST(x,n,p,cumulative)
x = Number of Successes = Discrete Random Variable = **number_s**
n = Total Steps/Trials = **trials**
p = Probability of Success = **probability_s**
cumulative = FALSE = 0 = probability mass function. Use to calculate exactly x or height of column in Column Chart
OR
cumulative = TRUE = 1 = probability mass function. Use to calculate <= x (everything from smallest up to and including x)
In Excel use: BINOM.DIST.RANGE(n,p,x_low,x_high)
n = Total Steps/Trials = **trials**
p = Probability of Success = **probability_s**
x_low = Number of Successes for first x = Discrete Random Variable = **number_s**. If you use this argument and NOT the [number2] argument, the function will calculate the probability for the x value, which is also the height of the column if you plot all x vales in a Column Chart.
x_high = Number of Successes for second x = Discrete Random Variable = **[number_s2]**. If you use this argument and the number_s argument, the function calculates the total probability between the first x and second x, inclusive.
Expected Value and Variance and Standard Deviation for a Binomial Distribution
E(x) = μ = mean n*p
Standard Deviation = σ = SQRT(n*p*(1-p))
Poisson Distribution
Poisson Probability Distribution can be used to estimate the number of occurrences over a specified interval of time or space
Discrete Random Variable = x = Number of occurrences over a specified interval of time or space.
Good for number of arrivals in waiting line situations over a certain time period (like 15 mins)
Good for number of repairs needed over a distance of road or pipe
Properties (Requirements) of a Poisson Experiment:
1) The probability of an occurrence is the same for any two intervals of equal length
2) The occurrence or nonoccurrence in any interval is independent of the occurrence or nonoccurrence in any other interval
3) Mean = Variance (observed from calculations made off of source data)
f(x) = P(x) = the probability of x occurrences in an interval = height of column when you plot column chart.
Discrete Random Variable = x = Number of occurrences over a specified interval of time or space.
No upper limit for X: 0,1,2,3... , but as x increases past the mean, the probability decreases and gest quite small.
Mean = μ = Calculated from past data
e = constant = 2.71828 = =EXP(1) in Excel
In Excel use: POISSON.DIST(x,mean,cumulative)
x = Number of occurrences over a specified interval of time or space = Discrete Random Variable
mean = μ = Calculated from past data
cumulative = 0 for exactly x OR 1 for less than or equal to x
Hypergeometric Distribution
Similar to Binomial Distribution except: 1) the trails are not independent AND 2) the probability of success changes from trial to trial.
In Excel use: HYPGEOM.DIST(x,n,population_s,number_pop,cumulative)
x = Number of Successes in Sample = Discrete Random Variable = **sample_s**
n = Trials/Steps in Experiment = Sample Size = **number_sample**
Number Successes in Pop. = **population_s**
Population Size = **number_pop**
cumulative = FALSE = 0 = probability mass function. Use to calculate exactly x
OR
cumulative = TRUE = 1 = probability mass function. Use to calculate <= x (everything from smallest up to and including x)