Inferential statistics

Dr. P. Kalpana, M.E., PhD.
Faculty of Mechanical Engineering
IIITDM Kancheepuram

What is statistics

- The science of collecting, organizing, presenting, analyzing, and interpreting data to assist in making more effective decisions
- Statistics is a science that helps us make better decisions in business and economics as well as in other fields
- Statistics teaches us how to summarize, analyze, and draw meaningful inferences from data that then lead to improve decisions
- These decisions that we make help us improve the running, for example, a department, a company, the entire economy, etc

Categories of statistics

Descriptive Statistics

✓Collect
✓Organize
✓Summarize
✓Display
✓Analyze

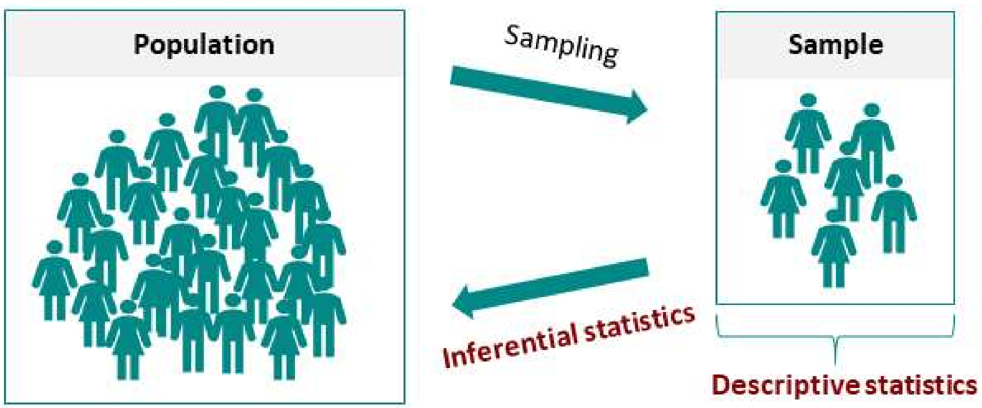
Inferential Statistics

✓Predict and forecast value of population parameters
✓Test hypothesis about value of population
parameter based on sample statistic
✓Make decision

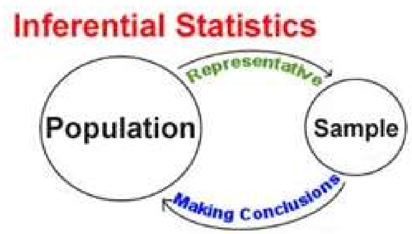
Inferential statistics

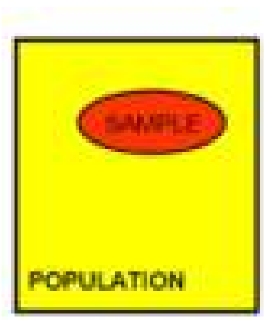
- Inferential statistics are used to draw conclusions about a population by examining the sample
- It provides the bases for prediction, forecasts and estimates that are used to transform the information into knowledge

Parameter

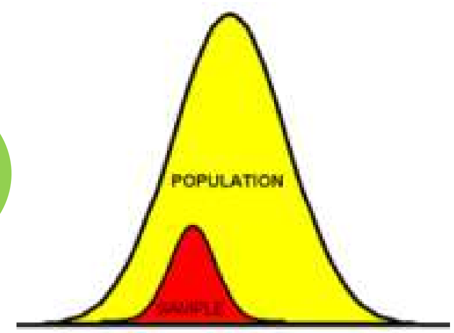


Sample Statistics





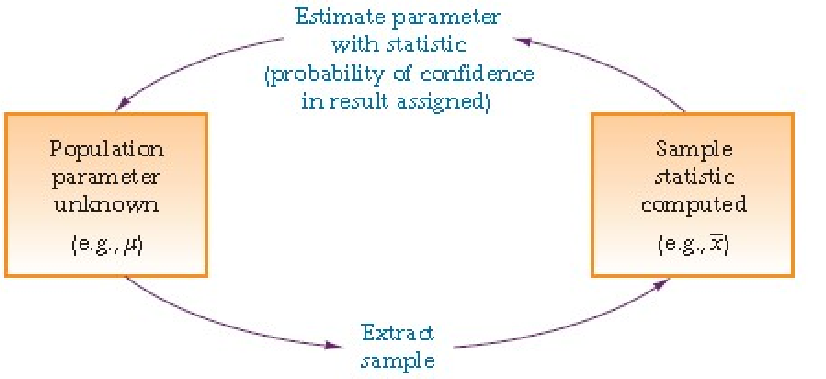
By constructing confidence interval s on population parameter or by setting up a hypothesis test on a population parameter



Probability plays a large part in this.

Inferential statistics

- Inferential statistics is used to make claims about the populations that give rise to the data we collect.
  - This requires that we go beyond the data available to us.
  - Consequently, the claims we make about populations are always subject to error; hence the term "inferential statistics" and not deductive statistics.
- Inferential statistics encompasses a variety of procedures to ensure that the inferences are sound and rational, even though they may not always be correct.
- Hence in short, inferential statistics enables us to make confident decisions in the face of uncertainty. At best, we can only be confident in our statistical assertions, but never certain of their accuracy.



Probability

- The study of obtaining a sample from a population is “probability.”
- To understand our level of confidence in the conclusions drawn from these statistical analyses, we first need to explore the role played by probability in inferential statistics.

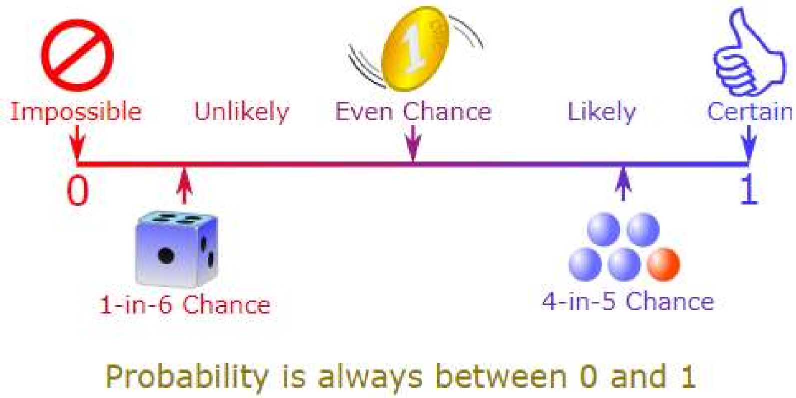
- **Random Experiment:**
- A **random experiment** is a process leading to two or more possible outcomes, without knowing exactly which outcome will occur.
- Examples of random experiments include the following:
  - A coin is tossed and the outcome is either a head or a tail.
  - A company has the possibility of receiving 0–5 contract awards.
  - A customer enters a store and either purchases a shirt or does not.
  - The number of persons admitted to a hospital emergency room during any hour cannot be known in advance.

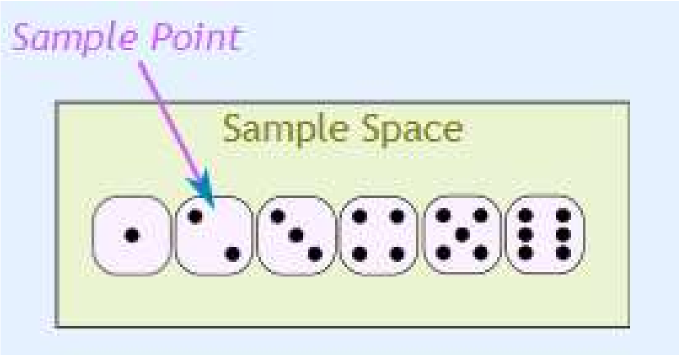
Probability

- comprehend the different way of assigning probability
- understand and apply in marginal union joint and conditional probabilities
- solving problem using laws of probability including law of addition, multiplication and conditional probability
- Bayes rule that to revise the probability

Probability

- The probability is the numerical measure of likelihood that an event will occur
- Probability is measured over the range from 0 to 1 ie $0 \leq P(A) \leq 1$ for any event A
- A probability of 0 indicates that the event will not occur, and a probability of 1 indicates that the
- event is certain to occur.
- Probability does not tell us exactly what will happen, it is just a guide.
- The sum of probabilities of all mutually exclusive and collectively exhaustive events is 1
  - $P(A)+P(B)+P(C)=1$
  - A,B, and C are mutually exclusive and collectively exhaustive events





The method of assigning probability

- Classical probability (Rules and Laws)
- Relative frequency probability (cumulated historical data)
- Subjective probability (personal intuition and reasoning)

- Classical probability
- It is the proportion of times that an event will occur, assuming that all outcomes in a sample space are equally likely to occur
- Dividing the number of outcomes in the sample space that satisfy the event by the total number of outcomes in the sample space determines the probability of an event
- Each outcome is equally likely
- Determined a priori before performing the experiment

$$\text{probability of event A} = \frac{N_A}{N} = \frac{\text{number of outcomes that satisfy the event}}{\text{total number of outcomes in the sample space}}$$

$$\boxed{P(A) = \frac{N_A}{N}}$$

N_A is the number of outcomes that satisfy the condition of event A

- N is the total number of outcomes in the sample space

Counting Possible outcomes

- To compute the **number of combinations of *n items taken x at a time***

$$C_x^n = \frac{n!}{x!(n-x)!} \qquad 0! = 1$$

- A practical difficulty that sometimes arises in computing the probability of an event is counting the numbers of basic outcomes in the sample space and the event of interest
- For some problems the use of permutations or combinations can be helpful
- 1. **Number of Possible Orderings**
  - x of objects that are to be placed in order
  - Each object may be used only once
  - The total number of possible ways of arranging *x objects in order is given by*

$$x(x-1)(x-2)\cdots(2)(1) = x!$$

  - where *x! is read "x factorial."*

Counting Possible outcomes

- ***Permutations***
- we have a number n of objects with which the x ordered (Arranged) boxes could be filled (with n > x)
- Each object may be used only once
- The number of possible orderings is called the number of *permutations of x objects chosen from n*
- It is denoted by the symbol nP_x

$n \quad (n-1) \quad (n-2) \quad \cdot \; \cdot \; \cdot \quad (n-x+2) \quad (n-x+1) \qquad (n-x)$ objects left over

$$\boxed{P_x^n = n(n-1)(n-2)\cdots(n-x+1)} = \frac{n!}{(n-x)!}$$

Example : Permutation

- Suppose that two letters are to be selected from A, B, C, D, and E and arranged in order. How many permutations are possible?
- The number of permutations, with n = 5 and x = 2, is as follows:

With out replacement

$$P_2^5 = \frac{5!}{3!} = 20$$

- These are

AB	AC	AD	AE	BC
BA	CA	DA	EA	CB
BD	BE	CD	CE	DE
DB	EB	DC	EC	ED

Order matters

With Replacement the number of permutation 5x5=25

Counting Possible Outcomes

- The number of different ways that x *objects can* be selected from n where no object may be chosen more than once *order is not important*
- The number of combinations, nC_x of x *objects chosen from n is the number of* possible selections that can be made.
- This number is

$$C_x^n = \frac{P_x^n}{x!} = \frac{n!}{x!(n-x)!}$$

- A personnel officer has 8 candidates to fill 4 similar positions. 5 candidates are men, and 3 are women. If, in fact, every combination of candidates is equally likely to be chosen, what is the probability that no women will be hired?
- The total number of possible combinations of 4 candidates chosen from 8

$$C_4^8 = \frac{8!}{4!4!} = 70$$

- The total number of possible combinations of 4 candidates chosen from 5 men (that no women will be hired)

$$C_4^5 = \frac{5!}{4!1!} = 5$$

- The probability that one of the 5 all-male combinations would be selected is 5/70 = 1/14

Relative Frequency Probability

- We use relative frequency to determine probabilities for a particular population
- It is the number of events in the population that meet the condition divided by the total number in the population
- It is the limit of the proportion of times that event A occurs in a large number of trials, n,
- Computed after performing the experiment

$$\frac{\text{Number of Times an Event Occurred}}{\text{Total Number of Opportunities for the Event to Occur}}$$

$$\boxed{P(A) = \frac{n_A}{n}}$$

Where,

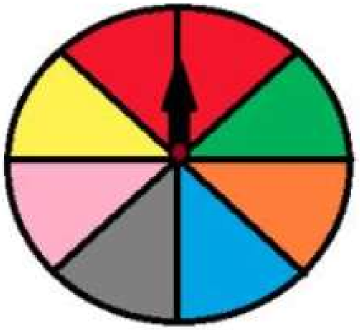
- n_A is the number of A outcomes and
- n is the total number of trials or outcomes.
- The probability is the limit as n becomes large (or approaches infinity)

Subjective Probability

- it comes from a person's intuitions or reasoning
- It expresses an individual's degree of belief about the chance that an event will occur.
- Subjective: different individuals may (correctly) assign different numeric probabilities to the same event
- These subjective probabilities are used in certain management decision procedures
- Useful for unique (single trial) experiments
    - New product introduction
    - Initial public offering of common stock
    - Site selection decisions
    - Sporting events

Probability Terminology

- Experiment
- event
- elementary event
- sample space
- union and intersections
- mutually exclusive events
- independent events
- collectively exhaustive events
- complimentary events



Experiment trial elementary event and event experiment

- ***Experiment***
  - a process that produces outcomes is experiment
  - more than one possible outcome
  - only one outcome per trial
- **trial**
  - one repetition of process
- **elementary event**
  - Even that cannot be decomposed or broken down into other events that is elementary events and
- **Event**
  - an outcome of an experiment
  - any subset of basic outcomes from the sample space
  - may be an elementary event
  - maybe aggregate of elementary event
  - usually represented by uppercase letter for example A, E_1.
  - The null event represents the absence of a basic outcome and is denoted by Ø

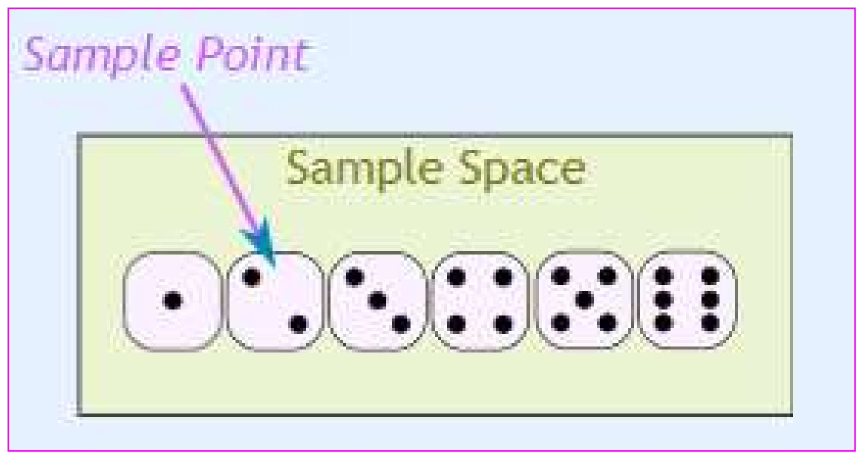
Example

- Experiment : randomly select without replacement 2 families from the residents of the small town
- Elementary event : the sample includes family A and C
- Event: each family in the sample has children in the household
- Event: the sample families own a total of 4 automobiles

Small Town Population		
Family	Children in household	Number of Automobiles
A	Yes	3
B	Yes	2
C	No	1
D	Yes	2

Sample Space

- The possible outcomes from a random experiment are called the **basic outcomes,**
- The set of all basic outcomes is called the **sample space.**
- **We use** the symbol S to denote the sample space.
- Methods of describing a sample space
  - Roaster or listing
  - Tree diagram
  - Set builder notation
  - Venn diagram



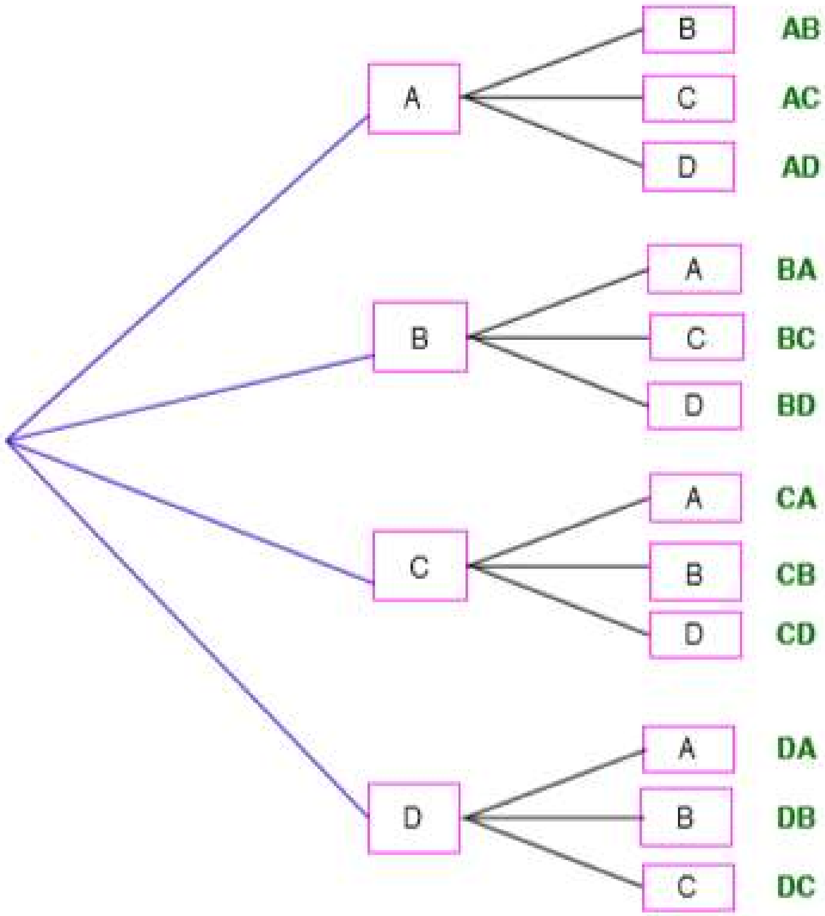
Sample Space Roaster Example

- Experiment: randomly select , with out replacement , two families form the residents of Small town
- Each ordered pair in the sample space is an elementary event

Family	Children in household	Number of Automobiles
A	Yes	3
B	Yes	2
C	No	1
D	Yes	2

Listing of sample space
(A B), (A C), (A D), (B A), (B C), (B D), (C A), (C B), (C D) (D A), (D B), (DC)

Sample Space Tree Diagram for Random Sample of two Families



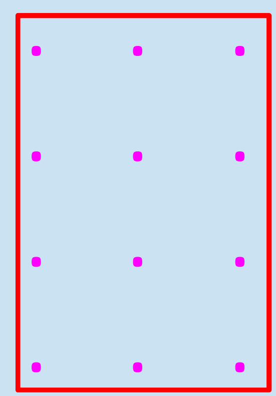
Sample Space: Set Notation for Random Sample of Two Families

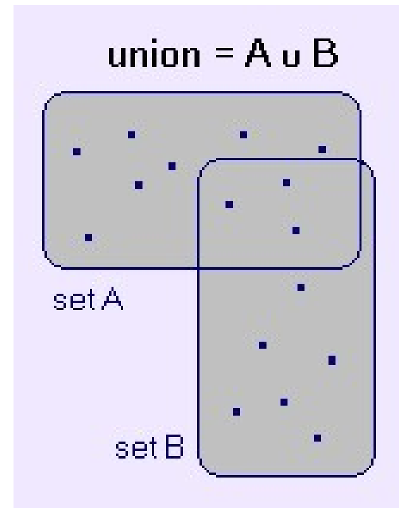
- S={(X,Y)|X is the family selected on the first draw, Y is the family selected on the second row
- Concise description of larger sample space}

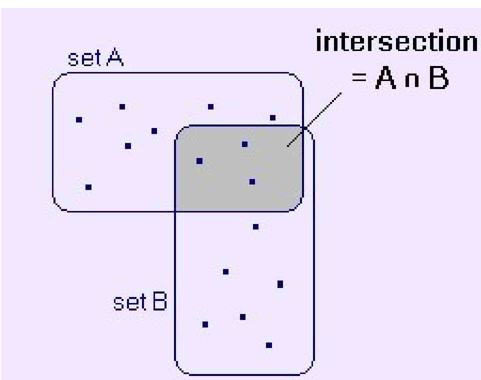
Listing of sample space

(A B), (A C), (A D),
(B A), (B C), (B D),
(C A), (C B), (C D)

Venn Diagram

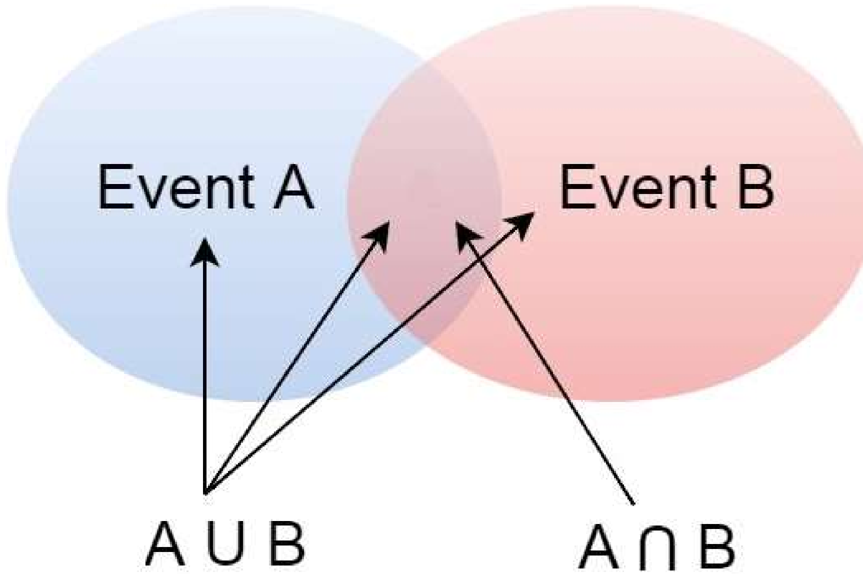


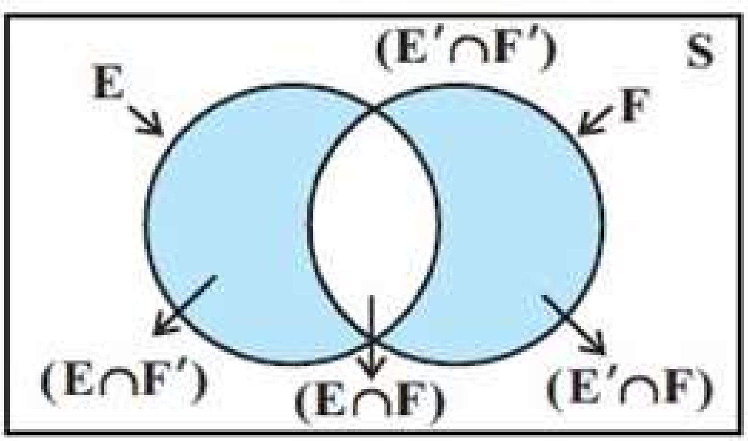




Union of Sets/Events

- The union of two set contain s an instance of each element of the two set
- A={1,3,5,7}
- B={2,3,4,5,6}
- A ∪ B={1,2,3,4,5,6,7}

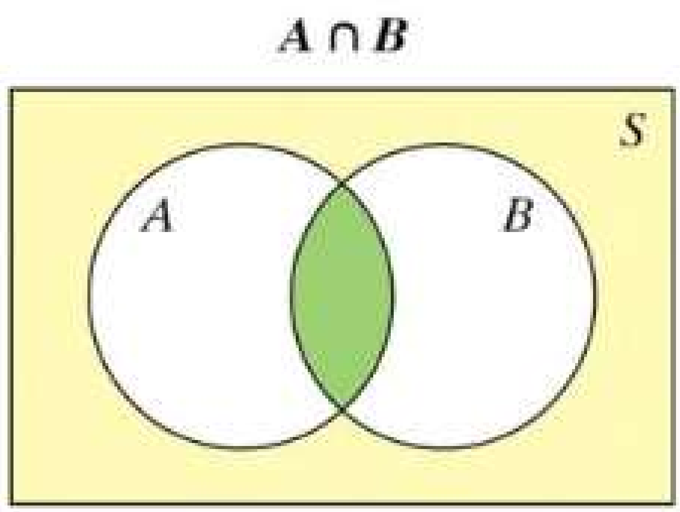




Intersection of Events

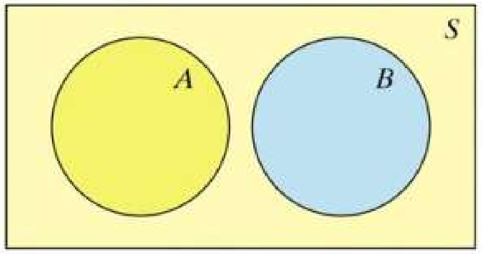
- Intersection the set of all basic outcomes in S that belong to all events
- For given K events E1, E2, . . . , EK, their intersection,
- E1 ∩ E2 ∩ . . . ∩ EK, is the set of all basic outcomes that belong to every E_i(i = 1, 2, . . . , K).
- The intersection of event A and B (A ∩B) is the set of all outcomes in both event A and B

- A={1,3,5,7}
- B={2,3,4,5,6}
- A ∩ B={3,5}



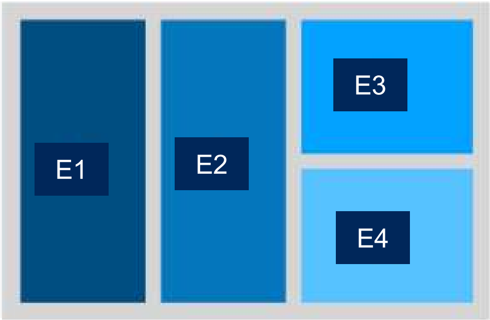
Mutually Exclusive

- If the events *A and B have no common basic outcomes, they are called mutually* exclusive, and their intersection, *A ∩ B, is said to be the empty set, indicating* that *A ∩ B has no members.*
- A={1,3,5,7}
- B={2,4,6}
- A ∩ B={ }
- P(A ∩ B)=0



Other Terms

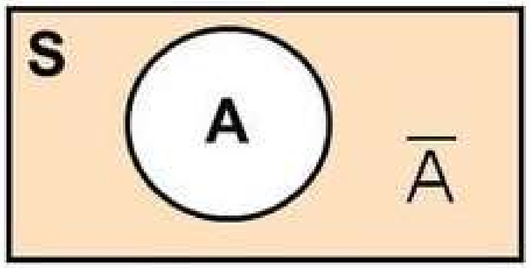
- **Collectively Exhaustive**
- contains all elementary events for an experiment
- Given the K *events* $E_1, E_2, \ldots, E_K$ *in the sample space, S, if* $E_1 \cup E_2 \cup \ldots \cup E_K = S$*, these K events are said to be* ***collectively exhaustive.***



Collectively Exhaustive

Not Collectively Exhaustive

- **Complementary Events**
- Let A be an event in the sample space, S.
- The set of basic outcomes of a random experiment belonging to S but not to A is called the **complement of A** and is denoted by $\bar{A}$.



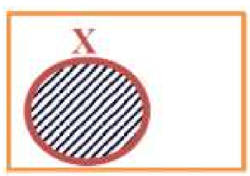
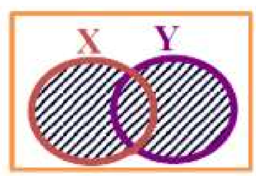
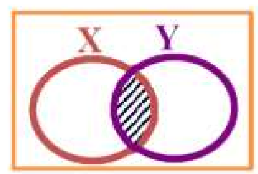
Probability rules

- **The mn rule for counting possibilities**
- **Sampling from a population of size 'N' with sample size 'n' is given by:**
- With Replacement: (N)^n
- Without replacement: (N!)/n!*(N-n)!
- **Complement Rule:**
- Let A be an event and A its complement. Then the complement rule is as follows:

$$P(\overline{A}) = 1 - P(A)$$

- **The general law of addition is:**
- For two events X and Y, P(XUY)=P(X)+P(Y)-P(X∩Y)
- **The general law of multiplication is:** For two events X and Y, P(X∩Y)=P(X).P(Y|X)=P(Y).P(X|Y)
- **Special Law of Multiplication for Independent Events:** If the events X and Y are independent, then, P(X)=P(X|Y) and P(Y)=P(Y|X). Hence, we get, P(X∩Y)=P(X).P(Y)

Types of Probability

Marginal	Union	Joint	Conditional
$P(X)$	$P(X \cup Y)$	$P(X \cap Y)$	$P(X \mid Y)$
The probability of **X** occurring	The probability of **X or Y** occurring	The probability of **X and Y** occurring	The probability of **X occurring given that Y** has occurred
			

Probabilities by using Contingency Table

- A contingency table is a type of table in a matrix format that displays the frequency distribution of the variables.
- Joint as well as marginal densities can be easily identified from the contingency tables.

Event	Event		Total
	B_1	B_2	
A_1	$P(A_1 \text{ and } B_1)$	$P(A_1 \text{ and } B_2)$	$P(A_1)$
A_2	$P(A_2 \text{ and } B_1)$	$P(A_2 \text{ and } B_2)$	$P(A_2)$
Total	$P(B_1)$	$P(B_2)$	1

Joint Probabilities

Marginal (Simple) Probabilities

$$P(A \cap B) = \frac{\text{number of outcomes satisfying A and B}}{\text{total number of elementary outcomes}}$$

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_k)$$

Where B_1, B_2, …, B_k are k mutually exclusive and collectively exhaustive events

Independent Events

- Occurrence of one event does not affect the occurrence of other event
- For two events X and Y, the conditional probability of X given Y is equal to the marginal probability of X.
- And also the conditional probability of Y given X is equal to the marginal probability of Y,
- ie, P(X|Y)=P(X) and P(Y|X)=P(Y).

- Conditional Probability:
- Let *A and B be two events. The* ***conditional probability of event A, given that*** event *B has occurred, is denoted by the symbol P(A | B) and is found to be* as follows:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \text{ provided that } P(B) > 0$$

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} \text{ provided that } P(A) > 0$$

Example 1 :Law of Addition

- A cell phone company found that 75% of all customers want text messaging on their phones, 80% want photo capability, and 65% want both. What is the probability that a customer will want at least one of these?
- What are the probabilities that a person who wants text messaging also wants photo capability and that a person who wants photo capability also wants text messaging?

$$P(A) = 0.75 \quad P(B) = 0.80 \quad \text{and} \quad P(A \cap B) = 0.65$$

- The required probability is as follows:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.75 + 0.80 - 0.65 = 0.90$$

	A	$\overline{A}$	
B	$P(A \cap B)$	$P(\overline{A} \cap B)$	$P(B)$
$\overline{B}$	$P(A \cap \overline{B})$	$P(\overline{A} \cap \overline{B})$	$P(\overline{B})$
	$P(A)$	$P(\overline{A})$	1.0

Example 1

- The probability that a person who wants photo capability also wants text messaging is the conditional probability of event *A, given event B is:*

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.65}{0.80} = 0.8125$$

- the probability that a person who wants text messaging also wants photo capability is as follows:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.65}{0.75} = 0.8667$$

	Text Messaging	No Text Messaging	
Photo	0.65	0.15	0.80
No Photo	0.10	0.10	0.20
	0.75	0.25	1.0

Example 2- Law of addition

- The client company data from the Decision Dilemma reveal that 155 employees worked one of four types of positions. Shown here again is the raw values matrix (also called a contingency table) with the frequency counts for each category and for subtotals and totals containing a breakdown of these employees by type of position and by sex.

Company Human Resource Data

		Sex		
		Male	Female	
Type of Position	Managerial	8	3	11
	Professional	31	13	44
	Technical	52	17	69
	Clerical	9	22	31
		100	55	155

Example 2

- if an employee of the company is selected randomly, what is the probability that the employee is female or professional worker?

$$P(F \cup P) = ?$$

$$P(F \cup P) = P(F) + P(P) - P(F \cap P)$$

Type of Position		Sex		
		Male	Female	Total
	Managerial	.052	.019	.071
	Professional	.200	.084	.284
	Technical	.335	.110	.445
	Clerical	.058	.142	.200
		.645	.355	1.000

P (F)=55/155 =0.335

P(P) =44/155 = 0.824

P (F ∩ P)=13 /155 =0.084

$$P(F \cup P) = .355 + .284 - .084 = .555$$

Type of Position		Sex		
		Male	Female	Total
	Managerial	.052	.019	.071
	Professional	.200	.084	.284
	Technical	.335	.110	.445
	Clerical	.058	.142	.200
		.645	.355	1.000

$$.019 + .084 + .110 + .142 + .200 = .555$$

Mutually exclusive Events

- If a worker is randomly selected from the company.
- what is the probability that the worker is either technical or clerical?
- What is the probability that the worker is either a professional or a clerical?
- **Solution:**
- Let T denote technical, C denote clerical, and P denote professional.
- The probability that a worker is either technical or clerical is

$$\boxed{P(T \cup C) = P(T) + P(C) = \frac{69}{155} + \frac{31}{155} = 0.645}$$

- The probability that a worker is either professional or clerical is

$$\boxed{P(P \cup C) = P(P) + P(C) = \frac{44}{155} + \frac{31}{155} = 0.484}$$

Example 3- Law of addition

- Shown here are the raw value matrix and corresponding probability matrix of the result of a national survey of 200 executives who are asked to identify the geographical location of their company and their company's industry type.

Suppose a respondent is selected randomly from these data.

a. What is the probability that the respondent is from the Midwest (F)?

b. What is the probability that the respondent is from the communications industry (C) or from the Northeast (D)?

c. What is the probability that the respondent is from the Southeast (E) or from the finance industry (A)?

Cross-Tabulation Table

		Geographic Location				
		Northeast D	Southeast E	Midwest F	West G	
Industry Type	Finance A	24	10	8	14	56
	Manufacturing B	30	6	22	12	70
	Communications C	28	18	12	16	74
		82	34	42	42	200

Example 3

PROBABILITY MATRIX

		Geographic Location				
		Northeast D	Southeast E	Midwest F	West G	
Industry Type	Finance A	.12	.05	.04	.07	.28
	Manufacturing B	.15	.03	.11	.06	.35
	Communications C	.14	.09	.06	.08	.37
		.41	.17	.21	.21	1.00

a. P (Midwest) = P (F) = .21

b. P (C D) = P (C) + P (D) - P (C D) = .37 + .41 - .14 = .64

c. P (E ´ A) = P (E) + P (A) - P (E ¨ A) = .17 + .28 - .05 = .40

Law of Multiplication

- **The general law of multiplication is:**
- For two events X and Y,

$$P(X \cap Y)=P(X).P(Y|X)=P(Y).P(X|Y)$$

- **Special Law of Multiplication for Independent Events:**
- If the events X and Y are independent, then, $P(X)=P(X|Y)$ and $P(Y)=P(Y|X)$. Hence, we get,

$P(X \cap Y)=P(X).P(Y)$

$$P(X \cap Y)=P(X).P(Y)$$

- Suppose that women obtain 54% of all bachelor’s degrees in a particular country and that 20% of all bachelor’s degrees are in business. Also, 8% of all bachelor’s degrees go to women majoring in business. Are the events “the bachelor’s degree holder is a woman” and “the bachelor’s degree is in business” statistically independent?

Law of Multiplication

- Let *A denote the event "the bachelor's degree holder is a woman" and B* denote the event "the bachelor's degree is in business."

$$P(A) = 0.54 \quad P(B) = 0.20 \quad P(A \cap B) = 0.08$$

$$P(A)P(B) = (0.54)(0.20) = 0.108 \neq 0.08 = P(A \cap B)$$

- *these events are not independent. The dependence can be seen from the conditional probability:*

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.08}{0.20} = 0.40 \neq 0.54 = P(A)$$

- *Thus, in the country of interest, only 40% of business degrees go to women, whereas women constitute 54% of all degree recipients.*

Law of Conditional Probability

- A conditional probability is the probability of one event, given that another event has occurred
- The conditional probability of X given Y is the joint probability of X and Y divided by the marginal probability of Y, ie,

$$\boxed{P(X|Y) = \frac{P(X \cap Y)}{P(Y)} = \frac{P(Y|X) \cdot P(X)}{P(Y)}}$$

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)} \longrightarrow$$ The conditional probability of A given that B has occurred

$$P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)} \longrightarrow$$ The conditional probability of B given that A has occurred

Where P(A and B) = joint probability of A and B

P(A) = marginal probability of A

P(B) = marginal probability of B

Conditional Probability

Of the cars on a used car lot, **70%** have air conditioning (AC) and **40%** have a CD player (CD). **20%** of the cars have both.

	CD	No CD	Total
AC	.2	.5	.7
No AC	.2	.1	.3
Total	.4	.6	1.0

Are the events AC and CD statistically independent?

What is the probability that a car has a CD player, given that it has AC?

i.e., we want to find P(CD | AC)

Conditional Probability

$P(AC \cap CD) = 0.2$

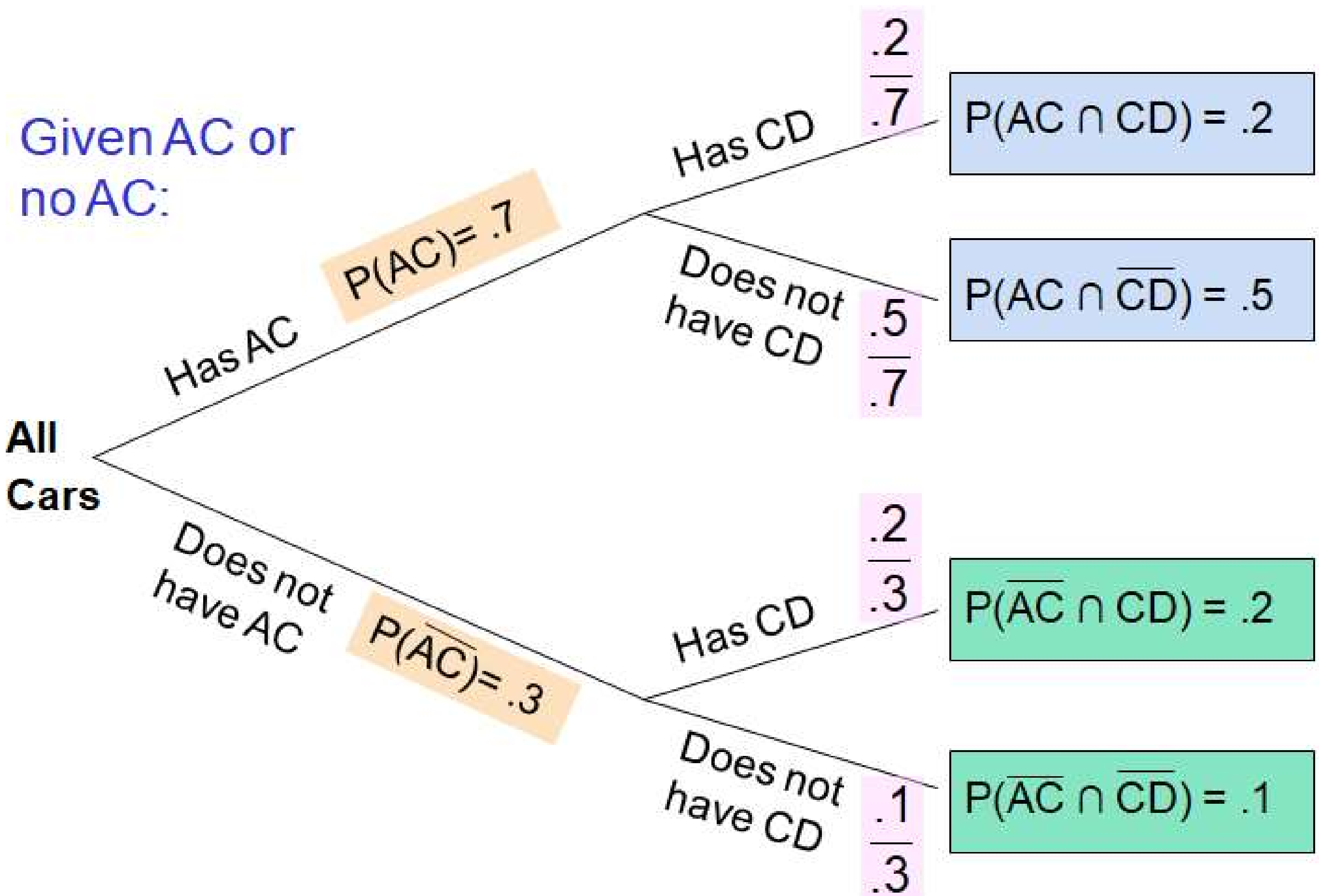
$$\left.\begin{array}{l} P(AC) = 0.7 \\ P(CD) = 0.4 \end{array}\right\} \quad P(AC)P(CD) = (0.7)(0.4) = 0.28$$

$P(AC \cap CD) = 0.2 \quad \neq \quad P(AC)P(CD) = 0.28$

So the two events are not statistically independent

$$P(CD \mid AC) = \frac{P(CD \text{ and } AC)}{P(AC)} = \frac{.2}{.7} = .2857$$

Conditional Probability Using a Tree Diagram



Bayes' Theorem

- Bayes' theorem (alternatively Bayes' law or Bayes' rule) describes the probability of an event, based on prior knowledge of conditions that might be related to the event.
- It is an extension to the conditional law of probabilities.
- allow revision of original probabilities with new information
- For example, if the probability that someone has cancer is related to their age, using Bayes' theorem the age can be used to more accurately assess the probability of cancer than can be done without knowledge of the age.

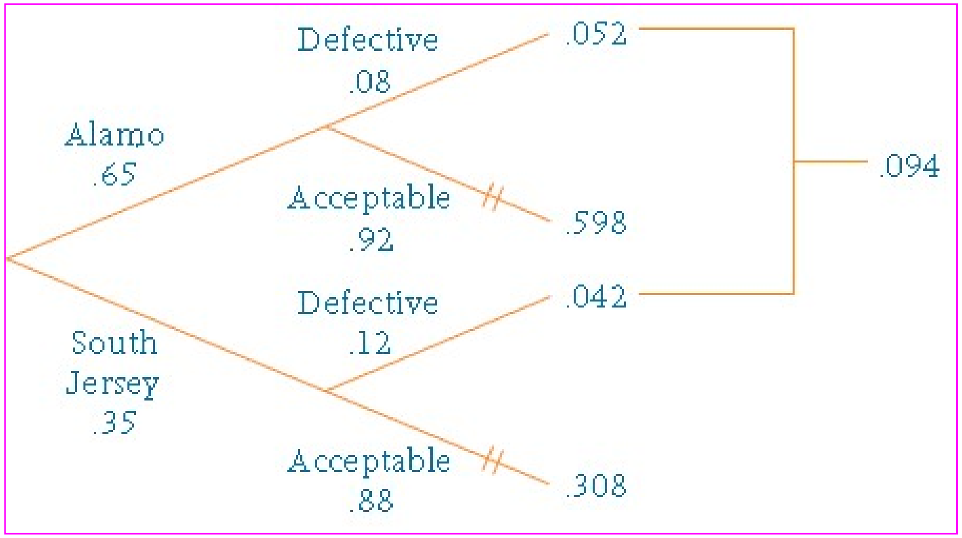
$$\boxed{P(X_i|Y) = \frac{P(X_i) \cdot P(Y|X_i)}{P(X_1) \cdot P(Y|X_1) + P(X_2) \cdot P(Y|X_2) + \cdots + P(X_n) \cdot P(Y|X_n)}}$$

Example: Ribbon Problem

- A particular type of printer ribbon is produced by only two companies, Alamo Ribbon Company and South Jersey Products.
- Suppose Alamo produces 65% of the ribbons and that South Jersey produces 35%.
- Eight percent of the ribbons produced by Alamo are defective and 12% of the South Jersey ribbons are defective.
- A customer purchases a new ribbon.
- What is the probability that Alamo produced the ribbon? What is the probability
- that South Jersey produced the ribbon?
- The ribbon is tested, and it is defective.
- Now what is the probability that Alamo produced the ribbon? That South Jersey produced the ribbon?

Bayesian Table for Revision of Ribbon Problem Probabilities

Event	Prior Probability $P(E_i)$	Conditional Probability $P(d\|E_i)$	Joint Probability $P(E_i \cap d)$	Posterior or Revised Probability
Alamo	.65	.08	.052	$\frac{.052}{.094} = .553$
South Jersey	.35	.12	.042	$\frac{.042}{.094} = .447$
			$P(\text{defective}) = .094$	



$$\text{Revised Probability: Alamo} = \frac{.052}{.094} = .553$$

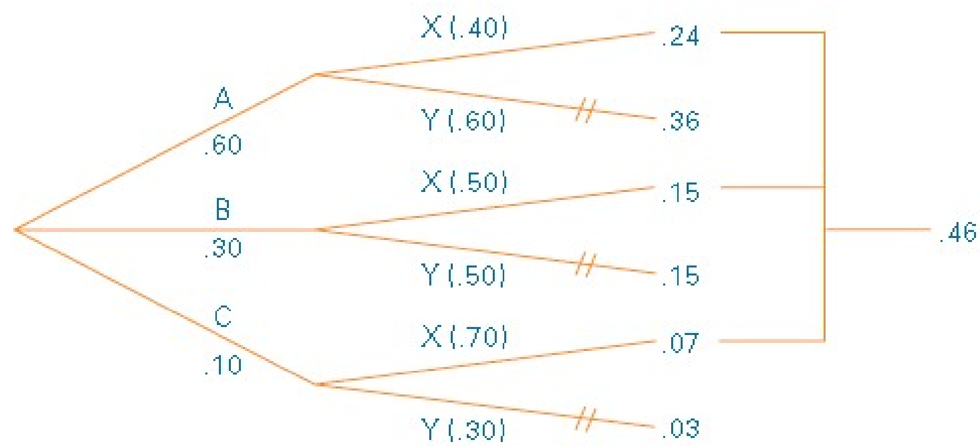
$$\text{Revised Probability: South Jersey} = \frac{.042}{.094} = .447$$

Example

- Machines A, B, and C all produce the same two parts, X and Y. Of all the parts produced,
- machine A produces 60%, machine B produces 30%, and machine C produces10%. In addition,
  - 40% of the parts made by machine A are part X.
  - 50% of the parts made by machine B are part X.
  - 70% of the parts made by machine C are part X.
- A part produced by this company is randomly sampled and is determined to be an X part. With the knowledge that it is an X part, revise the probabilities that the part came from machine A, B, or C.

Example

Event	Prior $P(E_i)$	Conditional $P(X \mid E_i)$	Joint $P(X \cap E_i)$	Posterior
A	.60	.40	$(.60)(.40) = .24$	$\frac{.24}{.46} = .52$
B	.30	.50	.15	$\frac{.15}{.46} = .33$
C	.10	.70	.07	$\frac{.07}{.46} = .15$
			$P(X) = .46$	



$$\text{Revised Probabilities: Machine A: } \frac{.24}{.46} = .52$$

$$\text{Machine B: } \frac{.15}{.46} = .33$$

$$\text{Machine C: } \frac{.07}{.46} = .15$$

Random Variable

- A random variable is a variable which contains the outcomes of a chance experiment.
- discrete values
  - Take one or more Countable values
    - number of apples in the basket
    - the number of defects in a batch of 50 items
    - Roll a die twice: the number of times 4 comes up
    - Toss a coin 5 times: the number of heads
- Continuous Values
  - take on values at every point over a given interval
    - the value of mass of objects
    - volume of liquid nitrogen in a storage tank
    - time between customer arrivals at a retail outlet
    - The yearly income for a family

Getting 1 Head as two

$X = 0, 1, 2, \ldots$

x	$f(x)$	$P(x)$
0	1	1/4
1	2	1/2
2	1	1/4
	4	

HH, TH, HT, TT

$P(x) \geq 0$

$\sum P(x) = 1$

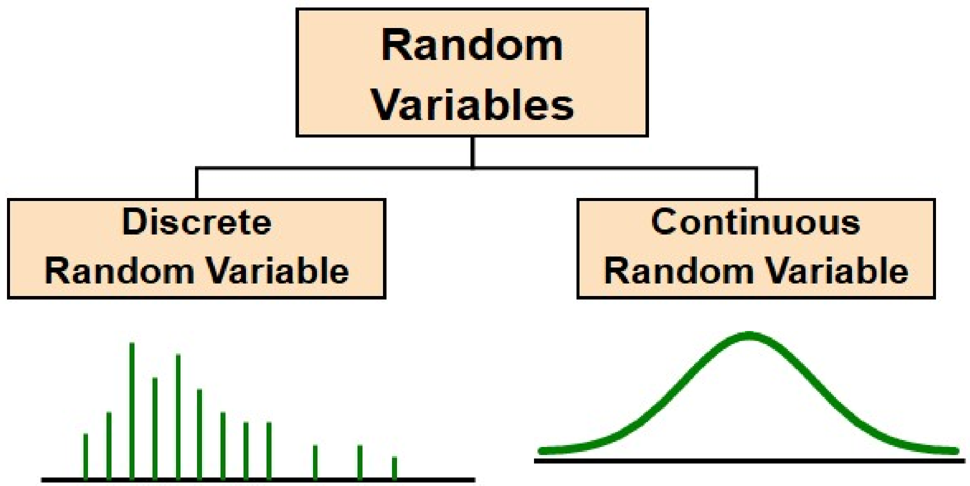
$P(x=1)$

$P(a<x<b)$

5.2 feet

5.10

.15



Probability distributions

- The probability distribution function or probability density function (PDF) of a random variable X means the values taken by that random variable and their associated probabilities
- Once continuous data are measured and recorded, they become discrete data
- Virtually all business data are discrete.
- For practical reasons, data analysis is facilitated greatly by using continuous distributions
- types of distributions
  - **discrete distributions** → Pmf
    - *constructed from discrete random variables, and*
  - ***continuous distributions,*** → Pdf
    - *based on continuous random variables*

Probability Distributions For Discrete Random Variables

- *P(x), of a discrete random variable X* represents the probability that *X takes the value x, as a function of x.*

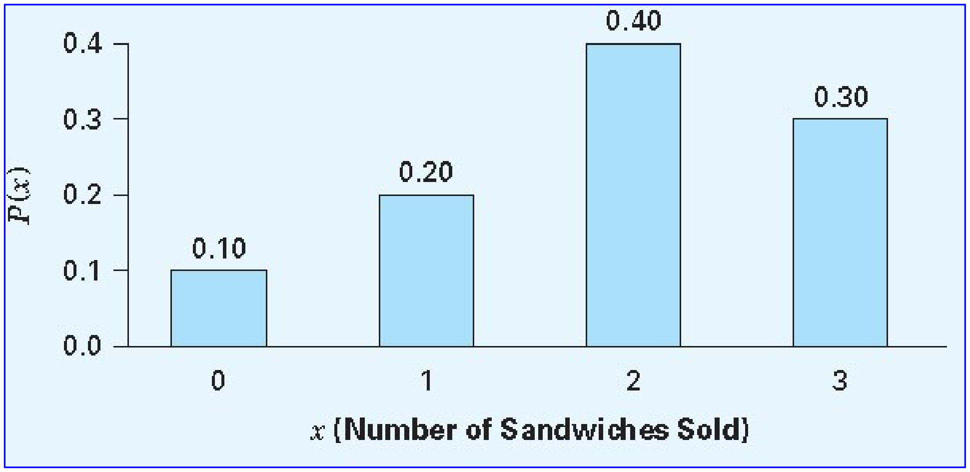
$$P(x) = P(X = x), \text{ for all values of } x$$

- **Number of Product Sales**
- Define and graph the probability distribution function for the number of sandwiches sold by a sandwich shop.
- Let the random variable *X denote the number of sales*
- The probability distribution of sales

x	$P(x)$
0	0.10
1	0.20
2	0.40
3	0.30

The probability of selling one sandwich is 0.20
The probability of selling two or more is 0.70

Graph of Probability Distribution



Examples:

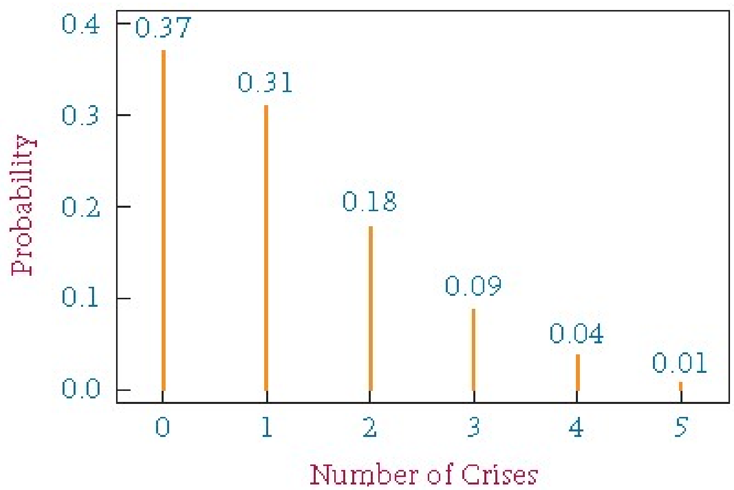
Discrete Distribution of Occurrence of Daily Crises

Number of Crises	Probability
0	.37
1	.31
2	.18
3	.09
4	.04
5	.01

$P(X \le 3) = 0.37+0.31+0.18+0.09$
0.95

$P(1 \le X \le 4) = 0.31+0.18+0.09+0.04$
0.62

$P(x=1) + P(x=2) + P(x=3) + P(x=4)$



Properties of Probability Distribution

- Let X be a discrete random variable with probability distribution P(x)

$$0 \leq P(X) \leq 1$$

- The individual probabilities sum to 1, that is, $\boxed{\sum_x P(x) = 1}$

- **Cumulative Probability distribution**

- The **cumulative probability distribution, *F*(*X*), *of a random variable X, represents***
- The probability that *X does not exceed the value x, as a function of x.*
- That is, $\boxed{F(X) = P(X \leq x)}$

- where the function is evaluated at all values of x.

➢ The CDFs always lies between 0 and 1. ie $0 \leq F(x_i) \leq 1$ for every xi

➢ if x_0 and x_1 are two numbers with $x_0 < x_1$, then $F(x_0) \leq F(x_1)$.

Automobile Sales

- Olaf Motors, Inc., is a car dealer in a small southern town.
- Based on an analysis of its sales history, the managers know that on any single day the number of cars sold can vary from 0 to 5.
- How can the probability distribution function shown in Table be used for inventory planning?

Probability Distribution Function for Automobile Sales

x	$P(x)$	$F(x)$
0	0.15	0.15
1	0.30	0.45
2	0.20	0.65
3	0.20	0.85
4	0.10	0.95
5	0.05	1.00

$P(x \le 1$

$P(x \le 5)$

Four cars in stock, satisfy customer 95% of the time.

Two cars are in stock, then 35% of the customers would not have their needs satisfied

Properties Of Discrete Random Variables

Expected Value

- The **expected value**, *E*[*X*] or μ , of a discrete random variable *X* is defined as

$$\boxed{E[X] = \mu = \sum_{x} xP(x)}$$

- where
- $E(x)$ = long-run average
- x = an outcome
- $P(x)$ = probability of that outcome

In the long run, the mean or expected number
of crises on a given Friday for this executive is 1.15 crises.
Of course, the executive will never have 1.15 crises.

Computing the Mean of the Crises Data

x	$P(x)$	$x \cdot P(x)$
0	.37	.00
1	.31	.31
2	.18	.36
3	.09	.27
4	.04	.16
5	.01	.05
		$\Sigma[x \cdot P(x)] = 1.15$
		$\mu = 1.15$ crises

Properties Of Discrete Random Variables

Variance and Standard Deviation

- The expectation of the squared deviations about the mean, $(X-\mu)^2$, is called the **variance**, denoted as σ^2 and given by

$$\boxed{\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2 P(x)}$$

- The variance of a discrete random variable X can also be expressed as

$$\boxed{\sigma^2 = E[X^2] - \mu^2 = \sum_x x^2 P(x) - \mu^2}$$

- The **standard deviation**, s, is the positive square root of the variance.

$$\boxed{\sigma = \sqrt{\Sigma[(x-\mu)^2 \cdot P(x)]}}$$

Variance and Standard Deviation

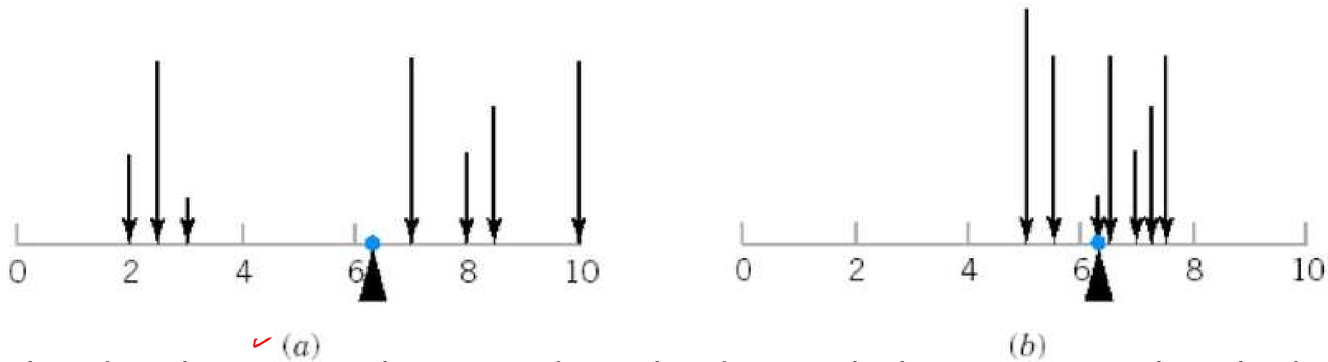
x	$P(x)$	$(x-\mu)^2$	$(x-\mu)^2 \cdot P(x)$
0	.37	$(0-1.15)^2 = 1.32$	$(1.32)(.37) = .49$
1	.31	$(1-1.15)^2 = .02$	$(0.02)(.31) = .01$
2	.18	$(2-1.15)^2 = .72$	$(0.72)(.18) = .13$
3	.09	$(3-1.15)^2 = 3.42$	$(3.42)(.09) = .31$
4	.04	$(4-1.15)^2 = 8.12$	$(8.12)(.04) = .32$
5	.01	$(5-1.15)^2 = 14.82$	$(14.82)(.01) = .15$
			$\Sigma[(x-\mu)^2 \cdot P(x)] = 1.41$

The variance of $\sigma^2 = \Sigma[(x-\mu)^2 \cdot P(x)] = 1.41$

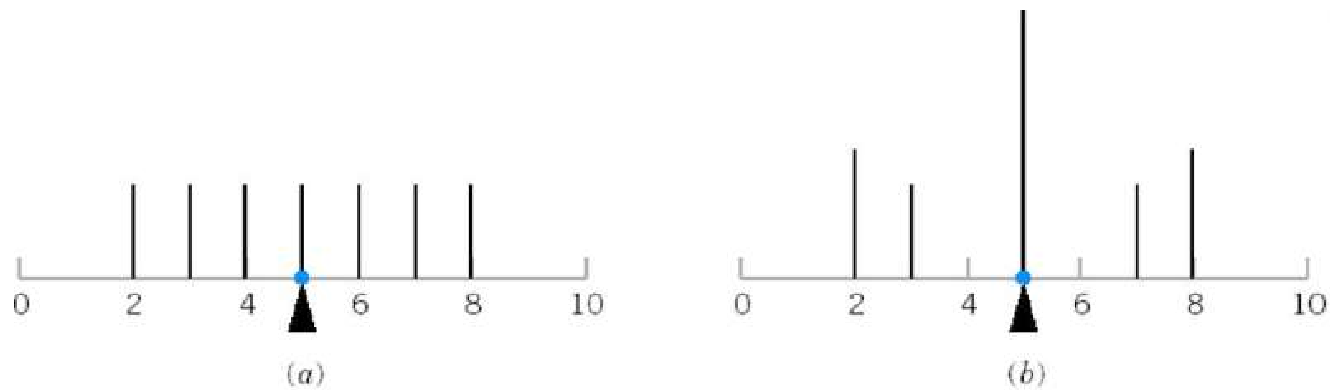
The standard deviation is $\sigma = \sqrt{1.41} = 1.19$ crises.

A probability distribution can be viewed as loading with the mean equal to the balance point.
Both pictures represent the same mean with different variance

Variance and Standard Deviation



A probability distribution can be viewed as a loading with the mean equal to the balance point. Parts (a) and (b) illustrate equal means, but Part (a) illustrates a larger variance



- The probability distribution illustrated in Parts (a) and (b) differ even though they have equal means and equal variances

Properties for Linear Functions of a Random Variable

- Let X be a random variable with mean μ_X and variance σ^2_X, and let a and b be any constant fixed numbers.

$\rightarrow Y = a + bX$

- Define the random variable Y as $a + bX$. Then, the **mean** and **variance of *Y*** are

$$\boxed{\mu_Y = E[a + bX] = a + b\mu_X}$$

$$\boxed{\sigma^2_Y = Var(a + bX) = b^2\sigma^2_X}$$

- so that the **standard deviation of *Y*** is

$$\boxed{\sigma_Y = |b|\sigma_X}$$

- Example:

A contractor is interested in the total cost of a project on which she intends to bid. She estimates that materials will cost \$25,000 and that her labor will be \$900 per day. If the project takes X days to complete, the total labor cost will be $900X$ dollars, and the total cost of the project (in dollars) will be as follows: $C = 25,000 + 900X$

Properties for Linear Functions of a Random Variable

- Using her experience the contractor forms probabilities of likely completion times for the project.
- a. Find the mean and variance for completion time *X*.
- b. Find the mean, variance, and standard deviation for total cost *C*.

Probability Distribution for Completion Times

Completion Time x (Days)	10	11	12	13	14
Probability	0.1	0.3	0.3	0.2	0.1

- The mean and variance for completion time X

$$\mu_X = E[X] = \sum_x xP(x)$$
$$= (10)(0.1) + (11)(0.3) + (12)(0.3) + (13)(0.2) + (14)(0.1) = 11.9 \text{ days}$$

$$\sigma_x^2 = E[(X - \mu_x)^2] = \sum_x (x - \mu_x)^2 P(x)$$
$$= (10 - 11.9)^2(0.1) + (11 - 11.9)^2(0.3) + \cdots + (14 - 11.9)^2(0.1) = 1.29$$

Properties for Linear Functions of a Random Variable

- The mean, variance, and standard deviation of total cost, *C*, are calculated as follows

$$\mu_C = E[25{,}000 + 900X] = (25{,}000 + 900\mu_X) = 25{,}000 + (900)(11.9) = \$35{,}710$$

$$\sigma_C^2 = Var(25{,}000 + 900X) = (900)^2\sigma_X^2 = (810{,}000)(1.29) = 1{,}044{,}900$$

$$\sigma_C = \sqrt{\sigma_C^2} = \$1{,}022.20$$

Covariance

- For two discrete random variables X and Y with $E(X) = \mu(x)$ and $E(Y) = \mu(y)$,
- The covariance between X and Y is defined as $Cov(XY) = \mu(xy) = E(X - \mu(x)) E(Y - \mu(y)) = E(XY) - \mu(x).\mu(y)$
- The covariance gives some information about how X and Y are statistically related

- In general, the covariance between two random variables can be positive or negative.
- If two random variables move in the same direction, then the covariance will be positive, if they move in the opposite direction the covariance will be negative.

- **Properties:**
- If X and Y are independent random variables, their covariance is zero. Since $E(XY) = E(X)E(Y)$
- $Cov(XX) = Var(X)$
- $Cov(YY) = Var(Y)$

Properties of Expected Value

$1. E(b) = b, b$ is a constant.

$2. E(X+Y) = E(X) + E(Y).$

$3. E\left(X/_{Y} \right) \neq \dfrac{E(X)}{E(Y)}.$

$4. E(XY) \neq E(X)E(Y)$ unless they are indpendendent.

$5. E(aX) = aE(X)$, a constant.

$6. E(aX+b) = aE(X) + b$, a and b are constants.

Properties of Variance

1. Var(constant) = 0
2. If X and Y are two independent random variables, then

   Var(X + Y) = Var(X) + Var (Y) and

   Var(X - Y) = Var(X) + Var (Y)
3. If b is a constant then Var(b+X) = Var(X)
4. If a is a constant then Var(aX) = a^2Var(X)
5. If a and b are constants then Var(aX+b) = a^2Var(X)
6. If X and Y are two independent random variables and a and b are constants then Var(aX+bY) = a^2Var(X) + b^2Var(Y)

If the random variables are not independent

$$\boxed{Var(X + Y) = \sigma_X^2 + \sigma_Y^2 + 2\,Cov(X, Y)}$$

Correlation Coefficient

- The covariance tells the sign but not the magnitude about how strongly the variables are positively or negatively related. The correlation coefficient provides such measure of how strongly the variables are related to each other.
- For two random variables X and Y with E(X) = μ(x) and E(Y) = μ(y), the correlation coefficient is defined as:

$$\boxed{\rho_{xy} = \frac{Cov(XY)}{\sigma_x \sigma_y} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}}$$

Binomial Distribution

- The Bernoulli model, which is a building block for the binomial
- The word *binomial* indicates, any single trial of a binomial experiment contains only two possible mutually exclusive and collectively exhaustive outcomes
- These two outcomes are labeled *success* or *failure*.

  p q

  $q = 1 - p$

  $\sum P(x) = 1$

  $\frac{1}{52}$

- This distribution is known as the *Bernoulli distribution*

  $\frac{1}{51}$

- **ASSUMPTIONS**

  H

- The experiment involves *n* identical trials.
- Each trial has only two possible outcomes denoted as success or as failure.
- Each trial is independent of the previous trials.
- The terms *p* and *q* remain constant throughout the experiment, where the term *p* is the probability of getting a success on any one trial and the term $q = (1 - p)$ is the probability of getting a failure on any one trial.

Binomial Distribution

- a random experiment can result in two possible mutually exclusive and collectively exhaustive outcomes, "success" and "failure," and that *P* is the probability of a success in a single trial.
- If n independent trials are carried out, the distribution of the number of resulting successes, x, is called the binomial distribution.
- Its probability distribution function for the binomial random variable X = x is as follows:

$$\boxed{\begin{aligned}&P(x \text{ successes in } n \text{ independent trials})\\ &= P(x) = \frac{n!}{x!(n-x)!}P^{x}(1-P)^{(n-x)}\ \textit{for}\ x = 0, 1, 2, \ldots, n\end{aligned}}$$

- p = the probability of getting a success in one trial
- $q = 1 - p$ = the probability of getting a failure in one trial

Mean and Variance of a Binomial Probability Distribution

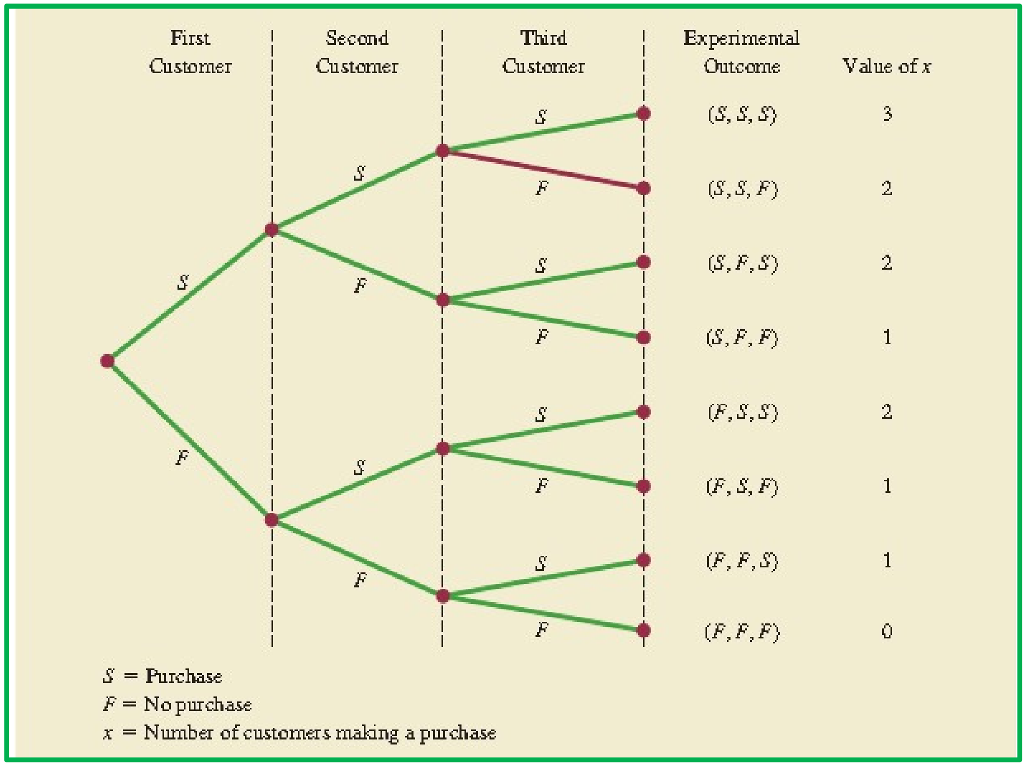
$$\boxed{\mu = E[X] = nP}$$

$$\boxed{\sigma_X^2 = E[(X - \mu_X)^2] = nP(1-P)}$$

Martin Clothing Store Problem

- Let us consider the purchase decisions of the next three customers who enter the Martin Clothing Store. On the basis of past experience, the store manager estimates the probability that any one customer will make a purchase is .30. What is the probability that two of the next three customers will make a purchase?

Tree Diagram for the martin clothing Store Problem



1 2 3

x	f(x)	P(x)
0	1	
1	3	
2	3	
3	1	

$P(x \geq 2)$

$P(x=2) + P(x=3)$

$P(x \geq 2)$

$P(x=2) + P(x=3) + \dots$

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

Martin Clothing Store Problem

- The number of ways of obtaining 2 successes in the *n* = 3 trials. From

$$\binom{n}{x} = \binom{3}{2} = \frac{3!}{2!(3-2)!} = \frac{(3)(2)(1)}{(2)(1)(1)} = \frac{6}{2} = 3$$

- Three of the experimental outcomes yield two successes such as (S, S, F), (S, F, S), and (F, S, S)

The number of ways of obtaining 3 successes in the n = 3 trials

$$\binom{n}{x} = \binom{3}{3} = \frac{3!}{3!(3-3)!} = \frac{3!}{3!0!} = \frac{(3)(2)(1)}{3(2)(1)(1)} = \frac{6}{6} = 1$$

- one experimental outcome with three successes is identified by (*S*, *S*, *S*)

The probabilities for all three experimental outcomes involving two successes follow

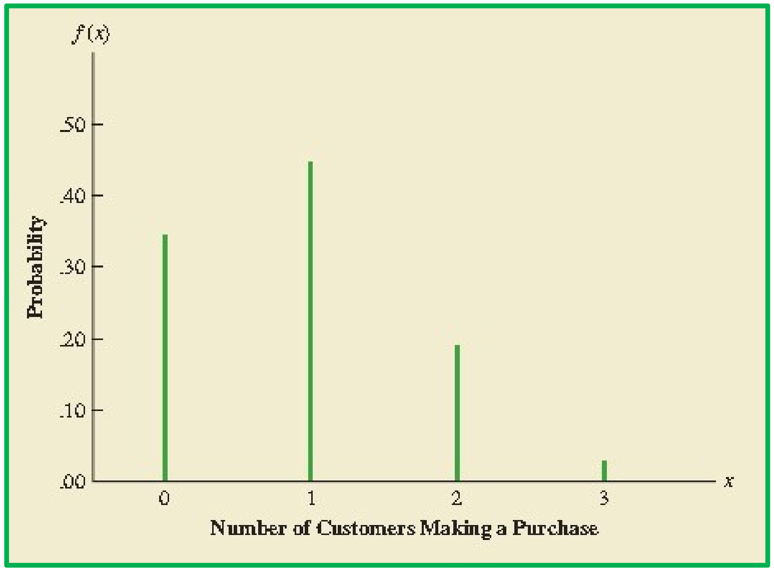
Trial Outcomes 1st Customer	2nd Customer	3rd Customer	Experimental Outcome	Probability of Experimental Outcome
Purchase	Purchase	No purchase	(S, S, F)	$pp(1-p) = p^2(1-p) = (.30)^2(.70) = .063$
Purchase	No purchase	Purchase	(S, F, S)	$p(1-p)p = p^2(1-p) = (.30)^2(.70) = .063$
No purchase	Purchase	Purchase	(F, S, S)	$(1-p)pp = p^2(1-p) = (.30)^2(.70) = .063$

Martin Clothing Store Problem

PROBABILITY DISTRIBUTION FOR THE NUMBER OF CUSTOMERS MAKING A PURCHASE

x	$f(x)$
0	$\frac{3!}{0!3!}(.30)^0(.70)^3 = .343$
1	$\frac{3!}{1!2!}(.30)^1(.70)^2 = .441$
2	$\frac{3!}{2!1!}(.30)^2(.70)^1 = .189$
3	$\frac{3!}{3!0!}(.30)^3(.70)^0 = .027$
	1.000

GRAPHICAL REPRESENTATIONFOR THE NUMBER OF CUSTOMERS MAKING A PURCHASE



Binomial Table

SELECTED VALUES FROM THE BINOMIAL PROBABILITY TABLE

EXAMPLE: *n* 10, *x* 3, *p* .40; *f* (3) = .2150

						p					
n	*x*	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
9	0	.6302	.3874	.2316	.1342	.0751	.0404	.0207	.0101	.0046	.0020
	1	.2985	.3874	.3679	.3020	.2253	.1556	.1004	.0605	.0339	.0176
	2	.0629	.1722	.2597	.3020	.3003	.2668	.2162	.1612	.1110	.0703
	3	.0077	.0446	.1069	.1762	.2336	.2668	.2716	.2508	.2119	.1641
	4	.0006	.0074	.0283	.0661	.1168	.1715	.2194	.2508	.2600	.2461
	5	.0000	.0008	.0050	.0165	.0389	.0735	.1181	.1672	.2128	.2461
	6	.0000	.0001	.0006	.0028	.0087	.0210	.0424	.0743	.1160	.1641
	7	.0000	.0000	.0000	.0003	.0012	.0039	.0098	.0212	.0407	.0703
	8	.0000	.0000	.0000	.0000	.0001	.0004	.0013	.0035	.0083	.0176
	9	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0008	.0020
10	0	.5987	.3487	.1969	.1074	.0563	.0282	.0135	.0060	.0025	.0010
	1	.3151	.3874	.3474	.2684	.1877	.1211	.0725	.0403	.0207	.0098
	2	.0746	.1937	.2759	.3020	.2816	.2335	.1757	.1209	.0763	.0439
	3	.0105	.0574	.1298	.2013	.2503	.2668	.2522	**.2150**	.1665	.1172
	4	.0010	.0112	.0401	.0881	.1460	.2001	.2377	.2508	.2384	.2051
	5	.0001	.0015	.0085	.0264	.0584	.1029	.1536	.2007	.2340	.2461
	6	.0000	.0001	.0012	.0055	.0162	.0368	.0689	.1115	.1596	.2051
	7	.0000	.0000	.0001	.0008	.0031	.0090	.0212	.0425	.0746	.1172
	8	.0000	.0000	.0000	.0001	.0004	.0014	.0043	.0106	.0229	.0439
	9	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0016	.0042	.0098
	10	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010

Mean and variance

- For the Martin Clothing Store problem with three customers, we can use the following equation to compute the expected number of customers who will make a purchase

$$E(x) = np = 3(.30) = .9$$

$$\sigma^2 = np(1 - p) = 3(.3)(.7) = .63$$
$$\sigma = \sqrt{.63} = .79$$

- Suppose that for the next month the Martin Clothing Store forecasts 1000 customers will enter the store. What is the expected number of customers who will make a purchase?

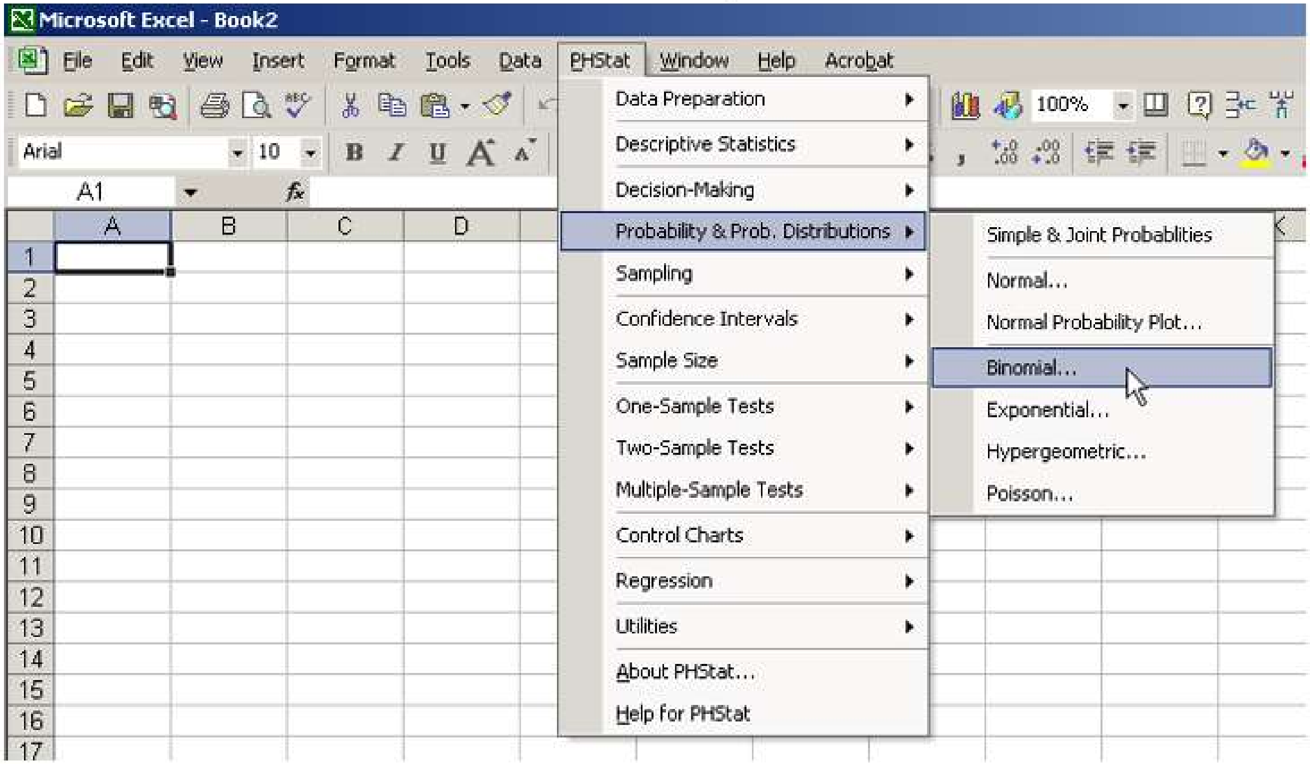
$$\mu = np = (1000)(.3) = 300.$$

- For the next 1000 customers entering the store, the variance and standard deviation for the number of customers who will make a purchase are

$$\sigma^2 = np(1 - p) = 1000(.3)(.7) = 210$$
$$\sigma = \sqrt{210} = 14.49$$

Using PHStat

Select PHStat / Probability & Prob. Distributions / Binomial…



Using PHStat

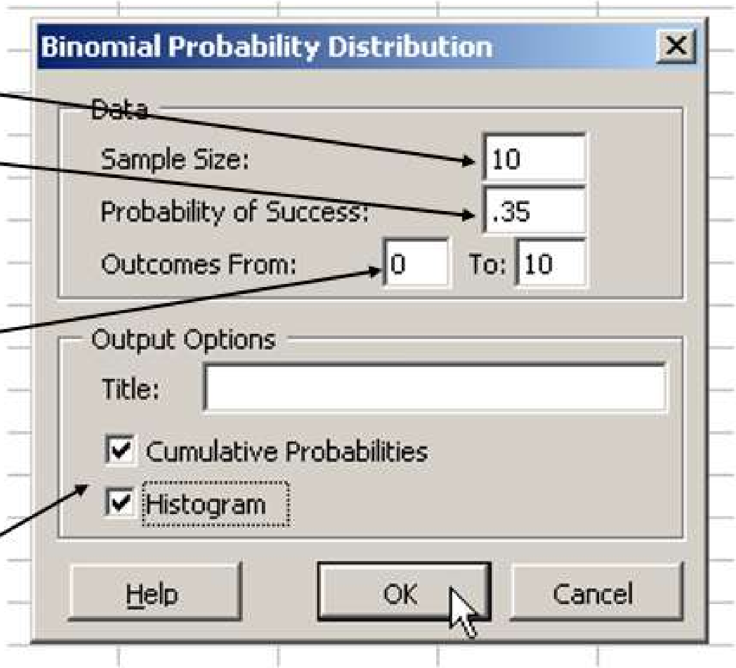
- Enter desired values in dialog box

Here: n = 10

p = .35

Output for x = 0 to x = 10 will be generated by PHStat

Optional check boxes for additional output



PHStat output

Binomial Probabilities

Data	
Sample size	10
Probability of success	0.35

Statistics	
Mean	3.5
Variance	2.275
Standard deviation	1.50831

Binomial Probabilities Table

X	P(X)	P(<=X)	P(<X)	P(>X)	P(>=X)
0	0.013463	0.013463	0	0.986537	1
1	0.072492	0.085954	0.013463	0.914046	0.986537
2	0.175653	0.261607	0.085954	0.738393	0.914046
3	0.25222	0.513827	0.261607	0.486173	0.738393
4	0.237668	0.751496	0.513827	0.248504	0.486173
5	0.15357	0.905066	0.751496	0.094934	0.248504
6	0.06891	0.973976	0.905066	0.026024	0.094934
7	0.021203	0.995179	0.973976	0.004821	0.026024
8	0.004281	0.99946	0.995179	0.00054	0.004821
9	0.000512	0.999972	0.99946	2.76E-05	0.00054
10	2.76E-05	1	0.999972	0	2.76E-05

$P(x = 3 \mid n = 10, P = .35) = .2522$

$P(x > 5 \mid n = 10, P = .35) = .0949$

Poisson Distribution

- It is another discrete distribution
- It has some similarities of Binomial distribution
- *focuses only on the number of discrete occurrences over some interval or continuum.*
- A Poisson experiment does not have a given number of trials (*n*) as a binomial experiment does
- describes the occurrence of *rare events*
- It has been referred to as the *law of improbable events*
- Each occurrence is independent of the other occurrences
- The occurrences in each interval can range from zero to infinity
- The expected number of occurrences must hold constant throughout the experiment *long-run average can be denoted by* λ
- Example:
- Number of telephone calls per minute
- Number of times a tire blows on a commercial airplane per week

$\boxed{P(x \geq 3)} = 1 - (P(x < 3)$

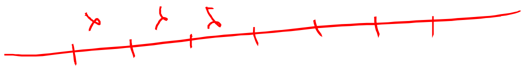
Poisson Distribution - Applications

- The number of failures in a large computer system during a given day
- The number of replacement orders for a part received by a firm in a given month
- The number of ships arriving at a loading facility during a 6-hour loading period
- The number of delivery trucks to arrive at a central warehouse in an hour
- The number of dents, scratches, or other defects in a large roll of sheet metal used to manufacture various component parts
- The number of customers to arrive for flights during each 10-minute time interval from 3:00 p.m. to 6:00 p.m. on weekdays
- The number of customers to arrive at a checkout aisle in your local grocery store during a particular time interval
- Number of telephone calls per minute
- Number of times a tire blows on a commercial airplane per week

We can use the Poisson distribution to determine the probability of each of these random variables, which are characterized as the number of occurrences or successes of a certain event in a given continuous interval (such as time, surface area, or length)

Assumptions of the Poisson Distribution

- Assume that an interval is divided into a very large number of equal subintervals so that the probability of the occurrence of an event in any subinterval is very small. The assumptions of a Poisson distribution are as follows:



- The probability of the occurrence of an event is constant for all subintervals.
- There can be no more than one occurrence in each subinterval.
- Occurrences are independent; that is, an occurrence in one interval does not influence the probability of an occurrence in another interval.

Poisson Distribution mean and Variance

- The random variable X is said to follow the Poisson distribution if it has the probability distribution

$$\boxed{P(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \text{ for } x = 0, 1, 2, \ldots}$$

- where
- $P(x)$ = the probability of x successes over a given time or space, given l
- λ = the expected number of successes per time or space unit, $\lambda > 0$
- $e \approx 2.71828$ (the base for natural logarithms)

- The mean and variance of the Poisson distribution are

$$\boxed{\mu_x = E[X] = \lambda \quad \text{and} \quad \sigma_x^2 = E\left[(X - \mu_x)^2\right] = \lambda}$$

Example:

- Bank customers arrive randomly on weekday afternoons at an average of 3.2 customers every 4 minutes. What is the probability of having more than 7 customers in a 4-minute interval on a weekday afternoon?

$\sum P(x)$ $1 - P(x \leq 7)$

$$\lambda = 3.2 \text{ customers/minutes}$$

$$x > 7 \text{ customers/4 minutes}$$

the solution requires obtaining the values of x = 8, 9, 10, 11, 12, 13, 14,. . . .∞

when the x values are away from λ the probabilities approaches 0

$$P(x = 8|\lambda = 3.2) = \frac{(3.2^8)(e^{-3.2})}{8!} = .0111$$

$$P(x = 9|\lambda = 3.2) = \frac{(3.2^9)(e^{-3.2})}{9!} = .0040$$

$$P(x = 10|\lambda = 3.2) = \frac{(3.2^{10})(e^{-3.2})}{10!} = .0013$$

$$P(x = 11|\lambda = 3.2) = \frac{(3.2^{11})(e^{-3.2})}{11!} = .0004$$

$$P(x = 12|\lambda = 3.2) = \frac{(3.2^{12})(e^{-3.2})}{12!} = .0001$$

$$P(x = 13|\lambda = 3.2) = \frac{(3.2^{13})(e^{-3.2})}{13!} = .0000$$

$$P(x > 7) = P(x \geq 8) = .0169$$

Example

- A bank has an average random arrival rate of 3.2 customers every 4 minutes. What is the probability of getting exactly 10 customers during an 8-minute interval?

$$\lambda = 3.2 \text{ customers}/\underline{\underline{4}} \text{ minutes}$$

18m

$$x = 10 \text{ customers}/8 \text{ minutes}$$

The intervals for lambda and the sample are different

- The intervals must be the same in order to use λ and x together in the probability formula.
- The right way to approach this dilemma is to adjust the interval for lambda so that it and x have the same interval.
- so lambda should be adjusted to an 8-minute interval

$$\lambda = 6.4 \text{ customers}/8 \text{ minutes}$$

$$\underline{\underline{x}} = 10 \text{ customers}/8 \text{ minutes}$$

$$\frac{(6.4)^{10} e^{-6.4}}{10!} = .0528$$

Poisson probability Distribution Table

SELECTED VALUES FROM THE POISSON PROBABILITY TABLES
EXAMPLE: $\mu = 10, x = 5; f(5) = .0378$

	μ									
x	9.1	9.2	9.3	9.4	9.5	9.6	9.7	9.8	9.9	10
0	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0000
1	.0010	.0009	.0009	.0008	.0007	.0007	.0006	.0005	.0005	.0005
2	.0046	.0043	.0040	.0037	.0034	.0031	.0029	.0027	.0025	.0023
3	.0140	.0131	.0123	.0115	.0107	.0100	.0093	.0087	.0081	.0076
4	.0319	.0302	.0285	.0269	.0254	.0240	.0226	.0213	.0201	.0189
5	.0581	.0555	.0530	.0506	.0483	.0460	.0439	.0418	.0398	.0378
6	.0881	.0851	.0822	.0793	.0764	.0736	.0709	.0682	.0656	.0631
7	.1145	.1118	.1091	.1064	.1037	.1010	.0982	.0955	.0928	.0901
8	.1302	.1286	.1269	.1251	.1232	.1212	.1191	.1170	.1148	.1126
9	.1317	.1315	.1311	.1306	.1300	.1293	.1284	.1274	.1263	.1251
10	.1198	.1210	.1219	.1228	.1235	.1241	.1245	.1249	.1250	.1251
11	.0991	.1012	.1031	.1049	.1067	.1083	.1098	.1112	.1125	.1137
12	.0752	.0776	.0799	.0822	.0844	.0866	.0888	.0908	.0928	.0948
13	.0526	.0549	.0572	.0594	.0617	.0640	.0662	.0685	.0707	.0729
14	.0342	.0361	.0380	.0399	.0419	.0439	.0459	.0479	.0500	.0521
15	.0208	.0221	.0235	.0250	.0265	.0281	.0297	.0313	.0330	.0347
16	.0118	.0127	.0137	.0147	.0157	.0168	.0180	.0192	.0204	.0217
17	.0063	.0069	.0075	.0081	.0088	.0095	.0103	.0111	.0119	.0128
18	.0032	.0035	.0039	.0042	.0046	.0051	.0055	.0060	.0065	.0071
19	.0015	.0017	.0019	.0021	.0023	.0026	.0028	.0031	.0034	.0037
20	.0007	.0008	.0009	.0010	.0011	.0012	.0014	.0015	.0017	.0019
21	.0003	.0003	.0004	.0004	.0005	.0006	.0006	.0007	.0008	.0009
22	.0001	.0001	.0002	.0002	.0002	.0002	.0003	.0003	.0004	.0004
23	.0000	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002
24	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001

Poisson Approximation to the Binomial Distribution

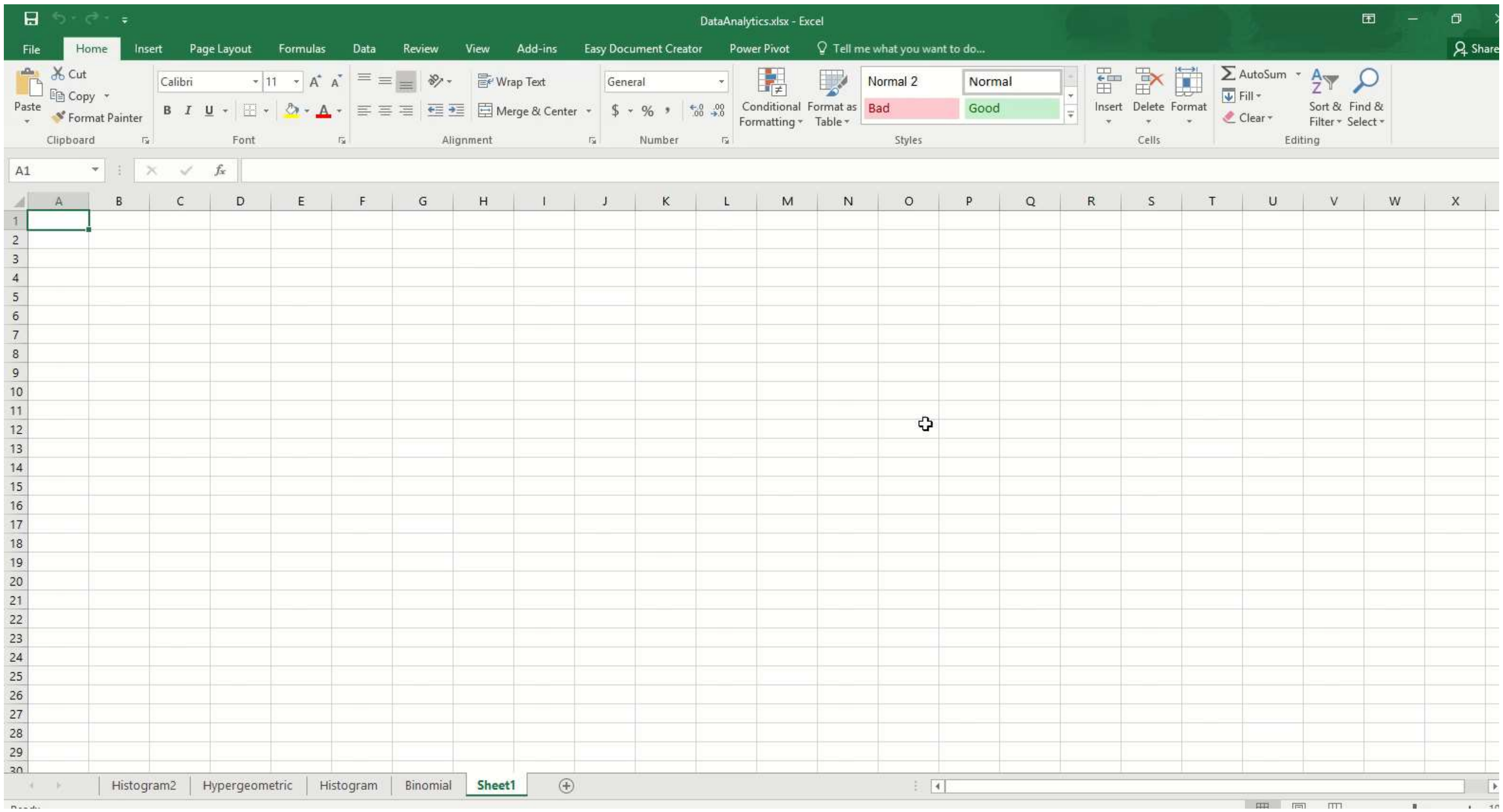
- Poisson probabilities can be directly computed from the binomial probability distribution by taking the mathematical limits as $P \to 0$ and $n \to \infty$.
- With these limits, the parameter $\lambda = nP$ is a constant that specifies the average number of occurrences (successes) for a particular time and/or space.
- As a rule of thumb, if n > 20 and np ≤7, the approximation is close enough to use the Poisson distribution for binomial problems

$\lambda = np$

- Examples:
- For a insurance company the probability that a single policy will result in a claim during the year is small (No of policies are more)
- If the probability that any one of them will break down in a single day is small (No of machines are more)

$$\boxed{P(x) = \frac{e^{-nP}(nP)^x}{x!} \text{ for } x = 0, 1, 2, \ldots}$$

Poisson distribution in PHStat



Hypergeometric Distribution

- To complement the types of analysis that can be made by using the binomial distribution
- Binomial distribution applies only to the experiments in which the trials are done with replacement (independent events)
- The hypergeometric distribution has the following characteristics:
- It is discrete distribution.
- Each outcome consists of either a success or a failure.
- Sampling is done without replacement.
- The population, N, is finite and known.
- The number of successes in the population, A, is known.

Binomial

$\frac{1}{2}$, $\frac{1}{2}$

with

$1 \rightarrow \frac{1}{52}$

$2 \rightarrow \frac{1}{51}$

$3 \rightarrow 1$ without

Hyper

Hypergeometric Distribution

Suc Failure

$$P(x) = \frac{{}_{A}C_{x} \cdot {}_{N-A}C_{n-x}}{{}_{N}C_{n}}$$

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

- where
- *N* = size of the population *n* = sample size *A* = number of successes in the population
- *x* = number of successes in the sample; sampling is done *without* replacement
- the binomial distribution - sampling with replacement.
- hypergeometric distribution—sampling without replacement, the probabilities change with each selection
- If the population is large (*N* > 10, 000) and the sample size is small (< 1%),

$n \to \infty$

- Then the change in probability after each draw is very small.
- In those situations the binomial is a very good approximation and is typically used.
- The mean and variance of a hypergeometric distribution are as follows.

$$E(x) = \mu = n\left(\frac{A}{N}\right)$$

$$\text{Var}(x) = \sigma^2 = n\left(\frac{A}{N}\right)\left(1 - \frac{A}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Example: quality control application

- Electric fuses produced by Ontario Electric are packaged in boxes of 12 units each. Suppose an inspector randomly selects three of the 12 fuses in a box for testing. If the box contains exactly five defective fuses, what is the probability that the inspector will find exactly one of the three fuses defective?

Solution:

n = 3 and N = 12. With A = 5 defective fuses in the box the probability of finding x = 1 defective fuse is

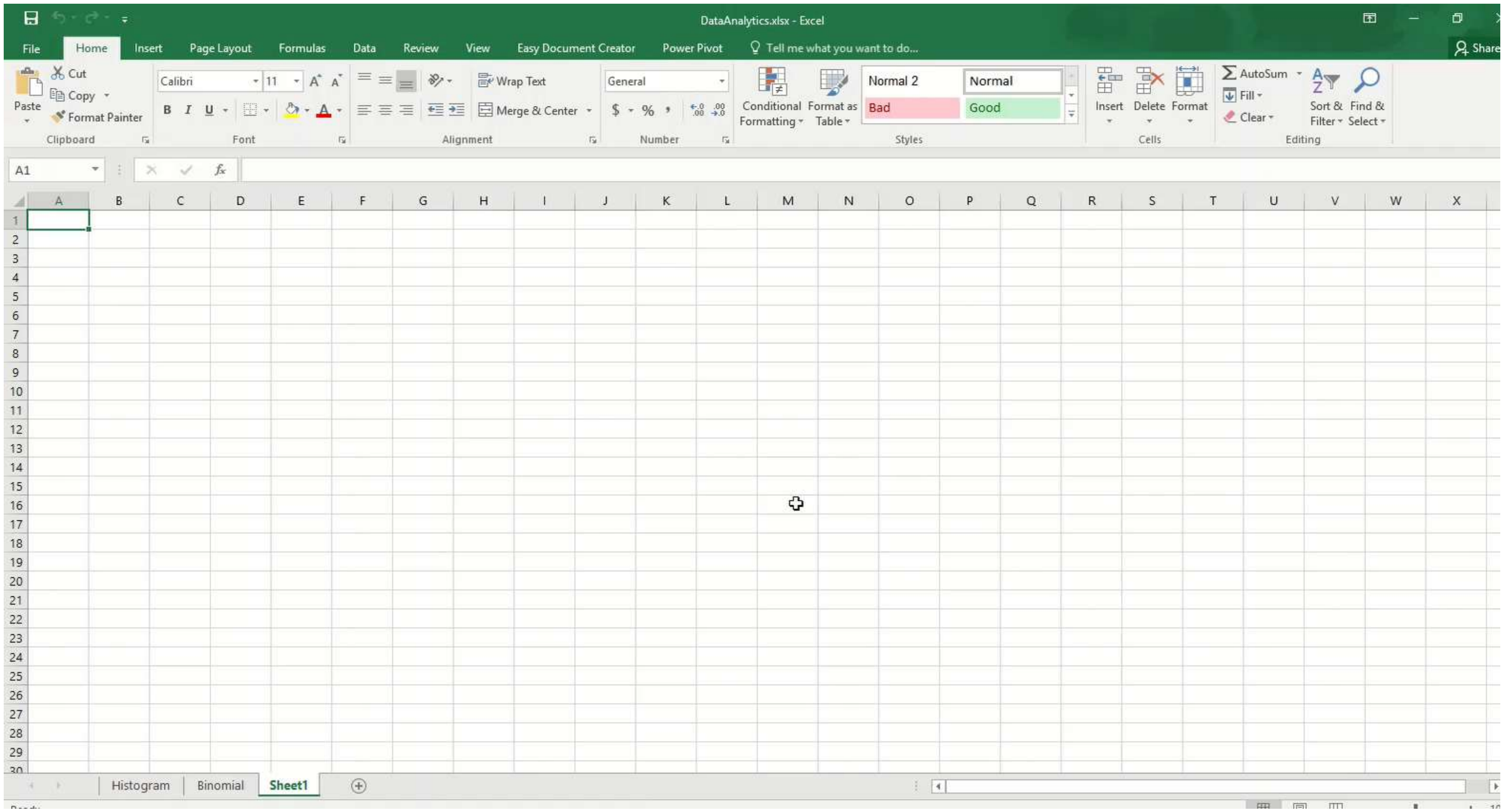
$$f(1) = \frac{\binom{5}{1}\binom{7}{2}}{\binom{12}{3}} = \frac{\left(\frac{5!}{1!4!}\right)\left(\frac{7!}{2!5!}\right)}{\left(\frac{12!}{3!9!}\right)} = \frac{(5)(21)}{220} = .4773$$

The probability of finding at least 1 defective fuse=1-f(0)

$$f(0) = \frac{\binom{5}{0}\binom{7}{3}}{\binom{12}{3}} = \frac{\left(\frac{5!}{0!5!}\right)\left(\frac{7!}{3!4!}\right)}{\left(\frac{12!}{3!9!}\right)} = \frac{(1)(35)}{220} = .1591$$

The probability of finding at least one defective fuse must be 1 - 0.1591=0.8409.

Hyper geometric distribution in PHStat



Continuous Probability Distributions

- A continuous random variable is a variable that can assume any value in an interval
  - –thickness of an item
  - –time required to complete a task
  - –temperature of a solution
  - –height, in inches
- Continuous distributions are constructed from continuous random variables
- The values are taken on for every point over a given interval
- Values are usually generated from experiments in which things are "measured" as opposed to "counted."
- The **cumulative distribution function**, *F(x)*, for a continuous random variable *X* expresses the probability that *X* does not exceed the value of *x*, as a function of *x*:

$P(x=1)$

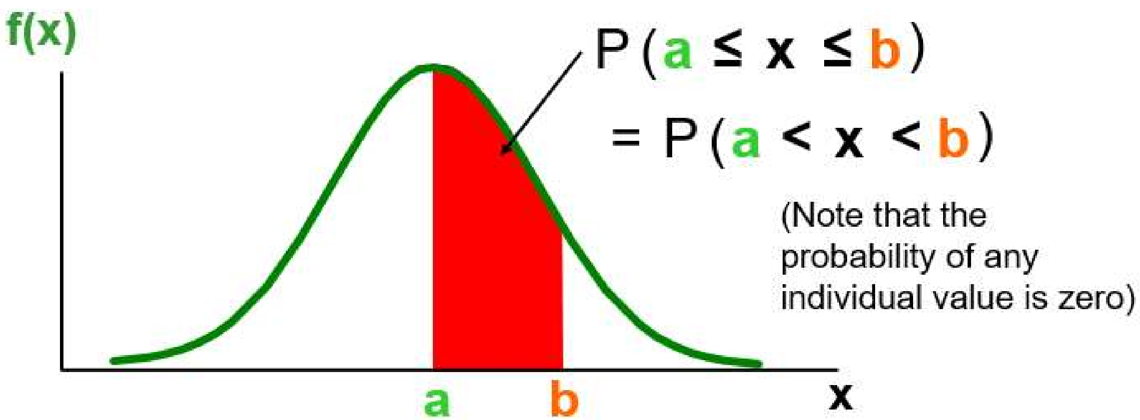
$$\boxed{F(x) = P(X \leq x)}$$

$$\boxed{f(x_0) = \int_{x_m}^{x_0} f(x)dx}$$

where x_m is the minimum value of the random variable x

Continuous Probability Distributions

- The probability density function, f(x), of random variable X has the following properties:
- f(x)>0 for all values of x
- probabilities of outcomes occurring between particular points is the area under the curve between those points.



- In addition, the entire area under the whole curve is equal to 1.

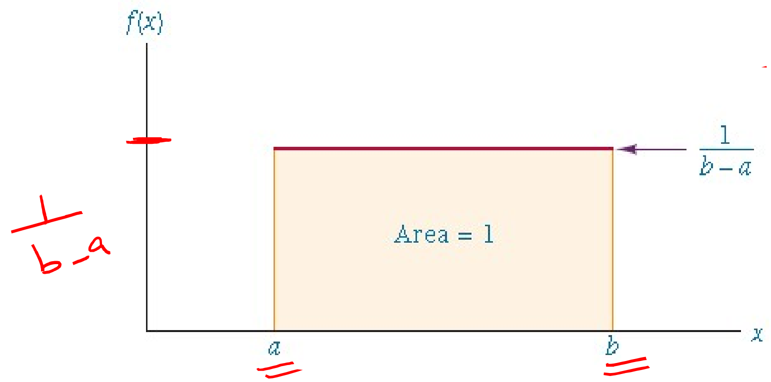
- The uniform distribution, the normal distribution, the exponential distribution, the *t* distribution, the chi-square distribution, and the *F* distribution.
- The probability of any particular value of the random variable is zero

Uniform distribution

rolling die

X	P(x)
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

- It is referred to as the **rectangular distribution**,
- It is a relatively simple continuous distribution
- The same height, or f (x), is obtained over a range of values.



- Area of Rectangle = (Length)(Height) = 1
- Length = (*b* - *a*)
- (*b* - *a*)(Height) = 1

$$\text{Height} = \frac{1}{(b-a)}$$

UNIFORM PROBABILITY DENSITY FUNCTION

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{elsewhere} \end{cases}$$

The mean and standard deviation of a uniform distribution

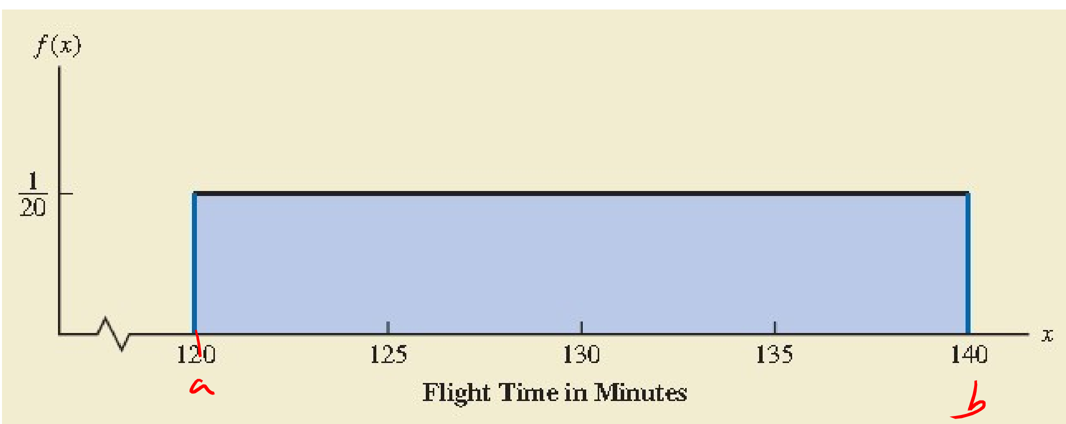
$$\mu = \frac{a+b}{2}$$

$$\sigma = \frac{b-a}{\sqrt{12}}$$

An important special case of these results is the standardized random variable

Example

- Consider the random variable *x* representing the flight time of an airplane traveling from Chicago to New York.
- Suppose the flight time can be any value in the interval from 120 minutes to 140 minutes.
- Because the random variable *x* can assume any value in that interval, *x* is a continuous rather than a discrete random variable.
- Let us assume that sufficient actual flight data are available to conclude that the probability of a flight time within any 1-minute interval is the same as the probability of a flight time within any other 1-minute interval contained in the larger interval from 120 to 140 minutes.



$\frac{1}{b-a}$

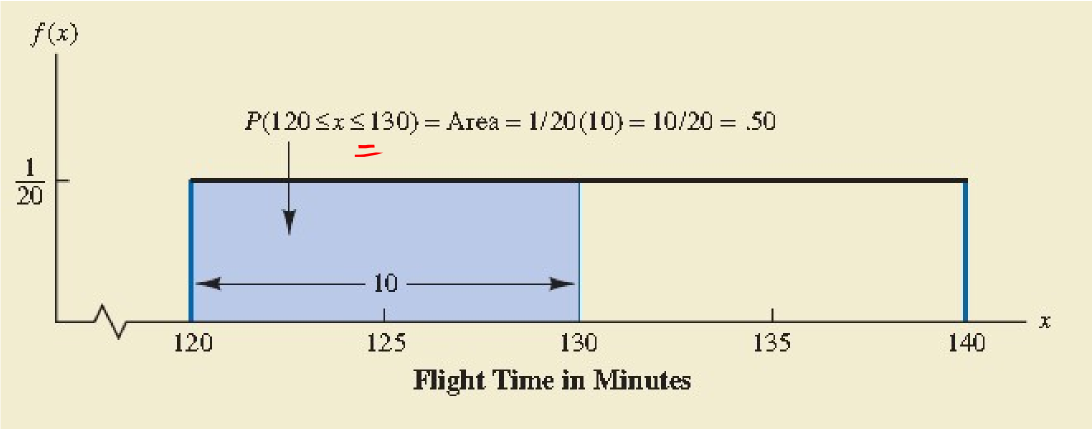
$$f(x) = \begin{cases} 1/20 & \text{for } 120 \leq x \leq 140 \\ 0 & \text{elsewhere} \end{cases}$$

$$E(x) = \frac{(120 + 140)}{2} = 130$$

$$\text{Var}(x) = \frac{(140 - 120)^2}{12} = 33.33$$

Example

- What is the probability that the flight time is between 120 and 130 minutes?
- That is, what is $P(120 \leq x \leq 130)$?
- Consider the area under the graph of $f(x)$ in the interval from 120 to 130.
- The area is rectangular, and the area of a rectangle is simply the width multiplied by the height.
- With the width of the interval equal to 130 -120 = 10 and the height equal to the value of the probability density function $f(x) = 1/20$,
- we have area = width x height =10(1/20)= 10/20 = 0.50

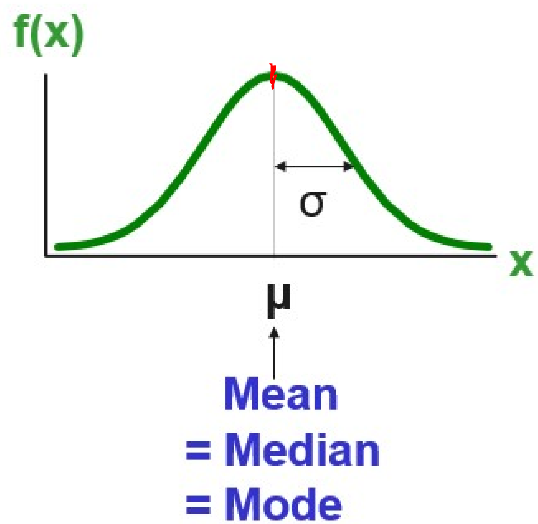


Normal Distribution

- Most often used for economics and business applications
- Many variables in business and industry also are normally distributed.
- There are many reasons for its wide application.
- The normal distribution closely approximates the probability distributions of a wide range of random variables
- Distributions of sample means approach a normal distribution, given a “large” sample size
- Computation of probabilities is direct and elegant
- The most important reason is that the normal probability distribution has led to good business decisions for a number of applications

Normal Distribution

- Probably the most widely known and used of all distributions is the **normal distribution**.
- The normal curve sometimes is referred to as the *Gaussian distribution* or the *normal curve of error or bell-shaped curve*
- The normal distribution exhibits the following characteristics.
  - –It is a continuous distribution.
  - –It is a symmetrical distribution about its mean.
  - –It is asymptotic to the horizontal axis.
  - –It is unimodal.
  - –It is a family of curves.
  - –Area under the curve is 1.
  - –Mean , median and mode are equal



Location is determined by the mean, μ

Spread is determined by the standard deviation, σ

The random variable has an infinite theoretical range: $+\infty$ to $-\infty$

Probability Density Function of the Normal Distribution

- The normal distribution is described or characterized by two parameters: the mean, μ , and the standard deviation σ
- The density function of the normal distribution is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2[(x-\mu)/\sigma]^2}$$

μ mean

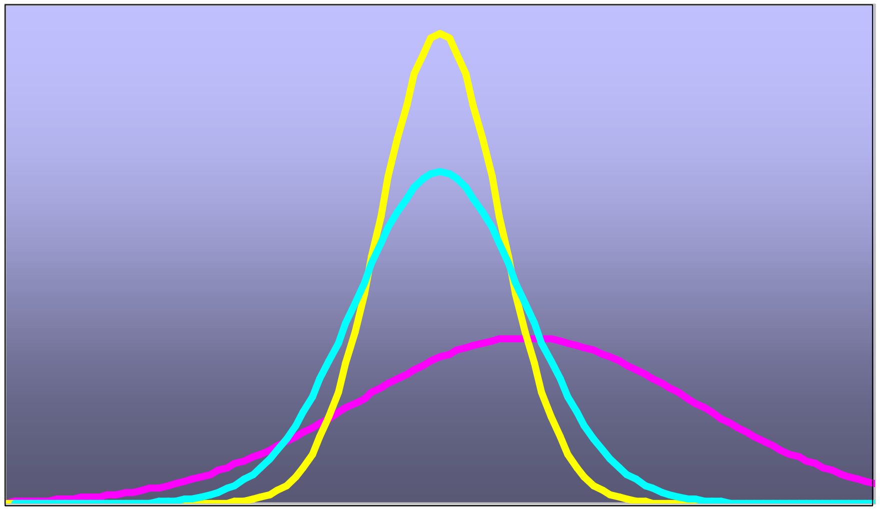
σ standard deviation

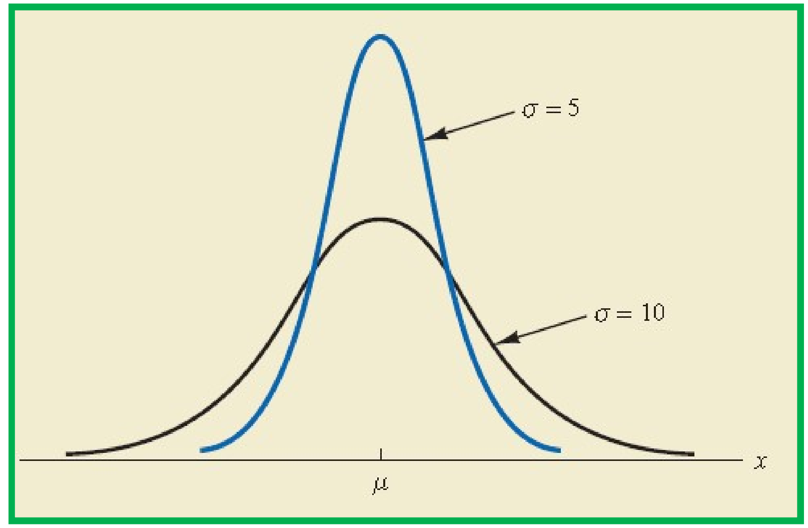
π 3.14159

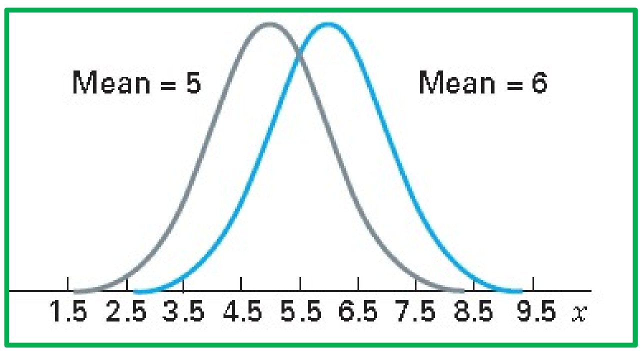
e 2.71828

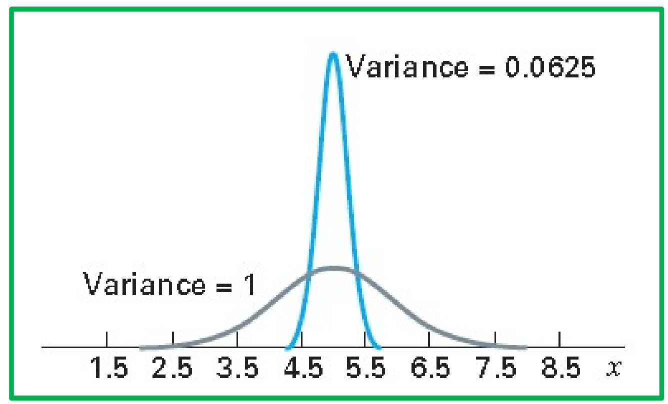
- It is represented as $\boxed{X \sim N(\mu, \sigma^2)}$

Many Normal Distributions







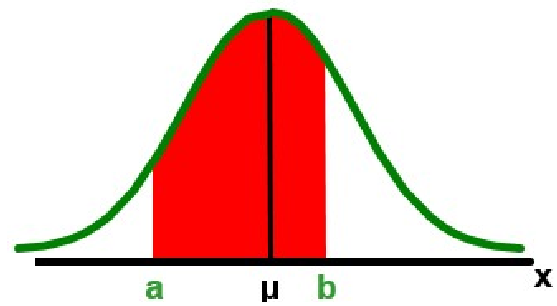


By varying the parameters μ and σ, we obtain different normal distributions

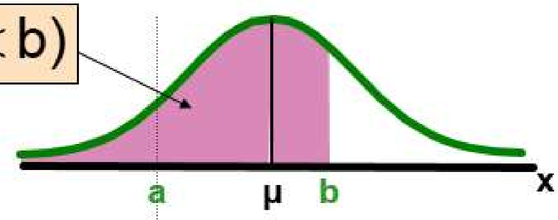
Finding Normal Probabilities

The probability for a range of values is measured by the area under the curve

$$\boxed{P(a < X < b) = F(b) - F(a)}$$



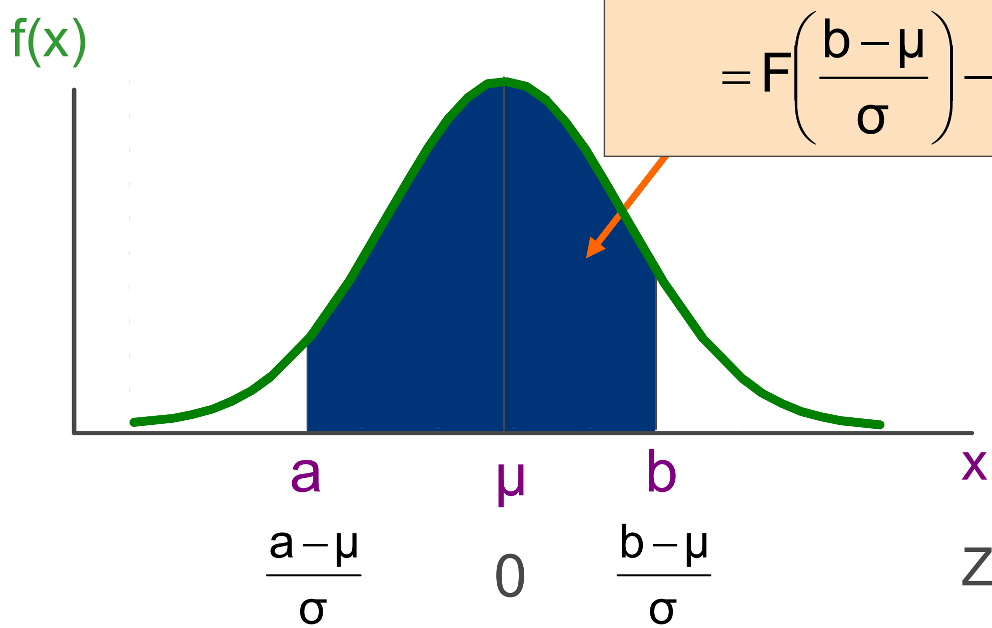
$$\boxed{F(b) = P(X < b)}$$



Finding Normal Probabilities

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right)$$

$$= F\left(\frac{b-\mu}{\sigma}\right) - F\left(\frac{a-\mu}{\sigma}\right)$$



Standard Normal Probability Distribution

- Using Integral Calculus to determine areas under the normal curve from this function is difficult and time-consuming
- a mechanism was developed by which all normal distributions can be converted into a single distribution: the z distribution.
- This process yields the standardized normal distribution

$$z = \frac{x - \mu}{\sigma}, \qquad \sigma \neq 0$$

- A ***z* score** is *the number of standard deviations that a value, x, is above or below the mean.*
- If the value of *x* is less than the mean, the *z* score is negative;
- if the value of *x* is more than the mean, the *z* score is positive;
- if the value of *x* equals the mean, the associated *z* score is zero.
- This formula allows conversion of the distance of any *x* value from its mean intostandard deviation units.

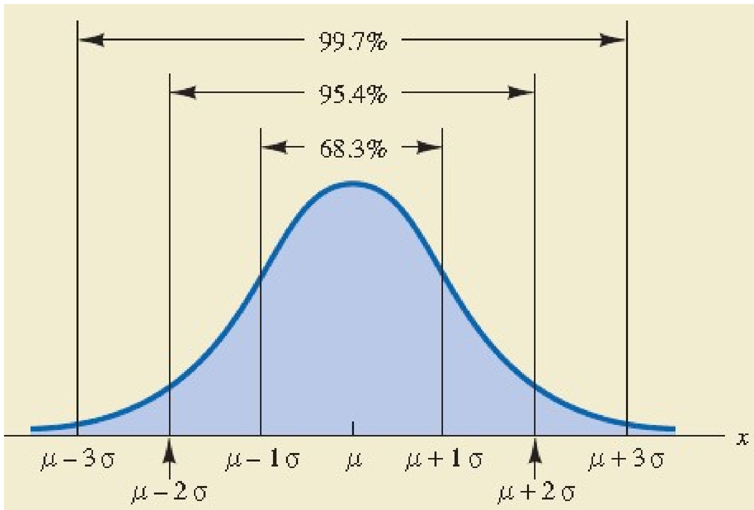
Standard Normal Probability Distribution

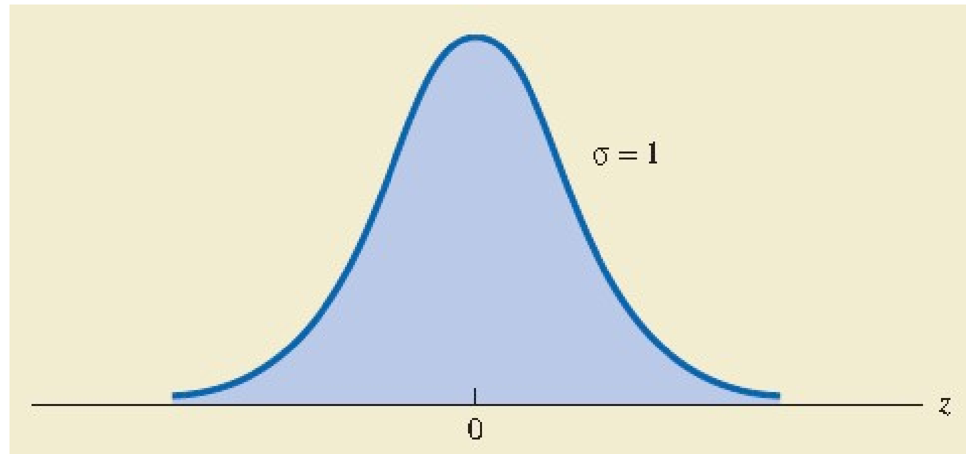
- A standard *z* score table can be used to find probabilities for any normal curve problem that has been converted to *z* scores
- The ***z* distribution** is *a normal distribution with a mean of 0 and a standard deviation of 1*
- Any value of *x* at the mean of a normal curve is zero standard deviations from the mean
- Any value of *x* that is one standard deviation above the mean has a *z* value of 1
- The empirical rule is based on the normal distribution in which about 68% of all values are within one standard deviation of the mean regardless of the values of μand σ .
- In a *z* distribution, about 68% of the *z* values are between *z* = -1 and *z* = +1.

STANDARD NORMAL DENSITY FUNCTION

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

Standard Normal Probability Distribution



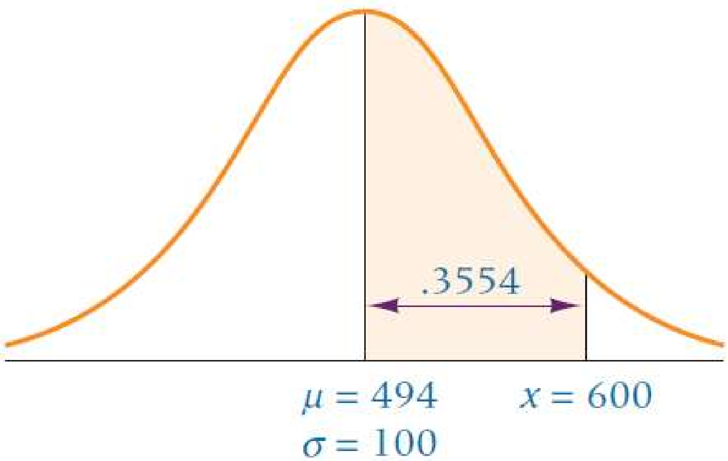


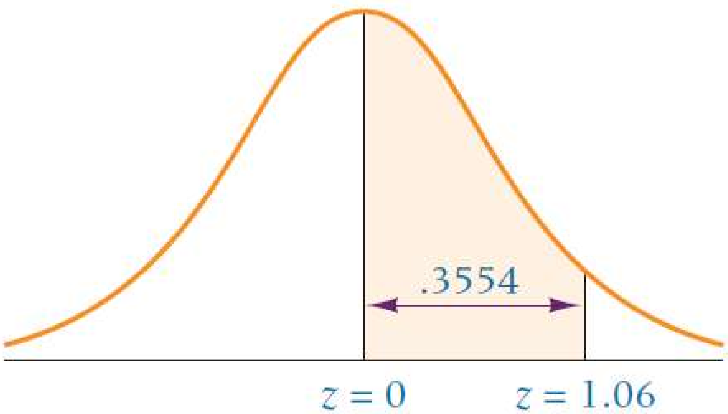
Example

- The Graduate Management Aptitude Test (GMAT), produced by the Educational Testing Service in Princeton, New Jersey, is widely used by graduate schools of business in the United States as an entrance requirement. Assuming that the scores are normally distributed, probabilities of achieving scores over various ranges of the GMAT can be determined. In a recent year, the mean GMAT score was 494 and the standard deviation was about 100.What is the probability that a randomly selected score from this administration of the GMAT is between 600 and the mean?

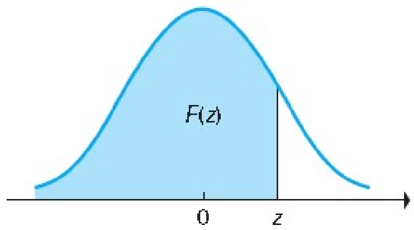
- x = 600

$$z = \frac{x - \mu}{\sigma} = \frac{600 - 494}{100} = \frac{106}{100} = 1.06$$





Cumulative Distribution function F(Z) of the standard Normal Distribution



z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441

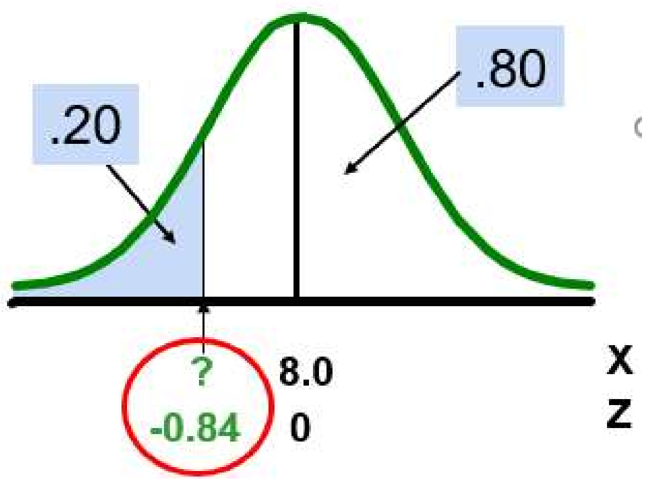
Finding the X value for a Known Probability

- Let X represent the time it takes (in seconds) to download an image file from the internet. Suppose X is normal with mean 8.0 and standard deviation 5.0. We have to find X such that 20% of download times are less than X.

Standardized Normal Probability Table (Portion)

z	F(z)
.82	.7939
.83	.7967
.84	.7995
.85	.8023

20% area in the lower tail is consistent with a Z value of -0.84



$$X = \mu + Z\sigma$$
$$= 8.0 + (-0.84)5.0$$
$$= 3.80$$

So 20% of the values from a distribution with mean 8.0 and standard deviation 5.0 are less than 3.80

Example:

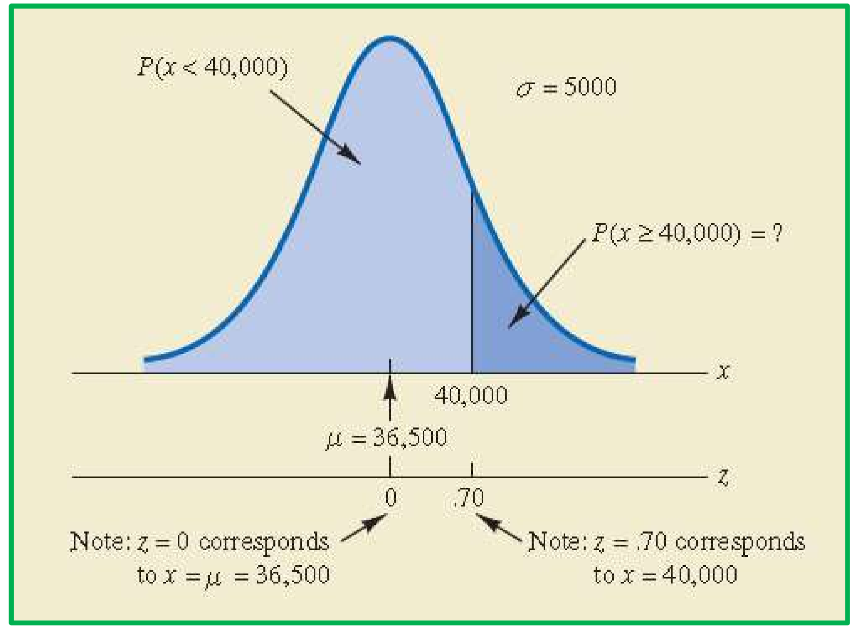
- Suppose the Grear Tire Company developed a new steel-belted radial tire to be sold through a national chain of discount stores.
- Because the tire is a new product, Grear's managers believe that the mileage
- guarantee offered with the tire will be an important factor in the acceptance of the product.
- Before finalizing the tire mileage guarantee policy, Grear's managers want probability information about x number of miles the tires will last.
- From actual road tests with the tires, Grear's engineering group estimated that the mean tire mileage is μ 36,500 miles and that the standard deviation is σ 5000. In addition, the data collected indicate that a normal distribution is a reasonable assumption.
- What percentage of the tires can be expected to last more than 40,000 miles?
- In other words, what is the probability that the tire mileage, x, will exceed 40,000?

Example:

- At *x* 40,000, we have

$$z = \frac{x - \mu}{\sigma} = \frac{40{,}000 - 36{,}500}{5000} = \frac{3500}{5000} = .70$$

- the area under the standard normal curve to the left of z = 0.70 is 0.7580.
- Thus, 1.000 -0 .7580 = 0.2420 is the probability
- that z will exceed 0.70 and hence x will exceed 40,000.
- We can conclude that about 24.2% of the tires will exceed 40,000 in mileage.



Example:

- Let us now assume that Grear is considering a guarantee that will provide a discount on replacement tires if the original tires do not provide the guaranteed mileage. What should the guarantee mileage be if Grear wants no more than 10% of the tires to be eligible for the discount guarantee?
- The area under the curve to the left of the unknown guarantee mileage must be .10.
- Z=-1.28 cuts off an area of .10 in the lower tail
- To find the value of x corresponding to z=-1.28,

$$z = \frac{x - \mu}{\sigma} = -1.28$$

$$x - \mu = -1.28\sigma$$

$$x = \mu - 1.28\sigma$$

- With μ = 36,500 and σ = 5000,
- x = 36,500 - 1.28(5000) = 30,100

- Thus, a guarantee of 30,100 miles will meet the requirement that approximately 10% of the tires will be eligible for the guarantee. Perhaps, with this information, the firm will set its tire mileage guarantee at 30,000 miles.

Normal Approximation of Binomial Approximation

- The shape of the binomial distribution is approximately normal if n is large
- When the number of trials becomes large, evaluating the binomial probability function by hand or with a calculator is difficult.
- In cases where $np \geq 5$, and $n(1- \ p) \geq 5$, the normal distribution provides an easy-to-use approximation of binomial probabilities.
- When using the normal approximation to the binomial, we set $\mu \ np$ and $\sigma = \sqrt{np(1-p)}$ in the definition of the normal curve.
- Standardize to Z from a binomial distribution

$$Z = \frac{X - E(X)}{\sqrt{Var(X)}} = \frac{X - np}{\sqrt{nP(1-P)}}$$

$$E(X) = \mu = nP$$

$$Var(X) = \sigma^2 = nP(1\text{-}P)$$

$$P(a < X < b) = P\left(\frac{a - nP}{\sqrt{nP(1-P)}} \leq Z \leq \frac{b - nP}{\sqrt{nP(1-P)}} \right)$$

Exponential Distribution

- Another useful continuous distribution is the exponential distribution.
- It is closely related to the Poisson distribution.
- Whereas the Poisson distribution is discrete and Describes random occurrences over some interval,
- the **exponential distribution** is *continuous and describes a probability distribution of the times between random occurrences.*
- The following are the characteristics of the exponential distribution.
  - It is a continuous distribution.
  - It is a family of distributions.
  - It is skewed to the right.
  - The x values range from zero to infinity.
  - Its apex is always at $x = 0$.
  - The curve steadily decreases as x gets larger.
- Examples:the time between arrivals at a car wash, the time required to load a truck, the distance between major defects in a highway, and so on.

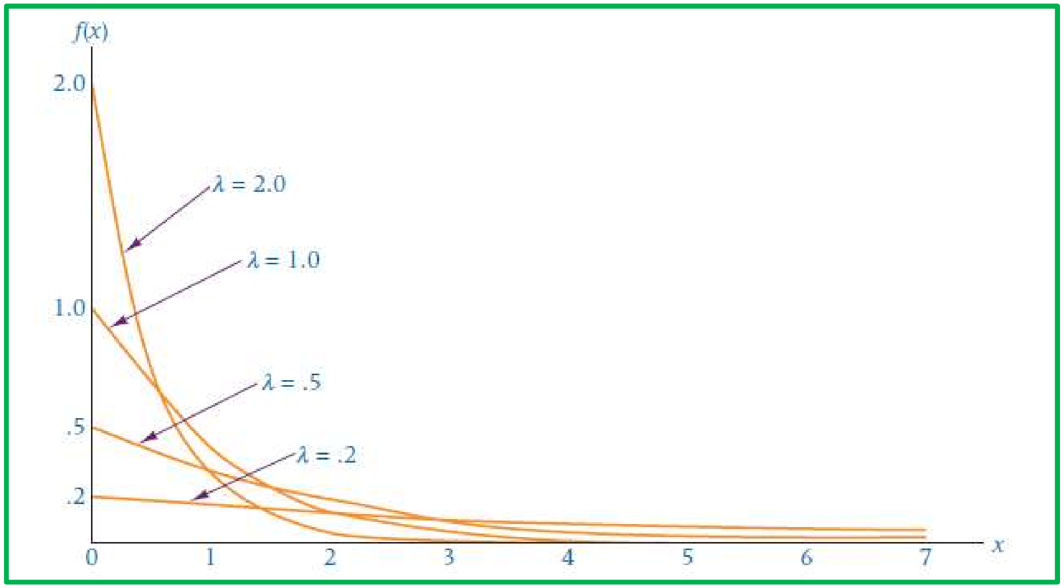
Exponential Distribution

- The exponential probability distribution is determined by the following

$$f(t) = \lambda e^{-\lambda t} \quad \text{for } t > 0$$

- where λ is the mean number of independent arrivals per time unit,
- *t* is the number of time units until the next arrival,
- *e* = 2.71828
- The cumulative distribution function is as follows

$$F(t) = 1 - e^{-\lambda t} \quad \text{for } t > 0$$



Example

- Service times for customers at a library information desk can be modeled by an exponential distribution with a mean service time of 5 minutes. What is the probability that a customer service time will take longer than 10 minutes?
- Solution Let t denote the service time in minutes. The service rate is $\lambda = 1/5 = 0.2$ per minute,

$$\begin{aligned} P(T > 10) &= 1 - P(T < 10) \\ &= 1 - F(10) \\ &= 1 - (1 - e^{-(0.20)(10)}) \\ &= e^{-2.0} = 0.1353 \end{aligned}$$

Thus, the probability that a service time exceeds 10 minutes is 0.1353.

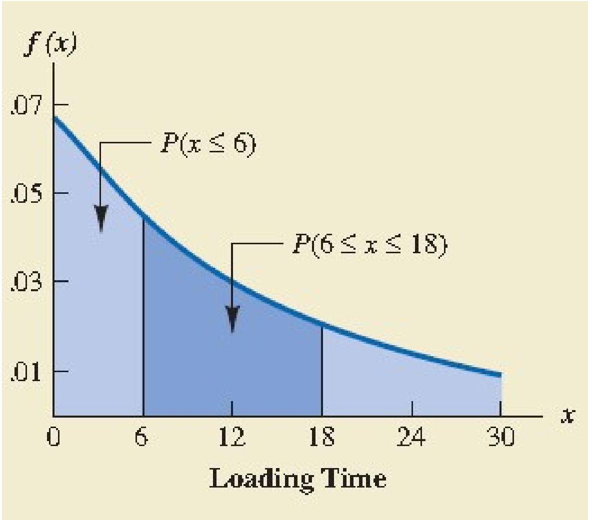
EXPONENTIAL DISTRIBUTION FOR THE SCHIPS LOADING DOCK EXAMPLE

- In the Ships loading dock example, the probability that loading a truck will take 6 minutes or less $P(x \leq 6)$ is defined to be the area under the curve

- For the Schips loading dock example, x loading time in minutes and μ 15 minutes

$$P(x \leq x_0) = 1 - e^{-x_0/15}$$

$$P(x \leq 6) = 1 - e^{-6/15} = .3297$$



- we calculate the probability of loading a truck in 18 minutes or less.

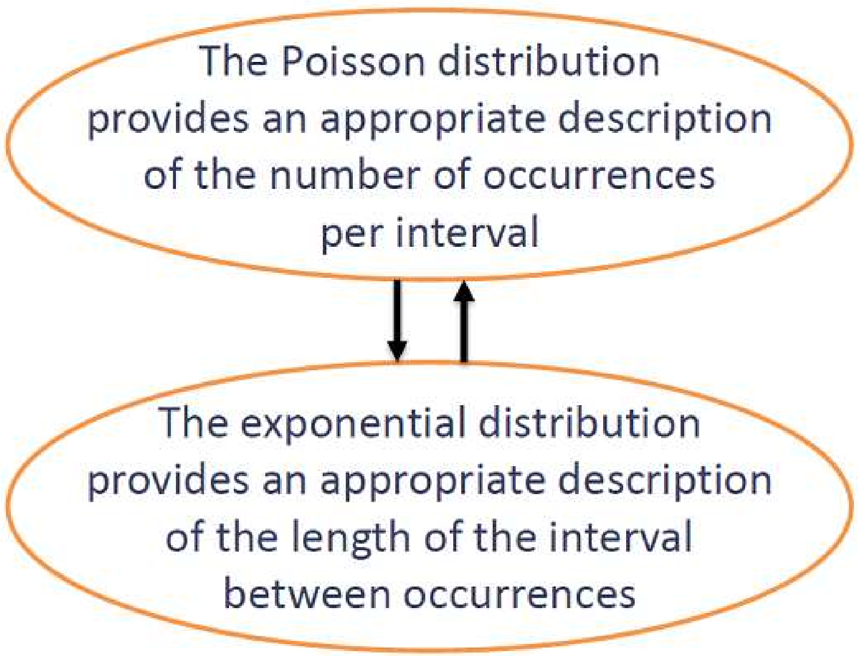
$$P(x \leq 18) = 1 - e^{-18/15} = .6988$$

Thus, the probability that loading a truck will take between 6 minutes and 18 minutes is equal to

0.6988 - 0.3297= 0.3691

Relationship between the Poisson and Exponential Distribution

Relationship between the Poisson and Exponential Distributions



Example

- To illustrate this relationship, suppose the number of cars that arrive at a car wash during one hour is described by a Poisson probability distribution with a mean of 10 cars per hour. The Poisson probability function that gives the probability of *x* arrivals per hour is

$$f(x) = \frac{10^x e^{-10}}{x!}$$

- Because the average number of arrivals is 10 cars per hour, the average time between cars arriving is

$$\frac{1 \text{ hour}}{10 \text{ cars}} = .1 \text{ hour/car}$$

- Thus, the corresponding exponential distribution that describes the time between the arrivals has a mean of μ= .1 hour per car; as a result, the appropriate exponential probability density function

$$\boxed{f(x) = \frac{1}{.1} e^{-x/.1} = 10e^{-10x}}$$

Python Demo for Distributions

- need to import the required libraries.

```
import scipy
import numpy as np
```

Binomial Distribution

```
from scipy.stats import binom
```

A survey found that 65% of all financial customers were satisfied with their primary financial institution. Suppose that 25 financial consumers are sampled and if survey result still holds true today, what is the probability that exactly 19 are very satisfied with their primary financial institution?

we can use binomial distribution since the outcome is either a success or failure

```
print(binom.pmf(k=19,n=25,p=0.65))
```

```
0.09077799859322791
```

Python Demo for Distributions

According to US census bureau, approximately 6% of all workers in Jackson, Mississippi, are unemployed. In conducting a random telephone survey in Jackson, what is the probability of getting two or fewer unemployed workers in a sample of 20?

```
binom.cdf(2,20,0.06)
```

```
0.8850275957378545
```

- **Poisson Distribution**

```
from scipy.stats import poisson
```

- ***Suppose bank customers arrive randomly on weekday afternoons at an average of 3.2 customers every 4 minutes. What is the probability of exactly 5 customers arriving in a 4 minute interval on a weekday afternoon?***

```
poisson.pmf(5,3.2)
```

```
0.11397938346351824
```

Python Demo for Distributions

- ***Bank customers arrive randomly on weekday afternoons at an average of 3.2 customers every 4 minutes. What is the probability of having more than 5 customers arriving in a 4 minute interval on a weekday afternoon?***

```
prob=poisson.cdf(5,3.2)
prob_more_than5=1-prob
prob_more_than5
```

```
0.10540810546917734
```

- ***A bank has an average arrival rate of 3.2 customers every 4 minutes. What is the probability of getting exactly 10 customers in a 8 minute interval?***
- for 4 minute interval, the average is 3.2 customers. So, for 8 minute interval, the average will be twice, ie, 6.4 customers.

```
poisson.pmf(10,6.4)
```

```
0.052790043854115495
```

Python Demo for Distributions

- **Uniform Distribution**

```
from scipy.stats import uniform
```

- ***Suppose the amount of time it takes to assemble a plastic module ranges from 27 to 39 seconds and the assembly time is uniformly distributed. Describe the distribution? What is the probability that a given assembly will take between 30 to 35 seconds?***

- Here we have an array that ranges from 27 to 39 that is uniformly distributed. So, we use *np.arange(starting value, end_value+1, increment value)* to create the array.

```
U=np.arange(27,40,1)
U
```

```
array([27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39])
```

- to find the mean of the values

```
uniform.mean(loc=27,scale=12)
```

```
33.0
```

Python Demo for Distributions

- to find the probability that the assembly will take 30 to 35 seconds

```
uniform.cdf(np.arange(30,36,1),loc=27,scale=12)
```

```
array([0.25      , 0.33333333, 0.41666667, 0.5       , 0.58333333,
       0.66666667])
```

- to find the probability of the interval, we need to subtract both the probability values.

```
Prob=0.66666667 - 0.25
Prob
```

```
0.41666667
```

Python Demo for Distributions

- ***According to the National Association of Insurance Commissioners, the average annual cost for automobile insurance in the United States in the recent years was $691. Suppose automobile insurance costs are uniformly distributed across US with a range of $200 to $1,182.What is the standard deviation of this uniform distribution?***

```
uniform.mean(loc=200,scale=982)
```

```
691.0
```

```
uniform.std(loc=200,scale=982)
```

```
283.4789821721062
```

- **Normal Distribution**

```
from scipy.stats import norm
```

- If the x value is 68, mean is 65.5 and standard deviation is 2.5, then, probability of obtaining a value less than or equal to 68 is

```
val,m,s=68,65.5,2.5
print(norm.cdf(val,m,s))
```

```
0.8413447460685429
```

Python Demo for Distributions

- If we have to find the cumulative value of x>value, ie, the probability of occurrence of values above 68, then, we subtract the above cdf value from one, ie,

```
print(1-norm.cdf(val,m,s))
```

```
0.15865525393145707
```

- If we have to find the cumulative value of probability of values in between an interval, ie, val1<x<val2, then,

```
print(norm.cdf(val,m,s)-(norm.cdf(63,m,s)))
```

```
0.6826894921370859
```

Python Demo for Distributions

- ***What is the probability of obtaining a score greater than 700 on a GMAT test that has a mean of 494 and a standard deviation of 100? Assume that GMAT scores are normally distributed.***
- here, we have to find the value of P(x>700),where mean =494 and standard deviation=100.

```
print(1-norm.cdf(700,494,100))
```

```
0.019699270409376912
```

- ***For the same GMAT examination, what is the probability of randomly drawing a score that is 550 or less?***

```
print(norm.cdf(550,494,100))
```

```
0.712260281150973
```

- ***What is the probability of randomly obtaining a score between 300 and 600 on a GMAT exam?***

```
print(norm.cdf(450,494,100)-norm.cdf(350,494,100))
```

```
0.2550348541262666
```

Python Demo for Distributions

- If we are given the area under the curve, then, we can find the z value,

```
norm.ppf(0.95)
```

```
1.6448536269514722
```

- **Hypergeometric Distribution**

```
from scipy.stats import hypergeom
```

- ***Suppose 18 major computer companies operate in US and 12 are located in California's Silicon Valley. If three computer companies are selected, what is the probability that one or more of selected companies are located in Silicon Valley?***
- Here, we can use hypergeomeric distribution. N is 18, n is 3, A is 12 and x is 1.

```
pval = hypergeom.sf(0,18,3,12)   #hypergeom.sf(x-1,N,n,A), sf=1-cdf
pval
```

```
0.9754901960784306
```

Python Demo for Distributions

- ***A western city has 18 officers eligible for promotion. 11 of 18 are Hispanic. Suppose only 5 of the police officers are chosen for promotion. If the officers for promotion had been chosen by chance alone, what is the probability that one or fewer of the five police officers would have been Hispanic?***

```
pval = hypergeom.cdf(1,18,5,11)
pval
```

```
0.04738562091503275
```

- **Exponential Distribution**

```
from scipy.stats import expon
```

- ***A manufacturing company has been involved in statistical quality control over several years. As part of the production process, parts are randomly selected and tested. From the records of these tests, it has been established that a defective part occurs in a pattern that is Poisson distributed on the average of 1.38 defects every 20 minutes during production runs. Use this information to determine the probability that less that 15 minutes will elapse between any two defects?***
- Here, time interval is 15/20=0.75 and mean is 1/1.38.

```
expon.cdf(0.75,(1/1.38))
```

```
0.025043397119053856
```


thank you
danke
謝謝
ngiyabonga
teşekkür ederim
gracias
спасибо
bedankt
dziękuję
obrigado
sukriya
kop khun krap
go raibh maith agat
mochchakkeram
tapadh leat
dank je
hvala
mauruuru
sagolun
taiku
arigatô
takk
dakujem
grazie
terima kasih
мерси
хвала
рахмат
Баярлалаа
ありがとう
감사합니다
ευχαριστώ
merci


நன்றி

Probability distribution

- A probability distribution is a mathematical function
- It provides the probabilities of occurrence of different possible outcomes in an experiment.
- It describes the ‘shape’ of a batch of numbers.
- The characteristics of a distribution can sometimes be defined using a small number of numeric descriptors called ‘parameters’ and the parameters change for each distribution.

- Distributions can serve as a basis for standardized comparison of empirical distributions.
- It can help in estimating confidence intervals for inferential statistics.
- Also, it forms a basis for more advanced statistical methods, like it can be used to understand the ‘fit’ between observed distributions and certain theoretical distributions.

Continuous Probability Distributions

- A continuous random variable is a variable that can assume any value on a continuum (can assume an uncountable number of values)
- Example: thickness of an item, time required to complete a task, temperature of a solution, height, etc.
- These can potentially take on any value, depending only on the ability to measure precisely and accurately.
- Uniform, Exponential as well as Normal distributions are few of the continuous distributions

- **Some Special Distributions**
  - **Discrete:** Binomial, Poisson, Hypergeometric
  - **Continuous:** Uniform, Exponential, Normal