PROBABILITY AND STATISTICS FOR SCIENTISTS AND ENGINEERS

PROBABILITY

- This is the possibility, chance or the expectation that a certain event may or may not happen.
- (Or probability value) is a number between 0 and 1 inclusive associated with the likelihood of occurrence of a given event. (We assign probability of 1 if we are certain that the event will happen, a probability of 0 if we are sure that the event will not happen, and a probability of 0.5 if we are only half – sure that the event will happen)

FERMAT – father of probability

SET THEORY

SET AND SET NOTATIONS:

Set

- is a collection of well– defined and distinguishable entities or
- is a collection of all things or objects in the universe or
- is a collection of objects of any kind.

Examples:
1. Set of students enrolled in Math 8A with code _______.
2. Set of subjects Mr. ________ enrolled this summer.

Elements

The elements of a given set are the objects or entities in a set.

Notation: Use capital letters to name a set.
Use small letters for the elements of a set.

$a \in A$ means <u>a</u> is an element of set A.
$a \notin A$ means <u>a</u> is not an element of set A.

Set Notation:

Two methods of naming the elements of a set.

1. <u>Roster Notation or Listing Method</u>
   – elements are enumerated.

Example:

$A = \{1,2,3,4\}$

$B = \{a, b, c, d, e\}$

$C = \{neutron, electron, proton\}$

2. Set–Builder Notation or Rule Method or Statement Method

– where the properties which must be satisfied by all elements of the set are specified

Examples:

$A = \{x | x \text{ is a positive integer less than } 5\}$

$(\text{read as : set A is a set of all elements of x such that x is a ...})$

$B = \{x | x \text{ is any of the first five letters of the english alphabet}\}$

$(\textit{read as : set} \text{ B is a set of all elements of x such that x is a ...})$

$C = \{x | x \text{ is a fundamental particle in an atom}\}$

Types of Sets

1. Finite Set

– set whose elements are countable or a set with a limited number of elements.

Example:

E = set of possible outcomes of a fair die when tossed

By Listing method: $E = \{1,2,3,4,5,6\}$

By Rule Method: $E = \{e | e \text{ is an outcome when a fair die is tossed}\}$

2. Infinite Set

– set whose elements are not countable or a set with no last element.

Example: R is an element of real numbers

$R = \{x | \text{x is a real number}\}$

3. Unit Set

– set containing only one element

Examples:

$A = \{a\}$
$B = \{x | x - 2 = 0\} = \{2\}$
$C = \{x | \text{x is Denis mother}\}$

4. Empty or Null Set
– set satisfied by no element.

Examples:
$A = \{\ \} = \varnothing$
$B = \{x | \text{x is an even number ending in 1}\}$

5. Subset of a given Set
– If every element of A is an element of S, then $A \subset S$ (reads A is a subset of S).
– B is a proper subset of A if every element of B is in A and if there is at least one element of A, which is not in B, then $B \subset A$.

Examples:

1.) $A = \{1,2,3,4,5,6,7\}$
$B = \{1,4,7\}$
$B \subset A$

2.) $A = \{a,b,c,d\}$ $\quad A \subset B, B \subset A$
$B = \{a,b,c,d\}$ $\quad C \subset A, C \subset B$
$C = \{a,b,c\}$ $\quad A \not\subset C, B \not\subset C$

6. Universal Set or Sample Space

– is the set containing all the possible elements of a given condition.
– is the general set or the biggest set in a discussion.
– the set of all possible outcomes or elements of a given statistical experiment.
– It is denoted by capital letter S or U

Sample Points – are the elements or the outcome of a given sample space.

Example:

1. On flipping a coin.
$S = \{H, T\}$ Only head or tail

2. On tossing a die.
$S = \{1,2,3,4,5,6\}$

7. Two sets are equal if and only if the elements of one set A are also the elements of the other set B or vice versa.

Example:

$$A = \{a, b, c, d, 1, 2\}$$
$$B = \{1, a, b, 2, c, d\}$$
$$A = B$$

Exercises:

1. Rewrite the following statements using set notation:
   a) x is an element of X;
   b) B is a subset of F;
   c) a does not belong to A;
   d) The set C is empty;

2. Let $\mathrm{A} = \{\mathrm{a}, \mathrm{b}, \mathrm{c}\}$. Which of the following statements are correct?

   a) $\mathrm{a} \subset \mathrm{A}$

   b) $\mathrm{a} \in \mathrm{A}$

   c) $\{\mathrm{a}\} \subset \mathrm{A}$

   d) $\{\mathrm{a}\} \in \mathrm{A}$

3. If $X = \{x | x^2 = 4, x > 3\}$, which of the following statements are correct?

$a) X = 0$

$b) X = \{0\}$

$c) X = \varnothing$

$d) X = \{\varnothing\}$

Complement of a set A with respect to U or S:

The set of all elements that are in U or S but not in A. Symbol A’

Example: $U = \{x | x \text{ is a positive integer}\}$
$A = \{x | x \text{ is a positive even integer}\}$
$A' = \{x | x \text{ is a positive odd integer}\}$

Unary Operation:

An operation performed on a single set such as getting the complement of a set.

Binary Operation:

An operation performed on two sets.

OPERATION ON SETS:

1. Union $(\cup)$ of sets

– The union of two sets A and B is a set consisting of the elements of A and/or elements of B.
– The set consisting of all elements which belong to A or to B or to both. Symbol $A \cup B$

Example:

1. $A \cup B = \{x | x \in A \text{ and/or } x \in B\}$
$A = \{a, b, c, d\}$
$B = \{c, d, e, f\}$
$A \cup B = \{a, b, c, d, e, f\}$

2. $A \cup B = \{x | x \in A \text{ and/or x } \in \text{B}\}$
$A = \{1,3,5,8,9,11\}$
$B = \{5,9,12,17\}$
$A \cup B = \{1,3,5,8,9,11,12,17\}$

2. Intersection $(\cap)$ of two sets

- the intersection of 2 sets A and B is a set containing the elements common to A and B.
- The set consisting of all elements that belong to both A and B. Symbol $\text{A} \cap \text{B}$

Note: Disjoint sets do not have an intersection; i.e. $\text{A} \cap \text{B} = \varnothing$

Example:

1. $A \cap B = \{x | x \in A \text{ and } \text{x} \in \text{B}\}$
$A = \{1,3,5,8,9,11\}$
$B = \{5,9,12,17\}$
$A \cap B = \{5,9\}$

3. Set Difference

- The difference of two sets A and B, written A – B, is defined to be the set of elements which are in A but not in B. The difference B – A is similarly defined to be the set of elements which are in B but not in A

$$A - B = \{x \in A; x \notin B\}$$
$$B - A = \{x \in B; x \notin A\}$$

Example:

Given :
$A = \{a, b, c, d\}$
$B = \{c, d, e, f\}$
$A - B = \{a, b\}$
$B - A = \{e, f\}$

4. Complement of A (A’) or ($\overline{\text{A}}$)

- set containing the elements of the universal set which are not elements of A.

$$A' = S - A \text{ or } \text{U} - \text{A}$$
$$\text{A} \cup \text{A'} = \text{U} = \text{S}$$

Example:

Given: $S = U = \{1,2,3,4,5,6\}$ Find: A'

$A = \{2,4,6\}$

$A' = \{1,3,5\}$

DE MORGAN'S LAWS

1. $(A \cap B)' = A' \cup B'$
2. $(A \cup B)' = A' \cap B'$
3. $A \cap \varnothing = \varnothing$
4. $A \cup \varnothing = A$
5. $A \cap A' = \varnothing$
6. $A \cup A' = U = S$
7. $U = \varnothing$
8. $\varnothing = U$

Example:

Given:

$A = \{x | x \text{ is an outcome when a fair die is tossed}\}$

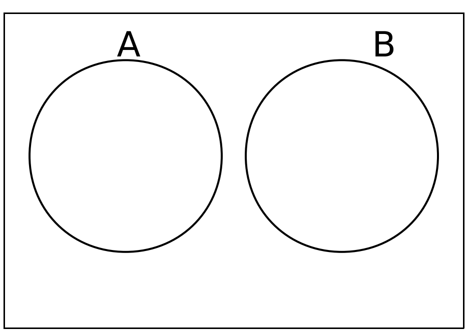
$B = \{y | y \text{ is an outcome when a fair coin is tossed}\}$

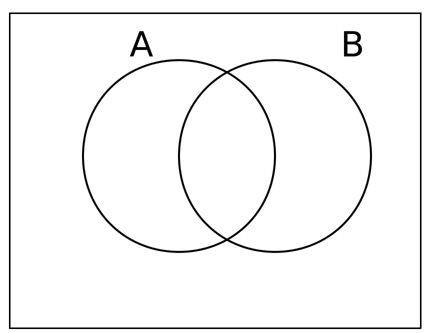
$C = \{z | z \text{ is the card drawn from a deck of card and z is an ace}\}$

Required: $A \cup B \cup C$

Venn diagram

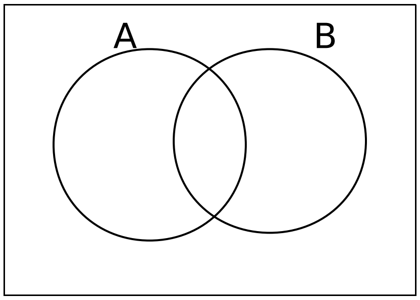
– A diagram consisting of a plane geometric figure to represent a set and set relations.

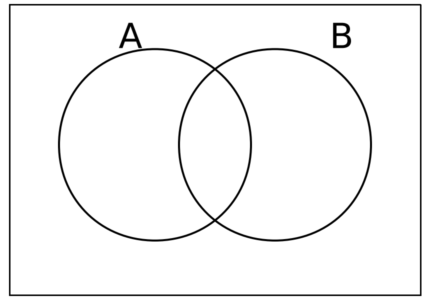




1. Represent $A \cup B$ for two disjoint sets .

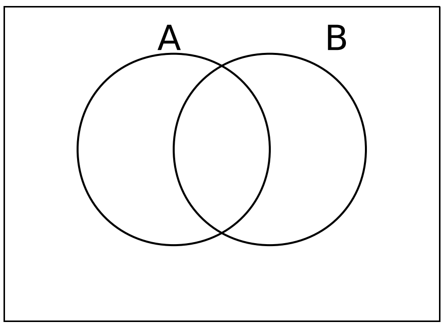
2. Represent $A \cup B$ for two sets which are not disjoint.

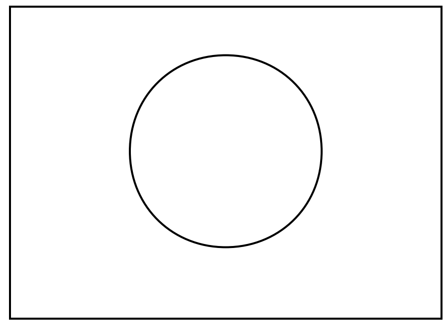




3. $A \cap B$

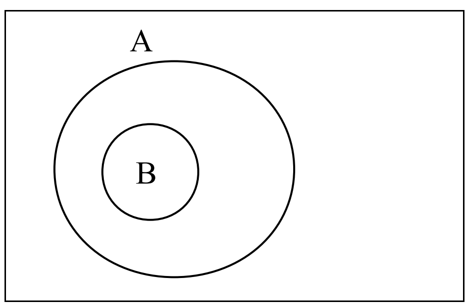
4. $A - B$





5. $B - A$

6. A'



7. $B \subseteq A$

Exercises:

1. Sixty five commuters in metro Manila were interviewed regarding the types of vehicles they take when they report to their respective jobs. The following data were gathered: 32 take the jeepney; 29 take the bus; 21 take the commuter train; 15 take the bus and the jeepney; 16 take the jeepney and the commuter train; 12 take the bus and the commuter train; 9 take the bus and the jeepney and the commuter train. Draw the Venn diagram and from the diagram gather additional information.

2. To join a certain club, a person must either be a mathematician or an engineer. Of the 25 members, 19 are engineers and 16 are mathematicians. How many persons in the club are both?

EXERCISES:

1. Let A, B, C be the subsets of Z^{+}(set of positive integers) defined by

$$A = \{1,2,3,4,5\}$$
$$B = \{3,5,7\}$$
$$C = \{2,4,6\}$$

Described the following subsets:

1. $A \cup B$

2. $A \cap C$

3. $A - B$

4. $B - C$

5. $A \cap (B \cup C)$

6. $(A \cap B) \cup (A \cap C)$

7. $A' \cap C'$

8. $(A \cup C)'$

2. If $S = \{0,1,2,3,4,5,6,7,8,9\}$
$A = \{0,2,4,6,8\}$
$B = \{1,3,5,7,9\}$
$C = \{2,3,4,5\}$
$D = \{1,6,7\}$

List the elements of the sets corresponding to the following events:

a) $A \cup C$

b) $A \cap B$

c) C'

d) $(C' \cap D) \cup B$

e) $(S \cap C)'$

f) $A \cap C \cap D'$

3. In a class of 40 students, 27 like Calculus and 25 like Chemistry. How many like both Calculus an d Chemistry.

4. A survey of 100 students reported that the number of those enrolled in various Mathematics subjects were

Algebra, Geometry and calculus	20
Algebra and Geometry	30
Algebra and Calculus	35
Geometry and Calculus	35
Algebra	70
Calculus	60

How many enrolled in Geometry?

5. There are 20 seniors serving the student council of the Cebu Institute of Technology this year. Of these, 3 have not served before, 10 served on the council in their junior years, 9 in their sophomore years, and 11 in their freshman years. There are 5 who served during both their sophomore and junior years, 6 during both their freshman and junior years, and 4 during both their freshman and sophomore years. How many seniors served on the student council during each of the four years in high school?

EXERCISES: (PAGE 16, 5th edition)

1. List the elements of each of the following sample spaces:
    a) The set of integers between 1 and 50 divisible by 8;

    b) The set $S = \{x \mid x^2 + 4x - 5 = 0\}$;

    c) The set of outcomes when a coin is tossed until a tail or three heads appear;

TREE DIAGRAM – is a schematic way of listing the elements of the sample space where the branches are the outcomes (sample points).

    d) The set $S = \{x \mid x \text{ is a continent}\}$;

e) The set $S: \{x \mid 2x - 4 \geq 0 \text{ and } x < 1\}$;

2. Which of the following events are equal?
   a) $A = \{1,3\}$
   b) $B = \{x \mid x \text{ is a number on a die}\}$;
   c) $C = \{x \mid x^2 - 4x + 3 = 0\}$;
   d) $D = \{x \mid x \text{ is the number of headwhen coins are tossed}\}$;

3. An experiment involves a tossing of a pair of dice, 1 green and 1 red, and recording the numbers that come up. If x equals the outcome on the green die and y the outcome on the red die, describe the sample space S: (a) by listing the elements (x, y); (b) by using the rule method.

4. An experiment consists of a die and then flipping a coin once if the number on the die is even. If the number on the die is odd, the coin is flipped twice. Using the notation 4H, for example, to denote the event that the die comes 4 and then the coin comes up heads, and 3HT to denote the event that the die comes up 3 followed by a head and then a tail on the coin, construct a tree diagram to show the 18 elements of the sample space.

TREE DIAGRAM – is a schematic way of listing the elements of the sample space where the branches are the outcomes (sample points).

5. Four students are selected at random from a chemistry class and classified as male or female. List the elements of the sample space S_1 using the letter M for "male" and F for "female". Define a second sample, S_2, where the elements represent the number of females selected.

6. For the sample space of exercise 3.
   a) List the elements corresponding to the event A that the sum is greater than 8.

   b) List the elements corresponding to the event B that a 2 occurs on either die.

c) List the elements corresponding to the event C that a number greater than 4 comes up on the green die.

d) List the elements corresponding to the event $A \cap C$

e) List the elements corresponding to the event $A \cap B$

f) List the elements corresponding to the event $B \cap C$

g) Construct a Venn diagram to illustrate the intersections and unions of the events A, B, C.

ASSIGNMENT #01
NAME: ______________________________ DATE: __________

1. Describe each of the following sets by listing their elements.

a) $A = \{x \mid x > 0, x^2 = 16\}$

b) $B = \{x \mid x > 0, x \text{ is an odd integer}\}$

c) Z denotes the set of integers
$C = \{x \in Z \mid 0 < x < 100\}$

d) Z^+ denotes the set of +integers
$C = \left\{\frac{1}{n} \middle| n \in Z^+\right\}$

2. Think of a way of describing each of the following sets by specifying a common property of the elements. (Note: there is no unique answer!)

a) $A = \{3,4,5,6,7,8\}$

b) $B = \{-\sqrt{2}, +\sqrt{2}\}$

c) $C = \{1, 10, 100, 1000, ...\}$

d) $E = \{..., -15, -10, -5, 0, 5, 10, 15, ...\}$

e) $F = \left\{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, ...\right\}$

THE FUNDAMENTAL PRINCIPLE OF COUNTING

If an event, (an event is a subset of a sample space) can occur in n_1 ways and after this has happened, another event can occur in n_2 ways, then the two events can happen in

$$n_1 \bullet n_2 \text{ ways}$$

Example 1: A room has two doors, in how many ways can one enter and leave the room?

Solution:

There are two activities involved here. One is entering the room, and there are two choices for doing this. Another is going out of the room, and there are also two choices for doing this. Suppose we call one door A, and the other, B. Let us explore some possible ways of solving this problem:

1. Listing method

The different entrance – exit pairs are AA, AB, BA, BB.

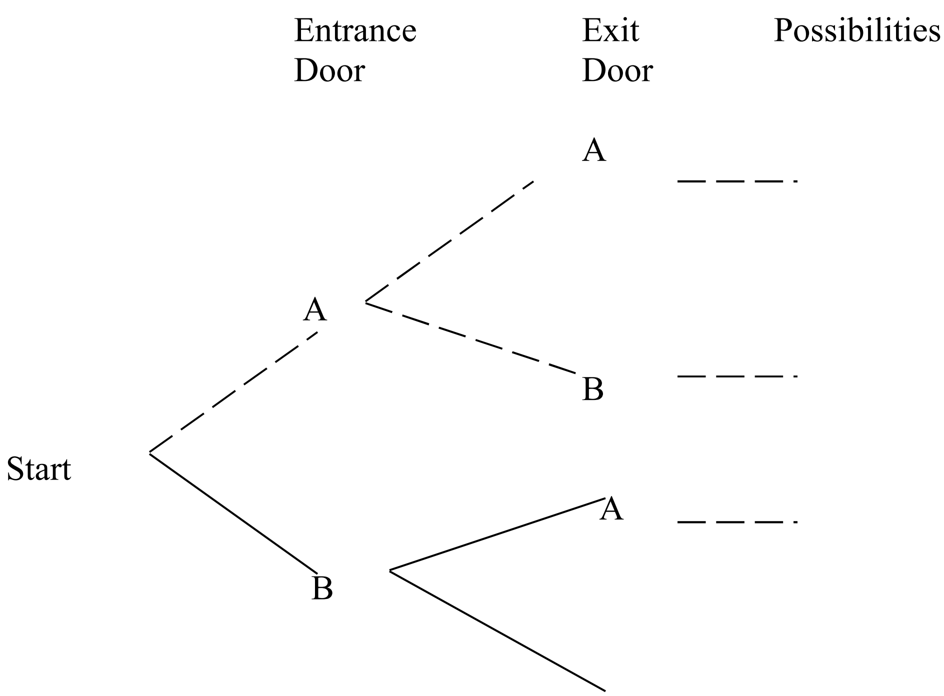
2. By using a table

Exit Door

	A	B
A	(A, A)	(A, B)
B	(B, A)	(B, A)

In the table presented above, the rows are labeled with choices for one activity. The column, on the other hand, are labeled with choices for doing the other activity. The entries on the cells, which are ordered pairs, represent the different choices for doing the sequence of activities, entering and then leaving.

3. By using Tree Diagram



B _ _ _ .

4. By using multiplication rule

   It can be argued that since there are two entrance doors and for each of these, two exit doors can be paired, then there are $2x2 = 4$ entrance – exit pairs

For a group of *k* things, if the first can be done independently in n_1 different ways, the second can be done independently in n_2 different ways, the third can be done independently in n_3 different ways, and so on, until the kth thing, then the total number of ways in which the k things can be done in the stated order is

$$n_1 \bullet n_2 \bullet n_3 .. n_k$$

Example 2: A pizza restaurant offers 6 kinds of meat toppings (beef, pepperoni, chorizo, etc.) and 4 kinds of vegetable toppings (onion, pepper, mushroom, etc.). A pizza comes in 3 different sizes (single, double, family) and in 3 different crust (thin, thick, pan). How many varieties of pizza can one select?

Solution: there are four activities involved here, namely:

1. selecting a meat (M) topping
2. selecting a vegetable (V) topping
3. selecting the size (S) and
4. selecting the crust (c)

Examples 3: In any banks, depositors can deposit or withdraw money any time through an automated teller. They just insert a card that contains some information into an automated teller machine and press their four-digit number code on a keyboard. The computerized machine reads the code in the card. It then processes the transaction in behalf of the depositor or based on the code. How many different four-digit number codes are possible?

Solution: There are four activities involved here:

1. choosing the first code
2. choosing the second code
3. choosing the third code
4. choosing the fourth code

Example 4: Cars have plate numbers for identification. A plate number consists of three letters followed by three numbers.

Exercises:

1. A high school senior who recently won a scholarship from a large manufacturing firm computes for the total number of ways in which he can choose a course and a school for his college studies. The manufacturing firm gave him a list of four technological courses and five schools to choose from.

2. A market researcher classifies a group of 200 respondents into three according to their highest educational attainment, into four according to their occupation, and into two according to their place of work. He wants to determine the number of classifications possible.

3. An office secretary tries to devise a coding scheme for certain records, using the digits 1 to 4. She wants to find the total number of codes of different digits if only three of the four digits are used.

4. A car dealer is interested in knowing how many choices a prospective buyer has, given five different models and six colors.

5. The employee in charge of five display windows of a department store has 10 designs, each design appropriate for each window. She wishes to determine the total number of arrangements possible for the ten designs.

6. Given the digits 0, 2, 5, 6, and 9,

   a) How many 3-digit numbers can be formed from these digits if no two digits are to be the same?

   b) Of the numbers formed in (a),

      b1) how many are even?

b2) how many are odd?

b3) how many are greater than 600?

c) How many numbers can be formed is a digit may be repeated?

PERMUTATIONS

A PERMUTATION is an arrangement of n different objects in a definite order.

The words rat, tar, and art are three different permutations for the letters a, r, and t. The three other permutations for these letters are rta, tra, and atr.

There are two different permutations for two different objects, six for three different objects, and twenty – four for four different objects. In general. the number of permutations for the n different objects, denoted by ${}_nP_n$, is

$$n(n-1)(n-2)...(3)(2)(1) \text{ or } n! \text{ (read "n factorial")}$$

A. PERMUTATIONS OF DIFFERENT THINGS:

The number of permutations of n objects taken $r(r \leq n)$ at a time is denoted by ${}_nP_r = n(n-1)(n-2)...(n-r+1)$

If no repetition occurs:

$$P(n,r) = \frac{n!}{(n-r)!}$$

If n = r

$$_nP_r = \frac{n!}{(n-r)!} = \frac{n!}{(n-n)!} = \frac{n!}{0!} = n!$$

Example 1: How many three digit number can be formed from the number 0, 1, 2, 3, 4, 5, 6, 7, 8, 9

Example 2: How many three digit numbers greater than 500 can be formed from the digits 1, 3, 4, 6, and 7 if

a) no digit is to be repeated

b) Repetition is allowed

B. PERMUTATIONS OF n THINGS NOT ALL DIFFERENT:

The permutation of “n” things taken “n” at a time in which “q” are alike, “r” are alike and so on:

$$P = \frac{n!}{q!r!}$$

Example 1: How many permutations can be made out of the letters in the word BESAVILLA?

Example 2: How many different signals each consisting of 6 flags hung in a vertical line can be formed from 4 identical red flags and 2 identical blue flags?

Example 3: In how many ways can 9 books, 6 Mathematics and 3 Design be arranged on a shelf if

a) Design books are not to be separated,

b) Mathematics books are not to be separated,

c) The two books are kept together.

Example 4: How many different ways can 5 people line up to pay their telephone bills at the Meralco office in any order?

C. CYCLICAL PERMUTATIONS OF “n” DIIFERENT THINGS TAKEN “n” AT A TIME IS $P = (n-1)!$

Example 1: How many ways can 6 people be seated in a round table?

Example 2: Four couples are to eat at a round table with the men and women alternating. If the hostess reserves a place for herself, in how many ways can she assign seats to the others?

Example 3: In how many ways can 8 people be seated at a round table?

Exercises: PERMUTATIONS

1. In how many ways can four boys and 3 girls be seated in a row of 5 chairs?

2. In how many ways can three of ten students participating in an interschool contest ranked first, second, and third?

3. List all the different permutations for the digits of the number 5,696. How many permutations would have been possible if four digits were different?

4. How many distinct permutations can be formed from the letters of the word STATISTICS?

5. How many ways can seven scientists be assigned to one triple and two double rooms?

6. The number or ways can 3 nurses and 4 engineers be seated on a bench with the nurses seated together.

7. In how many ways can 6 distinct books be arranged in a bookshelf?

8. What is the number of permutations of the letters in the word BANANA?

9. If 15 people won prizes in the state lottery (assuming that there are no ties), how many ways can these people win first, second, third, fourth, and fifth prizes?

10. Four different colored flags can be hung in a row to make coded signal. How many signals can be made is a signal consists of the display of one or more flags?

11. In how many ways can 4 boys and 4 girls be seated alternately in a row of 8 seats?

12. How many different ways can 5 boys and 5 girls form a circle with boys and girls alternate?

13. There are 4 balls of 4 different colors. Two balls are taken at a time and arranged in a definite order. For example, if a white and a red balls are taken, one definite arrangement is white first, red second, and another arrangement is red first, white second. How many such arrangements are possible?

14. In how many ways can a PSME Chapter with 15 directors choose a President, a Vice President, a Secretary, a Treasurer, and an Auditor, if no member can hold more than one position?

15. How many four-letter words beginning and ending with a vowel without any letter repeated can be formed from the word "personnel"

COMBINATIONS

Combination is the grouping of things whose arrangement is not important.

Suppose we have four objects denoted by A, B, C and D. There are twenty – four different permutations of four objects taken three at a time. Thus, if three of the four were chosen in succession, there would be twenty – four selections as shown below.

ABC	ABD	ACD	BCD
ACB	ADB	ADC	BDC
BAC	BAD	CAD	CBD
BCA	BDA	CDA	CDB

CAB	DAB	DAC	DBC
CBA	DBA	DCA	DCB

Without regard to the order, there are only four ways in which the three can be chosen from the four. These selections are called COMBINATIONS. The number of combinations of n objects taken r at a time is denoted by ${}_nC_r$. In the example, we find that for each combination, there are 3! or 6 different permutations. The total numbers of permutation can be written as

$$ {}_4P_3 = {}_4C_3 \bullet 3! $$

In general, if there are n different objects and we take r at a time,

$$ {}_nP_r = {}_nC_r \bullet r! $$

Dividing both sides of this equation by r!, we obtain

$$ {}_nC_r = \frac{{}_nP_r}{r!} $$

Therefore, the number of objects taken r at a time is

$$ {}_nC_r = \frac{n}{(n-r)!\,r!} $$

Example:

1. In how many ways can a committee of 4 be chosen from a group of 8 people?

2. In how many ways can we select 2 spades and 3 diamonds from a deck of 52 cards?

3. A box contains 5 red, 4 blue, and 3 white balls. In how many ways can we select 3 balls such that

   a) They are of different colors?

   b) They are all red?

   c) Two are blue and one is white?

   d) Exactly two are blue?

   e) None is blue?

4. Referring to the box in example 3, in how many ways can we select 4 balls such that at least two are red?

EXERCISES: MULTIPLICATION RULE:

1. From the digits 1, 2, 3, 4, 5, 7, and 8,
    a) How many numbers of four different digits each can be formed
    b) How many four digits odd numbers of this character can be formed?

2. In how many ways can 4 people seat themselves in 6 given seats?

3. Each of three departments in a store needs a secretary. There are 6 applicants for the positions. In how many ways can they be filled?

4. How many permutations (numbers) can be formed from the digits 2, 3, 4, 7, and 8 taken three at a time. Write out these permutations.

5. In how many ways can 5 people assume seats in a row of seven seats?

6. In how many ways can a man distribute gifts of $1, a sled, and a bicycle for 5 boys?

7. The captain of a baseball team assigns himself to the 4th place in the batting order. In how many ways can he assign the remaining places to his 8 teammates, if just three men are considered eligible for the 1st position?

Note: The word <u>and</u> means multiply
The word <u>or</u> means add

Combinations - is a set of things is a group of all or of any part of the things in this group.

- Combination of "n" different things taken "r" at a time.

$$ {}_nC_r = C(n,r) = \frac{{}_nP_r}{r!} = \frac{n!}{(n-r)!r!} $$

ex. How many combinations can be made out of the letters ABCD and E taken

(a) two at a time,

(b) four at a time

ex. How many triangles are determined by 8 points, no three of which are collinear?

- Combinations of Mutually Exclusive Events.

  When two sets of "h" ways and "k" ways, respectively are known to include no duplications, the total number of ways is (h + k)

  ex. A bag contains 7 black and 6 white balls. In how many ways can we draw from the bag groups of 5 balls involving at least 3 black balls?

PERMUTATIONS:

1. In how many ways can 5 differently colored marbles be arranged in a row?

2. In how many ways can 10 people be seated on a bench if only four seats are available?

3. It is required to seat 5 men and 4 women in a row so that the women occupy the even spaces. How many such arrangements are possible?

4. How many four digit numbers can be formed with the 10 digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 if:

    a) Repetitions are allowed,

    b) Repetitions are not allowed,

    c) the last digit must be zero and repetitions are not allowed?

5. Four different mathematics books, six different physics books, and two different chemistry books are to be arranged on a shelf. How many different arrangements are possible if:

    a) The book in each particular subjects must all stand together,

b) Only the mathematics books must stand together?

6. Five red marbles, two white marbles, and three blue marbles are arranged in a row. If all the marbles of the same color are not distinguishable from each other, how many different arrangements are possible?

7. In how many ways can seven people be seated at a round table if:

   a) They can sit anywhere,

   b) 2 particular people must not sit next to each other?

COMBINATIONS

1. In how many ways can 10 objects be split into two groups containing 4 and 6 objects respectively?

2. In how many ways can a committee of 5 people be chosen out of 9 people?

3. Out of 5 mathematicians and 7 physicists, a committee consisting of 2 mathematicians and 3 physicists is to be formed. In how many ways can this be done if:

b) Any mathematicians and any physicists can be included,

c) One particular physicists must be on the committee,

d) Two particular mathematicians cannot be on the committee?

4. From seven consonants and 5 vowels, how many words can be formed consisting of 4 different consonants and three different vowels? The words need not have meaning.

5. In how many ways can one make a selection of 4 black balls, 5 red balls, and 2 white balls from a box containing 8 black balls, 7 red balls, and 5 white balls.

6. How many different committees of 5 can be selected from 15 ilocanos and 10 cebuanos if:

a) It must contain 3 ilocanos and 2 cebuanos,

b) It must contain at least three ilocanos

PROBABILITY

We often say the words "likely", "uncertain", "maybe" and "impossible". What do we really mean when we say "it is likely to rain" or "it is impossible from her to be here" or "his chances are 50 – 50"?

We will try to make more precise what we mean when we talk about the chance of happening or occurring or not of something. The measure of the chance of an event taking place is called its PROBABILITY.

> "It is often said that a possibility is not necessarily a probability"
>
> This means that an event may be possible but not be probable. Example: It might be possible for a famous politician to win in an election but if he is not a candidate, this is not probable.

Probability (or probability value) – is a number between 0 and 1 inclusive associated with the likelihood of occurrence of a given event.

Definition of terms:

1. Experiment is defined as an activity, which can be done repeatedly under similar conditions, in which can result in an outcome.

   Example:

   1. Rolling a pair of dice and observing which numbers come on top. Rolling a die are experiments having six outcomes.

   2. Tossing a coin is an experiment involving two outcomes, head or tail.

   3. Drawing a card from a well – shuffled deck of playing cards are experiments having 52 outcomes.

2. Sample Space

   The set of all possible outcomes of an experiment is known as *Sample Space*, each particular outcome is a *Sample Point*

*Any subset of a sample space is an Event. “Selecting an ace at random” from a deck of 52 cards is an example of an event having four sample points. “Drawing a spade” is another event having thirteen sample points.
In general, AN EVENT IS SAID TO HAVE OCCURRED IF THE OUTCOME OF THE EXPERIMENT IS ONE OF THE SAMPLE POINTS CONTAINED IN THE EVENT; otherwise, the event is said not to have occurred.

Example:

1. Toss two coins. Find the sample space.

   Each coin has two sides. There are only three possible outcomes: both heads up, both tails up or one head and one tail up. Since either coin may come up with head and the other with tail, there are two ways of getting one head and one tail. Thus, the sample space is $\{HH, HT, TH, TT\}$.

2. The digits 1, 2, 3 and 4 are written on small slips of paper. Two slips are picked at random one after the other.

   a) What is the sample space if the second slip is picked without replacing the first slip that was taken?

   b) What is the sample space if the first slip is replaced before picking the second?

MEANING OF PROBABILITY

Consider an equiprobable space S (If any of the possible outcomes is equally likely to occur, the sample space is called an equiprobable space)take for example tossing a coin. What is the chance that a head turns up? We usually say 1 in 2. Or, if we choose one card from a deck of 52 cards, what is the chance that we draw a heart? We know that there are 13 cards in the heart suits. So, we say the chance of getting a heart is 13 in 52 of 1 in 4. We make this concept precise by the following definition.

DEFINITION:

The (theoretical) probability of an event is the ratio of the number of ways that an event can occur to the total number of outcomes when each outcome is equally likely to occur.

In symbols, the probability of an event E is

$$P(E) = \frac{n(E)}{n(s)} \qquad (1)$$

Where:

$n(E)$ represents the number of ways E can occur

$n(S)$ represents the number of possible outcomes in the sample space S.

NOTE: The probability we will be discussing is theoretical probability because we will obtain E by using the principles of counting and not by an actual experiment. If an actual experiment is done to determine $n(E)$, the probability obtained then will be empirical or experimental probability. The experimental probability is close to the theoretical probability when the number of trials performed is large enough.

Example:

1. Two coins are tossed. What is the probability of two heads appearing?

2. Two fair dice are rolled. Find the probability of each of the following events.

a) A = {both numbers are even}

b) B = {one number is three and the other is even}

c) C = {a sum less than 5 appears}

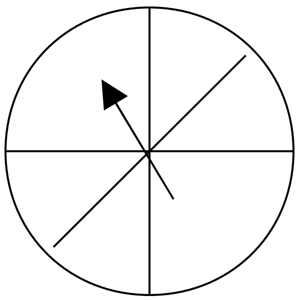
3. A school has 12 good runners of which are 4 girls. If two are chosen at random to represent the school, what is the probability of the event

a) A = {both chosen runners are girls

b) B = {both chosen runners are boys}

c) C = {one boy and one girl are chosen}

4. Spin the dial. What is the probability of



a) The event A that the pointer stops on an even number less than 6?

b) The event B that the pointer stops at 1?

Examples:

1. Two candidates, A and B, are running for public office. If the probability that A will win the election is .35, what is the probability that B will win?

2. Find the probability of getting an even number from a single toss of a die.

3. If a pair of dice is tossed, find the probability of obtaining a sum of seven.

4. If a card is drawn from an ordinary deck, find the probability of drawing

a) An ace

b) A spade

c) A face card

5. A box contains 5 red, 4 blue, and three white balls. If a ball is chosen at random, what is the probability that

   a) It is not red?

   b) It is not white?

6. With reference to the box in example #5, suppose three balls are drawn at random. What is the probability that

   a) They are of different colors?

   b) They are all red?

   c) Two are blue and one is white?

d) Exactly two are blue?

e) None is blue?

Mutually Exclusive and Non–mutually Exclusive Events

Mutually Exclusive Events

Two events are said to be mutually exclusive if there is no opportunity for them to occur simultaneously or if they have no common sample points.

*Drawing an ace and drawing a king from a deck of cards are mutually exclusive because it is not possible to obtain an ace which a king at the same time.

Non–mutually Exclusive Events

Events that can happen at the same time are non–mutually exclusive events. They are events, which have some sample points in common.

*the events, "drawing an ace and a spade", have one sample point in common since one of the four aces is a spade.

The ADDITION RULE

"$P(A \text{ or } B)$" Denote the probability of occurrence of events A or B. Some books emphasize the operation used in computing for this probability by using either: "$P(A \cup B)$" or "$P(A + B)$".

$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$ for any events A and B

If A and B are mutually exclusive events, $P(A \text{ and } B)$ is equal to zero. Therefore, the above formula is reduced to

$$P(A \text{ or } B) = P(A) + P(B)$$

If A_1, A_2, A_3, . . . A_k are mutually exclusive events, the probability that anyone of them will occur is

$$P(A_1 \text{ or } A_2 \text{ or } A_3 \text{ or} \ldots A_k) = P(A_1) + P(A_2) + P(A_3) + \ldots + P(A_k)$$

* What is the probability of drawing an ace or a king from an ordinary deck?

* If we select a ball at random from a box containing 5 red, 4 blue, and 3 white balls, what is the probability that it is either red or blue?

* What is the probability of an ace or spade?

Example:

1. If a card is drawn from an ordinary deck, find the probability of each of the following:

   a) Spade of face card

   b) Face card or red card

2. A retailer estimates that the probability that the price of a basic commodity will increase, decrease, or remain unchanged during a given month is .35, .10, and .55, respectively. What is the probability that the price of the commodity will change during the month?

3. An AB freshman student estimates that the probability that he will pass History 1 is .62; the probability that he will pass Sociology 1 is .50; and the probability that he will pass both subjects is .40. What is the probability that he will pass at least one of these two subjects?

4. The organizers of a seminar – workshop on business education observe that of the 50 participants, 32 are educators (faculty members and school administrators) and 26 are business executives of various companies. Moreover, 12 of the business executives are part – time professors. The rest are government officials. If one of the participants is chosen at random to head a committee, what is the probability that the one chosen is an educator or a business executive? Draw a Venn diagram to illustrate this situation.

Exercises:

1. Determine the probability of each of the following events:
    a) Drawing a spade from a deck of 52 playing cards;

    b) Drawing four spades in succession from a deck of 52 playing cards if after each card is drawn it is not replaced in the deck.

2. If a French book, a Spanish book, a German book, a Russian book, and an English book are placed at random on a shelf with space fro five books, what is the probability that the Russian and English books will be next to each other?

3. What is the probability of obtaining a sum of 7 or 5 if two dice are thrown?

4. A card is drawn at random from an ordinary deck of 52 playing cards. Find the probability that it is

a) An ace

b) A jack of hearts

c) A three of clubs or a six of diamonds

d) A heart

e) Any suits except hearts

f) A ten or a spade

g) Neither a four nor a club

5. Find the probability of a 4 turning up at least once in two tosses of a fair die.

6. A die is constructed so that a 1 or 2 occurs twice as often as a 5, which occurs 3 times as often as a 3, 4 or 6. If the die is tossed once, find the probability that

a) The number is even

b) The number is a perfect square

c) The number is greater than 4

7. If A and B are mutually exclusive events and P(A) = .3 and P(B) = .5, find

a) $P(A \cup B)$

b) $P(A')$

c) $P(A' \cup B)$

8. In a high school graduating class of 100 students, 54 studied math, 69 studied history, and 35 studied both math and history. If one of these students is selected at random, find the probability that

a) A student takes math or history

b) The student does not take either of these subjects

c) The students takes history but not math

9. Suppose that in a senior college class of 500 students it is found that 210 smoke, 258 drink alcoholic beverages, 216 eat between meals, 122 smoke and drink alcoholic beverages, 83 eat between meals and drink alcoholic beverages, 97 smoke and eat b/n meals, and 52 engage in all three of these bad health practices. If a member of this senior class is selected at random, find the probability that the student

   a) Smoke but does not drink alcoholic beverages

   b) Eats b/n meals and drinks alcoholic beverages but does not smoke

10. A pair of dice is tossed. Find the probability of getting

    a. A total of 8

    b. At most a total of 5

CONDITIONAL PROBABILITY

"Consider once again the TOSSING A DIE experiment". Let E denote the event of getting an even number and G the event of getting a number greater than 3.

Event E = (2, 4, 6)   Event G = (4, 5, 6)

P (E) = 3/6   and   P (G) = 3/6

Suppose that after tossing the die, we are told that G has occurred. What is the probability of E? "The information we have on the outcome of the experiment essentially reduces the number of possibilities". Before, we had six outcomes (1, 2, 3, 4, 5, 6), now we have only three (4, 5, 6). Since two of these correspond to the occurrence of E, we say that "THE PROBABILITY OF AN EVEN NUMBER, GIVEN A NUMBER GREATER THAN 3" IS 2/3

This is written as $P(E|G) = \frac{2}{3}$

"The probability that an event B occurs when it is known that some event A has occurred is called a CONDITIONAL PROBABILITY"

This is denoted by $P(B|A)$ which is read " the probability of B, given A." The probability of A, given B is written as $P(A|B)$

From the above example, it can be shown that for any events A and B,

$$P(B|A) = \frac{n(A+B)}{n(A)}$$

where:
$n(A+B) =$ number of sample points common to A and B (or number of ways in which A and B can occur together)
$n(A) =$ number of sample points in A

$$P(B|A) = \frac{n(A+B)}{n(A)} = \frac{\frac{n(A+B)}{N}}{\frac{n(A)}{N}} = \frac{P(A+B)}{P(A)} = \frac{P(A \cap B)}{P(A)} = \frac{P(A \text{ and } B)}{P(A)}$$

Commutativity or Conditional Probability

Note:
P (A or B) = P (B or A)
P (A and B) = P (B and A)
P (A/B) $\neq$ P (B/A)

If A and B are independent events
P (B/A) = P (B)
P(A/B) = P(A)

$P(A \cap B) = P(A)\, P(B)$

Example

1. The organizers of a seminar – workshop on business education observe that of the 50 participants, 32 are educators (faculty members and school administrators) and 26 are business executives of various companies. Moreover, 12 of the business executives are part – time professors. The rest are government officials. If one of the participants is chosen at random to head a committee, determine the probability that the person chosen to head the committee is an educator, given that he or she is a business executive.

2. An AB freshman student estimates that the probability that he will pass History 1 is .62; the probability that he will pass Sociology 1 is .50; and the probability that he will pass both subjects is .40. Determine the probability that if the student passed History 1, he will also pass sociology 1.

Exercises:

1. A class in advanced Physics is comprised of 10 juniors, 30 seniors, and 10 graduate students. The final grade showed that three of the juniors, 10 of the seniors, and 5 of the graduate students received an A for the course. If a student is chosen at random from this class and is found to have earned an A, what is the probability that he or she is a senior?

2. A random sample of 200 adults are classified below according to sex and the level of education attained

EDUCATION	MALE	FEMALE	TOTAL
Elementary	38	45	

Secondary	28	50
College	22	17

If a person is picked at random from this group, find the probability that

a) The person is a male, given that the person has a secondary education:

b) The person does not have a college degree, given that the person is a female.

3. In the senior year of a high school graduating class of 100 students, 42 studied math, 68 studied psychology, 54 studied history, 22 studied both math and psychology, 7 studied history but neither math nor psychology, 10 studied all three subjects, and 8 did not take any of the three. If a student is selected at random, find the probability that
Venn Diagram:

   a) A person enrolled in psychology takes all three subjects;

   b) A person not taking psychology is taking both history and math.

4. A pair of dice is thrown. If it is known that one die shows a 4, what is the probability that

    a) The other die shows a 5

    b) The total of both dice is greater than 7

5. A card is drawn from an ordinary deck and we are told that it is red. What is the probability that the card is greater than 2 but less than 9?

6. The probability that a married man watches a certain television show is o.4 and the probability that a married woman watches the show is 0.5. The probability that a man watches the show, given that his wife does is 0.7. Find the probability that

   a) A married couple watches the show

   b) A wife watches the show given that her husband does

   c) At least one person of a married couple will watch the show.

7. One bag contains 4 white balls and 3 black balls, and a second bag contains 3 white balls and 5 black balls. One ball is drawn at random from the second bag and is placed unseen in the first bag. What is the probability that a ball now drawn from the first is white?

8. Two cards are drawn in succession from a deck without replacement. What is the probability that

   a) Both are red;

   b) Both cards are greater than 3 but less than 8.

9. A lot of 20 articles contains 3 defective ones. If 2 articles are randomly selected, without replacement, what is the probability that

a) Both are defective?
b) Neither is defective?
c) The first is defective but the second is not?

10. Referring to the first problem, suppose that the 2 articles are randomly selected with replacement. What is the probability that

a) Both are defective?
b) Neither is defective?
c) The first is defective but the second is not?

BAYE'S RULE:

Let B = be the given sample space

$$P(A) = P(B_1 \cap A) + P(B_2 \cap A) + P(B_3 \cap A) + P(B_4 \cap A) + P(B_5 \cap A)$$

For Independents Events

$$P(A) = P(B_1)P(A/B_1) + P(B_2)P(A/B_2) + P(B_3)P(A/B_3) + P(B_4)P(A/B_4) + P(B_5)P(A/B_5)$$

For Dependent Events

$$P(B_k/A) = \frac{P(B_k)P(A/B_k)}{P(B_1)P(A/B_1) + P(B_2)P(A/B_2) + P(B_3)P(A/B_3) + P(B_4)P(A/B_4) + \ldots + P(B_k)P(A/B_k)}$$

Example:

1. In a certain assembly plant, three machines, B_1, B_2, B_3, make 30 %, 45%, and 25%, respectively, of the products. It is known from past experience that 2%, 3%, and 2% of the products made by each machine, respectively are defective. Now, suppose that a finished product is randomly selected. (a) What is the probability that it is defective?
(b) If a product were chosen randomly and found to be defective, what is the probability that it was made by machine B_3?

Solution:

Consider the following events:

A: the product is defective

B_1: the product is made by machine B_1

B_2: the product is made by machine B_2

B_3: the product is made by machine B_3

a) $P(A) =$

$$P(B_1)P(A/B_1) =$$
$$P(B_2)P(A/B_2) =$$
$$P(B_3)P(A/B3) =$$

$$P(A) = P(B_1)P(A/B_1) + P(B_2)P(A/B_2) + \ P(B_2)P(A/B_2) =$$
$$P(A) =$$

b) $P(B_3/A) = \dfrac{P(B_3)P(A/B_3)}{P(B_1)P(A/B_1) + P(B_2)P(A/B_2) + P(B_3)P(A/B_3)}$

2. A regional telephone company operates three identical relay stations at different locations. During a one-year period, the number of malfunctions reported by each station and the causes are shown below

	A	B	C
Problems with electricity supplied	2	1	1
Computer malfunctions	4	3	2
Malfunctioning electrical equipment	5	4	2
Caused by other human errors	7	7	5

Suppose that a malfunction was reported and it was found to be caused by other human errors. What is the probability that it came from station C?

3. In a certain region of the country it is known from past experience that the probability of selecting an adult over 40 years of age with cancer is 0.02. If the probability of a doctor correctly diagnosing a person with cancer as having the disease is 0.78 and the probability of incorrectly diagnosing a person without cancer as having the disease is 0.06. what is the probability that a person is diagnosed as having cancer?

4. Police plan to enforce speed limits by using radar traps at 4 different locations within the city limits. The radar traps at each of the locations L_1, L_2, L_3, and L_4 are operated at 40%, 30%, 20%, and 30% of the time, and if a person who is speeding on his way to work has probabilities of 0.2, 0.1, 0.5, and 0.2, respectively, of passing through these locations, what is the probability that he will receive a speeding ticket?

CHAPTER 3
RANDOM VARIABLES AND PROBABILITY DISTRIBUTION

A random event is an event whose occurrence or non – occurrence is dependent solely on chance factor.

A random variable is a function that associates a real number with each element in the sample space or these are functions, which assign a unique value for every element of the sample space.

2 Kinds of Random Sample Space

** If a sample space contains a finite numbers of possibilities or an unending sequence with as many elements as there are whole numbers, it is called a *discrete sample space*.

** If a sample space contains an infinite number of possibilities equal to the number of points on a line segment, it is called *continuous sample space.*

2 Kinds of Random Variables

1. A random variable is called a **DISCRETE RANDOM VARIABLE** it its set of possible outcomes is *COUNTABLE*.

2. When a random variable can take on values on a continuous scale, it is called **CONTINOUS RANDOM VARIABLE.** *(INFINITE AND NON – COUNTABLE*)

Examples of Discrete Random variable:
1. The # of students who are enrolled in M8A with a time 930 – 1030 (H704)
2. The # of chairs or tables found at 3^{rd} floor of OHB.

Examples of Continuous Random Variable:
1. Strands of hair in man's body.
2. Grains of sand in Boracay Island.

The set of ORDERED PAIRS [x, f(x)] is a probability function, probability mass function, or probability distribution of the discrete random variable X, if for each possible outcome x,

1) $f(x) > 0$
2) $\Sigma \ f(x) = 1$
3) $P(X = x) = f(x)$

The cumulative distribution F(x) of a discrete random variable X with probability distribution f(x) is given by

$$F(x)= P(X < x) = \sum_{t \geq x} f(t) \quad \text{for } -\infty < x < \infty$$

In dealing with continuous variable, f(x) is usually called the probability density function, or simply the density function of X.

The function f(x) is a probability density function for the continuous random variable X, defined over the set of real numbers R, if

1. $f(x) > 0$, for all $x \in R$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. $P(a < X < b) = \int_{a}^{b} f(x)dx$

The cumulative distribution F(x) of a continuous random variable X with density function f(x) is given by x

$$F(x) = P(X = x) = \int_{-\infty}^{x} f(t)dt \quad \text{for } -\infty < x < \infty$$

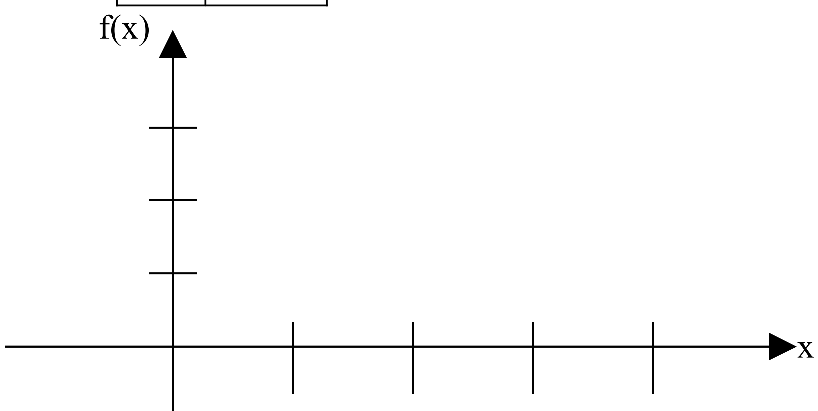
Histogram

A histogram is a graphic way of presenting data. It shows the distribution of the data. The standard format usually involves a vertical scale that delineates frequencies and a horizontal scale that delineates values of the data being represented. Each bar extends from its lower class boundary to its upper class boundary so that we can mark the class boundaries on the horizontal scale.

Probability Histogram

Instead of plotting the points (x, f(x)), we more frequently construct rectangles. Here the rectangles are constructed so that their bases of equal width are centered at each value x and their heights are equal to the corresponding probabilities given by f(x). The bases are constructed so as to leave no spaces between the rectangles.

F(x)	x
1/4	1
1/2	2
3/4	3



Example: Tossing a pair of coins: (fair coin)

Let x – be the random variable representing the number of heads

Tree diagram:

Probability Function Table:

Probability Distribution

Probability Histogram

Exercises:

1. Classify the following random variables as discrete or continuous:
X: The number of automobile accidents per year in Virginia
Y: The length of time to play 18 holes of golf
M: The amount of milk produced yearly by a particular cow
N: The number of eggs laid each month by a hen
P: The number of building permits issued each month in a certain city
Q: The weight of grain produced per acre

2. An overseas shipment of 5 foreign automobiles contains 2 that have slight paint blemishes. If an agency receives 3 of these automobiles at random, list the elements of the sample space S using the letters B and N for "blemished" and "non – blemished", respectively, then to each sample point assign a value x of the random variable X representing the number of automobiles purchased by the agency with paint blemishes.

Solution: Draw the tree diagram

3. Let W be a random variable giving the number of heads minus the number of tails in three tosses of a coin. List the elements of the sample space S for the three tosses of the coin and to each sample point assign a value w of W.

Solution: Draw the tree diagram

4. Determine the value of c so that each of the following functions can serve as a probability distribution of the discrete random variable X:

a) $f(x) = c\,(x^2 + 4)$ for $x = 0, 1, 2, 3$

b) $f(x) = c\ {}_2C_x\ {}_3C_{3-x}$ for $x = 0, 1, 2$

5. From a box containing 4 black balls and 2 green , each ball being replaced in the box before the next draw is made. Find the probability distribution for the number of green balls.

6. Find the probability distribution of the random variable W in exercise 3, assuming that the coin is biased so that the head is twice as likely to occur as a tail.

7. Find the probability distribution for the number of jazz records when 4 records are selected at random from a collection of 5 jazz records, 2 classical records, and 3 polka records. Express your results by means of a formula.

8. A shipment of 7 TV sets contains 2 defective sets. A hotel makes a random purchase of three of the sets. If X is the number of defective sets purchased by the hotel, find the probability distribution of X. Express the results graphically as a probability histogram.

9. Three cards are drawn in succession from a deck without replacement. Find the probability distribution for the number of spades.

10. Find the cumulative distribution of the random variable W in exercise 6. Using F(w), Find

c. $P(W > 0)$
d. $P(-1 \leq W < 3)$

11. Construct a graph of the cumulative distribution of exercise 10.

12. Find the cumulative distribution of the random variable X representing the number of defectives in exercise 8.

13. Construct a graph of the cumulative distribution of exercise 12.

14. The probability distribution of X, the number of imperfections per 10 meters of a synthetic fabric in continuous rolls of uniform width, is given by

x	0	1	2	3	4
F(x)	0.41	0.37	0.16	0.05	0.01

Construct the cumulative distribution of X.

15. a continuous random variable X that can assume values between x = 1 and x = 3 has a density function given by f(x) = 1/2

a) Show that the area under the curve is equal to 1
b) Find the P $(2 < x < 2.5)$
c) Find P $(x \leq 1.6)$

16. A continuous random variable X that can assume values between x = 2 and x = 5 has density function given by $f(x) = [2(1 + x)]/27$. Find

a) $P(X < 4)$
b) $P(3 < x < 4)$

17. The proportion of people who respond to a certain mail- order solicitation is a continuous random variable X that has the density function $f(x) = \begin{cases} \frac{2(x+5)}{5}, & 0 < x < 1 \\ 0, & e\,l\,s\,e\,w \end{cases}$

a) Show that $P(0 < X < 1) = 1$

b) Find the probability that more than 1/4 but less than 1/2 of the people contacted will respond to this type of solicitation.

18. For the density function of exercise 16, find F(x) and use it to evaluate P $(3 \leq X < 4)$

19. Consider the density function $f(x) = \begin{cases} k\sqrt{x}, 0 < x < 1 \\ 0, \text{ else} \end{cases}$

a) Evaluate k

b) Find F(x) and use it to evaluate P $(0.3 < X < 0.6)$

JOINT PROBABILITY DISTRIBUTION

- probability distribution that involves two or more random variables.

The function f(x, y) is a joint probability distribution of the discrete random variables X and Y if

a) $f(x, y) \geq 0$
b) $\sum_x \sum_y f(x, y) = 1$
c) $P(X = x, Y = y) = f(x, y)$,

The values of f(x, y) give the probability that outcomes x and y occur at the same time.

For any region A in the xy plane, $P[(x, y) \in A] = \sum\sum_A f(x, y)$

The function f(x, y) is a joint density function of the continuous random variables X and Y if

a) $f(x, y) \geq 0$ for all (x, y)
b) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\,dx\,dy = 1$
c) $P[(x, y) \in A] = \iint_A f(x, y)\,dx\,dy$

For any region A in the xy plane.

Example 1: Two refills for a ballpoint pen are selected at random from a box that contains 3 blue, 2 red, and 3 green refills. If X is the number of blue refills and Y is the number of red refills selected, find

a) the joint probability function f(x, y)
b) P[(X, Y)$\in$ A], where A is the region $\{(x, y) | x + y \le 1\}$

Example 2: A candy company distributes boxes of chocolates with a mixture of creams, toffees, and nuts coated in both light and dark chocolate. For a randomly selected box, let X and Y, respectively, be the proportions of the light and dark chocolates that are creams and suppose that the joint density function is

$$f(x,y) = \begin{cases} \frac{2}{5}(2x+3y), & 0 \le x \le 1, 0 \le y \le 1 \\ 0, & \text{elsewhere} \end{cases}$$

a) Verify condition : $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$

b) Find P[(X, Y)∈ A], where A is the region $\left\{(x,y) \middle| 0 < x < \frac{1}{2}, \frac{1}{4} < y < \frac{1}{2}\right\}$

MARGINAL DISTRIBUTION OF X AND Y

Given the joint probability distribution f(x, y) of the discrete random variables X and Y, the probability distribution **g(x)** of X alone is obtained by summing f(x, y) over the values of Y. Similarly, the probability distribution **h(y**) of Y alone is obtained by summing f(x, y) over the values of X.

The marginal distribution of X alone and Y alone are given by

For discrete case:

$$g(x) = \sum_{y} f(x, y) \quad \text{and} \quad h(y) = \sum_{x} f(x, y)$$

For continuous case:

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\,dy \quad \text{and} \quad g(y) = \int_{-\infty}^{\infty} f(x, y)\,dx$$

The term marginal is used because, in the discrete case, the value of g(x) and h(y) are just the marginal totals of the respective columns and rows when the values of f(x, y) are displayed in a rectangular table.

Example 3: Show that the column and row totals of the table in example 1 gives the marginal distribution of X and Y alone.

		x			
F(x, y)		1	2	3	row totals
	0	3/28	9/28	3/28	15/28
y	1	3/14	3/14		12/28
	2	1/28			1/28
column totals	5/14	15/28	3/28	1	

For the random variable x

$P(x = 0) = g(0) = \sum_{y=0}^{2} f(0, y) =$

$P(x = 1) = g(1) = \sum_{y=0}^{2} f(1, y) =$

$P(x = 2) = g(2) = \sum_{y=0}^{2} f(2, y) =$

For the random variable y

$P(y = 0) = h(0) = \sum_{x=0}^{2} f(x, 0) =$

$P(y = 1) = h(1) = \sum_{x=0}^{2} f(x, 1) =$

$P(y = 2) = h(2) = \sum_{x=0}^{2} f(x, 2) =$

Tabular form

Exercises: joint probability

1. Determine the value of c so that the following functions represent joint probability distributions of the random variables X and Y:

a) $f(x, y) = cxy$, for $x = 1, 2, 3$; $y = 1, 2, 3$.
b) $f(x, y) = c\,|x - y|$, for $x = -2, 0, 2$; $y = -2, 3$.

2. If the joint probability distribution of X and Y is given by $f(x, y) = \frac{x+y}{30}$, for x = 0, 1, 2, 3; y = 0, 1, 2
find:

a) P(X ≤ 2, Y = 1)
b) P(X > 2, Y ≤ 1)
c) P (X > Y)
d) P(X + Y = 4)

3. From a sack of fruit containing 3 oranges, 2 apples, and 3 bananas a random sample of 4 pieces of fruit is selected. If X is the number of oranges and Y is the number of apples in the sample, find

a) the joint probability distribution of X and Y
b) P[(X, Y) ∈ A], where A is the region given by $\{(x, y) | x + y \leq 2\}$

4. Let X, Y and Z have the joint have the probability joint density function

$$f(x,y,z) = \begin{cases} k x^2 y, & 0 < x < 1, 0 < y < 1, 0 < z < 2 \\ 0 & \text{elsewhere} \end{cases}$$

a) Find k
b) Find P(X < 1/4, Y > 1/2, 1< Z < 2).

5. Let X denote the number of times a certain numerical control machine will malfunction: 1, 2, or 3 times on a given day. Let Y denote the nuber of times a technician is called on an emergency call. Their joint probability distribution is given as:

			x		
F(x, y)		1	2	3	row totals
	1	0.05	0.05	0.1	
y	2	0.05	0.1	0.35	
	3	0	0.2	0.1	
column totals					

a) Evaluate the marginal distribution of X
b) Evaluate the marginal distribution of Y.
c) Find $P(Y=3|x=2)$

MATHEMATICAL EXPECTATION:

If x is a random variable with probability f(x), then the mathematical expectation of x, or **expected value** of x, or **mean or average** of x is given by

$$E(x)=\mu_X=\sum_x xf(x) \qquad \text{for discrete random variable, x}$$

$$E(x)=\mu_X=\int_{-\infty}^{\infty} xf(x)dx \qquad \text{for continuous random variable, x}$$

Exercises:

1. The probability distribution of the discrete random variable X is $f(x) = \binom{3}{x}\left(\frac{1}{4}\right)^{x}\left(\frac{3}{4}\right)^{3-x}$ for x = 0, 1, 2, 3. Find the mean of x.

2. A coin is biased so that the head is three times as likely to occur as a tail. Find the expected number of tails when this coin is tossed twice.

3. The probability distribution of X, the number of imperfections per 10 meters of a synthetic fabric in continuous rolls of uniform width, is given by

x	0	1	2	3	4

F(x)	0.41	0.37	0.16	0.05	0.01

Find the average number of imperfections per 10 meters of fabric.

4. An attendant at a car wash is paid according to the number of cars that pass through. Suppose the probabilities are 1/12, 1/12, 1/4, 1/4, 1/6, and 1/6, respectively , that the attendant receives \$7, \$9, \$11, \$13, \$15, or \$17 between 4:00 p.m. and 5:00 p.m. on any sunny Friday. Find the attendant's expected earnings for this particular period.

5. By investing a particular stock, a person can make a profit in one year of \$4000 with probability 0.3 or take loss of \$1000 with probability 0.7. What is this person's expected gain?

6. If a dealer's profit, in units of \$1000, on a new automobile can be looked upon as a random variable X having the density function $f(x) = \begin{cases} 2(1-x), & 0 < x < 1 \\ 0, & \text{elsewhere} \end{cases}$. Find the average profit per automobile.

7. The density function of coded measurements of pitch diameter of threads of a fitting is given by $f(x) = \begin{cases} \dfrac{4}{\pi(1+x^2)}, & 0 < x < 1 \\ 0, & \text{elsew} \end{cases}$.

Find the expected value of X.

8. What proportion of the people can be expected to respond to a certain mail – order solicitation if the proportion X has the density function $f(x) = \begin{cases} \frac{2(x+2)}{5}, & 0 < x < 1 \\ 0, & elsew \end{cases}$

VARIANCE AND COVARIANCE

The mean or expected value of a random variable X is of special importance in statistics because it describes where the probability distribution is centered. By itself, however, the mean does not give adequate distribution. We need to characterize the variability in the distribution. The most important measure of variability of a random variable X is obtained by $g(X) = (X - \mu)^2$, because of its importance in statistics, it is referred to as the variance of the random variable X or the variance of the probability distribution of X

σ^2 = Variance (sigma)
S.D = standard deviation
= $\sqrt{\sigma^2} = \sigma$

Let X be a random variable with probability distribution f(x) and mean μ The variance of X is

if X is discrete:

$$\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2 f(x)$$

if X is continuous:

$$\sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

The positive square root of the variance, σ, is called the standard deviation of X.

Exercises:

1. Let X be a random variable with the following probability distribution:

x	–2	3	5
f(x)	0.3	0.2	0.5

Find the standard deviation of X.

2. The random variable X, representing the number of errors per 100 lines of software codes, has the following probability distribution:

x	2	3	4	5	6
f(x)	0.01	0.25	0.4	0.3	0.04

Find the variance of X.

3. The dealer's profit, in units of $ 1000, on a new automobile is a random variable X having the density function $f(x) = \begin{cases} 2(1-x), & 0 < x < 1 \\ 0, & \text{elsewhere} \end{cases}$

Find the variance of X.

4. The proportion of people who respond to a certain mail order solicitation is a random variable X having the density function $f(x) = \begin{cases} \frac{2}{5}(x+2), & 0 < x < 1 \\ 0, & \text{elsewhere} \end{cases}$. Find the variance of X.

COVARIANCE

Let X and Y are random variables with joint probability distribution f (x, y). The COVARIANCE of X and Y is:

If X and Y are discrete:

$$\sigma_{xy} = E[(X-\mu_x)(Y-\mu_y)] = \sum_x \sum_y (x-\mu_x)(y-\mu_y) f(x,y)$$

Or

$$\sigma_{xy} = E(xy) - \mu_x \mu_y$$

If X and Y are continuous:

$$\sigma_{xy} = E[(X-\mu_x)(Y-\mu_y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_x)(y-\mu_y) f(x,y)\,dxdy$$

Or $\sigma_{xy} = E(xy) - \mu_x \mu_y$

Exercises:

1. Suppose that X and Y are independent random variables having the joint probability distribution

f(x, y)		x = 2	x = 4
	1	0.10	0.15
y	3	0.20	0.30
	5	0.10	0.15

Find:

a) Find the covariance in X and Y

b) Find the E (XY)

c) Find E (2x– 3y)

2. Let X represents the number that occurs when a red die is tossed and Y the number that occurs when a green die is tossed.

Find

a) E (X + Y)

b) E (X – Y)

c) E (XY)

d) σ_{xy}

SOME DISCRETE PROBABILITY DISTRIBUTION

A. Uniform Distribution

If the discrete random variable X assumes the values such as $x_1, x_2, x_3, \ldots x_n$ with equal possibilities, then the discrete uniform distribution is given by:

$$f(x;k) = \frac{1}{k}$$, $x_1, x_2, x_3, \ldots x_n$ and k is the # of outcomes

Each event has an equal probability of occurrence

Ex. In tossing a fair die

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

f(x; k) = f(x; 6) = 1/6

Ex. Flip of a coin

K = 2

P(H) = 1/2

P(T) = 1/2

F(x; 2) = 1/2

Mean and Variance of a Uniform Distribution

$$\mu(mean) = \frac{\sum_{x}^{k} x_i's}{k}$$

$$\sigma^2(\text{var}iance) = \frac{\sum_{x}^{k} (x_i - \mu)^2}{k}$$

Exercises:

1. An employee is selected from a staff of 10 to supervise a certain project by selecting a tag at random from a box containing 10 tags numbered from 1 to 10. Find the formula for the probability distribution of X representing the number on the tag that is drawn. What is the probability that the number drawn is less than 4.

2. A roulette wheel is divided into 25 sectors of equal area numbered from 1 to 25. Find a formula for the probability distribution of x, the number that occurs when the wheel is spin.

3. Find the mean and variance of the random variable x of exercise 1.

B. Binomial Distribution

A binomial random variable with parameter n and p represents the number of successes in n independent trials, when each trial is a success with probability p.

$$b(x;n;p) = \binom{n}{x}(p^n)(q)^{n-x}$$

where:

n – # of outcomes or cases
x – 0, 1, 2, 3, 4, . . .n
p – PROBABILITY OF SUCCESS
q – probability of failure

Exercises:

1. In a certain city district the need for money to buy drugs is given as the reason for 75 % of all thefts. Find the probability that among the next 5 thefts cases reported in this district

a) Exactly 2 resulted from the need for money to buy drugs
b) At most 3 resulted from the need for money to buy drugs.

Solution:
let x = the event that the need for money to buy drugs is the reason for theft.

2. A fruit grower claims that 2/3 of his peaches crop has been contaminated by the medfly infestation. Find the probability that among the 4 peaches inspected by this grower

a) All 4 have been contaminated;
b) Anywhere from 1 to 3 have been contaminated.

Solution:
Let x – event that a peach is contaminated

3. One prominent physician claims that 70 % of those with lung cancer are chain smokers. If his assertion is correct:

a) Find the probability that of 10 such patients recently admitted to a hospital, fewer than half are chain smokers.
b) Find the probability that of 20 such patients recently admitted to a hospital, fewer than half are chain smokers.

4. In testing a certain truck tire over a rugged terrain, it is found that 25 % of the trucks fail to complete the test run without a blowout. Of the next 15 trucks tested, find the probability that

a) From 3 to 6 have blowouts;
b) Fewer than 4 have blowouts;
c) More than 5 have blowouts.

Final Assignment
NAME:________________________________SCHEDULE: _____________

1. A nationwide survey of seniors by the University of Michigan reveals that almost 70% disapprove of daily pot smoking according to a report in Parade, September 14, 1980. If 12 seniors are selected at random and asked their opinion, find the probability that the number who disapprove of smoking pot daily is

a) Anywhere from 7 to 9;

b) At most 5;

c) Not less than 8.

2. According to a study published by a group of university of Massachusetts sociologist, approximately 60 % of the Valium users in the state of Massachusetts first took Valium for psychological problems. Find the probability that among the next 8 users interviewed from this state.

a) Exactly 3 began taking Valium for psychological problems.

B) At least 5 began taking Valium for problems that were not psychological.

3. The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what is the probability that

a) At least 10 survive

b) From 3 to 8 survive

c) Exactly 5 survive

4. Let the random variable X represent the number of automobiles that are used for official business purposes on any given workday. The probability distribution

for company A is given by:

x	1	2	3
f(x)	0.3	0.4	0.3

and for company B is given by:

x	0	1	2	3	4
f(x)	0.2	0.1	0.3	0.3	0.1

Show that the variance of the probability distribution for company B is greater than that of company A.

C. MULTINOMIAL DISTRIBUTION

- There are k possible outcomes out of n trials
- p_1, p_2, p_2, . . ., p_k are the probabilities of success of events 1, 2, 3, . . . k respectively.

$$\sum_{i=1}^{k} p_i = 1 \qquad \sum_{k=1}^{k} x_i = n$$

$$f(x_1, x_2, x_3, \ldots, x_k; n, p_1, p_2, p_3, \ldots p_k) = \frac{n!}{x_1! x_2! \ldots x_k} p_1^{x1} p_2^{x2} .. p_k^{xk}$$

Ercises:

1. A card is drawn from a well – shuffled deck of 52 playing cards, the result recorded, and the card replaced. If the experiment is repeated 5 times, what is the probability of obtaining 2 spades and 1 heart?

Let: x_1 = event a spade is taken = 2
X_2 = event a heart is taken = 1
X_3 = event a card other than spade/ heart is taken = 2
n = 5

2. According to the theory of genetics, a certain cross of guinea pigs will result in red, black, and white offspring in the ratio 8: 4: 4. Find the probability that among 8 offspring 5 will be red, 2 black, and 1 white.

Let: x_1 = offspring will be red = 5
X_2 = offspring will be black = 2
X_3 = offspring will be white = 1
n = 8

3. The probabilities are 0.4, 0.2, 0.3, and 0.1, respectively, that a delegate to a certain convention arrived by air, bus, automobile, or train. What is the probability that among 9 delegates randomly selected at this convention, 3 arrived by air, 3 arrived by bus, 1 arrived by automobile, and 2 arrived by train?

D. HYPERGEOMETRIC DISTRIBUTION

The probability distribution of the hyper geometric random variable X, the number of successes in a random sample of size n selected from N items of which k are labeled success and N–k labeled failure, is

$$h(x; N, n, k) = \frac{{}_{k}C_{x} \bullet {}_{N-k}C_{n-x}}{{}_{N}C_{n}}$$

where: x = 0, 1, 2, 3, . . ., n;
n = the # of items being dealt with
k = # of items labeled as success
N = total # of items

Exercises:

1. If 7 cards are being dealt from an ordinary deck of 52 playing cards, what is the probability that

a) Exactly 2 of them will be face cards?

b) At least one of them will be a queen?

2. To avoid detection at customs, a traveler has placed 6 narcotic tablets in a bottle containing 9 vitamin pills that are similar in appearance. If the customs official selects 3 of the tables at random for analysis, what is the probability that the traveler will be arrested for illegal possession of narcotics?

3. A homeowner plants 6 bulbs selected at random from a box containing 5 tulip bulbs and 4 daffodil bulbs. What is the probability that he planted 2 daffodil bulbs and 4 tulip bulbs?

4. From a lot of 10 missiles, 4 are selected at random and fired. If the lot contains 3 defective missiles that will not fired. What is the probability that

e. Al 4 will fire?

b) At most two will not fire?

E. MULTIVARIATE HYOERGEOMETRIC DISTRIBUTION

If N items can be partitioned into k cells $A_1, A_2, \ldots, A_k$ with $a_1, a_2, \ldots, a_k$ elements, respectively, then the probability distribution of the random variable $X_1, X_2, \ldots, X_k$, representing the number of elements selected from $A_1, A_2, \ldots, A_k$ in a random sample of size n, is

$$\sum_{i=1}^{k} x_i = n \qquad \sum_{k=1}^{k} a_i = N$$

$$f(x_1, x_2, x_3, \ldots, x_k; a_1, a_2, a_3, \ldots a_k, N, n) = \frac{{}_{a1}C_{x1} \bullet {}_{a2}C_{x2} \bullet \cdots {}_{ak}C_{xk}}{{}_{N}C_{n}}$$

Exercises:

1. Find the probability of being dealt a bridge hand of 13 cards containing 5 spades, 2 hearts, 3 diamonds, and 3 clubs.

2. A foreign student club lists as its members 2 Canadians, 3 Japanese, 5 Italians, and 2 Germans. If a committee of 4 is selected at random, find the probability that
a) All nationalities are represented;

b) All nationalities except the Italians are represented.

F. NEGATIVE BINOMIAL DISTRIBUTION

If repeated independent trials can result in a success with probability p and failure with probability q, then the probability distribution of the random variable x, the # of trial in which the k th success occurs is given by

$b^*(x,k,p) = \binom{x-1}{k-1} p^k q^{x-k}$ Where: x = k, k +1, k + 2 …

Exercises:

1. The probability that a person living in a certain city owns a dog is estimated to be 0.3. Find the probability that tenth person randomly interviewed in this city is the fifth one to own a dog.

Solution:

p = 0.3
q = 1 – 0.3 = 0.7
x = 10^{th}
k = 5^{th}

P =

2. A scientist inoculates several mice, one at a time, with a disease germ until he finds two that have contracted the disease. If the probability of contracting the disease is 1/6, what is the probability that 8 mice3 are required?

Solution:

$p = 1/6$
$q = 1 – 1/6 = 5/6$
$x = 8$
$k = 2$

P =

3. Find the probability that a person flipping a coin gets

a) The third head on the seventh flip;
b) The first head on the fourth flip;

Solution:

a) $p = 1/2$
$q = 1 – 1/2 = 1/2$
$x = 7^{th}$
$k = 3^{th}$

P =

b) $p = 1/2$
$q = 1 – 1/2 = 1/2$
$x = 4^{th}$
$k = 1^{st}$

P =

4. Suppose the probability is 0.8 that any given person will believe a tale about the transgressions of a famous actress. What is the probability that

a) The sixth person to hear this tale is the fourth one to believe it?
b) The third person to hear this tale is the first one to believe it?

Solution:

a) $p = 0.8$
$q = 1 - 0.8 = 0.2$
$x = 6^{th}$
$k = 4^{th}$

P =

a) $p = 0.8$
$q = 1 - 0.8 = 0.2$
$x = 3^{th}$
$k = 1^{st}$

P =

G. POISSON'S DISTRIBUTION

The probability distribution of the Poisson random variable X, representing the # of outcomes occurring in a given time interval or specified region is denoted by:

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

Where:

$x = 0,1,2,3,..., \infty$
$\mu = \lambda t$
$x = is$ the average # of outcomes per unit time
$e = 2.7182818$

Exercises:

1. In an inventory study it was determined that, on the average, demands for particular items at a warehouse were made 5 times per day. What is the probability that on a given day this item is requested?
   a) More than 5 times
   b) Not at all.

Solution:

a) $\mu = 5$ times
$x = 6, 7, 8\ldots \infty$

$$P(x > 5) = 1 - \sum_{0}^{5} (x;5)$$

b) $\mu = 5$ times
$x = 0$
$P(x = 0) =$

2. on the average of a certain intersection results in 3 traffic accidents per month. What is the probability that in any given month at this intersection

a) Exactly 5 accidents will occur?
b) Less than three accidents will occur?
c) At least two accidents will occur?

Solution:

a) $\mu = 3$
$x = 5$

$P(x = 5) =$

b) $\mu = 3$
$x = 0, 1, 2$

$$P(x < 3) = \sum_{0}^{2} P(x;3)$$

c) $\mu = 3$
$x = 2, 3, 4\ldots \infty$

$$P(x \geq 2) = 1 - \sum_{0}^{1} P(x;3)$$

3. A certain area of the United States is, on the average, hit by 6 hurricanes a year. Find the probability that in a given year this area will be hit by

a) Fewer than four hurricanes;
b) Anywhere from 6 to 8 hurricanes.

Solution:

a) $\mu = 6$
$x = 0, 1, 2, 3$

$P(x < 4) =$

b) $\mu = 6$
$x = 6, 7, 8$

$P(6 \leq x \leq 8) =$

H. GEOMETRIC DISTRIBUTION

If repeated independent trials can result in a success with a probability of p and a failure with probability q, then the probability distribution of the random variable x, representing the number of trials on which the 1st success occurs, is given by:

$g(x;p) = pq^{x-1}$ Where x = 1, 2, 3 …

Exercises:

1. The probability that a student pilot passes the written test for his private pilot's license is 0.7. Find the probability that a person passes the test
   a) On the third try;
   b) Before the fourth try.

Solution:

a) p = 0.7
   q= 0.3
   x = 3

$P(x = 3) =$

b) p = 0.7
   q= 0.3
   x = 1, 2, 3

$P(x < 4) =$

EMPIRICAL DISTRIBUTION

Raw Data

Raw data are collected data which have not been organized numerically. An example is the set of heights of 100 male students from an alphabetical listing of university records.

Table 1

HEIGHT (INCHES)	Number of Students (Frequency)
60-62	5
63-65	18
66-68	42
69-71	27
72-74	8

Total 100

Arrays

An array is an arrangement of raw numerical data in ascending or descending order of magnitude.

The difference between the largest and the smallest numbers is called the **RANGE**

$$R = Hv - Lv$$

For example, if the largest height of 100 male students is 74 inches and the smallest height is 60 inches, the range is 74 - 60 = 14 inches.

Frequency Distribution

When summarizing large masses of raw data it is often useful to distribute the data into classes or categories and to determine the number of individuals belonging to each class, called the CLASS FREQUENCY

A tabular arrangement of data by classes together with the corresponding class frequencies is a FREQUENCY DISTRIBUTION or FREQUENCY TABLE

Example: Table 1 is a frequency distribution of heights (recorded to the nearest inch) of 100 male students at XYZ University.

The first class category, for example, consists of heights from 60 to 62 inches and is indicated by the symbol 60 – 62. Since 5 students have heights belonging to this class, the corresponding **class frequency** is 5

Class Intervals and Class Limits

A symbol defining a class such as 60 – 62 in the above table is called a CLASS INTERVAL. The end numbers, 60 and 62, are called class limits; the smaller number 60 is the lower class limit and the larger number 62 is the upper class limit. The terms class and class interval are often used interchangeably, although the class interval is actually a symbol for the class.

A class interval which, at least theoretically, has either no upper class limit or no lower class limit indicated is called an open interval.

For example: referring to age groups of individuals, the class interval "65 years and over" is an open class interval.

Class Boundaries

If heights are recorded to the nearest inch, the class interval 60 – 62 theoretically includes all measurements from 59.5000… to 62.5000… inches. These numbers, indicated briefly by exact numbers 59.5 and 62.5, are called CLASS BOUNDARIES or TRUE CLASS LIMITS; the smaller number 59.5 is THE LOWER CLASS BOUNDARY and the larger number 62.5 is the UPPER CLASS BOUNDARY.

Size or Width of a Class Interval

The size or width of a class interval is the difference between the lower and the upper class boundaries and also referred to as the **class width, class size or class length**.

If all class intervals of a frequency distribution have equal widths, this common width is denoted by C. In such case C is equal to the difference between two successive upper class limits.

For the data in Table 1, for example, the class interval is c = 65.5 – 62.5 = 3

**Class Size – this is the width of the class interval

C = R / K,

Where k = 1 + 33 log N;
Where N = number of data

The Class Mark or Midpoint

The class mark is the midpoint of the class interval and is obtained by adding the lower and the upper class limits and dividing by 2. Thus, the class mark of the interval 60 – 62 is (60 + 62)/2 = 61. The class mark is also called the class midpoint.

The purpose of further mathematical analysis, all observations belonging to a given class interval are assumed to coincide with the class mark. Thus all heights in the class interval 60 – 62 inches are considered as 61 inches.
Relative Frequency Distributions

The relative frequency of a class is the frequency of the class divided by the total frequency of all classes and is generally expressed as a percentage.

For example: the relative frequency of all the class 66 – 68 in Table 1 is 42/100 = 42%.

The sum of the relative frequencies of all classes is clearly 1 or 100%

If frequencies in the above frequency table are replaced by corresponding relative frequencies, the resulting table is called a relative frequency distribution, percentage distribution, or relative frequency table.

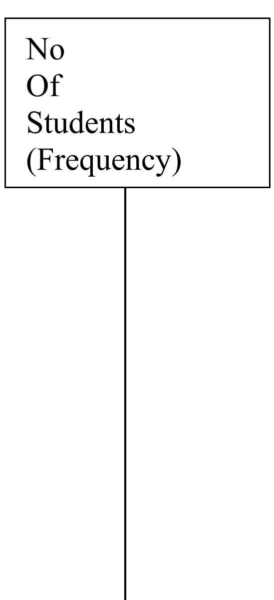
Graphical representations of relative frequency distributions can be obtained from the histogram or frequency polygon by simply changing the vertical scale from frequency to relative frequency, keeping exactly the same diagram. The resulting graphs are called relative frequency histogram or percentage histograms and relative frequency polygons or percentage polygons, respectively

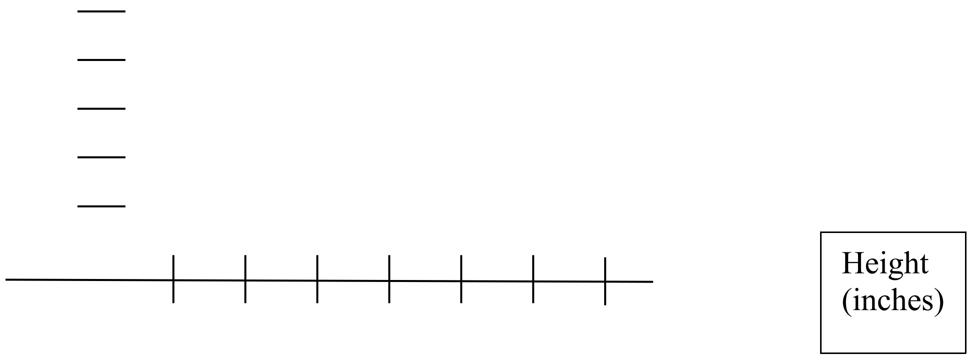
Histogram and Frequency Polygon

1. A histogram or frequency histogram consists of a set of rectangles having

   a) bases on a horizontal axis with centers at the class marks and lengths equal to the class interval sizes,
   b) Areas proportional to class frequencies.

2. A frequency polygon is a line graph of class frequency plotted against class mark. It can be obtained by connecting midpoints of the tops of the rectangles in the histogram.





Cumulative frequency Distribution. Ogives

The total frequency of all values less than the upper class boundary of a given class interval is called the *cumulative frequency* up to and including that class interval.

For example: The cumulative frequency up to and including the class interval 66 – 68 in Table 1 is 5 + 18 + 42 = 65, signifying that 65 students have heights less than 68.5

A table representing such cumulative frequencies is called cumulative frequency distribution, cumulative frequency table, or briefly a cumulative distribution, and is shown in Table 2 for the student height distribution

Height (inches)	Number of Students (Cumulative frequency ≥)	Height (inches)	Number of Students (Cumulative frequency ≤)
Less than 59.5	0	60 or more	100
Less than 62.5	5	62 or more	95
Less than 65.5	23	65 or more	77

Less than 68.5	65	68 or more	35
Less than 71.5	92	71 or more	8
Less than 74.5	100	74 or more	0

** Less than cumulative and greater than cumulative frequency

It is often desirable to cumulate, or add up, the absolute frequencies of a distribution to determine the number of observations that lie above (greater than) of below (less than) a class boundary. When the successive frequencies are added from the smallest to the largest class interval, we obtain a less than cumulative frequency distribution. When the frequencies are cumulated starting from that of the largest class interval the result is a greater than cumulative frequency distribution.

SUMMARY: Frequency Distribution

A. Frequency Distribution, defined

A frequency distribution is a tabular arrangement of data by classes or categories.

B. Parts of Frequency Distribution

1. Classes or categories or class intervals. These are groupings defined by a lower limit and an upper limit. Classes are preferred to be of uniform size and not overlapping.

2. Class frequency (f) - refers to the number of observations belonging to a class interval, or the number of items within a category.

3. Class boundaries - are more precise expressions of the class limits by at least .5 of their values. They are called the true class limits. The boundary is situated between the upper limit of one interval and the lower limit of the next interval.

4. Class midpoint or class mark (X_{mid} or X_m). It indicates the middle value of every class interval. It is obtained by taking the sum of the lower limit and the upper limit of the class interval and then dividing the sum by 2.

5. Cumulative frequency (fc). There are two types

a) Cumulative frequencies less or equal to the upper class limit (fc $\leq$ u.l).
This is obtained by taking the sum of the frequencies class by class, stating from the uppermost frequency down to the lowermost frequency.

b) Cumulative frequency greater than or equal to the lower class limits (fc ≥ l.u).

This is obtained by taking the sum of the class frequencies class by class, starting from the lowermost and upward to the uppermost frequency.

6. Relative Frequency (frel). It is in this calss frequency expressed a s a fractional part of the total frequency. The value may be in percent or in decimal form. The total of the frequency is one (1) or almost 1.

C. Steps in Constructing a Frequency distribution

1. Determine or decide on the number of class intervals or categories desired. The ideal number of class interval is somewhere between 5 and 15.

2. Determine the range by getting the difference between the highest and the lowest values in the data set.

3. Determine the class interval size (i) by following the formula.

$$i = \frac{highestdat\ avalue - lowestdata\ value}{numberofde\ siredclas\ \sin terval}$$

Class interval size I must be a whole number. Therefore do the necessary rounding procedures for i.

4. Determine or decide on the x-lowest value. This x-low ≤ lowest data value, this becomes the lower class limit of the first class interval.

5. Apply the class interval size to get all the lower limits and upper limits of all the class intervals.

6. Determine the class frequencies for each class interval by taking the sum of the tally column.

7. Show all the other columns of, class boundaries, class midpoints, frequency cumulative columns and that of the relative frequency.

D. Example

A research study on the benefits and privileges of employees of a certain company was conducted. One of the questions asked was their salary on an hourly basis. Response results were as follows.

120	140	161	148	130	156	165	146	142	173
133	150	149	139	143	151	138	150	158	148
180	170	124	161	137	120	147	149	152	142
138	153	168	142	147	118	167	129	130	159

Construct the frequency distribution table considering:

1. number class interval is 7
2. x-low = 118
3.

The frequency distribution table for the data above

Classes i=9	f	Class Boundaries	Class Midpoint Xmidpt	fc ≤ UL	fc ≥ LL	frel
118-126	4	117.5-126.5	122	4	40	.1
127-135	4	126.5-135.5	131	8	36	.1
136-144	9	135.5-144.5	140	17	32	.225
145-153	12	144.5-153.5	149	29	23	.3
154-162	5	153.5-162.5	158	34	11	.125
163-171	4	162.5-171.5	167	38	6	.1
172-180	2	171.5-180.5	176	40	2	.05
total	40					1

Some interpretations:

1. 9 under frequency indicate that there are 9 employees receiving salaries which range from P136 to P144 per hr.

2. 17 under fc ≤ UL means, there are 17 employees who receive salaries which are lesser or equal to the upper limit of P144/hr. These salaries are: P120, 140, 130, 142, 133, 139, 143, 138, 124, 137, 120, 142, 138, 142, 118, 129, & 130.

3. 11 under fc ≥ LL imply that there are 11 employees who receive salaries which are greater or equal to the lower limit P154/hr. they are: P161, 156, 165, 173, 158, 180, 170, 161, 168, 167, & 159.

4. .3 under relative frequency means there are 30% of the employees who receive salaries which range from P145 to P153/hr.

The Graphs of Data Gathered Based on a
Frequency Distribution Table

A graph aims to give a visual representation of the data gathered. The graphs are the histogram, the frequency polygon, the ogives, and the steam – leaf display.

1. The Histogram

It makes use of bar graphs which may be horizontal or vertical. If vertical, the x – axis scales the class boundaries and the y – axis scales the class frequencies. It emphasizes differences between classes within a distribution.

2. The Frequency polygon

This graph makes use of broken lines to connect the plotted points. The X – axis scales the class midpois while the y – ascales the class ies. The frequency polygon emphasizes the spread of the distribution

3. The Ogive

It makes use of a smooth curve. The x – axis scales boundaries and the y – axis scales the cumulative frequencies. There are two kinds: ogive less or equal to the upper limit (ogive $\leq$ UL) and the ogive greater or equal to the lower limit (ogive $\geq$ LL)

A) Ogive $\leq$ UL

It's graph starts from the lowest left corner, (P117.50, 0) and then smoothly goes upward to the uppermost right corner (P180.5, 40).

Interpretations:

(117.50, 0) means that no one from the 40 employees receive salaries which are lesser or equal to P117.50/hr, because the lowest salary is P118 per hr.

(180.50, 40) indicates that all the 40 employees receive salaries which are lesser than P180.50 per hr.

B) Ogive $\geq$ LL

The graphs starts from the uppermost left corner, (P117.50, 40) and smoothly goes downward to the lowest right corner (180.50, 0).

Interpretations:

(P117.50, 40) means that all the 40 employees receive salaries which are greater than P117.50 per hr.

(180.50, 0) indicates that none of the 40 employees receive salaries which are greater or equal to P180.50/hr.

The ogive graphs for the data are:

4. Steam –Leaf Display

The tens digit of the number is the leaf while the other digits serve as the stem. For this data, the stem – leaf display is shown below.

11	8								
12	0	4	0	9					
13	0	3	9	8	7	8	0		
14	0	8	6	2	9	3	8	7	9
15	6	0	1	0	8	2	3	9	
16	1	5	1	8	7				
17	3	0							
18	0								

Exercises:

1. A canteen recorded its daily sales of soft drinks S for fifty days as follows. The data is by the number of cases.

80	104	95	95	96	138	126	128	114	119

85	97	98	103	104	137	128	110	115	116
145	106	106	105	109	135	127	111	117	118
149	107	105	103	136	123	121	112	114	115
93	132	133	134	124	125	120	113	116	112

Required:

1. Show the frequency distribution table

   Use x-low = 80 cases and number of classes is 7.

2. Draw the 4 graphs

   a) Histogram

   b) frequency polygon

c) Ogives

d) Steam –leaf display

Final Assignment #2
Name: ___________________________________

1. The final grades in mathematics of 80 students at state University are recorded in the accompanying table:

68	84	75	82	68	90	62	88	76	93
73	79	88	73	60	93	71	59	85	75
61	65	75	87	74	62	95	78	63	72
66	78	82	75	94	77	69	74	68	60
96	78	89	61	75	95	60	79	83	71
79	62	67	78	85	76	65	71	75	97
65	80	73	57	88	78	62	76	53	74
86	67	73	81	72	63	76	75	85	77

With reference to this table
1. Find:
   a) the highest grade
   b) the lowest grade
   c) the range

d) the grades of the five lowest ranking students
e) the grade of the highest ranking students
f) the grade of the student ranking tenth highest
g) how many students received grades of 75 or higher
h) how many students received grades below 85
i) what percentage of students received grades higher than 65 but not higher that 85
j) Which grade did not appear at all?

Answers:

a)
b)
c)
d)
e)
f)
g)

h)

i)

j)

2. Show the frequency distribution table
use x-low = 53 and number of classes is 9

3. Draw the 4 graphs and label.

a) Histogram

b) Frequency distribution

c) Ogives

d) Steam – leaf display

2. In the following table the weights of 40 male students at State University are recorded to the nearest pound. Construct a frequency distribution. Draw the four graphs.

138	164	150	132	144	125	149	157
158	140	136	148	152	144	146	147
168	126	138	176	163	119	154	165
146	173	142	147	135	153	140	135
161	145	135	142	150	156	145	128

MEASURES OF AVERAGES OR MEASURES OF CENTRAL TENDENCIES

Definition:

It is a single number which describes the center or middle of a set of data

Types:

I. The **Mean**, Arithmetic Mean or Compound Average $\bar{x}$ or μ

$\bar{x}$ is sample arithmetic mean

μ is population arithmetic mean

A. Features of an arithmetic mean

- Every item is included in the computation

- Mean can take on a value that is not among the values of the data from which it is computed.

- Sum of the values of the deviations of the individual items from arithmetic mean is equal to zero.

- Trimmed mean is the mean of the remaining data after removing or eliminating the unusually very large data values from the original set of data. The purpose is to make the trimmed mean a better indicator of the central location of the data.

B. Computations of the Arithmetic Mean

DATA

A) Ungrouped data, $n < 30$

B) Grouped data, $n \geq 30$

1. Ungrouped Data

- Data considered as individual items.
- The mean for ungrouped data is computed by simply adding all the values and dividing the sum by the number of values.

Sample Mean, Mean $\bar{x} = \frac{\sum x}{n}$

Population Mean $\mu = \frac{\sum x}{N}$

Where $\sum$ = sum of

x = the individual data values

N = the number of items in the population data set

n = the number of items in the sample data set.

Example: Five light bulbs burned out after lasting for 867, 849, 840, 852, and 822 hours of continuous use. Find the mean.

$$\bar{x} = \frac{867 + 849 + 840 + 852 + 822}{5} = 846 hrs$$

2. Grouped data

- Data which are grouped into classes
- The mean of a grouped frequency distribution may be obtained in almost the same way as the mean of an ungrouped frequency distribution is computed.
- All the scores included in the class interval are represented by the midpoint or class mark of that class interval. This class mark is multiplied by its corresponding frequency; the product is summed, and divided by n or N depending on whether the data constitute a sample or a population.

a) Long Method formula:

Sample Mean, Mean $$\bar{x} = \frac{\sum fx_{mid}}{n}$$

Population Mean $$\mu = \frac{\sum fx_{mid}}{N}$$

Where: f = frequency

n = $\sum f$ = total frequency

x_{mid} = class mark or class midpoint

b) Deviation Method:

Use if i is the same for all classes

Sample Mean, Mean $$\bar{x} = \bar{x}' + \left(\frac{\sum fd'}{n}\right) i$$

Population Mean $$\mu = \mu' + \left(\frac{\sum fd'}{N}\right) i$$

Where: μ' = assumed population mean

$\bar{x}'$ = assumed sample mean

$$d' = deviationp\ rime = \frac{x_{mid} - \bar{x}'}{i} \ or \ \frac{\mu - \mu'}{i}$$

d' = difference of the X_{mid} from the column of class midpoints. It may be any of these values but preferably the value of the highest frequency and/or that value which occupies the middle most position.

Example: Grouped Data

The hourly salaries of 40 employees as grouped data are as follows. Find the arithmetic mean hourly salary.

i = 9	f	X_{mid}	fx_{mid}	d'	fd'
118-126	4	122	488	-3	-12
127-135	4	131	524	-2	-8
136-144	9	140	1260	-1	-9
145-153	**12**	**149**	1788	0	0
154-162	5	158	790	1	5
163-171	4	167	668	2	8
172-180	2	176	352	3	6
N or n	$\sum = 40$		$\sum = 5870$		$\sum = -10$

Using the long method:

$$\bar{x} = \frac{\sum fx_{mid}}{n}$$

$$\bar{x} = \frac{5870}{40} = 146.75$$

Using the deviation method:

$$\boxed{\bar{x} = \bar{x}' + \left(\frac{\sum fd'}{n}\right)i}$$

take $\bar{x}' = 149$

$$d' = \frac{140 - 149}{9} = -1$$

$$d' = \frac{131 - 149}{9} = -2$$

$$d' = \frac{122 - 149}{9} = -3$$

$$\bar{x} = 149 + \left(\frac{-10}{40}\right)9 = 146.75$$

3. Weighted Arithmetic Mean

- The arithmetic mean, in which each value is weighted according to its importance in the overall group.

$$\overline{x_w} = \frac{\sum w_i \bar{x}_i}{\sum w_i}$$

where: x = score or value

w = weight value of the data

Example:

In three separate weeks, a discount store chain sold 475,310 and 420 microwave ovens at average prices of $490, $520, and $495. What is the average price of the ovens sold?

$$\bar{x} = \frac{475(\$490) + 310(\$520) + 420(\$495)}{475 + 310 + 420} = \$499.46$$

II. The Median (Md)

This is also a measure of average which is also called Position Average. It is that data value which occupies the middle position when data values are arranged in an ascending or descending order which is called an array.

Computation of the Median (Md):

A. For Ungrouped Data:

Data must be an array which means data is arranged in an ascending order.

If data item is **odd** number, there is only one item occupying the middle position

$$Md = \left(\frac{n+1}{2}\right)^{th} value$$

Example: n = odd

Eleven large corporations reported that in 1997, they made cash donations to 9, 16, 11, 19, 11, 10, 13, 12, 6, 9, and 12 colleges. Find the median number of donations.

Array: 6 9 9 10 11 11 12 12 13 16 19 n = odd = 11
Median = (11 + 1)2 = 6^{th} value = 11 donations
Data 1: 7, 2, 3, 7, 6, 9, 10, 8, 9, 9, 10

If a data item is **even** number, there are two data values occupying the middle position. Therefore the median is the average of these two values.

Example: n = even

On ten days, a bank had 18, 13, 15, 12, 8, 3, 7, 14, 16 and 3 foreign currency transactions. Find the median.

Array: 3 3 7 8 12 13 14 15 16 18
Middle values are 12 and 13
Median = (12 + 13)/2 = 12.5

B. For Grouped Data:

$$Md = L_{Md} + \left(\frac{\frac{N}{2} - \sum fc<}{f_{Md}} \right) i$$

Where:

$\frac{N}{2} = \frac{\sum f}{2}$ = this value indicates the median class

L_{Md} = the lower class boundary of the median class

f_{Md} = the class frequency of the median class

$\sum fc<$ = the sum of all the frequencies of the classes above the median class (classes are arranged in an ascending order)

i = class size

Consider:

i = 9	f	fc<
118-126	4	4
127-135	4	8
136-144	9	17
145-153	12	29
154-162	5	34
163-171	4	38
172-180	2	40
	n = 40	

$$\frac{N}{2} = \frac{\sum f}{2} = \frac{40}{2} = 20^{th}\ term$$

Indicates 145 – 153 is the median class

$L_{Md} = 144.5$

$fc< = 17$

$f_{Md} = 12$

$i = 9$

$$Md = 144.5 + \left(\frac{20 - 17}{12} \right) 9 = 146.75$$

III. The Mode (Mo)

The Mode is the third type of measures of averages. It is the most frequently occurring value within the data set. Data sets may have only one mode or at times polymodal

Unimodal – one mode
Bimodal – 2 modes
Multimodal – more than 2 modes

A. For Ungrouped Data

Mo = the value or score which occurs with the greatest frequency

Ex:
Data 1: 1, 2, 2, 2, 3, 4, 5, 6, 7, 9
Mode = 2 unimodal

Data 2: 1, 2, 2, 2, 3, 4, 5, 6, 6, 7, 9

Mode = 2 and 6 bimodal

Data 3: 1, 2, 2, 2, 3, 4, 5, 6, 6, 7, 7

Mode = 2, 6, and 7 multimodal

B. Computation of the Mode for Grouped Data:

$$Mo = L_{Mo} + \left(\frac{A}{A+B} \right) i$$

Modal class is the class with the highest frequency.

L_{Mo} = the lower class boundary of the modal class

Let the classes be arranged in an ascending order. Therefore

A = is the difference between the frequencies of the modal class and the class preceding it.

B = is the difference between the frequencies of the modal class and the class succeeding it.

Consider

i = 9	f
118-126	4
127-135	4
136-144	9
145-153	**12**
154-162	5
163-171	4
172-180	2
	n = 40

Modal Class = 145-153

$L_{Mo} = 144.5$

$A = 12 - 9 = 3$

$B = 12 - 5 = 7$

$$Mo = 144.5 + \left(\frac{3}{3+7} \right) 9 = 147.20$$

MEASURES OF VARIATION, OR DISPERSION, OR DATA SPREAD

DISPERSION measures the scatter of observation away from their averages.

Measures of dispersion accompany averages to rate the effectiveness of the averages as representatives of their data, that is, as descriptions of central tendency.

Variations are useful in the analysis of variance, in quality control, and in correlation analysis.

Two or more sets of data may have the same mean but they may differ because of the variation in the data.

Smaller variation implies more uniform data, which may mean uniform uniformly low quality or uniformly high quality.

Types of Variation which are commonly used in statistics

1. Range is the difference between the smallest and the largest values in a statistical distribution.

2. Quartile Deviation is one half of the difference between the first and the third quartiles of the data.

Quartile deviation, Q.D., is a measure of dispersion which accompanies the median.

$$Q.D = \frac{Q_3 - Q_1}{2}$$

First Quartile Q_1 marks the upper limit of the 1st one-fourth of the data array.

Third quartile Q_3 marks the upper limit of three-fourths of the data.

Formulas for Quartile Deviation

1. For Ungrouped Data

Q_1 = the term which occupies the $\left(\frac{n}{4}\right)^{th}$ position of an array.

Q_3 = the term which occupies the $\left(\frac{3n}{4}\right)^{th}$ position of an array.

Example: Daily sales of a store for 6 days by hundred thousand pesos are: 2, 2.5, 1, 1.8, 4, 4.6.

Array: 1 1.8 2 2.5 4 4.6

$Q_1 = \frac{6}{4} = (1.5)^{th}\ place$ = 2nd number = 1.8

$Q_3 = \frac{3(6)}{4} = \frac{18}{4} = (4.5)^{th}\ place$ = 5th number = 4

$$Q.D. = \frac{4 - 1.8}{2} = 1.1 = 1.1(100{,}000) = 110{,}000\ pesos$$

2. for grouped data:

$$Q_1 = L_{Q_1} + \left(\frac{\frac{n}{4} - f_{<Q_1}}{f_{Q_1}}\right) i$$

$\frac{n}{4}$ indicates which class is Q_1 class

L_{Q_1} is the lower class boundary of Q_1 class

fQ_1 is the class frequency of Q_1 class

$fc<$ is the fcum< of all classes lesser than Q_1 class

$$Q_3 = L_{Q_3} + \left(\frac{\frac{3n}{4} - fc<}{fQ_3} \right) i$$

$\frac{3n}{4}$ indicates which class is Q_3 class

L_{Q_3} is the lower class boundary of Q_3 class

fQ_3 is the class frequency of Q_1 class

$fc<$ is the fcum< of all classes lesser than Q_3 class

Consider

i = 9	f	fc<
118-126	4	4
127-135	4	8
136-144	9	17
145-153	12	29
154-162	5	34

163-171	4	38
172-180	2	40
	n = 40	

For: $$Q_1 = L_{Q_1} + \left(\frac{\frac{n}{4} - f(c)}{fQ_1} \right) i$$

$\frac{n}{4} = \frac{40}{4} = 10th$ item which points to 136 – 144 as Q_1 class

$$L_{Q_1} = 135.5$$
$$fc <= 8$$
$$fQ_1 = 9$$
$$Q_1 = 135.5 + \left(\frac{10 - 8}{9} \right) 9 = 137.5$$

For: $$Q_3 = L_{Q_3} + \left(\frac{\frac{3n}{4} - f(c)}{fQ_3} \right) i$$

$\frac{3n}{4} = \frac{3(40)}{4} = 30th$ item which points to 154-162 as Q_3 class

$$L_{Q_3} = 153.5$$
$$fc <= 29$$
$$fQ_3 = 5$$
$$Q_1 = 153.5 + \left(\frac{30 - 29}{5} \right) 9 = 155.5$$
$$Q.D. = \frac{155.3 - 137.5}{2} = 8.9$$

The median and the quartile deviations are often used for open-ended distributions in which the mean cannot be computed.

3. Variance and Standard deviations

Provide numerical measures of how the data values tend to vary around the mean.

Small variance and small standard deviation imply that the data values cluster closely around the mean.

Large variance and large standard deviation imply that the data values are widely scattered around the mean.

Variance is in square units while standard deviation is the square root of the variance.

1. Ungrouped Data Formulas:

Population variance $\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$

Sample variance $s^2 = \frac{\sum (x_i - \bar{x})}{n-1}$

Example: n = 10

X	$x - \bar{x}$	$(x - \bar{x})^2$
3	0.5	0.25
0	-2.5	6.25
6	3.5	12.25
2	-0.5	0.25
2	-0.5	0.25
5	2.5	6.25
2	-0.5	0.25
0	-2.5	6.25
3	0.5	0.25
2	-2.5	0.25
$\sum = 25$		$\sum = 32.5$

As sample variance

$\sum x = 25$ $\bar{x} = 2.5$ $s^2 = \frac{32.5}{9} = 3.611$

3. Population Standard Deviation:

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}} \text{ or } \sigma = \sqrt{\frac{\sum x^2}{N} - \left(\frac{\sum x}{N}\right)^2}$$

4. Sample Standard Deviation:

$$s=\sqrt{\frac{\sum(x_i-\bar{x})}{n-1}} \qquad \text{or} \qquad s=\sqrt{\frac{n\left(\sum x^2\right)-\left(\sum x\right)^2}{n(n-1)}}$$

consider:

x	X^2
2	4
4	16
1	1
5	25
6	36
$\sum x=18$	$\sum x^2=82$

$$n=5$$

$$s=\sqrt{\frac{5(82)-(18)^2}{5(4)}}=2.07$$

means that the degree difference among the data values from the mean value is 2.07.

2. for grouped data Formulas:

1. Population Variance

$$\sigma^2=\frac{\sum f(x_{mid}-\mu)^2}{N}$$

Population standard deviation

$$\sigma=\sqrt{\frac{\sum f(x_{mid}-\mu)^2}{N}}$$

2. Sample variance

$$s^2 = \frac{\sum f(x_{mid} - \bar{x})^2}{n-1}$$

Sample standard deviation

$$s = \sqrt{\frac{\sum f(x_{mid} - \bar{x})^2}{n-1}}$$

If class size is uniform, deviation method

$$s = i\sqrt{\frac{\sum f(d')^2}{n-1} - \frac{\left(\sum fd'\right)^2}{n(n-1)}}$$

If class size is non-uniform

$$s = \sqrt{\frac{\sum fx^2{}_{mid}}{n-1} - \frac{\left(\sum fx_{mid}\right)^2}{n(n-1)}}$$

Example: find the standard deviation using the 2 formulas.

i = 9	f	X_{mid}	$(x_{mid})^2$	$f(x_{mid})^2$	$f(x_{mid})$	d’
118-126	4	122				
127-135	4	131				
136-144	9	140				
145-153	12	149				
154-162	5	158				
163-171	4	167				
172-180	2	176				
	n=40					

fd’	$(d')^2$	$f(d')^2$

Exercises:

1. A canteen recorded its daily sales of soft drinks S for fifty days as follows. The data is by the number of cases.

80	104	95	95	96	138	126	128	114	119
85	97	98	103	104	137	128	110	115	116
145	106	106	105	109	135	127	111	117	118
149	107	105	103	136	123	121	112	114	115
93	132	133	134	124	125	120	113	116	112

Solve for the mean, median, mode, quartile deviation, standard deviation, and coefficient of variation.

i=10	f	X_{mid}	fc<	fx_{mid}	d’	fd’

$(d')^2$	$f(d')^2$	$(fd')^2$	$(x_{mid})^2$	$f(x_{mid})^2$	$(fx_{mid})^2$

Solution:

SOME CONTINUOUS PROBABILITY DISTRIBUTION

The Normal Curve Distribution

Characteristics

1. Curve is symmetrically bell-shaped about $\bar{x}$ or μ

2. $\bar{x} = Md = Mo = \mu$

3. Tails or ends of the curve are asymptotic to the horizontal axis to allow the inclusion of extreme possible values.

4. Area of space under the whole curve is 1 or 100%. Therefore half of it has an area which is 0.5 or 50%.

5. The horizontal axis is subdivided equally into at least 3 standard scores at both sides of the $\bar{x}$.

a) Z values at the right of $\bar{x}$ are +, because the data values are greater than $\bar{x}$ value.

b) Z values at the left side of $\bar{x}$ are – because the data values are lesser than $\bar{x}$ value.

Steps in solving problems concerning Normal Distributions

1. Convert specified data values with x into standard scores z, z indicates the number of standard deviations distance of x value from the $\bar{x}$ value0

$$z = \frac{x-\mu}{s} = \frac{x-\bar{x}}{s} = \frac{x-\mu}{\sigma}$$

2. Find the area corresponding to the computed z-value using the table Appendix. All area values are + regardless of the sign of z. The area is between $\bar{x}$ and z.

3. The area needed is computed depending on the condition stated. The area needed may refer to situations of less than or equal to, more than or equal to, or between two given values.

Exercises:

1. Given a standard normal distribution, find the area under the curve which lies

a) to the left of $z = 1.43$
b) to the right of $z = -0.89$
c) between $z = -2.16$ and $z = -0.65$
d) to the left of $z = -1.39$
e) to the right of $z = 1.96$
f) between $z = -0.48$ and $z = 1.74$

2. Find the value of z if the area under a standard normal curve

a) to the right of $z = 0.3622$
b) to the left of z is 0.1131
c) between 0 and z, with $z > 0$ is 0.4838
d) between $-z$ and z, with $z > 0$, is 0.9500

3. Given the standard normal distribution, find the value of k such that

a) $P(z < k) = 0.0427$
b) $P(z > k) = 0.2946$
c) $P(-0.93 < z < k) = 0.7235$

4. Given a normal distribution with $\mu = 30$ and $\sigma = 6$, find

a) the normal – curve area to the left of x = 17
b) the normal – curve area to the left of x = 22
c) the normal – curve area between x = 32 and x = 41
d) the value of x that has 80 % of the normal curve area to the left
e) the two values of x that contain the middle 75% of the normal curve area

5. The loaves of rye bread distributed to local stores by a certain bakery have an average length of 30 centimeters and a standard deviation of 2 cm. assuming that the lengths are normally distributed, what %age of the loaves re

a) Longer than 31.7 cm?
b) Between 29.3 and 33.5 cm in length?
c) Shorter than 25.5 cm?

6. A soft–drink machine is regulated so that it discharges an average of 200 milliliters per cup. If the amount of drink is normally distributed with a standard deviation equal to 15 milliliters,

a) What fraction of the cups will contain more than 224 milliliters?
b) What is the probability that a cup contains between 191 and 209 milliliters?

Table A. 3 AREA UNDER THE NORMAL CURVE

Table A.3 AREAS UNDER THE NORMAL CURVE

Name:_________________

1. Consider the grouped data showing the hourly salaries. The distribution approaches a normal curve. You are required to

a) Compute for the arithmetic mean and the standard deviation value. Round the values to whole numbers
b) Show the normal curve correctly labeled.
c) Find the number of employees whose hourly rate are
   1) Equal to or greater than P154/hr.
   2) Equal to or less than P133/hr.
   3) Between P140/hr and P154/hr.
   4) Between P161/hr and P175/hr.

Accompany solutions of c1 to c4 with the correct normal curve diagrams showing the area needed.