Probability and Distributions

Anusha Illukkumbura
MSc. Business Statistics
B.A. Social Statistics

Acknowledgement

First Edition

First Edition June 2020
ISBN : 979-8651587940

Preface

Probability is a major and important topic discussed in statistics.

This book covers essential and important details related to

- Set Theory
- Permutation and Combination
- Probability Theory
- Random Variables
- Binomial Distribution
- Poisson Distribution
- Normal Distribution

This book can be used as a handbook and a self-study material. All the necessary theories related to probability and distributions are discussed lucidly.

Any suggestions to further improve the contents of this edition would be warmly appreciated.

Anusha Illukkumbura©
MSc. Business Statistics (University of Moratuwa, Sri Lanka)
B.A. Social Statistics (University of Kelaniya, Sri Lanka)
June 2020

Table of content

Examples

CHAPTER ONE: SET THEORY

1.1 Introduction

Set theory is a statistical theory, which is based on collection of well-defined elements. It is important to understand about set theory before learning probability.

1.2 Experiment

Experiment is a collection of well defined events which can be done for unlimited and repetitive times. Experiment is random and it has one or more possible outcomes. Experiment is also called as a trial. Outcomes of the trial are events.

Events

Set of results of an experiment are called events. It is the collection of one or more outcome of an experiment. A specific outcome of an experiment is called as a sample point; furthermore it is the most basic outcome. Each outcome of sample space is a sample point. Simple event is an event with possibility of only one outcome. When there is more than one possible outcome that event is called a compound or joint event. Favorable events are the number of desired outcomes in an experiment. When two or more events have an equal opportunity of happening, they are called equally likely events.

Incident of a child selecting a numbered ball out of a box

There are 100 balls numbered from 1 to 100 in a black bag. This bag belongs to a child named Peter. In this bag, balls are colored in 5 colors: red, blue, yellow, green and white.

Number 1-20 are colored in yellow, 21-40 in green, 41-60 in blue , 61-80 in red and 81-100 in white.

The child can take balls out of the bag for any number of times with replacement. Taking out the balls out of bag is the experiment.

Event: A child randomly takes out number 10 ball from the box. This is the event of taking number 10 out of the bag.

Sample point: number of the ball he took out for the *first time*. He can take only one ball for the first time. Therefore there is only one sample point in this event.

Simple event: A child takes out only one ball from the box.

Compound event: Child is taking out two balls simultaneously out of the bag.

Favorable events: Child wishes to take an even numbered ball out of the box. There are 50 even numbers between 1 to100. So the favorable event is taking out any of those 50 even numbers out of the box.

Equally likely events: there are 50 odd numbers and 50 even numbers from 1 to 100. Taking out an odd number or an even number are equally likely events.

Sample Space

Sample space is the set of all possible outcomes in an experiment. Sample space is represented by “S”.

All the balls numbered from 1 to 100 are included in sample space.

Event Space

Event space is the all possible sets and subsets of outcomes in an event. This is different from sample space. Members of sample space are sample points. But members of event space are sets and subsets.

Independent events and Dependent events

If one event doesn't have an effect on another event then those two events are independent. But when one event has an impact on another event it is called a dependent event.

The child takes one ball out of the box randomly without looking inside the bag. This doesn't have any effect from any other event. This is an independent event. He doesn't like the color of the ball. He puts it back; take another ball again without looking inside the bag (randomly). This is also an independent event as there are same amount of balls in the bag (100) as the child put it back (the first event doesn't affect the second event). This event is also called as an event with replacement.

Child is taking one ball out of the bag, he likes the color keep it with him, and without replacing it he takes another ball. This second event is dependent. The child doesn't replace the first ball. So there are only 99 balls in the bag in the second trial. The first event has influence on the outcome of the second event.

Mutually Exclusive Events

When two or more events cannot happen at the same time it is called mutually exclusive events. If A and B are two mutually exclusive events then $P(A \cap B) = 0$, which means intersection is null. (This concept will be discussed in coming sections.)

When the child takes only one ball out of the bag, there is no possibility that the number of this ball is both even and odd. Therefore the probability of taking out both even and odd number is zero. Therefore event of taking a ball with both even and odd numbers is a mutually exclusive event.

Collectively Exhaustive Events

At least one event out of set of events must happen.

When the child selects the ball, the ball should either be an even or an odd number. Therefore at least one outcome of odd or even numbered ball should happen. Taking a ball with an odd or an even number is a collectively exhaustive event.

1.3 Sets

Set is a well defined collection of elements. Each object of the set is an element.

Sets can be named using English capital letters (Ex: A,B,C).

In the example of child selecting a ball, set of all the numbers written on the balls inside the bag, has elements from 1 to 100.

This can be written as A= {1,2,3,...,99,100} , where A is the set of numbers written on balls inside the bag.

Element

Element is any member of a set.

Numbers from 1 to 100 are the members of the set of "numbers written on the balls inside the bag".

Mike, Saman, Nil, Ravi and Amal are five boys in a class who weight between 50kg to 60 kg. This group can be defined as a set; it can be named as "A".

Mike, Saman, Nil, Ravi and Amal are elements of set "A", but John who weights 70 kg is not an element of that set.

Element of a set can be written using element symbol "∈", Mike ∈ A, Saman ∈ A, Nil∈ A, Ravi∈ A and Author∈ A.

∉ is the symbol of "not an element of". When an event is not an element of a set, it is written as John ∉ A.

Null-set

When there is no element in a set, that set is called a null set.

The bag, child has only contained of balls. So getting cards out of that bag is not happening. Set of getting a card out of Peter's bag is a null set.

Sub Set

When all the elements of one set (Set A) is included in another set (Set B) , then set A is a sub set of set B. It is symbolized as, $A \subset B$.

In Peter's bag, there are 50 even numbers. Set of these even numbers are included in the set of numbers from 1 to 100.

A- Set of even numbers

B- Set of numbers from 1 to 100.

Therefore $A \subset B$.

Equal Sets

When two or more sets have the same elements, then those sets are called equal sets (A=B)

When $A \subset B$ and $B \subset A$ then they are A=B

Balls from number 1 to 20 of Peter's bag are colored in yellow. They are the only ones colored in the yellow in the bag.

A- Set of yellow colored balls in the bag

B- Set of all numbers from 1 to 20

In this case ,$A \subset B$ and $B \subset A$, therefore A=B

Universal Set

Set of all the possible elements of a set is called a universal set. This is represented by "U". This is similar to the sample space. Collection of all the colored balls with numbers written on, is the universal set of balls in Peter's bag.

Set Union

When there are two sets called A and B, the set includes the elements, which are belonged to either set A or B or both, is called set union. It is represented by $A \cup B$

A- Yellow colored balls in Peter's bag

B- Red colored balls in Peter's bag

A ∪ B –yellow **or** red colored balls in Peter's bag (There are 40 balls)

Set Intersection

When there are two sets called A and B, the set includes the elements, which are belonged to both A and B, is called set intersection. It is represented by

A ∩B

A- numbers from 11 to 50 in Peter's bag

B- Yellow colored balls in Peter's bag

A∩B –Set of both yellow colored **and** numbers from 11to 50 in Peter's bag

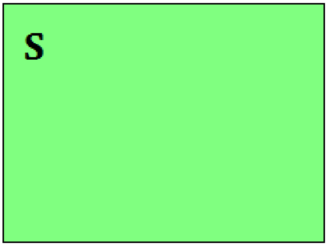
Questions with set union have the word "or" meaning of "or". Questions with intersection have the word "and" or its meaning.

Venn Diagrams

Venn is type of a diagram, which is used to illustrate the details of sets.

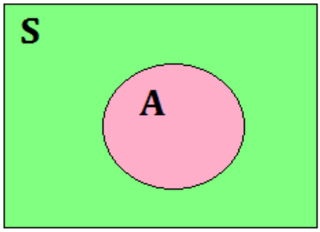
If S = Birds, A = Migratory birds and B = Hummingbirds

1.1 Venn Diagram



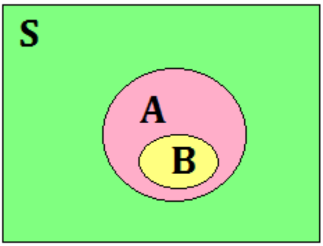
S is the universal set

1.2 Venn Diagram



A is a set

1.3 Venn Diagram

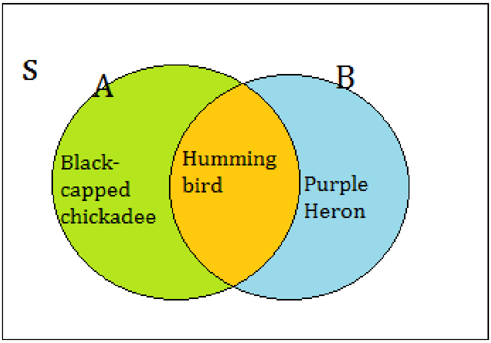


In this Venn diagram 1.3 ,
$B \subset A$

When A = Small Birds, B= Migratory birds

In Venn diagram 1.4, universal set (S) is birds. Black-capped chickadee is a non migratory small bird, so it belongs to set of small birds (A). Purple heron is a migratory big bird, so it belongs to set of migratory birds (B). Humming bird is both small and migratory bird so it belongs to the both category. Humming bird is an element of set intersection. If someone wants to take all the migrant or small or both small and migrant birds as a sample then that person should take the all elements of the set union of this example.

1.4 Venn Diagram



A- Small birds , Migrant birds

A = {Black-capped chickadee, Humming bird}

B= {Purple Heron, Humming bird}

A⋂B – Small and Migrant birds

A⋂B ={Humming birds}

A⋃B – Small or migrant birds

A⋃B = {Black-capped chickadee, Humming bird, Purple Heron}

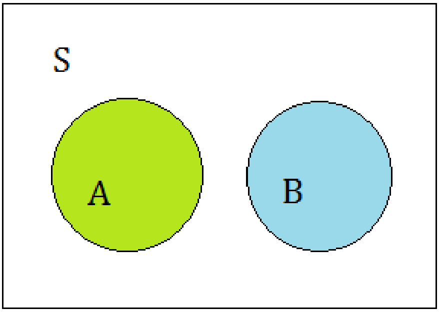
Disjoint Set

When there are no common elements among two sets, it is called a disjoint set.

It is represented by A⋂B =∅

When S= Animals , A= Birds and B = Centipedes then A⋂B =∅ (Venn diagram is illustrated in 1.5 Venn diagram of disjoint sets).

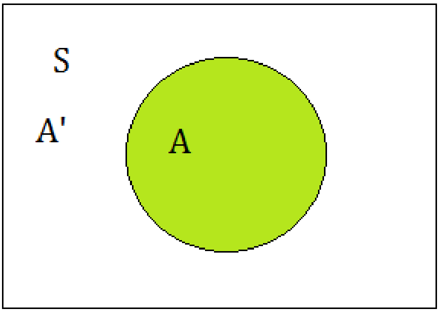
1.5 Venn Diagram of disjoint sets



Complement of a set

Complement of a set is the elements that are not included in a specific set, but they are included in universal set. It is represented by $A' = \{x \in S \mid x \notin A\}$.

1.6 Venn Diagram of complement of a set



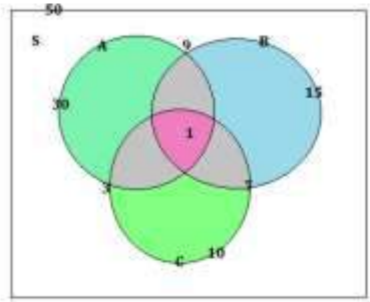
Number of elements in the set is called the **order of the set.** It is described by "n". Order of the set A is represented by n(A). If a set is countable and finite it is called **finite set** and if it is uncountable it is called **infinite set.** Number of students in a classroom is a finite set. Number of stars on the sky is infinite set.

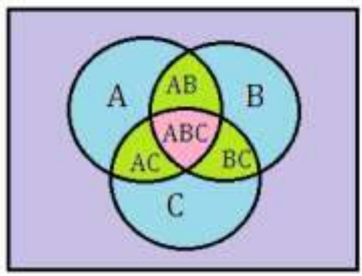
Example 1.1: Venn diagram

Bakery products manufacturing company tested 50 samples on chocolate cakes, coffee cakes and butter cakes using 50 customers. The company found that 30 of them liked chocolate cakes, 15 liked coffee cakes, 10 liked butter cakes, 7 liked both coffee and butter cakes and also 9 liked both coffee and chocolate cakes. 3 liked both chocolate and butter cakes. One liked all three types of cakes.

If A- People who like chocolate cakes
B- People who like coffee cakes
C- People who like butter cake

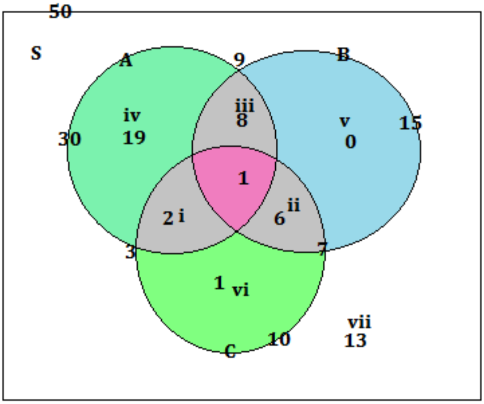
Above information can be illustrated in a Venn diagram as below





Calculate the missing information using the steps indicated by roman numbers (refer next page)

i. 3-1 =2
ii. 7-1 =6
iii. 9-1=8
iv. 30-8-2-1= 19
v. 15-8-1-6 =0
vi. 10-6-1-2
vii. 50-30-0-6-1 =13



Using the given information calculate

I) Number of people who like only two types of cake
= 8+6+2=16

II) Number of people who like only one type of cake
= 19+0+1 =20

III) Number of people who likes at least one type of cake
=8+2+6+1+19+0+1=37

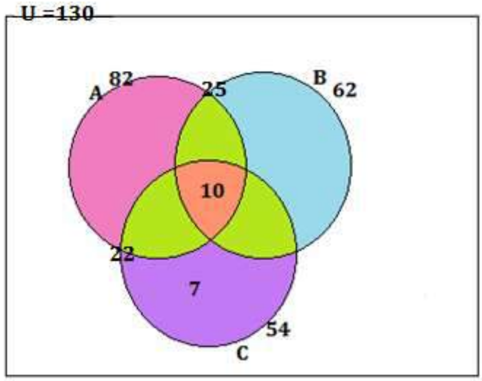
IV) Ones who like chocolate or coffee cake but not butter cake. = 19+ 8+0 =27

V) People who likes none of them
50-30-0-6-1 =13

Example 1.2: Venn diagram

130 tourists were questioned on three luxury restaurants in a popular tourist destination, "Sigiriya". Luxury restaurants are called A,B and C. 82 of them have visited A and 62 have visited B. 54 has visited C. 22 of them have visited both A and C , 25 of them have visited A and B.7 has visited C but not A or B.10 has visited all three restaurants.

This information can be drawn in a Venn diagram as below



<u>**How to calculate the missing information?**</u>

Calculation steps are presented in roman numbers

I. 22-10=12

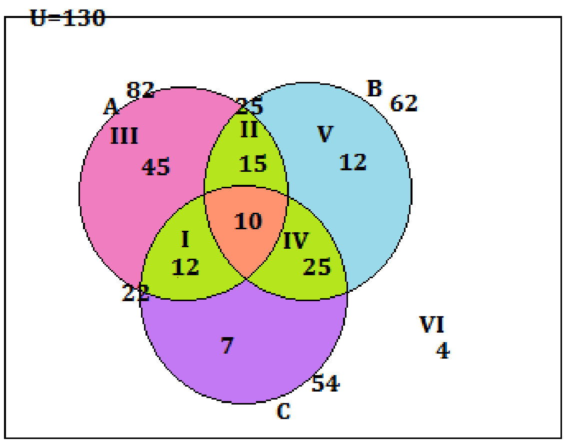
II. 25-10=15

III. 82- 12-10-15 =45

IV. 54-7-12-10=25

V. 62-15-10-25=12

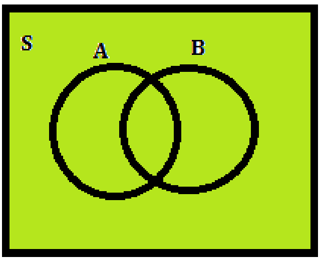
VI. 130 –82-12-25-7= 4



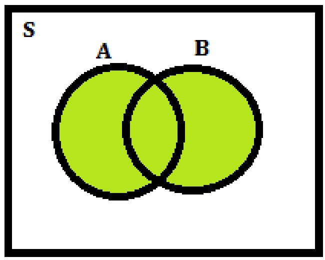
I) Number of people who have visited both B and C but not A =25
II) Number of people who have visited only B =12
III) Number of people who have visited only has visited A =45
IV) Number of visited who have not visited any of these three places =4

Below are some definitions in Set calculations illustrated using Venn diagrams

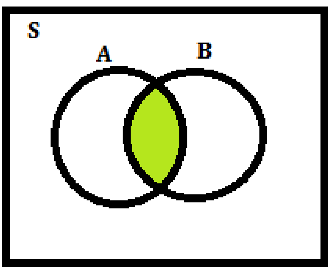
Universal Set



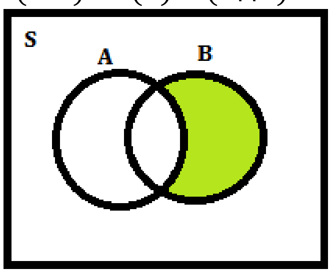
P(AUB)



P(A∩B)

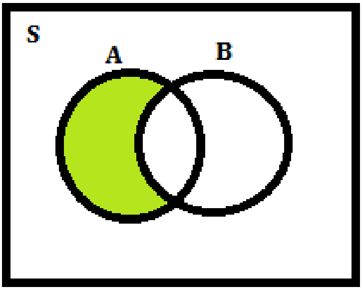


P(B-A)= P(B)- P(A∩B)

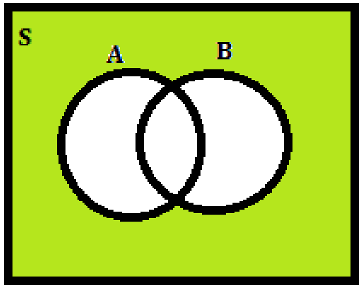


P(A-B)= P(A)- P(A∩B)

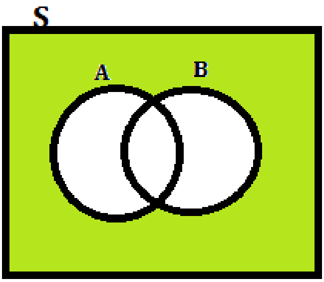
P(AUB)’



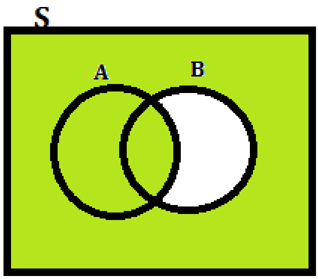
$A' \cap B' = 1-(A \cup B)$



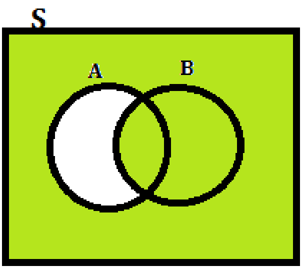
$(A \cup B')$

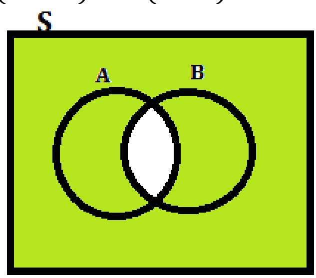


$(A' \cup B)$



$(A' \cup B') = 1- (A \cap B)$





1.4 Factorial Notation

Factorial notation is appeared as an exclamation mark following the number. Example: 10!, 21!.

5! is similar to multiplication of all the consecutive integers from 1 to 5.

Example: 5! = 5*4*3*2*1 =120

n! is similar to multiplication of all the consecutive integers from 1 to n.
Example: n! = 1 *2*3*…*n

Example 1.3: Factorial Notation
Solve below questions

I) 2!7!
= (1*2)* (1*2*3*4*5*6*7)
= 2* 5040
=1080

II) $\frac{8!}{4!}$
$=\frac{(1*2*3*4*5*6*7*8)}{(1*2*3*4)}$
$=\frac{40,320}{24}$
=1680

III) $\frac{8!+3!}{4!-2!}$
$= \frac{40,320+6}{24-2}$
$= \frac{40326}{22}$
= 1833

IV) $\frac{1}{4!}+\frac{2}{3!}$
$=\frac{1}{(1*2*3*4)}+\frac{2}{1*2*3}$
$=\frac{1}{24}+\frac{1}{3}$
= 0.042 + 0.33
=0.372

V) $\frac{\frac{5!}{2!}}{4!*3!}$
$= \frac{\frac{1*2*3*4*5}{1*2}}{(1*2*3*4)*(1*2*3)}$
$= \frac{3*4*5}{24*6}$
$= \frac{5}{12}$

Example 1.4: Factorial Notation
Write below numbers in factorial notation

I) 12

$$= \frac{12!}{11!}$$

II) 8*7*6

$$= \frac{8!}{5!}$$

III) $\frac{7*6*5}{2*3*4}$

$$= \frac{7!}{4!}$$

1.5 Permutation

"r" number of items which are taken out of total of "n" number of items that can be arranged in an ordered way is permutation.

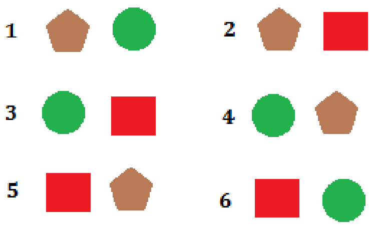
It can also be interpreted as an ordered arrangement of "r" numbers of different elements that are selected from total of "n" numbers of elements. Order is very important in permutation. r should be less than or equal to n.

It is represented by ${}^{n}P_{r}$, $P_{n,r}$ and P (n,r).

$${}^{n}P_{r} = \frac{n!}{(n-r)!}$$

Let's see how many permutations we can arrange by selecting 2 out of 3 geometrical shapes

Permutations



Below is the mathematical calculation.

$${}^{n}P_{r} = \frac{n!}{(n-r)!}$$

$$^{3}P_{2} = \frac{3!}{(1)!}$$

$$^{3}P_{2} = 3*2*1 = 6$$

Example 1.5: Permutation

I) There are 12 books. A student was asked to select three of them randomly. How many different ways the student can select the books?

$$^{12}P_{3} = \frac{12!}{9!}$$

$$^{3}P_{2} = \frac{12*11*10*9!}{9!}$$

$$^{3}P_{2} = 12*11*10$$

$$^{3}P_{2} = 12*11*10$$

$$^{3}P_{2} = 1320$$

II) 20 competitors participate in a race. How many ways are there that the competitors can win gold, silver and bronze medals?

Gold medal can be won in 20 ways

Silver medal can be won in 19 ways

Bronze medal can be won in 18 ways

Therefore at one race the ways of models can be won is 20*19*18 =6840

III) There are 4 doors to a cinema. In how many ways can a customer enter the room through a door and leave the room by a different door?

There are 4 doors and if a customer can enter the supermarket in 4 ways. Now one door is used. Now he has to use other 3 doors to leave the super market without using the first door. Therefore the calculation is 4*3= 12

There are twelve ways that a customer can enter and leave the supermarket using different doors.

IV) How many passwords can be formed by re-arranging the letters of the word STATISTICS . Meaning of the word is not relevant.

There are 10 letters in the word. There are 3 letters S and T. There are 2 "I"s. There is only one "A"and C

$$\frac{10!}{3!\,3!\,2!\,1!\,1!}$$ When this number is solved answer is 50400.

This type of problem solving is called MISSISSIPI rule in statistics.

V) How many passwords can be formed by re-arranging letters of the word SMILE such that S and E occupy the first and last positions respectively? Meaning of the word is not relevant.

Here S and E are in fixed places. Only permutation of M,I,L should be calculated.

1* 3*2*1*1 = 6

S is fixed so; it is used in only one way. Second letter can be selected out of 3 letters. Third letter can be selected out of 2 letters. For fourth letter only one choice is left. Fifth letter E can also be used in only one way.

VI) How many passwords can be formed by re-arranging letters of the word SMILE such that S and E occupy the first and last position? Meaning of the word is not relevant.

In this arrangement it is told S and E can be used as first or last letters. So there are two choices for the first letter. But for the last letter there is only one letter is left out of S and E.

2* 3*2*1*1 = 12

VII) 6 teachers, the principle and the vice principle are to be seated around a circular table. If the principle should sit between a teacher and the vice principle, in how many ways can they be seated?

There is only one principle and a vice principle, principle should be in between a teacher and a vice principle.

Take them as one set. . there are 6 sets now (five of them consists of single teachers) . A teacher and Vice principle can be seated in two ways ; teacher and vice principle or the other way round (2 ways) . Teachers can be seated in 6! Ways. Therefore, 6! *2 = 720*2 = 1440

VIII) Find the number of permutations of the letters of the word 'PROBLEM' such that the vowels always occur in odd places.

There are 7 different letters in the word PROBLEM

There are 6 constants and 2 vowels.

There are four odd places and three even places.

Number of ways 2 vowels can appear in 4 different places = ${}^4P_2$ = 12.

After 2 vowels take 2 places, 5 more spaces are left. They can be arranged in 5! =120 ways.

Therefore, total number of permutations possible is 120*12 = 1440

1.6 Combination

Any subsets that can be created by selecting “r” number of different items out of total of “n” numbers of items are called combination. Order is not important in permutation. “r” should be less than or equal to “n”.

It is represented by ${}^{n}C_{r}$, $C_{n,r}$ $\binom{n}{r}$and C (n,r)

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

Let’s see how many combinations we can create by selecting 2 out of 3 geometrical shapes

Combinations

Below is the mathematical calculation.

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

$$ {}^{3}C_{2} = \frac{3!}{2!\,(1)!} $$

$$ {}^{3}C_{2} = 3 $$

Since the order does not matter in combination, only 3 combinations can be selected.

Example 1.6: Combinations

I) There are 3 books and 4 magazines. A student is told to choose 1 book and 2 magazines. How many ways can he select 1 book and 2 magazines?

$${}^{n}C_{r} * {}^{n}C_{r} = \frac{3!}{1!\,(2)!} * \frac{4!}{2!\,(2)!}$$

${}^{3}C_{1} * {}^{4}C_{2} = 3 * 6$

${}^{3}C_{1} * {}^{4}C_{2} = 18$ (There are total of 18 ways)

II) There is a set of 6 blue balls and 3 red balls, a child is told to select 5 such that at least 3 of them are blue balls. How many selections can he make?

There are 9 balls, 5 balls are to be selected, but there should be at least 3 blue balls. So these 5 balls can consists of 3 blue balls and 2 red balls (${}^{6}C_{3} * {}^{3}C_{2}$), 4 blue balls (${}^{6}C_{4} * {}^{3}C_{1}$) and 1 red balls and only 5 blue balls (${}^{6}C_{5} * {}^{3}C_{0}$).

$=({}^{6}C_{3} * {}^{3}C_{2}) + ({}^{6}C_{4} * {}^{3}C_{1}) + ({}^{6}C_{5} * {}^{3}C_{0})$.

$=(20 * 3) + (15 * 3) + (6 * 1)$.

$= 60+45+6$

$=111$ (There are total of 111 ways)

III) Three directors and one CEO should be selected out of 7 director candidates and 3 CEO candidates of a company. How many ways can 3 directors and 1 CEO be selected?

${}^{7}C_{3} * {}^{3}C_{1}$

$= 35 * 3$

$=105$ (There are 105 ways)

IV) Passwords can be created using 3 numbers and 7 letters of English language. How many combinations of passwords can be created?

$^{10}C_3 * {}^{26}C_7$

=120*657800

=78936000

Example 1.7: Permutation & Combination

"PCIKTIVVATIKG" is a word in a language. In how many rearrangements of the letters can be done with no two 'I's appear together?

There are 13 letters including one P ,C ,A and G, two K, T and V s and three I s.

First remove the three I s. There will be 10 letters. Calculate the permutation of them it will be $10!/(2!*2!*2!) = 453{,}600$

There will be 11 spaces where three of the I s can be added between other 10 letters.

1 P 2 C 3 I 4 K 5 T 6 I 7 V 8 V 9 A 10 T 11 I 12 K 13 G

Find out how can be three "I"s be placed in 11 space using combination $^{11}C_3 = 165$

So the total will be $165*453{,}600 = 74{,}844{,}000$.

Example 1.8: Permutation & Combination

Lottery has a 3 digit vowels and one number.

I) A person can win if those vowels and the number appear in any order. How many ways can it happen?

$^{6}C_3 * {}^{10}C_1$

=20*10

=200

II) If the vowels should appear in first three places and the number should appear the last, how many ways can it happen?

$^{6}P_{3} * {}^{10}P_{1}$

$=6*5*4*10$

$=1200$

Read more at
https://www.amazon.com/gp/product/B08BCR2B4Z?notRedirectToSDP=1&ref_=dbs_m_mng_rwt_calw_tkin_tpbk&storeType=ebooks

Author’s other books:

Probability Questions and Answers

Introduction to Regression Analysis

Introduction to Hypothesis Testing

Introduction to Categorical Data Analysis

Introduction to Time Series Data Analysis

“If you wish more statistics books for affordable price, please leave a review on Amazon stating the feedback and also the topics you wish to have books on.”