Statistical and Mathematical Methods for Data Analysis

Dr. Faisal Bukhari

Punjab University College of Information Technology (PUCIT)

Textbooks

❑ **Probability & Statistics for Engineers & Scientists**, Ninth Edition, Ronald E. Walpole, Raymond H. Myer

❑ **Elementary Statistics: Picturing the World,** 6th Edition, Ron Larson and Betsy Farber

❑ **Elementary Statistics**, 13th Edition, Mario F. Triola

Three things cannot be long hidden: the sun, the moon, and the truth

Buddha

Textbooks

❑**Probability & Statistics for Engineers & Scientists**, Ninth Edition, Ronald E. Walpole, Raymond H. Myer

❑**Elementary Statistics: Picturing the World, 6th Edition, Ron Larson and Betsy Farber**

❑**Elementary Statistics, 13th Edition, Mario F. Triola**

Reference books

❑ **Probability and Statistical Inference, Ninth Edition,** Robert V. Hogg, Elliot A. Tanis, Dale L. Zimmerman

❑ **Probability Demystified**, Allan G. Bluman

❑ **Schaum's Outline of Probability,** Second Edition, Seymour Lipschutz, Marc Lipson

❑ **Python for Probability, Statistics, and Machine Learning,** José Unpingco

❑ **Practical Statistics for Data Scientists: 50 Essential Concepts,** Peter Bruce and Andrew Bruce

❑ **Think Stats: Probability and Statistics for Programmers,** Allen Downey

References

Readings for these lecture notes:

Probability & Statistics for Engineers & Scientists, Ninth Edition, Ronald E. Walpole, Raymond H. Myer

Elementary Statistics, 10th Edition, Mario F. Triola

http://mathworld.wolfram.com/CircularPermutation.html

https://www.zero-factorial.com/whatis.html

http://mathworld.wolfram.com/CircularPermutation.html

http://doubleroot.in/lessons/permutations-combinations/circular-permutations-examples/#.WbeY04-cHIU

http://www.onlinemathlearning.com/combinations.html

These notes contain material from the above resources.

Counting Sample Points

❑In many cases, we shall be able to solve a probability problem by counting the number of points in the **sample space without actually listing each element**.

❑The **fundamental principle of counting**, often referred to as the **multiplication rule**, is stated in Rule 2.1.

Rule 2.1: If an operation can be performed in n_1 **ways**, and if for **each of these ways a second operation** can be performed in n_2 **ways**, then **the two operations** can be performed together in $n_1 n_2$ **ways**.

❑**Example :** How many **sample points** are there in the **sample space** when a **pair of dice** is thrown once?

Solution :

❑The first die gives us, n_1 = 6 ways.

❑For each of these 6 ways, the second die gives us, n_2 = 6 ways.

❑Therefore, the pair of dice give us $n_1 n_2$ = (6)(6) = 36 possible ways.

Example : If a **22-member club** needs to elect a **chair** and a **treasurer**, how many **different ways** can these **two** to be elected?

Solution :

❑For the **chair position**, we have n_1 = **22 ways**

❑For the **treasurer** position, for each of those **21 possibilities,** we have n_2 = **21ways**

❑Total number of ways = $n_1 \times n_2$ = **22 × 21 = 462**

❑**Rule 2.2:** If an operation can be performed in n_1 ways, and **if for each of these a second operation** can be performed in n_2 **ways**, and for each of the **first two a third operation** can be performed in n_3 **ways**, and so forth, then the sequence of k operations can be performed in $n_1 n_2 \cdots n_k$ **ways**.

❑**Example:** Sam is going to assemble a computer by himself. He has the choice of chips from **two brands**, a **hard drive** from **four**, **memory** from **three**, and an **accessory bundle** from **five** local stores. How many **different ways** can Sam order the parts?

Solution :

n_1 = 2 **(No of brands)**

n_2 = 4 **(No of hard drives)**

n_3 = 3 **(No of memory sticks)**

n_4 = 5 **(No of accessory bundles)**

❑Total number of ways = **$n_1 \times n_2 \times n_3 \times n_4$ = 2 × 4 × 3 × 5** = 120

Permutation

Permutation: A **permutation** is an arrangement of all or part of a set of objects.

OR

An arrangement of a set of *n objects* in a **given order** is called a *permutation* of the objects (taken all at a time).

Example: Consider the three **letters *a, b,*** and *c*.

The possible **permutations** are *abc, acb, bac, bca, cab,* and *cba*.

Permutation

❑ **Definition** For any **non-negative integer *n*, *n*!,** called **"*n* factorial,"** is defined as $n! = n(n - 1) \cdots (2)(1)$, with special case **0! = 1**.

❑ **Theorem 2.1:** The number of permutations of ***n* objects** is ***n*!.**

❑ **Example** The number of permutations of the **four letters *a*, *b*, *c*,** and ***d*** will be **4! = 24**.

Why 0! one?

The idea of the factorial (in simple terms) is used to compute the number of permutations (combinations) of arranging a set of **n numbers**.

n	Number of permutations (n!)	Visual examples
1	**1**	{1}
2	**2**	{1, 2}, {2, 1}
3	**6**	{1, 2, 3}, {1, 3, 2}, {2, 1, 3}, {2, 3, 1}, {3, 1, 2}, {3, 2, 1}
⋮	⋮	⋮
0	**1**	{}

Why 0! one?

$n! = n \times (n - 1)!$

$\Rightarrow (n - 1)! = \dfrac{n!}{n}$

Substitute 1, we get

$\Rightarrow (1 - 1)! = \dfrac{1!}{1}$

$\Rightarrow 0! = 1$

Permutations Rule (When items are all different)

❑**Theorem 2.2:** The number of permutations of **n** **distinct objects** taken **r** at a **time** is $_nP_r = \dfrac{n!}{(n-r)!}$, where **r** $\leq$ **n**

Permutations Rule (When items are all different)

1. There are *n different* items available. (This rule does not apply if some of the items are identical to others.)

2. We select *r* of the *n* items (without replacement).

3. We consider rearrangements of the same items to be different sequences. (The permutation of *ABC* is different from *CBA* and is counted separately)

Permutations Rule (When items are all different)

❑ If the preceding **requirements are satisfied**, the number of **permutations (or sequences)** of *r items* selected from *n different* available items **(without replacement)** is $_nP_r = \dfrac{n!}{(n-r)!}$, where $r \le n$.

Example : In one year, **three awards (research, teaching, and service)** will be given to a class of **25** graduate students in a statistics department. If each student can receive at **most one award**, how many possible selections are there?

Solution : Since the awards are **distinguishable**, it is a permutation problem. The total number of sample points is

$$_{25}P_3 = \frac{25!}{(25-3)!} = \frac{25!}{22!} = (25)(24)(23) = 13,800.$$

Example Clinical Trial of New Drug When testing a new drug, Phase I involves **only 8 volunteers**, and the objective is to assess the drug's safety. To be very cautious, you plan to treat the **8 subjects in sequence**, so that any particularly adverse effect can allow for stopping the treatments before any other subjects are treated. If **10 volunteers** are available and **8 of them are to be selected**, how **many different sequences of 8 subjects** are possible?

Solution We have *n = 10 different subjects* available, and we plan to **select *r = 8*** of them **without replacement**. The number of different sequences of arrangements is found as shown

$$_{10}P_8 = \frac{10!}{(10-8)!} = \frac{10!}{2!} = 1,814,400$$

Example Find the number of ways of forming four-digit codes in which no digit is repeated.

Solution

To form a four-digit code with no repeating digits, you need to select 4 digits from a group of 10,

so **n = 10** and **r = 4**.

$$_{10}P_4 = \frac{10!}{(10-4)!} = \frac{10!}{6!} = 5040$$

So, there are **5040** possible four-digit codes that do not have repeating digits.

Example How many 4-digit numbers are there with no digit repeated?

Solution

Total number of ways = $9 \times 9 \times 8 \times 7$

$= 4536$

Example Forty-three race cars started the **2013 Daytona 500**. How many ways can the cars finish first, second, and third?

Solution

You need to select three race cars from a group of 43, so n = 43 and r = 3. Because the order is important, the number of ways the cars can finish first, second, and third is

$$_{43}P_3 = \frac{43!}{(43-3)!} = \frac{10!}{6!} = 74{,}046.$$

Theorem 2.3: The number of **permutations** of n **objects** arranged in a circle is $(n - 1)!$.

**Example** In how many ways can **6 people** be seated at a **round table**?

Solution

Here n = 6 (total number of people)

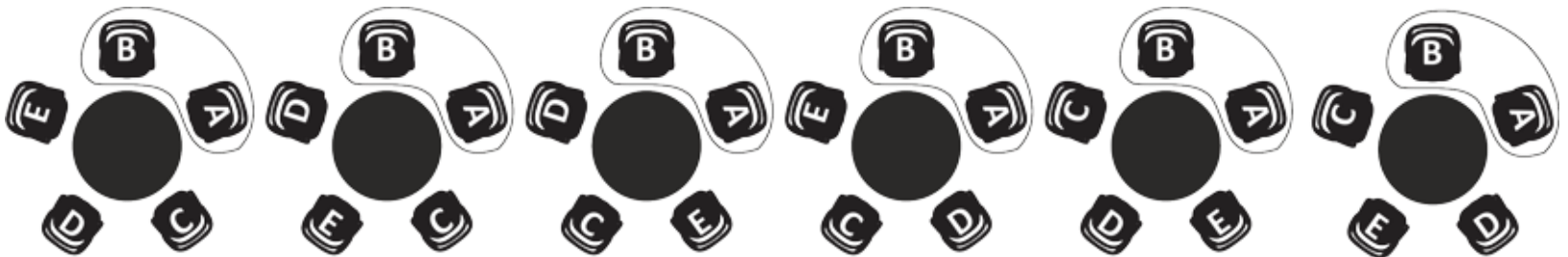The total number of ways = (6 − 1)! = 120

Example Find the number of ways in which **5 people A, B, C, D, E** can be seated at **a round table**, such that

a.   **A and B** must always sit together.

b.   **C and D** must not sit together.

❑**Solution a.** If we wish to seat **A and B** together in all arrangements, we can consider these **two as one unit**, along with 3 **others**. So effectively we've to arrange **4 people** in a circle. The number of ways = **(4 − 1)! = 6**
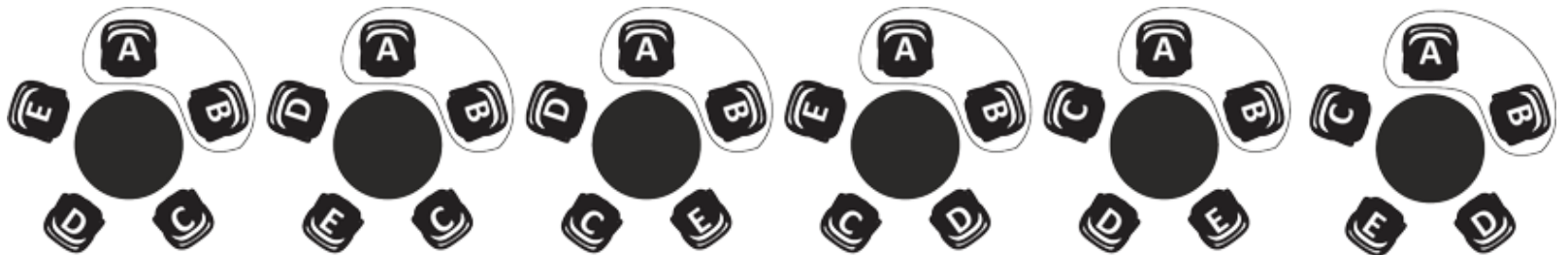
But in each of these arrangements, **A** and **B** can themselves interchange places in **2 ways**.

Therefore, the total number of ways will be **6 × 2 = 12.**

5 people A, B, C, D, E

❑**b.** The total number of ways will be **(5 – 1)!** or **24**.

❑ Similar to **a.** above, the number of cases in which **C** and **D** are seated together, will be **12**.

❑ Therefore the required number of ways = **24 – 12 = 12.**

Permutations when repetition is allowed

❑ So far we have considered **permutations** of **distinct objects**. That is, all the objects were completely different or distinguishable.

❑ Obviously, if the letters **b and c** are both equal to **x**, then the **6 permutations** of the **letters a, b, and c** become **axx, axx**, *xax, xax*, **xxa,** and **xxa**, of which only **3** are distinct.

❑ Therefore, with **3 letters**, **2 being the same**, we have **3!/2! = 3** distinct permutations.

Permutations Rule (When some items are identical to others)

❏ **Theorem 2.4:** The number of **distinct permutations** of *n* **things** of which n_1 are of one kind, n_2 of a second kind, *. . . ,* n_k of a *k*th kind is

$$\frac{n!}{n_1!n_2! \cdots nk!}$$

where $n_1 + n_2 + n_3 + ... + n_k = n$.

Permutations Rule (When some items are identical to others)

Requirements

1. There are **_n_ items available**, and some items are **identical** to others.

2. We select all of the **_n_ items (without replacement).**

3. We consider **rearrangements of distinct items** to be **different sequences**.

If the preceding requirements are satisfied, and if there are **n_1 alike**, **n_2 alike**, . . . , **n_k alike**, the number

$$\frac{n!}{n_1! n_2! \cdots n_k!}$$

Example A building contractor is planning to develop a subdivision. The subdivision is to consist of **6 one-story houses**, **4 two-story houses**, and **2 split-level houses**. In how many **distinguishable** ways can the houses be arranged?

Solution

The total number of arrangements = $\dfrac{\boldsymbol{12!}}{\boldsymbol{6!\ 4!\ 2!}}$

$$= 13{,}860$$

distinguishable ways

Example Calculate the number of distinguishable permutations of the letters **AAAABBC**.

Solution

The total number of arrangements = $\dfrac{7!}{4!\,2!\,1!}$

$= 105$

distinguishable ways

❑**Example** In a college football training session, the defensive coordinator needs to have **10 players** standing in a row. Among these **10 players**, there are **1 freshman**, **2 sophomores**, **4 juniors**, and **3 seniors**. How many **different ways** can they be arranged in a row if only their **class level** will be distinguished?

❑Solution

n = 10 (Total number of players)

n_1 = 1 (Total number of freshman)

n_2 = 2 (Total number of sophomores)

n_3 = 4 (Total number of juniors)

n_4 = 3 (Total number of seniors)

The total number of arrangements = $\dfrac{10!}{1!\ 2!\ 4!\ 3!}$

$= 12,\ 600.$

Theorem 2.5: The number of ways of **partitioning** a set of *n* **objects** into *r* **cells** with n_1 **elements** in the first cell, n_2 **elements** in the second, and so forth, is

$$\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1! n_2! \cdots n_r!}$$

where $n_1 + n_2 + \cdots + n_r = n$

Example : In how many ways can **7 graduate** students be assigned to **1 triple** and **2 double** hotel rooms during a conference?

***Solution* :**

$$\frac{n!}{n_1! n_2! \cdots nk_!}$$

Here n = 7

$n_1 = 3$

$n_2 = 2$

$n_3 = 2$

The total number of possible partitions would be

$$\frac{7!}{3!\ 2!\ 2!} = 210$$

Theorem 2.6: The number of combinations of *n* **distinct objects** taken *r* at a time is

$$_nC_r = \frac{n!}{r!(n-r)!} \text{ , where } r \leq n.$$

Combinations Rule

Requirements

1. There are *n different* items available.

2. We select *r* of the *n* **items** (**without replacement**).

3. We consider rearrangements of the **same items** to **be the same**. (The combination *ABC* is the same as *CBA*.)

If the preceding requirements are satisfied, the number of **combinations** of *r* **items** selected from *n* **different items** is

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

Combination vs. Permutations

❑The **2-permutations** of the letters **A, B, C, and D** are*:*

AB, AC, AD, BA, BC, BD, CA, CB, CD, DA, DB, DC.

❑The combinations of two out of these four letters are:

AB, AC, AD, BC, BD, CD.

(Since the elements of a combination are unordered, **BA** is not viewed as being distinct from **AB**.)

Example : A young boy asks his mother to get **5** Game-Boy^{TM} cartridges from his collection of **10 arcade** and **5 sports games**. How many ways are there that his mother can get **3 arcade** and **2 sports** games?

Solution **:** The number of ways of **selecting 3** cartridges **from 10** is

$$_{10}C_3 = \frac{10!}{3!(10 - 3)!} = 120$$

The number of ways of selecting **2 cartridges** from **5** is $_5C_2 = \frac{5!}{2!(5 - 2)!} = 10$

❑ Using the multiplication rule (Rule 2.1) with $n_1 = $ **120 and** $n_2 = $ **10**, we have **(120)(10) = 1200 ways**.

❑**Example:** In a lottery, each ticket has **5 one-digit numbers 0-9** on it.

a) You win if your ticket has the **digits in any order.** What are your changes of winning?

b) You would win only if your ticket has the digits in **the required order.** What are your chances of winning?

Solution:

There are 10 digits to be taken 5 at a time.

a) The number of ways of **selecting 5** tickets **from 10** is

$$_{10}C_5 = \frac{10!}{5!(10-5)!} = \frac{10 \times 9 \times 8 \times 7 \times 6}{5!} = 252$$

∴The chances of winning are **1 out of 252 (0.0040 or 0.3968 %**

❑b) Since the order matters, we should use permutation instead of combination.

$$_{10}P_5 = \frac{10!}{(10 - 5)!} = \frac{10!}{5!} = 10 \times 9 \times 8 \times 7 \times 6 = \textbf{30240}$$

∴ The chances of winning are 1 out of 30240 (or **0.0033 %)**