NOTAS DE CLASE

PROBABILIDAD Y ESTADÍSTICA

{Con ejemplos de programación}

/* ****** **Jaime E. Montoya M.** ****** */

NOTAS DE CLASE

PROBABILIDAD Y ESTADÍSTICA

{Con ejemplos de programación}

```
/**
* Versión 1.0
* Fecha: 2025, semestre 2
* Licencia software: GNU GPL
* Licencia doc: GNU Free Document License (GNU FDL)
*/


class Author {
    String name = "Jaime E. Montoya M.";

    String profession = "Ingeniero Informático";

    String employment = "Docente y desarrollador";

    String city = "Medellín - Antioquia - Colombia";

    int year = 2025;
}
```

Tabla de contenido

Introducción

Este documento es un complemento a las clases presenciales y virtuales, y está basado en la bibliografía del curso, así como de otras fuentes adicionales que se indican a lo largo del texto, además de la experiencia del autor en su función docente en las áreas de ciencias básicas. No se pretende reemplazar los textos guías con este manual, sino servir de ayuda didáctica y apoyo académico a los estudiantes.

La guía incluye, además de los conceptos teóricos, ejemplos, gráficas, desarrollos en clase, y al final de cada capítulo, unas preguntas y ejercicios que permitan reforzar los conceptos y promover la práctica y el estudio de los conceptos vistos.

Al final de este manual, se indican fuentes y referencias adicionales que el estudiante puede consultar. Las notas al pie de página contienen enlaces a lecturas complementarias.

El apéndice de este texto presenta distintas aplicaciones de la estadística en programación; los desarrollos son presentados en distintos lenguajes de programación tales como C/C++, PHP, Python, Java, Javascript y C#, entre otros, así como en pseudocódigo y PSeInt.

Capítulo 0. Preliminares: aritmética y álgebra

Los conjuntos numéricos

La idea de conjunto se emplea bastante en matemáticas, por ser un concepto básico, no tiene una definición formal. Más adelante, se hablará un poco más acerca de los conjuntos y las operaciones que podemos realizar con ellos, por ahora veremos los conjuntos de números más conocidos.

Símbolos comunes en notación de conjuntos

Existen varios símbolos muy utilizados particularmente en el tratamiento de conjuntos.

$\in$: $Pertenece.\ Ejemplo:\ a \in A.\ Se\ lee:\ 'a'\ pertenece\ (es\ elemento\ de)\ a\ 'A'$
$\notin$: $No\ pertenece.\ Ejemplo:\ a \notin A.\ Se\ lee:\ 'a'\ no\ es\ elemento\ 'A'$
$\forall_x$: $Cuantificador\ universal.\ Ejemplo: \forall_x\ x > 9.\ Se\ lee:\ Para\ todo\ x\ se\ cumple\ que\ x\ es\ mayor\ que\ 9$
$\exists_x$: $Cuantificador\ existencial.\ Ejemplo:\ \exists_x x > 9.\ Se\ lee:\ Existe\ al\ menos\ un\ x\ tal\ que\ x\ es\ mayor\ que\ 9$
$\exists!_x$: $Cuantificador\ existencial$ único. $Se\ lee:\ Existe\ exactamente\ un\ x$
$\subset$: $Contenido\ en,\ o\ es\ subconjunto\ de.\ Ejemplo:\ A \subset B.\ Se\ lee:\ A\ es\ subconjunto\ de\ B$
$\subseteq$: $Es\ subconjunto\ de\ o\ igual\ al\ conjunto.\ Ejemplo:\ A \subseteq B.\ Se\ lee:\ A\ es\ subconjunto\ o\ igual\ a\ B$

El conjunto de los números naturales **N**

Los números naturales surgen de la necesidad del hombre por contar. Es un conjunto infinito y lo simbolizamos así:
$\aleph = \{1, 2, 3, 4, 5, ...\}$

El conjunto de los números enteros **Z**

Los números enteros incluyen a los números naturales, esto es a los números positivos, a sus respectivos negativos y al cero. Es un conjunto infinito y lo simbolizamos así:

$Z = \{..., -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, ...\}$

Podemos decir entonces que $N \subseteq Z$

El conjunto de los números racionales **Q**

Este conjunto incluye a los números enteros, las fracciones como $\frac{3}{4}$, $-\frac{9}{5}$, los números decimales conmensurables como *2.33, -5.99* y los números decimales inconmensurables periódicos como *0.333333..., 6.778877887788...* Es un conjunto infinito y lo simbolizamos así:

$$Q = \{\frac{p}{q} \mid p \in Z \ \wedge q \in Z;\ q \neq 0\}$$

Por tanto $N \subseteq Z \subseteq Q$

El conjunto de los números irracionales **Q'**

Los números que no son racionales, se denominan irracionales, y son decimales inconmensurables que no son periódicos. Ejemplos de número irracionales, son $\pi = 3.141592...$, $\sqrt{2}$, $e = 2.718281...$, $\sqrt[3]{7}$., etc. Es un conjunto infinito y lo simbolizamos así:

$$Q' = \{x \in R \mid x \notin Q\}$$

El conjunto de los números reales **R**

El conjunto de los números reales está conformado por los números racionales y los irracionales. Ejemplos de número reales son todos los que hemos visto anteriormente. Es un conjunto infinito y lo simbolizamos así:

$$\Re = Q \cup Q'$$

El conjunto de los números complejos **C**

El conjunto de los números complejos está conformado por números de la forma

$$z = a + bi;\ donde\ a,\ b \in R;\ i = \sqrt{-1}\ es\ la\ unidad\ imaginaria$$

Ejemplos de número complejos son: $-54, 45 + 3i, 9 - 2i$. Decimos que los reales están estrictamente contenidos en los complejos:

$$R \subset C$$

<u>Observación</u>
En este curso nos limitaremos al trabajo dentro del conjunto de los números reales **R**.

Propiedades fundamentales de los números reales

El estudio de las propiedades nos permite entender para qué fines sirven, reconocer sus implicaciones y poder derivar o concluir otras cosas de ellas, en otras palabras, cómo trabajar y usarlas en distintas situaciones. Vemos cuáles son:

Sean a, b, c número reales y vamos a aplicar las operaciones de suma y multiplicación.

Observación
La suma $+$ y la multiplicación $\times$ son operaciones fundamentales, mientras que la resta $-$ y la división $\div$ son derivaciones de éstas. Estas operaciones son *binarias*, en el sentido que involucra dos *operandos*.

Propiedad asociativa

La suma y la multiplicación son asociativas, esto es, los operandos se pueden agrupar de cualquier forma.

$a + (b + c) = (a + b) + c = a + b + c$
$a.(b.c) = (a.b).c = a.b.c$

Propiedad conmutativa

La suma y la multiplicación son conmutativas, es decir, el orden de los sumandos o los factores no altera ni la suma ni el producto, respectivamente.

$a + b = b + a$
$a.b = b.a$

Módulo de la suma y el producto. Elementos neutros

El módulo o elemento neutro, es un valor que al ser computado, no altera el resultado. El módulo o elemento neutro en la suma es el cero (0), ya que al sumar cualquier número por éste, obtenemos el mismo número. Similarmente sucede en la multiplicación, donde el uno (1) es el elemento neutro o módulo de esta operación. Los números 0 y 1 también son llamados *elementos identidad* para la suma y la multiplicación, respectivamente.

$a + 0 = 0 + a = a$
$a \times 1 = 1 \times a = a$

Inversos

Todo número real **a** tiene su *inverso aditivo* **-a** (llamado también el negativo de **a**) y satisface:

$a + (-a) = (-a) + a = 0$

Similarmente, todo número **a** distinto de cero tiene un único *inverso multiplicativo* $\mathbf{a^{-1}}$ que satisface:

$a.(a^{-1}) = (a^{-1}).a = 1,\ si\ a \neq 0$

Cabe recordar que $a^{-1} = \frac{1}{a}$

Propiedad distributiva

$a.(b + c) = a.b + a.c$
$(b + c).a = b.a + c.a$

Resta y división

La suma y la multiplicación son operaciones básicas, mientras que la resta y la división derivan de éstas. La resta es la suma de un inverso aditivo y la división es la multiplicación por un recíproco. Tenemos entonces.

$a - b = a + (-b)$
$\frac{a}{b} = a \div b = a.b^{-1} = a.\frac{1}{b}$

Algoritmo

Aunque es un término bastante empleado tanto en matemáticas como en las ciencias computacionales, es común encontrar variantes en la definición. Sin embargo, el consenso general en ciencias, permite definir un ***algoritmo*** como *un conjunto de pasos finitos y ordenados que buscan la solución de un problema*. Este nombre al parecer tuvo influencia en el matemático persa Al-Juarismi, que en latín antiguo se conocía como *Algorithmi*.

En la antigüedad hubo desarrollos de procesos algorítmicos para resolver problemas, entre ellos se encuentra uno de los más famosos conocido como *Algoritmo de Euclides* para hallar el *Máximo Común Divisor (MCD*) de dos enteros.

Orden y Valor absoluto

Orden

En términos *geométricos*, decimos que un número **a** es menor que otro número **b**, si **a** se encuentra a la izquierda de **b** en la recta numérica. Recordemos que todo número real, excepto el 0, es negativo o positivo; entonces, si tenemos estos números, decimos que:

$a < b \; si \; b - a > 0 \, , esto \; es, b - a \; es \; positivo$

Ejemplo 0.1

1. $-3 <-2 \; ya \; que \; -2-(-3) =-2+3=1>0 \; positivo$
   $Observe \; además, \; que \; -2 \; está \; a \; la \; derecha \; de \; -3 \; en \; la \; recta \; numérica$
2. $5 < 11 \; ya \; que \; 11 - 5 = 6 > 0 \; positivo \; (5 \; está \; a \; la \; izquierda \; de \; 11 \; en \; la \; recta \; numérica)$

Valor absoluto

En términos *geométricos*, se define el **valor absoluto** de un número como la distancia que hay desde el cero (0) hasta dicho número en la *recta numérica*; en términos prácticos, la función valor absoluto tiene como fin convertir un número a positivo. Se simboliza encerrando el número entre dos barras: **| a |** donde ***a*** es cualquier número real.

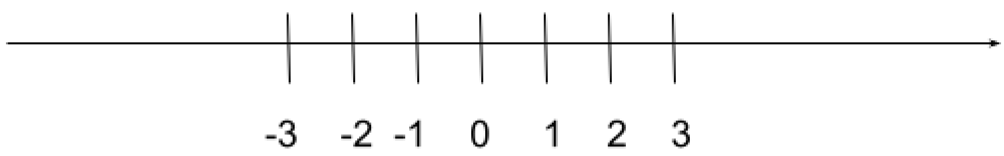
Estrictamente, la función valor absoluto de un número se define así:

$|a| = a \; si \; a \geq 0 \; ó \; |a| = -a, \; si \; a < 0$

Ejemplo 0.2

Observando la recta numérica, podemos ver que la distancia desde el 0 hasta el 3 y desde el 0 hasta el -3 es la misma, esto es, 3 unidades; aplicando la definición de valor absoluto, obtenemos de igual manera 3:

$|3| = 3$
$|-3| = -(-3) = 3$



Propiedades (teoremas y corolarios) del valor absoluto

Sean $a, b \in R$ Se cumplen las siguientes propiedades para el valor absoluto.

1. $||a|| = |a|$
2. $|-a| = |a|$
3. $|a.b| = |a|.|b|$
4. $\left|\frac{a}{b}\right| = \frac{|a|}{|b|}\ si\ b \neq 0$
5. $|a|^2 = a^2$
6. $|x| < a\ si\ y\ solo\ sí\ -a < x < a$
7. $|x| \leq a\ si\ y\ solo\ sí\ -a \leq x \leq a$
8. $|x| > a\ si\ y\ solo\ sí\ x > a\ o\ x < -a$
9. $|x| \geq a\ si\ y\ solo\ sí\ x \geq a\ o\ x \leq -a$
10. $|a + b| \leq |a| + |b|\ (desigualdad\ triangular)$

Ejemplo 0.3

Encuentre el valor absoluto de las siguientes cantidades

1. $|5| = 5$
2. $|-9| = -(-9) = 9$
3. $|-254| = -(-254) = 254$
4. $|0| = 0$
5. $|-1| = 1$

Números pares e impares

Un número ***par*** es un *número entero,* tal que puede escribirse como:

$2k,\ con\ k \in Z$

Decimos que los números pares son exactamente divisibles por *2* y también son múltiplos de *2*.

El conjunto de los números pares es infinito:

$P = \{..., -6, -4, -2, 0, 2, 4, 6, ...\}$

Ejemplo 0.4

1. $0 = 2 \times 0,\ 0 \in Z$
2. $2 = 2 \times 1,\ 1 \in Z$
3. $4 = 2 \times 2,\ 2 \in Z$
4. $6 = 2 \times 3,\ 3 \in Z$

5. $-8 = 2 \times (-4), \ -4 \in Z$

Un número ***impar*** es un *número entero,* tal que puede escribirse como:

$2k + 1, \, con\, k \in Z$

El conjunto de los números impares también es infinito:

$I = \{..., -7, -5, -3, -1, 1, 3, 5, 7, ...\}$

Ejemplo 0.5

1. $1 = 2 \times 0 + 1, \, 0 \in Z$
2. $3 = 2 \times 1 + 1, \, 1 \in Z$
3. $5 = 2 \times 2 + 1, \, 2 \in Z$
4. $7 = 2 \times 3, \, 3 \in Z$
5. $-9 = 2 \times (-5) + 1, \ -5 \in Z$

Propiedades de las operaciones entre números pares e impares

Sean p_1, p_2 dos números pares e i_1, i_2 dos números impares. Se cumplen las siguientes operaciones entre ellos:

1. $p_1 + p_2 = 2n$
2. $p_1 . p_2 = 2n$
3. $p_1 + i_1 = 2n + 1$
4. $p_1 . i_1 = 2n$
5. $i_1 + i_2 = 2n$
6. $i_1 \times i_2 = 2n + 1$

Demostración

1. $p_1 + p_2 = 2a + 2b = 2(a + b) = 2c = 2n$
2. $p_1 . p_2 = 2a \times 2b = 2(a. 2b) = 2c = 2n$
3. $p_1 + i_1 = 2a + 2b + 1 = 2(a + b) + 1 = 2c + 1 = 2n + 1$
4. $p_1 . i_1 = 2a(2b + 1) = 2a \times 2b + 2a = 2(a. 2b) + 2a = 2(c + a) = 2n$
5. $i_1 + i_2 = 2a + 1 + 2b + 1 = 2(a + b + 1) = 2c = 2n$
6. $i_1 \times i_2 = (2a + 1)(2b + 1) = 2a \times 2b + 2a + 2b + 1 = 2(a. 2b) + 2a + 2b + 1$
   $= 2(2ab + a + b) + 1 = 2c + 1 = 2n + 1$

Nota: Paridad del cero
El cero (0) es un número par que cumple con las propiedades comentadas arriba.

Números primos

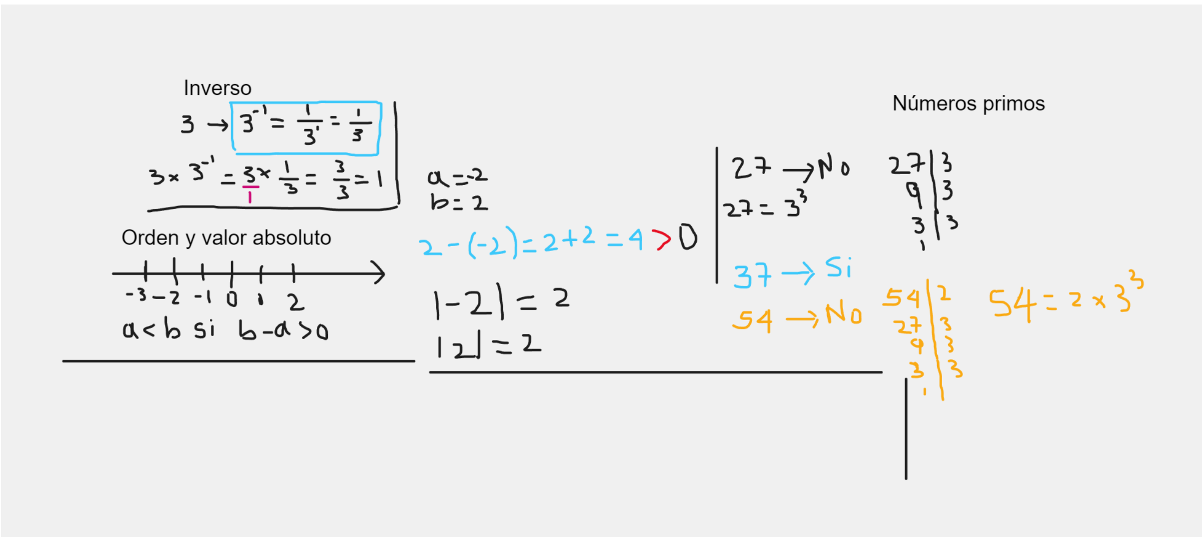
Un número primo es un número entero positivo que tiene solo dos divisores exactos distintos: el mismo número y la unidad.

Ejemplo 0.6
Los siguientes son algunos números primos

2, 3, 5, 7, 11, 13, 17, 23, 29, 31,...

Tablero virtual

Ilustración en clase virtual del valor absoluto y los números primos



Sumatoria, Productoria y Factorial

Sumatoria

Es una notación matemática utilizada para representar la suma de varios términos, o incluso infinitos, como es común encontrar en el cálculo, lo que simplifica la escritura de sumas grandes.
La sumatoria se representa con la letra griega Σ (*sigma* mayúscula).

Nota
La sumatoria es una suma, por tanto podemos utilizar ambos términos sin problema en este contexto.

Notación

$$\sum_{i=m}^{n} a_i = a_m + a_{m+1} + a_{m+2} + ... + a_n$$

Esto se lee: Sumatoria de *i*, desde *i* igual a *m* hasta *n*, de *a* sub *i*

Se debe cumplir, además, que:
$m \leq n$

Si $m = n$, entonces $\sum_{i=m}^{m} a_i = a_m$

Por definición, si $m > n$, entonces $\sum_{i=m}^{n} a_i = 0$

El número de términos a sumar es, por tanto: $n - m + 1$

Ejemplo 0.7

1. $\sum_{i=1}^{5} i = 1 + 2 + 3 + 4 + 5 = 15.\ Términos\ a\ sumar:\ 5 - 1 + 1 = 5$
2. $\sum_{i=3}^{8} i = 3 + 4 + 5 + 6 + 7 + 8 = 33.\ Términos\ a\ sumar:\ 8 - 3 + 1 = 6$
3. $\sum_{i=1}^{3} i^2 = 1^2 + 2^2 + 3^2 = 1 + 4 + 9 = 14$
4. $\sum_{i=1}^{\infty} i = 1 + 2 + 3 + 4 + 5 + ... = \infty.\ Términos\ a\ sumar: infinitos$

Productoria

Es una notación matemática utilizada para representar el producto de varios términos, o incluso infinitos, simplificando la escritura de grandes multiplicaciones.
La sumatoria se representa con la letra griega Π (*pi* mayúscula).

Nota
La productoria es una multiplicación, por tanto podemos utilizar ambos términos sin problema en este contexto.

Notación

$$\prod_{k=m}^{n} a_k = a_m \times a_{m+1} \times a_{m+2} \times \ldots \times a_n$$

Esto se lee: Productoria de *k*, desde *k* igual a *m* hasta *n*, de *a* sub *k*

Se debe cumplir, además, que:
$m \leq n$

Si $m = n$, entonces $\prod_{k=m}^{n} a_k = a_m$

Por definición, si $m > n$, entonces $\prod_{k=m}^{n} a_k = 1$

El número de términos a multiplicar es, por tanto: $n - m + 1$

Ejemplo 0.8

1. $\prod_{k=1}^{5} k = 1 \times 2 \times 3 \times 4 \times 5 = 120.\ Términos\ a\ multiplicar\colon\ 5 - 1 + 1 = 5$
2. $\prod_{k=5}^{8} k = 5 \times 6 \times 7 \times 8 = 1680.\ Términos\ a\ multiplicar\colon\ 8 - 5 + 1 = 4$
3. $\prod_{k=1}^{3} k^2 = 1^2 \times 2^2 \times 3^2 = 1 \times 4 \times 9 = 36$
4. $\prod_{k=1}^{\infty} k = 1 \times 2 \times 3 \times 4 \times 5 \times \ldots = \infty.\ Términos\ a\ multiplicar\colon infinitos$

Factorial

El *factorial* de un número entero positivo, se define como el producto de los números desde uno hasta dicho número. Se simboliza acompañando al número del signo de cierre de exclamación (***!***).

$n! = 1 \times 2 \times 3 \times ... \times (n - 2) \times (n - 1) \times n$
Y se lee: *n* factorial o factorial de *n*.

Dado que conocemos la notación de productoria, podemos escribir:

$n! = \prod_{k=1}^{n} k$

El factorial de *n* $(n \in Z_+ \; - enteros\, positivos\, -)$ también se define de manera *recursiva*:

$n! = 1, \; si\, n = 0 \; o \; n = 1$
$n! = n(n - 1)!, \; si\, n > 1$

Esto se conoce como un proceso *recurrente* (*recursivo*), ya que se llama a sí mismo hasta llegar a un estado básico, el cual determina cuándo debe dejar de ser recurrente el proceso para no caer en un ciclo infinito; para el caso del factorial, sus estados básicos son 0 y 1.

Observe que por definición tenemos que:

$0! = 1$

Ejemplo 0.9

1. $5! = 5 \times 4! = 5 \times 4 \times 3! = 5 \times 4 \times 3 \times 2! = 5 \times 4 \times 3 \times 2 \times 1! = 5 \times 4 \times 3 \times 2 \times 1 = 120$
2. $4! = 4 \times 3 \times 2 \times 1 = 24$
3. $7! = 5040$

Potencias, exponentes y radicales

Potencia

Potencia de una expresión es la misma expresión o el resultado de tomarla como factor dos o más veces. La potencia es una operación basada en multiplicaciones sucesivas, donde se toma un número llamado *base* y se multiplica tantas veces indique otro número llamado *exponente*.

$a^b = c, \; donde$
$a: \; base$
$b: \; exponente$
$c: \; potencia$

Exponentes

Los exponentes están ligados a unas reglas que nos facilitan el trabajo de cálculo con potencias. Aunque las reglas se aplican en general a todos los números, los exponentes fraccionarios tienen una interpretación que veremos más adelante.

Exponentes enteros

Veamos las principales reglas para trabajar con exponentes enteros.

1. $a^{m}.a^{n} = a^{m+n}$
2. $(a^{m})^{n} = a^{m.n}$
3. $a^{0} = 1\ si\ a \neq 0,\ o\ lo\ que\ es\ equivalente\ a\ decir:\ a^{-n} = \frac{1}{a^{n}},\ si\ a \neq 0$
4. $\frac{a^{m}}{a^{n}} = a^{m-n},\ si\ a \neq 0$
5. $(a.b)^{n} = a^{n}.b^{n}$
6. $(\frac{a}{b})^{n} = \frac{a^{n}}{b^{n}},\ si\ b \neq 0$

Exponentes fraccionarios

Los exponentes fraccionarios tienen una interpretación propia, tal y como se enuncia en el texto Álgebra Elemental "*Toda cantidad elevada a un exponente fraccionario equivale a una raíz cuyo índice es el denominador del exponente y la cantidad subradical la misma cantidad elevada a la potencia que indica el numerador del exponente*"[1]. Esto es:

$$a^{\frac{m}{n}} = \sqrt[n]{a^{m}}$$

Radicales

Recordemos las partes de un radical:

$\sqrt[n]{a^{m}} = b\ ,\ donde$

$\sqrt{\ }\ símbolo\ \ o\ símbolo\ radical$

$n:\ índice\ del\ radical\ (si\ no\ aparece,\ se\ sobreentiende\ que\ es\ 2\ o\ raíz\ cuadrada)$

$a^{m}:\ cantidad\ subradical$

$b:\ raíz$

[1] Álgebra de Elemental. Aurelio Baldor. Interpretación del exponente fraccionario. Página 402

Propiedades de los radicales

$\sqrt[n]{a.b.c} = \sqrt[n]{a}.\sqrt[n]{b}.\sqrt[n]{c}$

$\sqrt[n]{\frac{a}{b}} = \frac{\sqrt[n]{a}}{\sqrt[n]{b}}\ si\ b \neq 0$

Del álgebra sabemos que el símbolo $\sqrt{a},\ a \geq 0$, se define como el único *x no negativo* tal que $x^2 = a$

Ejemplo 0.10

$\sqrt{4} = 2; \sqrt{0} = 0; \sqrt{\frac{9}{25}} = \frac{3}{5}$

Nota

$\sqrt{4} \neq -2,\ aun\ cuando\ (-2)^2 = 4,\ ya\ que\ \sqrt{4}\ denota\ \text{ú}nicamente\ la\ ra\text{í}z\ positiva\ de\ 4$

De la definición de $\sqrt{a}$ se deduce que:

$\sqrt{x^2} = |x|$

Ejemplo 0.11

$\sqrt{5^2} = |5| = 5;\ \sqrt{(-3)^2} = |-3| = 3$

Tablero virtual

Ilustración en clase virtual de operaciones con exponentes, radicales y conceptos teóricos.

$a^b = c$ ← exponente ← potencia ↑ base

$2^3 = 2\times2\times2 = 8$

$3 + 2^3 \times 3 - 4 \div 2$

$3 + 2\hat{}3 \times 3 - 4/2$

$3 + 8 \times 3 - 4/2$

$3 + 24 - 4/2$

$3 + 24 - 2 \rightarrow 27 - 2$

$25 //$

$2 = 3 \rightarrow$ Falso

$4 \leq 5 = \rightarrow$ Verdadero

-2 -1 0 1 2 3 4 5 6

$3^5 = 3\times3\times3\times3\times3 = 243$

$8\hat{}3 = 512$

$10^1 = 10$

$30^0 = 1$

$-3^2 = -3\times3 = -9$

$(-7)^2 = (-7)(-7) = 49$

$-9^3 = -729$

$(-4)^3 = -64$

$4\hat{}(1/2) = \sqrt{4} = 4^{1/2}$

$\sqrt[3]{27} = 27^{1/3}$

$\boxed{\sqrt[n]{a^m} = a^{m/n}}$

$(ab)^n = a^n b^n$

$(2\times3)^2 = 2^2 \times 3^2$

$6^2 = 4 \times 9$

$36 = 36$

$(a^n)^m = a^{mn}$

$\sqrt[3]{3^3} = 3^{3/3} = 3^1 = 3$

$\boxed{\sqrt[5]{45} = 45^{1/5}}$

$= (3^2 \times 5)^{1/5}$

$= (3^2)^{1/5} \times (5)^{1/5}$

$= 3^{2/5} \times 5^{1/2}$

45	3
15	3
5	5
1	

$45 = 3^2 \times 5$

El divisor o denominador debe ser diferente de 0

$\frac{a}{b}$ si $b \neq 0$ | $a \div b$, a/b, $a \lfloor b$

$\frac{15}{30} = 15/30 = 15 \div 30 = 15 \lfloor 30$

Numerador

Denominador

Divisor

Dividendo

$\frac{\cancel{15}}{\cancel{30}} = \frac{1}{2} = 0.5$

Radicales con índice par

$\sqrt{4} = \pm 2 \rightarrow 2^2 = 4$

$(-2)^2 = 4$

$\sqrt{-4} = \sqrt{(-1)(4)} = \sqrt{-1} \cdot \sqrt{4} = \pm 2\sqrt{-1}$

$\sqrt{-1} = i \rightarrow$ Unidad imaginaria

$\sqrt[3]{27} = \sqrt[3]{3^3} = 3^{3/3} = 3^1 = 3$

$\sqrt[3]{-27} = -3$

Conjuntos

El concepto de ***conjunto*** es muy utilizado en matemáticas, pero *no se define*, así como no se definen otros conceptos como ***punto***, ***recta*** o ***plano***, por ejemplo. Se suelen usar sinónimos: *agrupación, clase, reunión, colección*.

Los conceptos matemáticos que no se definen se conocen como *primitivos*, y entre ellos también se encuentra ***elemento*** y ***pertenencia***.

Un *conjunto* puede o no contener *elementos*, y éstos no se repiten, esto es, todos son distintos; se les suele nombrar con letras minúsculas, mientras que las letras mayúsculas se usan para los nombres de los conjuntos.

El orden de los elementos en el conjunto no es relevante en general.

Un conjunto puede ser *finito* si se conoce el número de elementos que contiene, en caso contrario se dice que es un conjunto *infinito*.

Representación de conjuntos

Hay dos formas de representar un conjunto:

- Diagramas de Venn
- Elementos encerrados entre llaves

<u>Ejemplo</u>

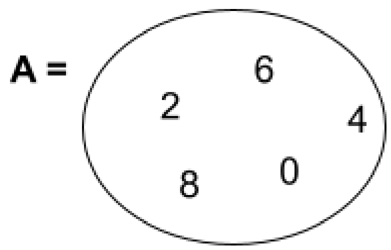
Sea el conjunto **A** formado por los números pares del 0 al 8.

El conjunto es finito y es fácil de representar en cualquier forma:

Representación con llaves

$A = \{0, 2, 4, 6, 8\}$

Diagrama de Venn



Cuando el número de elementos es muy grande, así sea un número finito, el diagrama de Venn deja de ser práctico y conviene utilizar la notación de llaves, poniendo puntos suspensivos.

<u>Ejemplo</u>

Sea **B** el conjunto de los números naturales del 1 al 100

$B = \{1, 2, 3, 4, 5,..., 100\}$

Si el conjunto es infinito, se puede utilizar esta forma escribiendo un número significativo de elementos de tal forma que permitan observar la regla que cumplen éstos.

Ejemplo
Sea C el conjunto formado por los positivos múltiplos de 3

$C = \{3, 6, 9, 12, 15, 18, 21,...\}$

Esta forma de escribir los conjuntos, se conoce por ***extensión***. La otra forma de escribir un conjunto es por ***comprensión***, que veremos más adelante.

Conjunto vacío

En un conjunto que carece de elementos. Se simboliza de cualquiera de las siguientes formas:

- $\varnothing$
- $\{\}$

Cardinalidad de un conjunto | |

Es el número de elementos de un conjunto.

Ejemplo
$Sea\ A = \{a,\ b, c, d, e\}$
$|A| = 5$

Símbolos comunes en notación de conjuntos

Como vimos en la Unidad 0, existen varios símbolos muy utilizados en el tratamiento de conjuntos. Veamos de nuevo cuales son:

$\in$: $Pertenece.\ Ejemplo:\ a \in A.\ Se\ lee:\ 'a'\ pertenece\ (es\ elemento\ de)\ a\ 'A'$
$\notin$: $No\ pertenece.\ Ejemplo:\ a \notin A.\ Se\ lee:\ 'a'\ no\ es\ elemento\ 'A'$
$\forall_x$: $Cuantificador\ universal. Ejemplo: \forall_x\ x > 9.\ Se\ lee:\ Para\ todo\ x\ se\ cumple\ que\ x\ es\ mayor\ que\ 9$
$\exists_x$: $Cuantificador\ existencial.\ Ejemplo:\ \exists_x x > 9.\ Se\ lee:\ Existe\ al\ menos\ un\ x\ tal\ que\ x\ es\ mayor\ que\ 9$
$\exists!_x$: $Cuantificador\ existencial$ único. $Se\ lee:\ Existe\ exactamente\ un\ x$
$\subset$: $Contenido\ en,\ o\ es\ subconjunto\ de.\ Ejemplo:\ A \subset B.\ Se\ lee:\ A\ es\ subconjunto\ de\ B$
$\subseteq$: $Es\ subconjunto\ de\ o\ igual\ al\ conjunto.\ Ejemplo:\ A \subseteq B.\ Se\ lee:\ A\ es\ subconjunto\ o\ igual\ a\ B$

Variables en conjuntos

Es un símbolo (x, y, z, etc.) que representa un elemento no definido o especificado de un conjunto dado.

Las variables son útiles para escribir los conjuntos por comprensión, expresado mediante alguna regla que deben cumplir éstas.

Ejemplo

$Sea\, x > 9$

Esta expresión no es una proposición, ya que al no conocer el valor de la *variable* **x**, no es posible determinar un valor de verdad para ésta, a no ser que diéramos un valor específico a la variable x. Esta expresión nos dice que x es cualquier número (no sabemos cuál) mayor que 9.

Podemos definir el conjunto **Q** que cumplan la regla que todos los elementos son números mayores a 9, lo cual puede escribirse así:

$Q = \{x \,/\, x > 9,\ x \in \aleph\}$

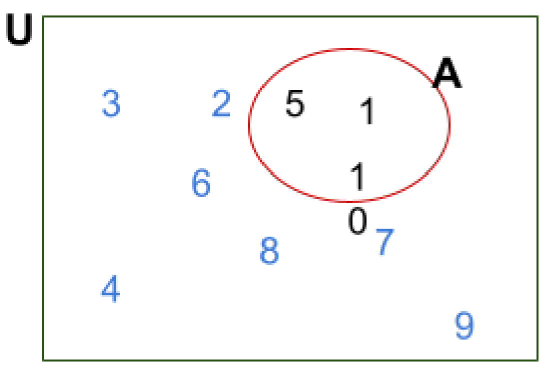
Esta forma de escribir un conjunto se conoce por ***comprensión***, y se lee: "**Q** es el conjunto compuesto cualquier número **x**, tal que ***x*** es mayor que nueve y **x** está en los naturales".

Conjunto Referencial o Universal

Es el conjunto formado por todos los elementos de estudio en un contexto dado. Se representa por **U**.

Ejemplo

Sea el conjunto $U = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ y el conjunto $A = \{1, 5, 10\}$



El conjunto A puede ser escrito así:

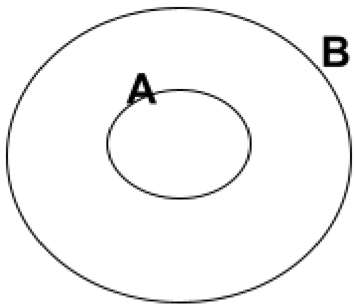
$A = \{x \,/\, x = 1 \lor x = 5 \lor x = 10,\ x \in U\}$

Subconjuntos

A es subconjunto de B sí y sólo sí todos los elementos de A son elementos de B. Se simboliza: $A \subset B$

Ejemplo

Representación gráfica de un subconjunto: $A \subset B$



Propiedades de los subconjuntos

1. El conjunto vacío es subconjunto de todo conjunto, esto es, si A es un conjunto, entonces $\{\} \subset A$
2. Todo conjunto es subconjunto de sí mismo; es decir, si A es un conjunto, entonces $A \subset A$
3. Si A es subconjunto de B, no necesariamente B es subconjunto de A
4. Transitividad: Si A es subconjunto de B y es subconjunto de C, entonces A es subconjunto de C

Igualdad de conjuntos

Dos conjuntos A y B son iguales sí y sólo sí todo elemento de A es elemento B. Se escribe: $A = B$

Operaciones entre conjuntos

A través de varias operaciones con conjuntos, es posible crear nuevos conjuntos. Veamos cuáles son.

Unión ∪

La ***unión*** de dos conjuntos A y B, es el conjunto formado por los elementos comunes y no comunes de A y B:

$A \cup B = \{x \,/\, x \in A \vee x \in B\}$

Intersección ∩

La ***intersección*** de dos conjuntos A y B, es el conjunto formado por los elementos comunes entre A y B:

$A \cap B = \{x \,/\, x \in A \wedge x \in B\}$

Diferencia −

La ***diferencia*** entre el conjunto A y el conjunto B, es el conjunto formado por los elementos comunes de A que no están en B:

$A - B = \{x \,/\, x \in A \wedge x \notin B\}$

Diferencia simétrica Δ

La ***diferencia simétrica*** entre el conjunto A y el conjunto B, es el conjunto formado por los elementos que pertenecen al conjunto A o al conjunto B y que no sean comunes; se puede expresar de las siguientes formas:

$A \Delta B = \{x \,/\, x \in (A \cup B) \wedge x \notin (A \cap B)\}$

$A \Delta B = (A \cup B) - (A \cap B)$
$A \Delta B = (A - B) \cup (B - A)$

Complemento de un conjunto

Sea A un subconjunto de U. El ***complemento*** de A, simbolizado A', es el conjunto formado por los elementos de U que no están en A.

$A' = \{x \,/\, x \in U \wedge x \notin A\}$
$A' = U - A$

<u>Ejemplo</u>
Realice las siguientes operaciones sobre los conjuntos `U = {1, 2, 3, a, b, c, d}`, `A = {1, 2, 3}`, `B = {a, b}`: unión, intersección, diferencia, diferencia simétrica, complemento de A

$A \cup B = \{1, 2, 3, a, b\}$
$A \cap B = \{\,\}$
$A - B = \{1, 2, 3\}$
$A \Delta B = \{1, 2, 3, a, b\}$
$A' = \{a, b, c, d\}$

Conjunto potencia o conjunto de partes P

Sea A un conjunto. El conjunto ***potencia*** de A está formado por todos los subconjuntos del conjunto A.

$P(A) = \{X / X \subset A\}, \ donde\ X\ es\ cualquier\ subconjunto\ de\ A$

El número de subconjuntos (elementos) del conjunto potencia está dado por 2^n, donde **n** es el número de elementos del conjunto; de ahí que reciba este nombre.

<u>Ejemplo</u>

$Sea\ A = \{a, b\ c\}$

$El\ número\ de\ subconjuntos\ de\ A\ es\colon\ |P(A)| = 2^3 = 8$

$P(A) = \{\{\}, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c,\}, \{b, c\}, \{a,\ b,\ c\}\}$

Producto cartesiano ×

El ***producto cartesiano*** entre los conjuntos A y B es otro conjunto formado por los ***pares ordenados*** entre A y B.

$A \times B = \{(a,\ b) / a\ \in A \wedge b \in B\}$

Un par ordenado (a, b) está formado por dos elementos donde importa el orden, esto quiere decir que $(a, b) \neq (b, a)$

El ***plano cartesiano*** es un claro ejemplo que ilustra pares ordenados, donde la primera coordenada o parámetro indica el eje **x** y la segunda el eje **y**.

<u>Ejemplo</u>

$Sea\ la\ pareja\ ordenada\ del\ plano\ cartesiano\ (x, y) = (5, -4). Punto\ ubicado\ en\ el\ cuarto\ cuadrante$

$Ahora\colon\ (y,\ x) = (-4, 5).\ Punto\ ubicado\ en\ el\ segundo\ cuadrante$

El número de elementos de este conjunto está dado por el producto entre el número de elementos de A y el número de elementos de B.

$|A \times B| = |A| * |B|$

<u>Ejemplo</u>

$Sea\ A = \{a, b\},\ B = \{1\ ,\ 2,\ 3\}$

$Cardinalidad\ de\ |A \times B| = |2| \times |3| = 2 \times 3 = 6\ elementos$

$A \times B = \{\{(a, 1)\}, \{(a, 2)\}, \{(a, 3)\}, \{(b, 1)\}, \{(b, 2)\}, \{(b, 3)\}\}$

$B \times A = \{\{(1, a)\}, \{(2, a)\}, \{(a, 2)\}, \{(1, b)\}, \{(2, b)\}, \{(3, b)\}\}$

Observe que $A \times B \neq B \times A$

Operadores básicos en matemáticas y computación

Un *operador* es un símbolo usado en matemáticas para representar una operación a realizar, la cual puede ser unaria (con un *operando*) o binaria (con dos *operandos*). En la aritmética y en el álgebra se cuenta, entre otros, con varios operadores elementales; cada operador tiene una *prioridad* asignada, lo cual significa que los de mayor prioridad, se ejecutarán primero. Se dividen en tres grupos, de los cuales se muestra su representación matemática, así como algorítmica.

Operadores aritméticos

Utilizados para realizar cálculos (operaciones) matemáticos

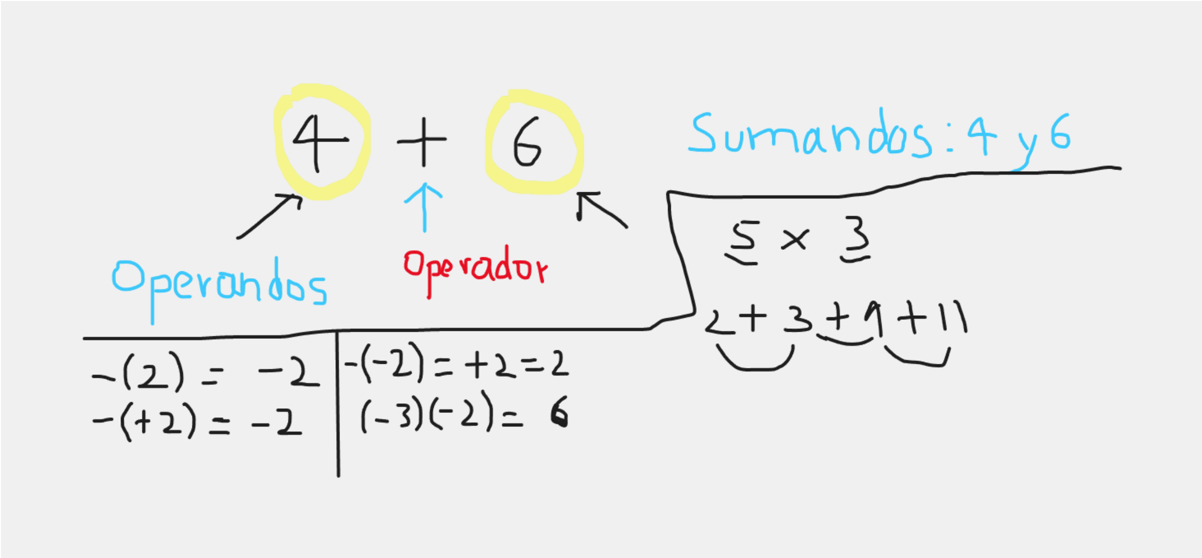
Nombre Operador	**Símbolo matemático y algorítmico (computacional)**	**Prioridad**
Negación aritmética unaria	$-$	Alta
Potencia	x^y ∧ **	Alta - media
Raíz cuadrada	$\sqrt{x}$ |/ $raiz2(x)$ $sqrt(x)$	
Multiplicación	$\times$ * . $()()$	Media
División	$\div$ / $\frac{a}{b}$ a|div	
Módulo	% mod	
División entera	\ div //	
Suma	$+$	Baja
Resta	$-$	

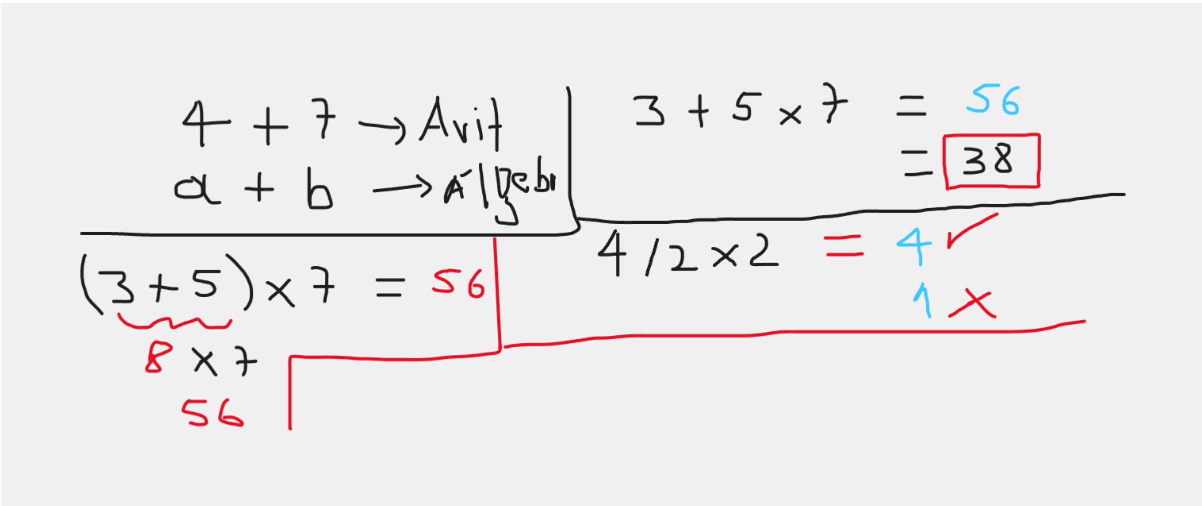
Notas

- La negación aritmética es una operación *unaria* que consiste en negar el símbolo del número (operando). Ejemplo: $-(+2),\ -(8),\ -(-5),\ -94$
- La prioridad se refiere al orden en que los operadores se efectúan en una expresión aritmética: los de mayor prioridad se efectúan primero.
- Las operaciones encerradas entre paréntesis se efectúan primero, por lo que tienen mayor prioridad. Los paréntesis modifican la prioridad de los operadores en una expresión.
- Si hay dos operadores de igual prioridad, se ejecuta primero el que se encuentre más a la izquierda, esto es, se sigue el orden de izquierda a derecha.

Tablero virtual

Ilustración en clase virtual acerca de las partes que intervienen en una expresión matemática y prioridad de los operadores.





Operación de módulo

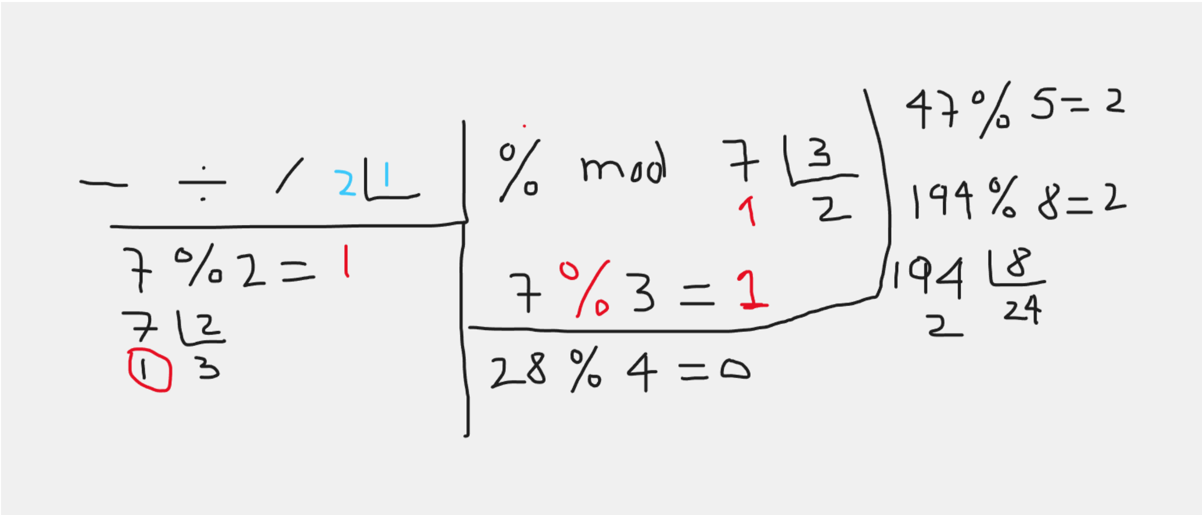
Es una división entera que devuelve el residuo de ésta. Se representa con el símbolo ***%*** ó la palabra ***mod***, entre otros usados.

Ejemplo

```
1. 7 % 5 = 2
2. 17 % 2 = 1
3. 48 mod 4 = 0
4. 57 % 6 = 3
5. 49 mod 5 = 4
6. 9 mod 20 = 9
7. 55 % 11 = 0
```

Tablero virtual

Ilustración en clase virtual del uso del operador de módulo.



Operadores relacionales o de comparación

Permiten realizar comparaciones entre dos expresiones; el resultado de una comparación es un valor booleano (verdadero (V) o falso (F); 1 ó 0, respectivamente); esto son:

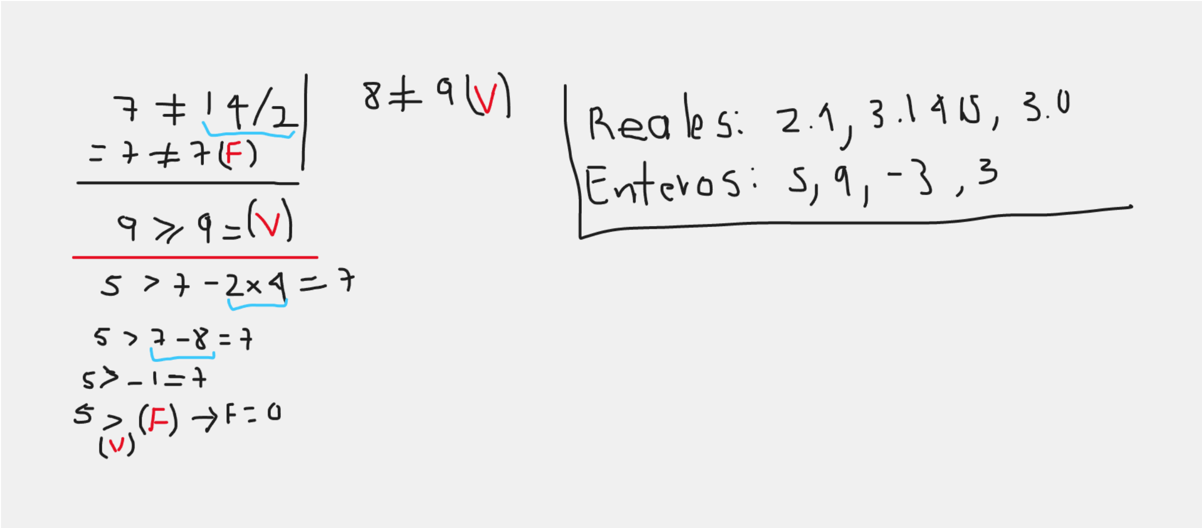
Nombre Operador	**Símbolo**	**Prioridad**
Igual	$=$ $==$	Alta
Diferente	$\neq$ $<>$ $!=$	
Mayor que	$>$	Media
Menor que	$<$	
Mayor o igual que	$\geq$ $>=$	Baja
Menor o igual que	$\leq$ $<=$	

Ejemplo

1. $8 \neq 9 \rightarrow (V)$
2. $9 \geq 9 \rightarrow (V)$
3. $7 \neq 14 \div 2 \rightarrow (F)$
4. $9 \times 2 \leq 50 \div 10 \rightarrow (F)$
5. $-8 = 8 \rightarrow (F)$

Tablero virtual

Ilustración en clase virtual de operaciones usando operadores de comparación.



Operadores lógicos o booleanos

Permiten conectar expresiones de comparación y realizar operaciones lógicas. El valor devuelto (verdadero o falso) depende del conectivo lógico utilizado, según las leyes del álgebra proposicional y booleana; estos son algunos:

Nombre Operador	Símbolo	Prioridad
Negación lógica unaria	$\neg$ $\sim$ $-$ $!$ no not	Alta ↓ (mayor a menor)
Conjunción	$\wedge$ $\&\&$ $\bullet$ y and	
Disyunción	$\vee$ $\|\|$ $+$ o or	
Disyunción exclusiva	$\veebar$ $\oplus$ W $'o\ bien'$ $xor\ \ eor$	

Al llegar a la sección de *lógica matemática* realizaremos operaciones con estos operadores y se ampliará más este tema.

Notación algorítmica

Al trabajar con computadores y lenguajes de programación, algunos símbolos matemáticos son difíciles de obtener desde los caracteres estándar del teclado. Es por ello que las expresiones matemáticas deben ser reescritas cuando las llevamos a un lenguaje de programación utilizando para ello la notación algorítmica típica de la informática.

Para ello, nos basaremos en los operadores vistos anteriormente y su prioridad, así como en las propiedades del álgebra para los números reales y observando cuáles de estos operadores pueden ser usados en un lenguaje determinado. En lógica de programación, el tema de los operadores puede flexibilizarse, pero siempre manteniendo la notación algorítmica. Veamos algunos ejemplos.

Ejemplo
Escribir en notación algorítmica las siguientes expresiones matemáticas

1. $ab + 3ac^3$

2. $\sqrt[3]{b^2} + \frac{a}{3}$
3. $\frac{a-2b+3c}{\sqrt{2}}$
4. $a \geq 0 \wedge b \neq (4 + 2ab^3) \vee [\neg(a + 2 < b) \wedge (-9 = c)]$

<u>Solución</u>

1. $a * b + 3 * a * c \wedge 3$
2. $b \wedge (2/3) + a / 3$
3. Veamos varias formas de escribir esta expresión
    a. $(a - 2 * b + 3 * c) / 2 \wedge (1/2)$
    b. $(a - 2 * b + 3 * c) / 2 \wedge 0.5$
    c. $(a - 2 * b + 3 * c) / raizc(2);\ donde\ raizc()\ es\ una\ función$
4. Veamos cómo escribir esta expresión que incluye todos los operadores. Para la conjunción podemos usar: **y**, **and**, ó **&&**, que son admitidos en lógica de programación y algunos lenguajes; análogamente para la disyunción podemos usar: o, or ó ||. Por último, podemos usar para la negación: **no**, **not** ó **!**. Recordemos que los operadores lógicos trabajan como conectivos.

    $a >= 0\ \&\&\ b <> (4 + 2 * a * (b \wedge 3))\ ||\ (!\ (a + 2 < b)\ \&\&\ (-9 = c))$

<u>Nota</u>
Observe que por la prioridad de los operadores, no es necesario usar paréntesis en algunas expresiones, a no ser que se quiera modificar ésta.

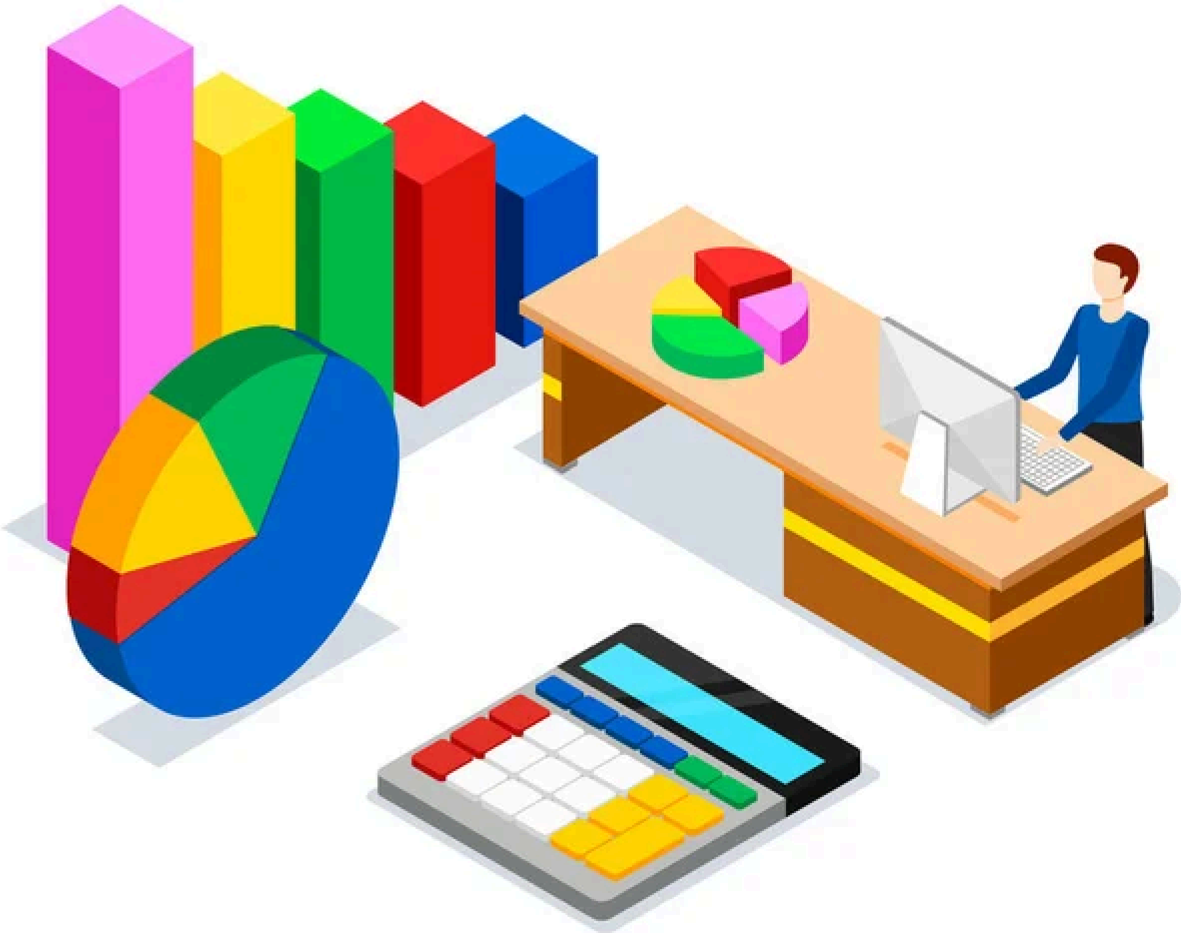
Preguntas

1. Describa los operadores más comunes utilizados en matemáticas e informática
2. ¿Qué es un operador matemático?
3. ¿Cuál es la operación inversa de la suma?
4. ¿Cuál es la operación inversa de la multiplicación?
5. ¿Cuál es la operación inversa de la potencia?
6. ¿Qué es un número primo?
7. ¿Cómo se interpreta el orden y el valor absoluto?
8. ¿Qué devuelve una operación de comparación?
9. ¿Cuáles son los operadores fundamentales?
10. ¿Cuáles son los valores de verdad de las constantes booleanas?
11. ¿A qué hace referencia la prioridad y cómo puede alterarse?
12. ¿Todos los números primos son impares?
13. ¿Cuáles son las propiedades de los números reales?
14. ¿Cuáles son las partes de una potencia y una raíz?
15. ¿Cuáles son los módulos de la suma y el producto?
16. ¿Cuántos números primos hay entre 1 y 25, entre 26 y 50, 51 y 75, 76 y 100? ¿Cómo se comporta el patrón, disminuyen o aumentan en cada intervalo?

Ejercicios

1. Realice los siguientes cálculos para encontrar el valor numérico de cada expresión si $a = 3,\ b = 4,\ c =- 1,\ d =- 5,\ e = 2$
    a. $ab^2 + 3c - \sqrt{b}$
    b. $5e - 2bcd + 4(d^3 - 2a + c) + a \% e$
    c. $(\frac{b}{e} + \frac{e}{c}) - 6(\frac{c^2}{a})$
    d. $\sqrt[3]{d^6} + \frac{a+b+c}{e} + e - b \% 2$
    e. $\sqrt{ab - a} + 5d \div c - a^4 be$
2. Reescriba cada expresión del punto 1) en notación algorítmica o informática
3. Teniendo en cuenta los resultados encontrados en el punto 1), determine el valor lógico de las siguientes comparaciones (las letras corresponden a resultados encontrados en los literales del punto 1), no a los valores numéricos dados allí)
    a. a > b
    b. ab = cd / e
    c. d ^ 2 <= 5ae - b/2
    d. d <> 2ea
    e. 3/c >= 6be + 4a
4. Determine el valor absoluto para cada resultado encontrado en el punto 1)
5. Determine si los siguientes números son primos o no
    a. 154
    b. 96
    c. 63
    d. 37
    e. -13
6. Calcule: $\sum_{i=1}^{7} (i + 2)^2$; $\prod_{j=3}^{6} \sqrt{(j - 1)}$; 8!; $\sum_{i=1}^{4} i!$
7. Demuestre que las siguientes expresiones en valor absoluto son válidas:
    a. |abcd| = |a|.|b|.|c|.|d|
    b. |abc| = |ab|.|c|
    c. |ab| / |cd| = |a|.|b| / (|c|.|d|)
    d. ||a|| = |a|
    e. |ab| / |c| = |a||b| / |c| = |a| / |c| * |b|
8. Sean $p_1,\ p_2$ dos números pares e $i_1,\ i_2$ dos números impares. Demuestre que:
    a. $2p_1$ *es par*
    b. $2i_1$ *es par*
    c. $(i_1)^3$ *es impar*
    d. $(p_1 + i_1)(p_2 + i_2)$ *es impar*
    e. $p_1(i_1 + i_2) + (i_2)^2$ *es impar*

Parte I. Estadística Descriptiva



Capítulo 1. Introducción a la estadística descriptiva

La investigación científica

Es un proceso sistemático y riguroso que utiliza el método científico para generar nuevo conocimiento, verificar hipótesis o responder preguntas sobre el mundo que nos rodea. Se basa en la observación, el análisis, la interpretación y la experimentación, con el objetivo de obtener resultados válidos y confiables.

La investigación científica se define como un proceso ordenado y sistemático de estudio y análisis, utilizando métodos específicos para abordar problemas y generar nuevo conocimiento. Implica la observación, análisis, interpretación y formulación de hipótesis, que se someten a prueba a través de experimentos o investigaciones. Su objetivo principal es la adquisición de conocimiento que pueda ser aplicado en el mundo real.

Características de la investigación científica

Es un proceso sistemático y riguroso
La investigación científica sigue un orden lógico y estructurado, desde la identificación del problema hasta la presentación de resultados. En éste se utilizan métodos y técnicas específicas para asegurar la validez y confiabilidad de los resultados.

Utiliza el método científico
Se basa en la observación, formulación de hipótesis, experimentación, análisis de datos y conclusiones.

Permite la generación de conocimiento
Busca ampliar la comprensión de fenómenos, teorías o problemas existentes, o generar nuevos conocimientos.

Realiza la verificación de hipótesis
Permite comprobar o refutar afirmaciones o suposiciones sobre la realidad.

Permite la resolución de problemas
Puede utilizarse para encontrar soluciones a problemas prácticos o teóricos.

Es objetiva
Busca resultados imparciales, libres de influencias personales.

Es empírica
Se fundamenta en la observación y experimentación.

Es falsable
Las hipótesis pueden ser sometidas a prueba y refutadas.

Es reproducible
Los resultados deben poder ser replicados por otros investigadores.

Es progresiva
El conocimiento científico se construye gradualmente, con base en investigaciones previas.

Tipos de investigación científica

Investigación pura o básica

Se enfoca en la generación de conocimiento teórico sin una aplicación práctica inmediata.

Investigación aplicada

Busca soluciones a problemas específicos utilizando conocimientos científicos previos.

Breve historia de la estadística

La estadística, como disciplina, tiene raíces que se remontan a la antigüedad, vinculadas a la necesidad de los estados de contar y registrar información sobre sus poblaciones y recursos. A lo largo de la historia, ha evolucionado desde simples recuentos hasta convertirse en una ciencia rigurosa con aplicaciones en diversas áreas.

Antigüedad (antes del 3000 a.C)

La necesidad de realizar censos y registros para la administración, la guerra y la economía impulsó los primeros intentos de recopilación y análisis de datos sobre población, impuestos y agricultura. Civilizaciones como Sumeria, Babilonia, Egipto, China, Grecia y Roma destacaron en esta etapa. Un ejemplo notable es el del Rey Yao en China que ordenó recolectar datos sobre comercio e industria, según los escritos de Confucio.

Edad media. Nacimiento de la probabilidad

Aunque hubo menos avances significativos en comparación con la antigüedad, algunos registros como el Domesday Book en Inglaterra (siglo XI) muestran la continuidad de la recopilación de datos sobre la propiedad de la tierra. Girolamo Cardano (1501–1576): En su obra Liber de Ludo Aleae, analizó matemáticamente los juegos de azar, anticipando la teoría de la probabilidad.

Blaise Pascal y Pierre de Fermat desarrollaron formalmente la teoría de la probabilidad en el contexto de los juegos de azar.

Renacimiento y edad moderna

La estadística comenzó a adquirir un carácter más formal, con el desarrollo de métodos de conteo y análisis más sofisticados. El término "estadística" proviene del alemán *Statistik* y se popularizó con el trabajo de Gottfried Achenwall en el siglo XVIII, quien la definió como la "ciencia de las cosas pertenecientes al Estado". Se usaba principalmente para censos, impuestos y planificación estatal.

Siglos XIX–XX. Formalización matemática

La estadística se desarrolló como una disciplina matemática, con la contribución de figuras como Adolphe Quetelet, Francis Galton, y Karl Pearson, quienes introdujeron conceptos como el "hombre promedio" y métodos de análisis más rigurosos. La aplicación de la estadística se extendió a diversas áreas, incluyendo la biología, la medicina, la psicología y las ciencias sociales; desarrollaron conceptos como regresión, correlación y diseño experimental. Florence Nightingale usó gráficos estadísticos para mejorar la sanidad pública.

Siglo XXI. La era digital y del Big Data

La estadística se ha integrado con la informática, dando lugar a la estadística computacional y al análisis de grandes volúmenes de datos (Big Data). Actualmente se aplica en salud, economía, política, deportes y muchas disciplinas más.

¿Qué es la estadística?

La estadística es una rama de las matemáticas que se ocupa de la recopilación, organización, análisis, interpretación y presentación de datos numéricos. Se utiliza para obtener información sobre un conjunto de datos, realizar inferencias y tomar decisiones basadas en las evidencias.

La estadística hace con los datos tareas como:

- **Recopilarlos**: Obtener información relevante de una población o muestra.
- **Organizarlos**: Clasificar y estructurar la información para facilitar su análisis.
- **Analizarlos**: Calcular medidas estadísticas, identificar patrones y tendencias.
- **Interpretarlos**: Extraer conclusiones significativas a partir de los resultados obtenidos.
- **Presentarlos**: Comunicar los hallazgos de manera clara y efectiva, a menudo a través de gráficos y tablas.

La estadística es una herramienta fundamental en diversas disciplinas, como la medicina, las ciencias básicas (física, biología, química), la psicología, la economía, la ingeniería y la investigación de mercados. Permite analizar fenómenos complejos, tomar decisiones informadas y realizar predicciones sobre el comportamiento de eventos futuros.

Algunos ejemplos de aplicación de la estadística:

- **Investigación biológica**: Estudiar el comportamiento de un microorganismo
- **Investigación médica**: Estudiar la eficacia de un nuevo medicamento.
- **Mercadeo**: Analizar las preferencias de los consumidores para dirigir campañas publicitarias.
- **Economía**: Predecir el comportamiento del mercado de valores.
- **Ciencias sociales**: Estudiar la opinión pública sobre temas políticos.

Estadística descriptiva

Es la rama de la estadística que se ocupa de resumir, organizar y presentar datos de manera clara y concisa para facilitar su comprensión y análisis. No realiza inferencias más allá de los datos observados, sino que se enfoca en describir las características principales de un conjunto de datos; utiliza diversas herramientas para resumir grandes conjuntos de datos en medidas como la media, mediana, moda, desviación estándar, etc. Para organizar los datos, los presenta en tablas, cuadros, gráficos y diagramas para facilitar su interpretación.

La estadística descriptiva se centra en describir las características de una muestra o población, sin hacer generalizaciones a otros grupos cuando hace el análisis y trabaja tanto con variables cualitativas como cuantitativas.

Con esta rama de la estadística disponemos de una poderosa herramienta para:

- Resumir y organizar información numérica de manera que sea más accesible y fácil de entender mediante la comprensión de grandes conjuntos de datos.
- Identificar patrones y tendencias mediante gráficos y tablas que ayudan a visualizar los datos y a identificar relaciones entre variables.
- Facilitar la toma de decisiones usando la información descriptiva analizada en los experimentos estadísticos

Estadística inferencial

Es una rama de la estadística que se enfoca en utilizar datos de una muestra para hacer generalizaciones, predicciones o inferencias sobre una población más grande. En lugar de simplemente describir los datos de la muestra, como lo hace la estadística descriptiva, la estadística inferencial busca sacar conclusiones sobre la población completa a partir de la información obtenida de una parte de ella.

La estadística inferencial utiliza técnicas como:

- **Pruebas de hipótesis**: Se utilizan para determinar si existe suficiente evidencia estadística para rechazar una hipótesis nula (una afirmación sobre la población).
- **Estimación**: Se utiliza para estimar parámetros poblacionales (como la media o la proporción) basándose en datos muestrales.
- **Intervalos de confianza**: Se utilizan para calcular un rango de valores dentro del cual se espera que se encuentre el parámetro poblacional con un cierto nivel de confianza.

Conceptos básicos

Población

Es el conjunto completo de elementos que son objeto de estudio. Estos elementos pueden ser individuos, objetos, eventos o cualquier otra cosa que comparta características comunes y sobre la cual se quiere obtener información.

Características de la población estadística

- **Unidad de análisis**: Cada elemento individual dentro de la población se considera una unidad de análisis.
- **Tamaño**: El número de elementos que componen la población se denomina tamaño de la población. Este tamaño puede ser finito (un número limitado de elementos) o infinito (un número ilimitado o extremadamente grande).
- **Heterogeneidad**: Las poblaciones pueden ser homogéneas (todos los elementos son similares) o heterogéneas (los elementos presentan diferencias entre sí).
- **Finitud**: Una población es finita si se puede contar o determinar el número exacto de elementos que la componen, como el número de estudiantes en una escuela.
- **Infinitud**: Una población es infinita si no se puede determinar el número exacto de elementos, como el número de granos de arena en una playa.

Ejemplo
Algunos ejemplos de poblaciones estadísticas, son:

- Todos los habitantes de una ciudad.
- Los árboles de un bosque.
- Los resultados de lanzar una moneda infinitas veces.
- Los clientes de una empresa.
- Los productos fabricados por una empresa en un día.

Importancia de la población estadística

El concepto de población estadística es fundamental en estadística porque permite definir el universo de estudio y establecer los límites de la investigación. Al definir claramente la población, se puede seleccionar una muestra representativa para obtener datos y realizar análisis que permitan inferir conclusiones sobre el comportamiento o las características de la totalidad de la población. La población estadística se diferencia de una muestra, que es una parte de la población que se utiliza para realizar análisis y sacar conclusiones sobre el conjunto total, como veremos a continuación.

Muestra

Es un subconjunto representativo de una población estadística, seleccionada para su estudio. En lugar de analizar a toda la población, se analiza una muestra para obtener conclusiones que puedan generalizarse a toda la población.

Importancia del uso de muestras

- **Costo**: Estudiar toda la población puede ser muy costoso y llevar mucho tiempo.
- **Viabilidad**: A veces es imposible analizar toda la población (por ejemplo, si se trata de poblaciones enormes).
- **Eficiencia**: Las muestras permiten obtener resultados más rápidamente y con menos recursos.

Muestras probabilísticas

Cada miembro de la población tiene una probabilidad conocida de ser seleccionado. Algunos tipos de muestreo probabilístico, son:

- **Aleatorio simple**: Se selecciona una muestra de la población de manera que cada miembro tenga las mismas posibilidades de ser elegido.
- **Sistemático**: Se elige un elemento de la población y luego se seleccionan los siguientes a intervalos regulares.
- **Estratificado**: La población se divide en estratos o grupos, y se toma una muestra aleatoria de cada estrato.
- **Conglomerados**: La población se divide en grupos (conglomerados), y se seleccionan algunos grupos completos para el estudio.

Muestras no probabilísticas

La probabilidad de selección no es conocida. Algunos tipos de muestreo probabilístico, son:

- **Por conveniencia**: Se selecciona la muestra basándose en la facilidad de acceso a los individuos.
- **Intencional o de juicio**: Se selecciona la muestra basándose en el criterio del investigador.
- **Por cuotas**: Se seleccionan individuos hasta cumplir con cuotas predefinidas de características específicas.

- **Bola de nieve**: Se utiliza para poblaciones difíciles de alcanzar, donde los participantes iniciales recomiendan a otros participantes.

Ejemplo

Algunos ejemplos de muestras, son:

- Para conocer la intención de voto de una ciudad, se puede encuestar a una muestra representativa de sus habitantes.
- Para probar la calidad de un producto, se puede analizar una muestra de la producción.
- Para conocer la prevalencia de una enfermedad en una población, se puede realizar un estudio sobre una muestra de personas.

Datos

Un dato es un elemento básico de información, un símbolo que representa un hecho, una condición, un valor o una descripción, sin un significado inherente por sí solo. Los datos pueden ser números, letras, símbolos, o cualquier otra representación que pueda ser interpretada y procesada. En esencia, los datos son la materia prima que, al ser procesada y contextualizada, se transforma en información.

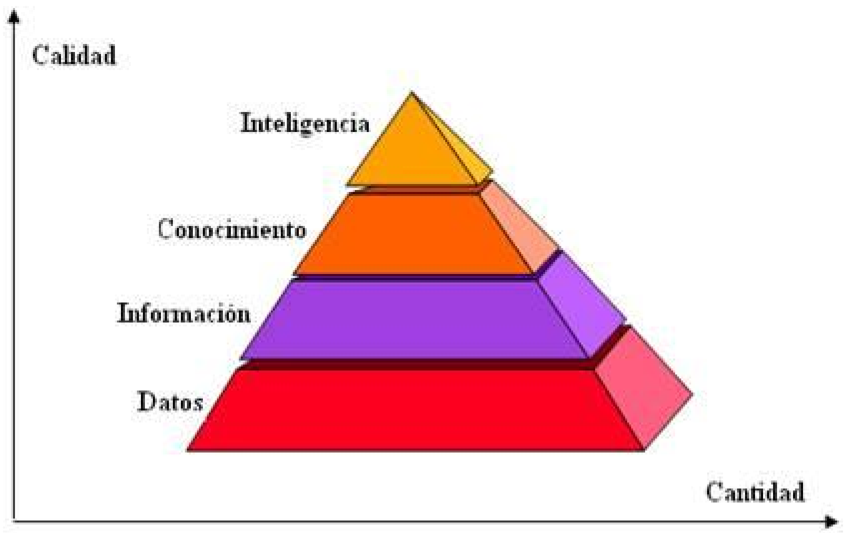


Figura. Pirámide de la información[2]

Los datos son el producto de las observaciones, mediciones y valores que se obtienen al recopilar datos sobre algún fenómeno o característica de interés; pueden ser de naturaleza cuantitativa (numérica) o cualitativa (descriptiva) y se utilizan para analizar, interpretar y obtener conclusiones sobre la información que representan.

[2] Imagen tomada de: Pirámide de la Información. | Download Scientific Diagram

Datos cuantitativos

Son aquellos que se expresan en forma numérica y pueden ser medidos. Los datos cuantitativos se dividen en:

- **Discretos**: Solo pueden tomar valores enteros, como el número de hijos de una familia o el número de autos en un estacionamiento.
- **Continuos**: Pueden tomar cualquier valor real dentro de un intervalo de valores, como la altura de una persona o la temperatura.

Datos cualitativos

Son aquellos que describen características o atributos y no se expresan en números. Los datos cualitativos se dividen en:

- **Nominales**: No tienen un orden específico, como el color de ojos o el estado civil.
- **Ordinales**: Tienen un orden o jerarquía, como el nivel de satisfacción (bajo, medio, alto) o la clasificación en una competencia.

Variables estadísticas

Una variable estadística es una característica o atributo que se puede medir u observar en un individuo u objeto, y que puede tomar diferentes valores. Estas variables son la base para analizar y comprender datos, y se clasifican en cualitativas y cuantitativas.

Variables cualitativas

No se expresan numéricamente, sino que describen cualidades o categorías. A su vez, se dividen en:

- **Nominales**: Los valores no tienen orden entre sí (ej: color de ojos, marca de motos).
- **Ordinales**: Los valores tienen un orden o jerarquía (ej: nivel de satisfacción, grado de escolaridad).

Variables cuantitativas

Se expresan numéricamente, indicando cantidades. Se dividen en:

- **Discretas**: Pueden tomar valores aislados (ej: número de hijos, cantidad de errores en una prueba).
- **Continuas**: Pueden tomar cualquier valor dentro de un intervalo (ej: altura, peso, temperatura).

Ejemplo
Algunos ejemplos de variables estadísticas, son:

- Cualitativa nominal: El estado civil de una persona (soltero, casado, viudo, etc.).

- Cualitativa ordinal: La calificación de un estudiante (insuficiente, suficiente, bien, notable, sobresaliente).
- Cuantitativa discreta: El número de libros en una biblioteca.
- Cuantitativa continua: La estatura de un grupo de personas.

Experimentos estadísticos

Un experimento estadístico es un proceso controlado para obtener datos, ya sean numéricos o no numéricos, y estudiar las variables involucradas en dicho experimento. Se caracteriza por tener resultados posibles conocidos de antemano, resultados que son impredecibles individualmente pero predecibles estadísticamente en muchas repeticiones, y la posibilidad de repetirse bajo condiciones similares. Estos experimentos se utilizan para analizar datos, predecir resultados y tomar decisiones.

En un experimento estadístico hay conceptos a tener presente:

Espacio muestral

Es el conjunto de todos los posibles resultados de un experimento aleatorio. Por ejemplo, al lanzar una moneda, el espacio muestral es {Cara, Sello}.

Suceso elemental

Es cada uno de los resultados individuales de un experimento. Por ejemplo, al lanzar un dado, cada uno de los números {1, 2, 3, 4, 5, 6} es un suceso elemental.

Suceso compuesto

Es un subconjunto del espacio muestral, es decir, un conjunto de sucesos elementales. Por ejemplo, sacar un número par al lanzar un dado es un suceso compuesto {2, 4, 6}.

Experimento aleatorio

Es un experimento en el que no se puede predecir el resultado exacto cada vez, aunque se repita bajo las mismas condiciones; se utiliza para estudiar situaciones donde el azar juega un papel importante, como por ejemplo lanzar una moneda o un dado.

Experimento determinista

Es un experimento donde el resultado es predecible y siempre el mismo bajo las mismas condiciones. Por ejemplo, la reacción química entre hidrógeno y oxígeno para formar agua.

Recolección de datos

Es el proceso de recopilar y medir información de diversas fuentes para obtener datos relevantes sobre un tema específico, con el objetivo de analizarlos y utilizarlos para responder preguntas, evaluar resultados y tomar decisiones informadas. Las fuentes pueden ser encuestas, entrevistas, observaciones, registros administrativos y pruebas o evaluaciones. Además, se pueden utilizar fuentes secundarias como bases de datos existentes, documentos y estudios previos. Este proceso puede incluir la recopilación de datos tanto cuantitativos (numéricos) como cualitativos (descriptivos).

El proceso de recolección debe tener presente:

- **Definir el objetivo**: Es fundamental establecer claramente qué información se necesita recolectar y por qué.
- **Identificar las fuentes de datos**: Determinar de dónde se obtendrán los datos, ya sea a través de encuestas, entrevistas, experimentos, bases de datos, etc.
- **Seleccionar los métodos de recolección**: Elegir las técnicas más adecuadas para obtener la información deseada, como cuestionarios, observación, análisis de documentos, etc.
- **Diseñar los instrumentos de recolección**: Elaborar herramientas como cuestionarios, guiones de entrevista, formatos de observación, etc., para estandarizar el proceso.
- **Recopilar los datos**: Aplicar los métodos y utilizar los instrumentos diseñados para obtener la información requerida.
- **Organizar y analizar los datos**: Clasificar, resumir y analizar los datos recopilados para extraer conclusiones significativas.

La recolección de datos es crucial para la toma de decisiones informadas en diversos ámbitos, como la investigación científica, la gestión empresarial, la planificación estratégica, etc., ya que permite:

- Comprender fenómenos y tendencias
- Evaluar el desempeño de procesos y productos
- Identificar problemas y oportunidades
- Tomar decisiones basadas en evidencia
- Desarrollar nuevas estrategias y productos

Capítulo 2. Representación tabular de muestras

Representación tabular simple de una muestra

Consiste en organizar los datos en una tabla, utilizando filas y columnas para facilitar su lectura, comprensión y análisis. Hay tablas sencillas, como otras más elaboradas que permiten organizar y a su vez extraer datos más fácilmente, como las tablas de frecuencias que resume la información agrupándola según sus clases y su frecuencia, la cual indica cuántas veces se repite cada valor o categoría.

Ejemplo

Se entrevistaron 15 estudiantes en la universidad donde se preguntó cuál es la nota final más alta obtenida en la universidad, encontrando los siguientes resultados que se deben tabular: 5.0, 4.2, 3.9, 4.0, 3.8, 4.6, 4.8, 4.9, 4.8, 4.6, 4.5, 4.4, 4.0, 4.5, 4.7

Persona	**Mejor nota**
1	5
2	4,2
3	3,9
4	4
5	3,8
6	4,6
7	4,8
8	4,9
9	4,8
10	4,6
11	4,5
12	4,4
13	4
14	4,5
15	4,7

Tabla: Tabulación de datos de estudiantes

Moda: multimodal: `4, 4.6, 4.8 y 4.5`, cada una con dos apariciones.
Nota promedio: `4.45`
Varianza: `0.15`
Desviación: `0.38`

Distribución de frecuencias

Es una tabla en que se disponen las modalidades de la variable por filas. En las columnas se dispone el número de ocurrencias por cada valor, porcentajes, etc. La finalidad de las agrupaciones en frecuencias es facilitar la obtención de la información que contienen los datos.

Ejemplo
Se quiere conocer si un grupo de individuos está a favor o en contra de la exhibición de imágenes violentas por televisión, para lo cual han recogido los siguientes datos:

```
X = {2, 1, 5, 3, 3, 2, 3, 1, 4, 2, 4, 2, 3, 2, 3, 4, 3, 3, 1, 2}
```

Reglas para la codificación
1: en contra
2: muy en contra
3: indiferente
4: muy a favor
5: a favor

La inspección de los datos originales no permite responder fácilmente a cuestiones como cuál es la actitud mayoritaria del grupo, y resulta bastante más difícil determinar la magnitud de la diferencia de actitud entre hombres y mujeres.

Podemos hacernos una mejor idea si disponemos en una tabla los valores de la variable acompañados del número de veces (**frecuencia**) que aparece cada valor:

X_i	Frecuencia (f)
1	3
2	6
3	7
4	3
5	1
Total	**20**

Tabla: Distribución de frecuencias preferencias de personas

Donde x_i representa el la variable general que toma valores de 1 a 5, y f la frecuencia con que se repite.

La distribución de frecuencias de los datos muestra que la actitud mayoritaria de los individuos del grupo estudiado es indiferente.

La interpretación de los datos ha sido facilitada porque se ha reducido el número de datos a examinar: en lugar de los 20 datos originales, la tabla contiene sólo 5 valores de la variable y 5 frecuencias.

Otro tipo de tabla se conoce como **tabla de datos agrupados en intervalos** y permite interpretar los datos con mayor sencillez. En esta tabla se acostumbra a ordenar los intervalos del menor al mayor, ya que obedece a una representación lógica del rango, donde el primer intervalo contendrá el mínimo valor y el último intervalo el mayor valor. De igual forma se puede observar claramente cómo se distribuyen las frecuencias a lo largo del rango de los datos, desde los valores más bajos hasta los más altos, así como identificar rápidamente el intervalo modal.

Esta distribución de datos es útil cuando son muchos, ya que la clasificación anterior no permite una clara interpretación de éstos. Generalmente las tablas incluyen varías columnas que se describen a continuación.

Frecuencia absoluta

Es la cantidad de observaciones que pertenecen a cada grupo o intervalo. También se interpreta como la cantidad de veces que se repite un suceso. Se simboliza f_i.

Frecuencia relativa

Es el número de ocurrencias dividido por el total de datos, y se simboliza fr_i o p_i, es decir, es el porcentaje que representa el grupo de datos sobre la muestra en valor decimal.

Frecuencia absoluta acumulada

Es el total de frecuencias de los valores iguales o inferiores al de referencia, y se simbolizan f_a o n_a. La frecuencia acumulada también es definida incluyendo al valor de referencia. En otras palabras, resulta de sumar las frecuencias absolutas de una clase o grupo de la muestra (o población) con la anterior o las anteriores. Por ejemplo, para calcular la frecuencia absoluta acumulada del tercer grupo se suman las frecuencias absolutas del primer, segundo y tercer grupo.

Frecuencia relativa acumulada

Es el total de frecuencias relativas de los valores iguales o inferiores al de referencia; es similar a la operación anterior, pero sumando los porcentajes, y se simboliza f_r o p_a

Ejemplo

Para el conjunto de datos del ejemplo anterior, crear las columnas para mostrar las frecuencias absolutas, relativas, absolutas acumuladas y relativas acumuladas.

X_i	f_i	fr_i	f_a	fr_a
1 - 2	9	0.45	9	0.45
3	7	0.35	16	0.80
4 - 5	4	0.2	20	1.00
Total	**20**	**1.00**		

Tabla: Distribución de frecuencias preferencias de personas

Podemos observar que es más fácil leer los datos e interpretarlos. Por ejemplo, encontramos que las personas que están en contra y muy en contra (intervalo 1 - 2) representan el 45% de la muestra. También que a lo sumo el 55% seleccionó las opciones 3, 4 y 5.

Ejemplo

Se consultaron las edades de un grupo de personas adultas que pueden votar y se encontraron los siguientes resultados:

Adulto 1 (joven) (entre 18 y 28 años): 35
Adulto 2 (entre 29 y 44 años): 20
Adulto 3 (entre 45 y 59 años): 27
Adulto 4 (mayor) (60 o más años): 18

Organizar los datos en una tabla de frecuencias agrupada por intervalos

X_i	f_i	fr_i	f_a	fr_a
[18 - 28]	35	0.35	35	0.35
[29 - 44]	20	0.20	55	0.55
[45 - 59]	27	0.27	82	0.82
[60 - +∞)	18	0.18	100	1.00
Total	**100**	**1.00**		

Tabla: Distribución de frecuencias edades de personas

Podemos observar que el intervalo modal es el primero con un 35% de personas "Adulto 1 (joven)" que pueden votar y que más de la mitad de la población estimada es menor de 45 años.

Capítulo 3. Medidas estadísticas

Son herramientas que permiten resumir y analizar conjuntos de datos numéricos. Estas medidas se clasifican en tres categorías principales: medidas de tendencia central, medidas de dispersión, medidas de posición y medidas de forma.

Medidas de tendencia central

Estas medidas están diseñadas para proporcionar alguna medida cuantitativa de dónde está el centro de los datos en una muestra.

Media de la muestra (media aritmética)

Esta medida es un promedio aritmético; esto es, si la muestra está compuesta de `n` datos (elementos), entonces la media es la suma de estos datos dividido entre el total de éstos, es decir, entre `n`.

$$\overline{x} = \sum_{i=1}^{n} \frac{x_i}{n} = \frac{x_1 + x_2 + x_3 + ... + x_n}{n}$$

Propiedades de la media aritmética

- **Suma de desviaciones igual a cero**: La suma de las diferencias entre cada valor de un conjunto de datos y su media aritmética es siempre cero. Matemáticamente: $\sum_{i=1}^{n} (x_i - \overline{x}) = 0$ donde los x_i son los valores individuales y $\overline{x}$ la media aritmética.
- **Unicidad**: Para un conjunto de datos dado, solo existe una media aritmética única.
- **Extremos**: La media aritmética siempre se encuentra entre el valor más pequeño y el valor más grande del conjunto de datos.
- **Sensibilidad a cambios**: Si se modifica cualquier valor en el conjunto de datos, la media aritmética también cambiará.
- **Afectada por valores extremos**: La media aritmética es muy sensible a valores atípicos o extremos, lo que puede afectar su representatividad de la distribución de los datos.
- **Suma y multiplicación**: Si se suma o multiplica cada valor de un conjunto de datos por una constante, la media aritmética también se sumará o multiplicará por esa misma constante.
- **Mínimo de desviaciones cuadráticas**: La suma de los cuadrados de las desviaciones de cada valor con respecto a la media aritmética es mínima, es decir, es menor que la suma de los cuadrados de las desviaciones con respecto a cualquier otro valor.

Mediana de la muestra

Su propósito es reflejar la tendencia central de la muestra de manera que no esté influida por los valores extremos.

$$\widehat{x} = x_{(n+1)/2}\ si\ n\ es\ impar$$

ó

$$\widehat{x} = \frac{x_{n/2}+x_{n/2+1}}{2}\ si\ n\ es\ par$$

La media está influida de manera considerable por los valores extremos de la muestra, mientras que la mediana hace énfasis en el verdadero "centro" del conjunto de datos.

Media geométrica

Es una medida de tendencia central que se usa para calcular el promedio de tasas de crecimiento porcentual. Se calcula como la raíz n-ésima del producto de los n datos del conjunto. Es útil en finanzas, economía y otros campos donde se trabaja con porcentajes o índices relativos.

Esta medida de tendencia central puede utilizarse para mostrar los cambios porcentuales en una serie de números. Tiene una amplia aplicación en los negocios y en la economía, debido a que con frecuencia se está interesado en establecer el cambio porcentual en las ventas en el producto interno bruto o en cualquier serie económica. Se define como la raíz n-ésima del producto de los `n` datos.

Si los datos de la muestra son $x_1, x_2, x_3, \ldots, x_n$, tenemos que la media geométrica `G` está dada por:

$$G = \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[n]{x_1 \times x_2 \times x_3 \times \ldots \times x_{n_i}}\ si \prod_{i=1}^{n} x_{i_i} \geq 0 \text{ para } n \text{ par}$$

Media armónica

Es otra medida de tendencia central que se utiliza cuando se promedian datos que representan velocidades, tiempos o cocientes. Se calcula como el recíproco de la media aritmética de los recíprocos de los datos. Puede ser más robusta que la media aritmética en la presencia de valores atípicos grandes en el conjunto de datos.

Esta medida se emplea para promediar variaciones con respecto al tiempo tales como productividades, tiempos, rendimientos, cambios, etc.

La media armónica `H` está dada por:

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}+\dots+\frac{1}{x_n}},\ si\ x_i \neq 0\ y\ \sum_{i=1}^{n} \frac{1}{x_i} \neq 0$$

Aplicaciones de la media armónica

Precio promedio
Si se compran varios tipos de productos con distintas cantidades de unidades de cada tipo, pero gastando en ellos igual cantidad de dinero, el precio promedio por unidad es igual a la media armónica de los precios por unidad de cada tipo de producto.

Rendimiento promedio de producción
En un grupo puede haber operarios con distinta velocidad para producir un artículo. Si cada una de estas personas tiene que elaborar igual cantidad de artículos, el promedio de velocidad de rendimientos de tal grupo, es igual al promedio armónico de las velocidades de rendimiento de cada uno de los operarios que lo integran.

Por ejemplo, sean $v_1,\ v_2,\ \dots,\ v_n$ las velocidades de rendimiento de cada uno de los operarios de una empresa de producción. Aunque sean distintas las velocidades de cada operario, producen igual cantidad de productos, por tanto el promedio de velocidad de rendimiento del grupo es:

$H = n / (1/v_1 + 1/v_2 + \dots 1/v_n)$ donde n es el número de operarios.

Moda de la muestra

Es otra medida de tendencia central que indica cuál es el valor que más se repite en un conjunto de datos. No hay un cálculo específico, simplemente se identifica el valor con mayor frecuencia. Un conjunto de datos puede tener una o más modas, o incluso ninguna si ningún valor se repite.

Características

- Fácil de identificar: No requiere cálculos complejos, solo observar la frecuencia de los valores[3].
- Puede ser unimodal, bimodal, multimodal o amodal: Un conjunto de datos puede tener una moda (unimodal), dos modas (bimodal), más de dos modas (multimodal) o simplemente no tener moda, esto es, cuando ningún valor se repite en el conjunto de datos (amodal).

[3] A pesar de que en conjuntos pequeños la moda puede ser fácilmente identificable, no lo será así en conjuntos de muchos datos. El algoritmo para encontrar la moda en un conjunto de datos aunque no es el más complejo, tampoco es trivial, y tiene un nivel de dificultad mayor que los algoritmos para encontrar la media y la desviación, más aún, si se trata de encontrar una multimoda.

- Se puede aplicar a datos cualitativos y cuantitativos: A diferencia de la media y la mediana, que sólo se aplican a datos numéricos, es posible encontrar la moda en conjuntos de datos cualitativos.

Ventajas

- Es útil para datos nominales (categorías).
- En general, es fácil de entender y calcular.
- Puede ser informativa incluso en distribuciones con datos atípicos.

Desventajas

- Puede no ser única, lo que dificulta su interpretación.
- Puede no ser representativa de la distribución de los datos, especialmente si hay varias modas o la distribución es muy dispersa.

Medidas de variabilidad o dispersión

Son parámetros estadísticos que indican la mayor o menor separación de los datos de una distribución con respecto a su valor central, generalmente la media, es decir, muestran qué tan dispersos o agrupados están los datos en torno a un valor promedio.

Estas medidas sirven, entre otras cosas, para:

- Comprender la distribución de los datos: Permiten visualizar si los datos están muy concentrados alrededor de la media o si se distribuyen ampliamente.
- Evaluar la representatividad de la media: Si los datos están muy dispersos, la media puede no ser un buen resumen de la distribución, mientras que una baja dispersión indica que la media es más representativa.
- Comparar diferentes conjuntos de datos: Permiten comparar la variabilidad de diferentes grupos de datos, incluso si están expresados en diferentes unidades.
- Evaluar riesgos e incertidumbre: En contextos como la inversión o la investigación de mercados, la dispersión ayuda a evaluar el riesgo asociado a los datos.

Rango

Es la diferencia entre el valor máximo y el valor mínimo de un conjunto de datos. Permite obtener una idea de la dispersión de los datos: cuanto mayor es el rango, más dispersos estarán los datos (sin considerar la afectación de los valores extremos). El rango, también es llamado **amplitud** o **recorrido de medida**. El rango R se expresa como:

$$R = X_{max} - X_{min}$$

Varianza

Mide la dispersión de los datos respecto a la media, calculando el promedio de las diferencias al cuadrado entre cada dato y la media. La varianza estadística es una medida que cuantifica la dispersión de un conjunto de datos, es decir, qué tan alejados o agrupados están los valores con respecto a su media o promedio. Se calcula como la media de los cuadrados de las desviaciones de cada dato con respecto a la media. Una varianza alta indica que los datos están muy dispersos, esto es, la media no es un buen representante de los datos individuales, mientras que una varianza baja indica que están agrupados cerca de la media, dando a entender que la media sí representa bien a los datos individuales. La varianza σ^2 ó S^2 se puede calcular para la población entera o para una muestra, y está dada por:

Varianza poblacional

Calcula la dispersión de los datos de toda una población. Se calcula sumando las diferencias al cuadrado entre cada dato y la media poblacional, dividida por el número total de elementos en la población (n).

$$\sigma^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Varianza muestral

Estima la varianza de la población a partir de una muestra, utilizando la media muestral. Al usar la media muestral (en lugar de la verdadera media poblacional que no conocemos), los cálculos de la varianza muestral tienden a ser un poco menores que la varianza real de la población. La solución a esto es conocida como **Corrección de Bessel** y consiste dividir entre $n-1$ (tamaño de la muestra) para ajustar este sesgo. Este ajuste aumenta ligeramente el valor de la varianza, lo que proporciona una estimación más precisa e insesgada de la varianza de la población.

<u>Grados de libertad</u>
El uso de $n-1$ se relaciona con la idea de que, al estimar la media muestral, se pierde un "grado de libertad" de los datos. Una vez que se calcula la media muestral, los puntos de datos restantes no pueden variar independientemente; están restringidos para ajustarse a esa media.

$$\sigma^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

<u>**Nota**</u>
Al trabajar generalmente con muestras, utilizaremos lógicamente la varianza muestral.

Desviación Estándar o típica

Es la raíz cuadrada de la varianza, y se expresa en las mismas unidades que los datos originales. Es una medida estadística que indica qué tan dispersos están los datos con respecto a su media o promedio. En otras palabras, nos dice si los valores individuales de un conjunto de datos tienden a estar cerca de la media o si están muy dispersos. Una desviación estándar baja indica que los datos están muy agrupados alrededor de la media, mientras que una desviación estándar alta indica una mayor dispersión. La desviación σ está dada por:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}$$

Rango Intercuartil (RIC)

Es la diferencia entre el tercer cuartil y el primer cuartil, y es útil para analizar la dispersión en la parte central de los datos, ignorando los valores atípicos.

Medidas de posición

Las medidas de posición se utilizan para describir la posición que un dato específico posee en relación con el resto de los datos cuando están en orden por categorías. Los **cuartiles**, **deciles** y **percentiles** son las medidas de posición más populares.

Cuartiles

Son valores de la variable que dividen los datos ordenados en cuartos; cada conjunto de datos tiene tres cuartiles. El primer cuartil, Q_1, es un número tal que a lo sumo el 25% de los datos son menores en valor que dicho valor y a lo sumo el 75% son mayores. El segundo cuartil, Q_2, es la mediana (50%). El tercer cuartil, Q_3, es un número tal que a lo sumo el 75% de los datos son valores menores que dicho valor y a lo sumo el 25% son mayores.

Datos clasificados en orden ascendente

25%	25%	25%	25%

L_i Q_1 Q_2 Q_3 L_s

Deciles

Son valores de la variable que dividen los datos ordenados en diez partes iguales (9 divisiones).

Datos clasificados en orden ascendente

10% 10% 10% 10% 10% 10% 10% 10% 10% 10%

L_i D_1 D_2 D_3 D_4 D_5 D_6 D_7 D_8 D_9 L_s

Percentiles

Son los valores de la variable que dividen un conjunto de datos clasificados en 100 subconjuntos iguales; cada conjunto de datos tiene 99 percentiles. El i-ésimo percentil, P_i, es un valor que a lo sumo el `i%` de los datos son menores que el valor de P_i y a lo sumo `(100 - i)%` de los datos son mayores.

Datos clasificados en orden ascendente

1% 1% 1% 1% ... 1% 1% 1% 1%

L_i P_1 P_2 P_3 ... P_{96} P_{97} P_{98} P_{99} L_s

Para hallar los cuartiles de un conjunto de datos, primero se ordenan los datos de menor a mayor y luego se calcula la posición de cada cuartil. La fórmula para calcular cada cuartil es:

`Q = k * (n + 1) / 4`

Donde `k` es el número del cuartil: 1 para el primero, 2 para el segundo y 3 para el tercero y `n` es el total de datos.

El resultado de la fórmula indica la posición del cuartil. Si el resultado es un número entero, el cuartil es el valor en esa posición. Si el resultado es un número decimal, se interpola entre los valores adyacentes, por ejemplo, si la posición es 2.75, se calcula un promedio ponderado[4] entre el valor en la posición 2 y el valor en la posición 3, con pesos de 0.75 y 0.25 respectivamente:

```
((2 * 0.75) + (3 * 0.25)) / (0.75 + 0.25) = (1.5 + 0.75) / 1 = 
2.25
```

Ejemplo

Se tienen los siguientes datos:
1, 1, 1, 2, 3, 3, 4, 4, 5, 9, 9, 19, 20, 20

Encontrar:

- El valor de cada cuartil
- El octavo decil
- Los percentiles 42, 50 y 87

[4] El promedio ponderado es un tipo de promedio que da más importancia o "peso" a algunos valores de un conjunto de datos que a otros, a diferencia del promedio simple (o no ponderado), donde todos los valores tienen la misma importancia. Se calcula multiplicando cada dato por su peso, sumando estos productos y luego dividiendo el resultado entre la suma de todos los pesos. Se utiliza para obtener una representación más precisa de los datos, especialmente en finanzas (costos de inventario, portafolios de inversión), educación (calificaciones) y análisis de datos.

Medidas de forma

Las medidas estadísticas de forma son coeficientes que describen la forma de una distribución de datos, indicando si ésta es simétrica o sesgada y su nivel de apuntamiento o concentración de datos alrededor de la media. Las dos medidas principales son la **asimetría**, que mide el grado de simetría (o falta de ella) de los datos, y la **curtosis** (o apuntamiento), que describe qué tan puntiaguda o plana es la distribución y la concentración de datos en el centro.

Capítulo 4. Gráficos estadísticos

Un gráfico estadístico expresa una relación entre un conjunto de datos cuantitativos, obtenidos normalmente a partir de una investigación de campo o experimentos estadísticos, aunque también pueden ser obtenidos de otras fuentes, como bases de datos, documentos, etc. Entre los gráficos más comunes, se tienen:

Líneas: se utilizan para mostrar cómo cambia una variable con el correr del tiempo. En este tipo de gráficos, un conjunto de puntos es conectado por medio de líneas rectas que, en conjunto, muestran la dinámica más o menos regular del comportamiento de una variable en relación con otra. Por ejemplo, se pueden utilizar para mostrar cómo ha variado la temperatura promedio de una ciudad en los últimos cinco años.

Barras

Circulares

Líneas

Dispersión

Los gráficos son una herramienta fundamental en la estadística, pues condensan una gran cantidad de información en un espacio reducido, facilitando así su lectura y comprensión. Existen múltiples tipos de gráficos, en los que se puede transmitir información de cualquier tipo. Por ejemplo: los resultados de las elecciones, las ventas de una empresa, los resultados de un censo poblacional.

Ejercicios

1. Se realizó una encuesta a un grupo de personas en varias ciudades acerca de las preferencias audiovisuales para entretenerse y se encontraron los siguientes resultados:

Ciudad	**TV**	**Internet**	**Radio**	**Total**
Medellín	240	470	120	
Bello	200	350	95	

Copacabana	150	260	110	
Girardota	170	300	80	
Barbosa	185	280	120	
Total				

Realice lo siguiente:

a. Determine el tamaño de la muestra
b. Determine el total de encuestados por ciudad y por preferencia
c. Indique cuáles son las variables y de qué tipo son éstas
d. Indique el porcentaje de personas que usan cada medio en general y por ciudad
e. Encuentre las medidas de tendencia central para toda la muestra, para cada medio y para cada ciudad
f. Encuentre las medidas de dispersión para toda la muestra, para cada medio y para cada ciudad
g. Cree un gráfico de barras para representar los datos
h. Cree gráficos circulares para indicar los porcentajes que usan cada medio por ciudad
i. Cree gráficos circulares para indicar los porcentajes que usan en cada ciudad por medio
j. Utilice Excel para resolver este problema
k. (Para los de sistemas) Utilice un lenguaje de programación que permita el ingreso de datos cuantitativos y el respectivo cálculo de las medidas de tendencia central y de dispersión

2. En una investigación que está realizando la universidad, se selecciona un grupo de 20 estudiantes de distintas carreras a los que se les pregunta su edad y estatura. Determine y encuentre:
    a. Cuál es la población y cuál es la muestra
    b. Cuál es el tamaño de la población y de la muestra
    c. Cuáles son los tipos de datos y variables: ¿cuantitativas o cualitativas, discretas o continuas, nominales u ordinales?
    d. Las medidas de tendencia central de los datos medidos
    e. Las medidas de variabilidad de los datos medidos
    f. La gráfica de líneas de los datos por separado

g. Utilice Excel para resolver este problema
h. (Para los de sistemas) Utilice un lenguaje de programación que permita el ingreso de datos cuantitativos y el respectivo cálculo de las medidas de tendencia central y de dispersión

3. Los votos obtenidos por varios candidatos estudiantes en las elecciones al consejo directivo de la universidad que actualmente cuenta con 1500 estudiantes, fueron los siguientes:
   Candidato 1: 632
   Candidato 2: 101
   Candidato 3: 344
   Candidato 4: 715
   Candidato 5: 283
   En blanco: 94
   Anulados: 20

   a. Determine el tamaño de la muestra y de la población
   b. Determine el candidato ganador
   c. Indique cuáles son las variables y de qué tipo son éstas
   d. Indique el porcentaje de votos por cada candidato
   e. Encuentre las medidas de tendencia central para toda la muestra
   f. Encuentre las medidas de dispersión para toda la muestra
   g. Cree un gráfico de barras para representar los datos
   h. Cree un gráfico circulares para indicar los porcentajes de votación
   i. Utilice Excel para resolver este problema
   j. (Para los de sistemas) Utilice un lenguaje de programación que permita el ingreso de datos cuantitativos y el respectivo cálculo de las medidas de tendencia central y de dispersión

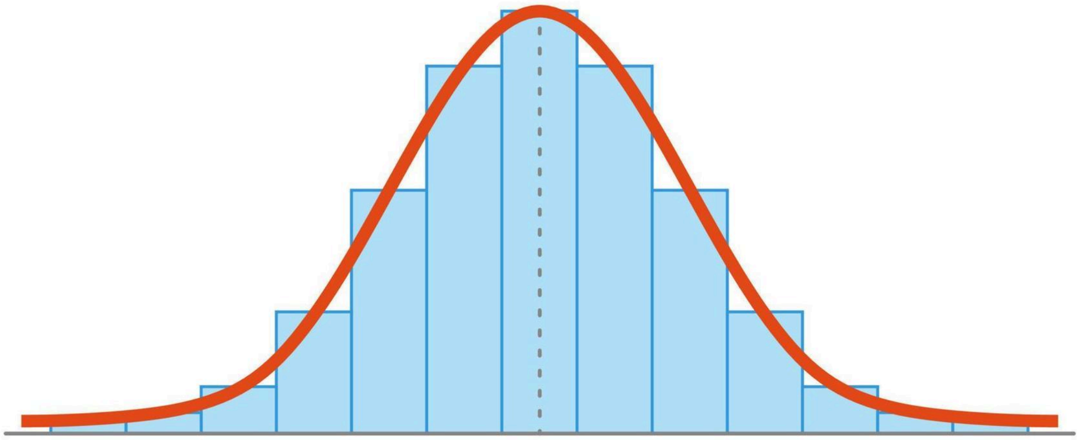
4. Un estudio en una universidad basado en el efecto de un hongo sobre un tipo de árbol, arrojó los siguientes resultados. Se plantaron 10 árboles en un invernadero y se transfirieron minerales. Una muestra fue tratada con nitrógeno y la otra sin él. Los pesos de los tallos se registraron al final de 140 días:

Sin nitrógeno	**Con nitrógeno**
0.32	0.26
0.53	0.43
0.28	0.47
0.37	0.49
0.47	0.52

0.43	0.75
0.36	0.79
0.42	0.86
0.38	0.62
0.43	0.46

a. Grafique los datos
b. Identifique las variables indicando su tipo
c. Utilice gráficos circulares para mostrar el porcentaje de cada medida
d. Calcule las medidas de tendencia central de cada muestra
e. Calcule las medidas de dispersión de cada muestra

Parte II. Estadística Inferencial



Capítulo 5. Probabilidad

Espacio muestral

Es el conjunto de todos los resultados posibles de un experimento estadístico. Se representa con el símbolo S. En la terminología de conjuntos, es el equivalente al conjunto universal del experimento.

Punto muestral

Es cada uno de los resultados de un espacio muestral, también conocido como **elemento** o **miembro** de éste.

Ejemplo
Si S es el espacio muestral de todos los posibles resultados cuando se lanza una moneda al aire es:

$S = \{C, T\}$, donde C corresponde a "cara" y T a "sello o cruz"

Ejemplo
Si S es el espacio muestral de todos los posibles resultados cuando se lanza una dado al aire y se observa la cara superior, tenemos que:

$S_1 = \{1, 2, 3, 4, 5, 6\}$, donde C corresponde a "cara" y T a "sello o cruz"

Si sólo interesa saber si el número es par o impar, se tiene que el espacio muestral es:

$S_2 = \{par, impar\}$

Nota
Interesan los espacios muestrales que más información aportan. Por ejemplo, en el ejemplo anterior, S_1 aporta más información que S_2, pues si sabemos qué elemento ocurre en S_1 podemos saber el resultado en S_2, sin embargo, al contrario, no es posible, esto es, si conocemos cuál elemento sucede en S_2, no podemos saber que elemento ocurre en en S_1.

Diagramas de árbol

Es una herramienta útil para listar los elementos del espacio muestral de un experimento estadístico de forma sistemática.

Ejemplo

Se lanza una moneda y después se lanza una segunda vez si sale cara. Si sale sello en el primer lanzamiento, se lanza un dado una vez. Listar los elementos del espacio muestral mediante un diagrama de árbol.

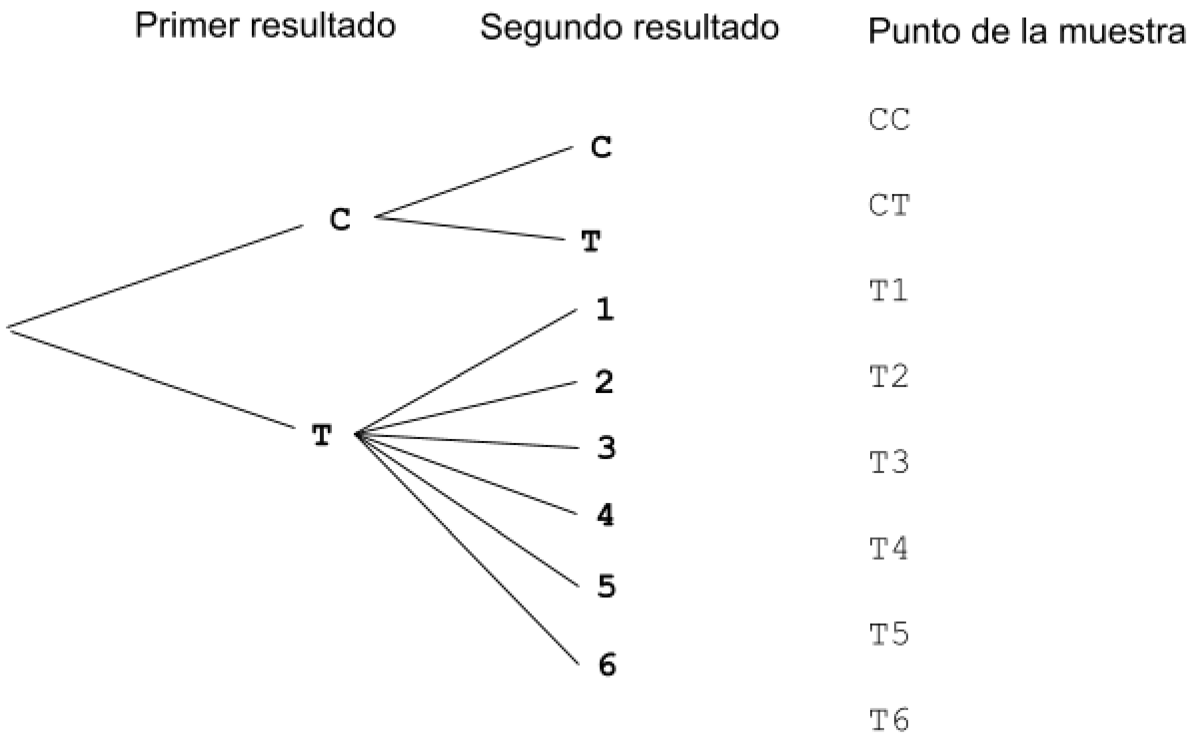


Figura. Diagrama de árbol para el lanzamiento de una moneda y un dado

Las trayectorias a través de las ramas muestran todos los puntos muestrales:

S = {CC, CT, T1, T2, T3, T4, T5, T6}

Ejemplo

Se seleccionan tres artículos de forma aleatoria de un proceso de fabricación. Los artículos se clasifican como “defectuoso” (D) y “no defectuoso” (N). Construir el diagrama de árbol del experimento que muestre las distintas posibilidades de sacar los tres artículos

El espacio muestral está dado por:

S = {DDD, DDN, DND, DNN, NDD, NDN, NND, NNN}

Para espacios muestrales grandes o infinitos es más conveniente expresar una **regla** o **enunciado**, describiendo sus elementos por **compresión** en lugar de hacerlo por **extensión**.

Ejemplo

El espacio muestral son todas las ciudades del mundo con una población superior a un millón de habitantes.

El espacio muestral se puede expresar así:

S = {x|x es una ciudad con más de 1000000 de habitantes}

Si el espacio muestral el el conjunto de todos los puntos sobre la frontera y el interior de una circunferencia de radio r y centro en el origen, entonces:
$S = \{(x, y) | x^2 + y^2 \leq r^2\}$

Evento

Es un subconjunto de un espacio muestral.

En muchos experimentos es posible que haya interés sólo en algunos eventos que ocurren. Por ejemplo, al lanzar un dado puede interesar el evento de obtener sólo número pares, esto es A = {2, 4, 6}, o estar interesados en saber cuando sale más de un artículo defectuoso, es decir B = {DDD, DDN, DND, NDD}.

Nota
Dado que el espacio muestral es un conjunto (estadístico) y un evento un subconjunto, podemos aplicar las distintas operaciones entre conjuntos matemáticos:

- **Unión $A \cup B$**: elementos comunes y no comunes de A y B
- **Intersección $A \cap B$**: elementos comunes a A y B. Si $A \cap B = \varnothing$ entonces significa que los eventos son **mutuamente excluyentes** o **disjuntos**.
- **Diferencia $A - B$**: elementos que están en A pero no en B
- **Complemento A'** : lo que le falta a A para ser igual a S

Conteo de puntos de la muestra

En ocasiones se desea evaluar la ocurrencia de ciertos eventos cuando se lleva a cabo un experimento y en muchos casos hay que ser capaces de resolver un problema mediante el conteo del número de puntos en el espacio muestral sin tener que listar cada elemento.

Regla de multiplicación

Teorema
Si una operación puede llevarse a cabo en n_1 formas, y si para cada una de éstas se puede realizar una segunda operación en n_2 formas, entonces las dos operaciones pueden ejecutarse en $n_1 n_2$ formas.

Ejemplo

¿Cuántos puntos muestrales hay en el espacio muestral cuando se lanza una vez un par de dados?

Solución

El primer dado puede caer de $n_1=6$ formas posibles. Para cada una de estas formas, el segundo dado también puede caer de $n_2=6$ formas posibles. Por tanto, ambos dados pueden caer de $n_1 n_2 = 6 \times 6 = 36$ maneras posibles.

La regla de multiplicación puede extenderse a k operaciones: $\prod_{i=1}^{k} n_i = n_1 n_2 n_3 \dots n_k$

Permutación

Es un arreglo de todo o parte de un conjunto de objetos.

Las permutaciones son útiles cuando se quiere conocer todas las posibles ordenaciones de la colección de objetos de la muestra. Por ejemplo, se puede querer conocer de cuántas formas se pueden ubicar los estudiantes en un salón de clase o sacar dos billetes de lotería de 20 disponibles. Estos casos son ejemplos de permutaciones.

Teorema

El número de permutaciones de n objetos distintos es $n!$.

Ejemplo

Se desea ordenar las letras x, y, z de todas las formas posibles: xyz, xzy, yxz, yzx, zxy, zyx, esto es hay $3! = 3 \times 2 \times 1 = 6$ formas de ordenar las tres letras, esto es, seis permutaciones.

Ejemplo

Se tienen las letras a, b, c, d. El número de arreglos posibles es $4! = 24$. Si se quiere saber el número de permutaciones posibles al tomar las cuatro letras, dos a la vez, tenemos: ab, ac, ad, ba, bc, bd, ca, cb, cd, da, db, dc. De aquí se ve que para la primera posición hay $n_1 = 4$ letras disponibles y para la segunda posición hay $n_2 = 3$ letras disponibles, por tanto hay: $n_1 n_2 = 4 \times 3 = 12$ permutaciones.

En general, n objetos distintos tomados de r a la vez se pueden arreglar en:

$n(n - 1)(n - 2) \dots (n - r + 1)$ formas posibles. Este producto se expresa de la siguiente forma en el teorema que se enuncia a continuación.

Teorema

El número de permutaciones de n objetos distintos tomados de r a la vez es

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

Ejemplo

Del ejemplo anterior se concluye que: $P^4_2 = \frac{4!}{(4-2)!} = \frac{4!}{2!} = \frac{24}{2} = 12$ formas posibles

Ejemplo
Se sacan dos billetes de lotería de 20 para un primer y segundo premio. Encontrar el número de puntos muestrales en el espacio S.

Solución
n = 20; r = 2; por tanto:
$P^n_r = \frac{n!}{(n-r)!} = P^{20}_2 = \frac{20!}{(20-2)!} = \frac{20\times19\times18!}{18!} = 20 \times 19 = 380$ que es el número de puntos de la muestra.

Teorema
El número de permutaciones de n objetos distintos arreglados en un círculo es $(n - 1)!$.

Teorema
El número de permutaciones distintas de n cosas de las que n_1 son de una clase, n_2 de una segunda clase, ..., n_k de una k-ésima clase es

$$\frac{n!}{n_1!n_2!n_3!..n_k!}$$

Ejemplo
¿De cuantas formas diferentes se pueden organizar 3 focos rojos, 4 amarillos y 2 azules en una serie navideña con 9 portalámparas?

El número total de arreglos es: $9! / (3! \times 4! \times 2!) = 1260$

Celdas

Con frecuencia interesa el número de formas de dividir un conjunto de n objetos en r subconjuntos que se denominan celdas. Se consigue una partición si la intersección de todo par posible de los r subconjuntos es el conjunto vacío y si la unión de todos los subconjuntos da el conjunto original. El orden de los elementos dentro de una celda no es importante.

Teorema
El número de formas de partir un conjunto de n objetos en r celdas con n_1 elementos en la primera celda, n_2 en la segunda, y así sucesivamente, es:

$$(n,\ (n_1, n_2, n_3,..., n_r)) = \frac{n!}{n_1!n_2!n_3!..n_r!},\ donde\ n_1 + n_2 + n_3 +...+ n_r = n$$

Ejemplo
Sea $V = \{a, e, i, o, u\}$ el conjunto de las vocales. Las particiones posibles en dos celdas en las que la primera celda contiene 4 elementos y la segunda uno, son:

{(a, e, i, o), (u)}, {(a, e, i, u), {o}}, {(a, e, o, u), (i)}, {(a, i, o, u), (e)}, {(e, i, o, u), (a)}

Vemos que hay 5 formas de partir el conjunto V, lo cual se obtiene aplicando el teorema anterior: (5, (4, 1)) = 5! / (4!1!) = 5.

Ejemplo
¿En cuántas formas se pueden asignar siete personas a una habitación triple y a dos dobles?

Solución
Son 7 personas y se deben crear las particiones 3, 2, 2, por tanto, el número total de particiones posibles es: (7, (3, 2, 2)) = 7!/(3!2!2!)= 210 y vemos que 3 + 2 + 2 = 7.

Ejemplo
Para organizar el comedor, se dispone de 4 manteles, 2 juegos de cubre manteles y 5 candelabros ¿De cuántas formas se puede organizar la mesa?

Solución
Son 11 cosas agrupadas en 3 clases diferentes:
$n = 11;\ n_1 = 4;\ n_2 = 2;\ n_3 = 5.$
El número de formas (arreglos) de organizar la mesa es 11!/(4!2!5!)= (11*10*9*8*7*6*5!)/(4!2!5!) = 6930.

Combinación

Es el número de formas de *seleccionar* r objetos de n sin importar el orden. Una combinación es realmente una partición con dos celdas, una celda contiene los r objetos seleccionados y la otra contiene (n - r) objetos restantes.

Teorema
El número de combinaciones de n objetos distintos tomados de r a la vez es

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

Ejemplo
De 4 químicos y 3 físicos encuentre el número de comités que se pueden formar que consistan de dos químicos y un físico.

Solución
Para seleccionar dos químicos de 4:
(4, 2) = 4!/(2!(4 - 2)!) = 4!/(2!2!) = 24/4 = 6

Para seleccionar un físico de 3:

$(3, 1) = 3!/(1!(3 - 1)!) = 3!/(1!2!) = 6/2 = 3$

Al aplicar la regla de la multiplicación para $n_1 = 6$, $n_2 = 3$, se tiene que podemos formar $n_1n_2 = 6 \times 3 = 18$ comités.

Ejemplo

Un grupo de 15 personas va a participar en distintas actividades: música, canto y baile ¿De cuántas formas se pueden organizar las personas si sólo pueden escoger dos actividades?

Solución

Tenemos una combinación. Las 15 personas solo pueden escoger dos de las tres actividades; es decir, cada persona escoge: música-canto, música-baile o canto-baile, o sea, es una combinación de 3 elementos tomados de a 2:

${}_3P_2 = C(3, 2) = 3!/(2!(3-2)!) = 3!/(2!1!) = 3*2!/(2!1!) = 3$ formas posibles para participar.

Ahora, cada persona puede escoger tres formas de participar y son 15, se tiene que:

$3 * 15 = 45$ formas de organizar a los participantes.

Ejemplo

En un concurso hay tres modalidades: música, canto y baile; los concursantes pueden escoger participar en una, dos o las tres modalidades ¿De cuántas formas puede participar una persona?

Solución

Tenemos 3 actividades: música (m), canto (c) y baile (b). Cada persona puede participar en una, dos o tres actividades:

Para participar en sólo una actividad se tiene que puede seleccionar 1 de 3:

$C(3, 1) = 3! / (1!(3-1)!) = 3*2! / 2! = 3$

Para participar en dos actividades se tiene que puede seleccionar 2 de 3:

$C(3, 2) = 3! / (2!(3-2)!) = 3! / 2! = 3$

Para participar en las tres actividades se tiene que puede seleccionar 3 de 3:

$C(3, 3) = 3! / (3!(3-3)!) = 3! / (3!0!) = 1$

Ahora, sumando las combinaciones, tenemos que cada persona puede participar de $3 + 3 + 1 = 7$ formas diferentes.

Observe que una persona puede escoger las opciones: m, mc, mb, mcb, c, cb, b, cuyo número corresponde con las 7 formas de que dispone una persona.

Probabilidad de un evento

Se evalúa por medio de un conjunto de números reales denominados **pesos** o **probabilidades** que van de 0 a 1. Para cada punto del espacio muestral se asigna una probabilidad de tal forma que la suma de todas las probabilidades es 1.

Probabilidad de un evento A

La probabilidad de un evento A es la suma de los pesos de todos los puntos muestrales en A, por tanto:

$0 \leq P(A) \leq 1$; $P(\varnothing) = 0$ y $P(S) = 1$

Ejemplo
Se lanza una moneda dos veces. ¿Cuál es la probabilidad de que ocurra al menos una cara?

Solución
El espacio muestral del experimento es:
S = {CC, CT, TC, TT}
Se asume que la moneda no está cargada y asignamos por tanto una probabilidad de w a cada punto muestral y como la suma de todos los pesos debe ser uno, tenemos:
$4w = 1$; $w = 1/4$
Si A es el evento de que al menos caiga una cara, se tiene que:
A = {CC, CT, TC}
$P(A) = 1/4 + 1/4 + 1/4 = 3/4$

Ejemplo
Se carga un dado de tal forma que sea el doble de probable de obtener un número par que uno non. Si E es el evento de que ocurra un número menor a 4 en un solo lanzamiento, encuentre P(E)

Solución
S = {1, 2, 3, 4, 5, 6}; E = {1, 2, 3}
Sea w la probabilidad de cada impar, por tanto $2w$ es la probabilidad de cada par
$P(S) = 1 = w + 2w + w + 2w + w + 2w \Rightarrow 1 = 9w \Rightarrow w = 1/9$
$P(E) = 1/9 + 2/9 + 1/9 = 4/9$

Ejemplo
Supongamos que en el ejemplo anterior se considera el evento A de que salga un número par y otro evento B de que salga un número divisible por 3. Encontrar la probabilidad de que salga un número par o divisible por 3 y la probabilidad de que salga un número par divisible por 3

Solución
Escribamos los eventos A y B
A = {2, 4, 6}; B = {3, 6}

Para encontrar la probabilidad de que salga un número par o divisible por 3 debemos encontrar $P(A \cup B)$.

$A \cup B = \{2, 3, 4, 6\}$

$P(A \cup B) = 2/9 + 1/9 + 2/9 + 2/9 = 7/9$

Para encontrar la probabilidad de que salga un número par y divisible por 3 debemos encontrar $P(A \cap B)$.

$A \cap B = \{6\}$

$P(A \cap B) = 2/9$

Teorema

Si un experimento puede tener como resultado cualquiera de N diferentes resultados igualmente probables, y si exactamente n de estos resultados corresponden al evento A, entonces la probabilidad del evento A es:

$$P(A) = n/N$$

Bibliografía

Walpole, Myers. Probabilidad y Estadística para Ingenieros. Pearson Educación. Sexta Edición. México, 1999.

Historia de la estadística: Qué es, sus etapas y evolución

Historia de la estadística: desde sus orígenes hasta la actualidad

Historia de la estadística - Wikipedia, la enciclopedia libre

Breve Historia de la Estadistica I - John Graunt (1620-1674) Breve historia del desarrollo de la - Studocu.

Historia de la Estadística

Población y muestra

Reseña histórica breve historia de la estadística | PDF | Historic Site and Landmark Tours

Antecedentes históricos de la estadística y algunos conceptos básicos..

3 Distribución de frecuencias

Distribución de Frecuencias - Qué es y cómo se utiliza

UNIDAD II MEDIDAS DE TENDENCIA CENTRAL

Rango (estadística) - Wikipedia, la enciclopedia libre

Varianza - Wikipedia, la enciclopedia libre

What Is Variance in Statistics? Definition, Formula, and Example