UNIVERSIDAD AUTÓNOMA DE CIUDAD JUÁREZ

ESTADÍSTICA DESCRIPTIVA

Principios básicos de estudio

Mateo Fabián Itza Ortiz
Angélica María Escárcega Ávila
José María Carrera Chávez
Ernesto Orozco Lucero
Diana Marcela Beristain Ruiz

UNIVERSIDAD AUTÓNOMA DE CIUDAD JUÁREZ

Juan Ignacio Camargo Nassar
Rector

Daniel Constandse Cortez
Secretario General

Salvador David Nava Martínez
Director del Instituto de Ciencias Biomédicas

Jesús Meza Vega
Director General de Comunicación Universitaria

Universidad Autónoma de Ciudad Juárez

ESTADÍSTICA DESCRIPTIVA

Principios básicos de estudio

MATEO FABIÁN ITZA ORTIZ
ANGÉLICA MARÍA ESCÁRCEGA ÁVILA
JOSÉ MARÍA CARRERA CHÁVEZ
ERNESTO OROZCO LUCERO
DIANA MARCELA BERISTAIN RUIZ

Avenida Plutarco Elías Calles 1210, Fovissste Chamizal, C. P. 32310
Ciudad Juárez, Chihuahua, México
Tels. +52 (656) 688 2100 al 09

Primera edición, 2024
Disponible en: elibros.uacj.mx

ESTE LIBRO FUE DICTAMINADO POR PARES ACADÉMICOS MEDIANTE EL SISTEMA DOBLE CIEGO

Estadística descriptiva: principios básicos de estudio / Mateo Fabián Itza Ortiz, Angélica María Escárcega Avila, José María Carrera Chávez, Ernesto Orozco Lucero, Diana Marcela Beristain Ruiz.– Primera edición – Ciudad Juárez, Chihuahua, México: Universidad Autónoma de Ciudad Juárez, 2024. 137 páginas; 22 centímetros.

ISBN: 978-607-520-512-0

Contenido: Prefacio.— Historia de la Estadística.— Método estadístico.— Funciones o Teoría de Conjuntos.— Sistema de coordenadas o plano cartesiano.— Gráficas.— Distribución de frecuencias.— Medidas de centralización o de posición.— Medidas de tendencia central de datos no agrupados.— Medidas de dispersión de datos no agrupados.— Medidas de posición no central de datos no agrupados.— Medidas de tendencia central de datos agrupados.— Medidas de dispersión de datos agrupados.— Coeficiente de Curtosis.— Medidas de posición no central de datos agrupados.— Distribución binomial (n, p).— Distribución normal (μ, σ).— Distribución normal reducida (tipificada).— Referencias.— Anexo.

1. Estadística descriptiva — Problemas, ejercicios, etc.
2. Estadística descriptiva — Ciencias de la salud — Problemas, ejercicios, etc.

LC – QA276.12 E77 2024

La edición, diseño y producción editorial de este documento estuvieron a cargo de la Dirección General de Comunicación Universitaria, a través de la Subdirección de Editorial y Publicaciones de la Universidad Autónoma de Ciudad Juárez.

Coordinación editorial:
Mayola Renova González
Cuidado editorial:
Subdirección de Editorial y Publicaciones
Diseño de portada y diagramación:
Karla María Rascón González

Índice

Prefacio

La Estadística es una herramienta que usamos de manera cotidiana en nuestra vida. Solamente al tomar un curso de Estadística podemos entender y empezar a diferenciar los conceptos básicos de esta disciplina científica y, desde luego, su importancia, uso y aplicación, que van desde describir eventos hasta la inferencia.

El presente libro fue diseñado para encaminar al alumno de Ciencias Veterinarias o Ciencias de la Salud en el estudio de la *Estadística Descriptiva* con conceptos y principios básicos, así como ejercicios relacionados con esta disciplina.

Los autores ofrecen al lector de manera práctica y sencilla un contenido que requiere para aprender esta división de la Estadística. Además, en este material se exponen historia, conceptos y procedimientos, así como algunos errores básicos cometidos al momento de analizar datos estadísticos. Esta edición complementa y detalla temas agregados a la carrera de Medicina Veterinaria y Zootecnia, que seguramente fortalecerán el desempeño académico del estudiante.

Asimismo, se han agregado conceptos y temas relacionados con la *Estadística Descriptiva*, tanto de datos agrupados como no agrupados; además, la obra puede ser considerada como indispensable en su biblioteca personal para consulta o enseñanza.

Mateo Fabián Itza Ortiz

Historia de la Estadística

Desde los comienzos de la civilización han existido formas sencillas de Estadística. El *homo sapiens* utilizaba representaciones gráficas y otros símbolos en pieles, rocas, palos de madera y paredes de cuevas, para contar el número de personas, animales o ciertas cosas. Hacia el año 3000 a. C. los babilonios usaban ya pequeñas tablillas de arcilla, para recopilar datos sobre la producción agrícola y de los géneros vendidos o cambiados mediante trueque. Los egipcios analizaban los datos de la población y la renta del país mucho antes de construir las pirámides en el siglo XXXI a. C. Los libros bíblicos de Números y Crónicas incluyen, en algunas partes, trabajos de Estadística. El primero contiene dos censos de la población de Israel y el segundo describe el bienestar material de las diversas tribus judías. En China existían registros numéricos similares con anterioridad al año 2000 a. C. Los griegos clásicos realizaban censos cuya información se utilizaba hacia el año 594 a. C. para cobrar impuestos.

El Imperio Romano fue el primer gobierno que recopiló una gran cantidad de datos sobre la población, superficie y renta de todos los territorios bajo su control. Durante la Edad Media solo se llevaron a cabo algunos censos exhaustivos en Europa. Los reyes carolingios Pipino el Breve y Carlomagno ordenaron hacer estudios minuciosos de las propiedades de la Iglesia en los años 758 y 762, respectivamente.

Después de la conquista normanda de Inglaterra en 1066, el rey Guillermo I de Inglaterra encargó un censo.

"La palabra estadística a menudo nos trae a la mente imágenes de números apilados en grandes arreglos y tablas, de volúmenes de cifras relativas a nacimientos, muertes, impuestos, poblaciones, ingresos, deudas, créditos y así sucesivamente" (Huntsberger y Billingsley, 1983). La Estadística es mucho más que solo números apilados y gráficas: es una ciencia con tanta antigüedad como la escritura y es por sí misma auxiliar de otras disciplinas científicas, por ejemplo, en estudios de mercados, medicina, ingeniería, física, etcétera. La falta de un estudio estadístico conllevaría a un "caos generalizado", es decir, dejaría a los estudios sin información vital a la hora de tomar decisiones.

La Estadística que conocemos hoy en día debe gran parte de su realización a los trabajos matemáticos de hombres que desarrollaron la Teoría de las Probabilidades, con la cual se adhirió la Estadística a las ciencias formales.

La historia de la Estadística está resumida en tres grandes etapas o fases:

1. **Censos:** desde el momento en el que se constituye una autoridad política la idea de inventariar de una forma, más o menos regular, la población y las riquezas existentes en el territorio está ligada a la conciencia de soberanía y los primeros esfuerzos administrativos.
2. **De la descripción de los conjuntos a la aritmética política:** las ideas mercantilistas extrañan una intensificación de este tipo de investigación.
3. **Estadística y cálculo de probabilidades:** el cálculo de probabilidades se incorpora rápidamente como un instrumento de análisis extremadamente poderoso para el estudio de los fenómenos económicos y sociales, y en general para el estudio de fenómenos "cuyas causas son demasiado complejas para conocerlos totalmente y hacer posible su análisis" (Ruiz, 2004).

Método estadístico

El conjunto de métodos que se emplean para medir las características de la información, resumir los valores individuales y analizar los datos, a fin de extraer el máximo de información, se llama métodos estadísticos.

Los métodos de análisis para la información cuantitativa se pueden dividir en seis pasos:

1. Definición del problema;
2. Recopilación de la información existente;
3. Obtención de la información original;
4. Clasificación;
5. Presentación; y
6. Análisis.

Utilidad e importancia del método estadístico

Los métodos estadísticos tradicionalmente se utilizan para propósitos descriptivos, así como para organizar y resumir datos numéricos. Las técnicas estadísticas se aplican de manera amplia en mercadotecnia, contabilidad, control de calidad y otras actividades; estudios de consumidores; análisis de resultados en deportes; administradores de instituciones; educación; organismos políticos o médicos; y por otras personas que intervienen en la toma de decisiones.

El proceso seguido en el estudio estadístico de cierta característica o variable puede subdividirse en tres pasos sucesivos:

a) **Recopilación de los datos:** una vez planteada la prueba o encuesta oportuna y recogidos los datos que correspondan, el primer análisis que efectuaremos es el del tipo de variable que pretendemos estudiar (cualitativa o cuantitativa; discreta o continua). Esto condicionará en gran medida su posterior tratamiento.

b) **Organización de los datos:** una vez determinado el modo de agrupamiento de las observaciones procederemos a su recuento, construyendo una tabla de frecuencias. Posteriormente podremos visualizar tales frecuencias de forma gráfica.

c) **Análisis final:** la obtención de muy diversas conclusiones respecto de la variable estudiada, se podrá realizar con auxilio de los diferentes parámetros estadísticos (centralización, posición, dispersión, etcétera).

Errores estadísticos comunes

En el momento de recopilar los datos que serán procesados se es susceptible de cometer errores, así como durante el cómputo. No obstante, hay otros errores que no tienen nada que ver con la digitación y que no son tan fácilmente identificables. Algunos de estos son:

» *Sesgo*: es imposible ser completamente objetivo o no tener ideas preconcebidas antes de comenzar a estudiar un problema, pues existen muchas maneras en las que una perspectiva o estado mental pueden influir en la recopilación y el análisis de la información. En estos casos se dice que hay un sesgo cuando el individuo da mayor peso a los datos que apoyan su opinión que a aquellos que la contradicen. Un caso extremo de sesgo sería la situación donde primero se toma una decisión y después se utiliza el análisis estadístico para justificar la decisión ya tomada.

- *Datos no comparables*: el establecer comparaciones es una de las partes más importantes del análisis estadístico, pero es extremadamente importante que tales comparaciones se hagan entre datos que sean comparables.
- *Proyección descuidada de tendencias*: la proyección simplista de tendencias pasadas hacia el futuro es uno de los errores que más ha desacreditado el uso del análisis estadístico.
- *Muestreo incorrecto*: en la mayoría de los estudios sucede que el volumen de la información disponible es tan inmenso (N = Tamaño de la población) que se hace necesario estudiar muestras (*n* = Tamaño de la muestra), para derivar conclusiones acerca de la población a la que pertenece la muestra. Existen metodologías para calcular el tamaño de la muestra (*n*) que podemos consultar en libros de Estadística Inferencial. Si la muestra (*n*) se calcula correctamente tendrá básicamente las mismas propiedades que la población (N) de la cual fue extraída; pero si el tamaño del muestreo se calcula incorrectamente entonces puede suceder que los resultados no signifiquen nada.

Definición de Estadística

Podemos encontrar definiciones amplias del concepto de Estadística y todas encierran un concepto básico de obtención, orden, análisis y presentación de un conjunto de datos. Así que podemos definir la estadística como una herramienta de las matemáticas cuyo objetivo es reunir información cuantitativa o cualitativa de forma científica de individuos, grupos, hechos, etcétera, que son ordenados, procesados y analizados como un conjunto de datos, con el fin de obtener predicciones y explicaciones sobre fenómenos observados con conclusiones relevantes conocidas como significativas. También, se puede decir que la estadística es la ciencia de los datos y que su *objetivo* principal es mejorar la comprensión de los hechos a partir de la información disponible.

Bajo el término *Estadística Descriptiva*, se engloban las técnicas que nos permitirán realizar un análisis elemental de las observaciones experimentales observadas.

La Estadística Descriptiva se subdivide en dos bloques:

1. *Estadística primaria*: se obtiene de un grupo de observaciones experimentales. Este apartado nos enseña a ordenar adecuadamente, de modo que se ofrezca una información lo más clara posible.
2. *Estadística derivada o secundaria*: con los datos "observados" se realizan cálculos, obteniendo así medidas conocidas como medidas de tendencia central o de dispersión.

División de la Estadística

La Estadística se ocupa de reunir, organizar y analizar datos numéricos (también datos categóricos), y ayuda a resolver problemas con el diseño de experimentos y toma de decisiones; entonces, la estadística se divide para su estudio en dos grandes categorías:

1. Estadística Descriptiva; y
2. Estadística Inferencial.

- *Estadística Descriptiva*: consiste en la presentación de datos (cuantitativos o cualitativos) en forma gráfica o ilustrativa (cuadros o tablas) y el cálculo de medidas descriptivas. Comprende cualquier actividad relacionada con los datos y está diseñada para resumir o describirlos sin factores pertinentes adicionales; esto es, sin intentar inferir nada que vaya más allá de los datos como tales.
- *Estadística Inferencial*: se deriva de muestras u observaciones hechas acerca de una parte de un conjunto numeroso de elementos y esto implica que su análisis requiera de generalizaciones que van más allá de los datos. La Estadística Inferencial investiga o analiza una población partiendo de una muestra de la población (*n*) tomada.

Definición de variable

Una variable estadística es cada una de las características o cualidades que poseen los elementos de una población, que pueden tomar *valores* que varían de individuos a individuos o de grupos a grupos. Una definición menos ortodoxa define a la variable como un carácter alfanumérico, que puede tomar cualquier *valor* que puede fluctuar y cuya variación es susceptible de adoptar diferentes valores que pueden medirse u observarse. Por ejemplo, la variable es 2a y su valor puede ser cualquier valor numérico: 2a = 7; A = 4; b = 8. Se tiene que precisar que casi siempre es un valor numérico y aunque algunas veces se observan y describen (variable cualitativa), estas son consideradas o conocidas como *campos*.

Tipos de variables estadísticas

» *Población y muestra*: se llama *población* al conjunto formado por todos los elementos cuyo conocimiento nos interesa. En la estadística el tamaño de la población se simboliza con la letra mayúscula N y a cada uno de los elementos, se les llama individuos o unidad experimental cuando se trata de un diseño. Es importante mencionar que un grupo de individuos también puede ser llamado unidad experimental. Generalmente, en la estadística no se emplea toda la población, debido a que representaría gastos o procedimientos cuantiosos innecesarios con el empleo del 100% de los individuos; para ello, se calcula un "tamaño representativo" de dicha población, lo cual es menos costoso. El tamaño de la *muestra* se simboliza con la letra minúscula *n* y es un subconjunto, individuo o unidad experimental tomado de N, es decir, es un número "limitado" que fue extraído de una población con el objetivo de reducir el número de individuos participantes y obtener una probabilidad o significancia estadísticamente confiable. La probabilidad o significancia obtenida será estadísticamente la misma que la de toda la población, pero con un menor número de individuos calculado previamente con base en la metodología del tamaño de la muestra (*n*) para una población finita o infinita.

Es importante mencionar que con base en lo anterior las variables que hacen referencia a la población, se representan con una letra mayúscula y las correspondientes a una muestra, con una letra minúscula.

- *Variables*: para el conocimiento de una población estadística, se debe analizar a cada uno de sus individuos como población, o a su muestra. Este análisis no puede ser exhaustivo, pues se deberá seleccionar una característica —o varias— y ver cómo esta se manifiesta en cada uno de los individuos. Por ejemplo, de una población de gallinas de traspatio las características de estudio pueden ser: color, tipo de cresta, longitud del tarso, longitud del pico, peso corporal, entre otras.

Entonces se pueden distinguir básicamente dos tipos o categorías de variables: las que poseen una cualidad (cualitativas) o aquellas que se pueden cuantificar (cuantitativas), y que, a su vez, se subdividen en grupos que poseen características específicas.

Existen diferentes tipos de variables (Figura 1).

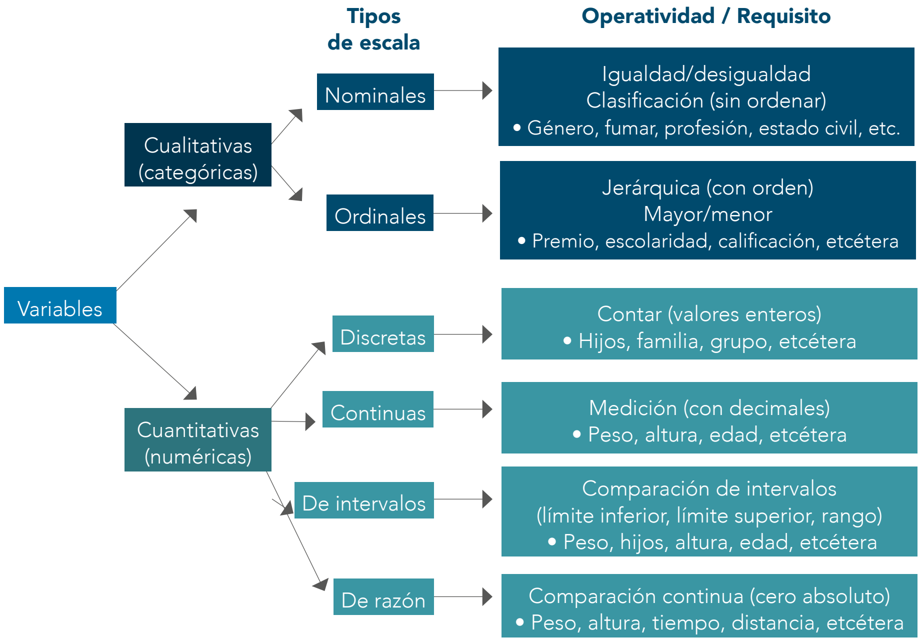


Figura 1. Tipos de variables

Fuente: elaboración propia.

a) Variable cualitativa: se refiere a características o cualidades que no pueden ser medidas con números. Podemos distinguir dos tipos:

Variable cualitativa nominal: presenta modalidades no numéricas que no admiten un criterio de orden y que utilizan escalas sin valor cuantitativo; son esencialmente clasificadas como "nombres". Por ejemplo, la variable nominal sexo (hembra/macho) se clasifica como "dicotómica" (tiene dos respuestas), y la variable nominal estado civil posee las siguientes modalidades: soltero, casado, separado, divorciado y viudo, conocida como "nominal con orden".
Variable cualitativa ordinal o variable cuasicuantitativa: presenta modalidades no-numéricas, en las que existe un orden que comúnmente son proporciones de ideas no numéricas.

Por ejemplo, la nota en un examen: aprobado, notable, sobresaliente; el puesto conseguido en una prueba deportiva: 1.º, 2.º, 3.º...; las medallas de una prueba deportiva: oro, plata y bronce.

b) **Variable cuantitativa:** es la que se expresa mediante un número; por tanto, se pueden realizar operaciones aritméticas con ella. Podemos distinguir cuatro tipos:

Variable discreta: es aquella que toma valores aislados, es decir, no admite valores intermedios entre dos valores específicos. Por ejemplo, el número de lechones nacidos vivos: 13, 12, 11, 10,…

Variable continua: es aquella que puede tomar valores comprendidos entre dos números (decimales o fracciones). Por ejemplo, la altura en metros de 5 caballos: 1.73, 1.82, 1.77, 1.69 y 1.75. Se recomienda el uso de dos decimales, pero también se podrían emplear tres decimales.

Variable de intervalos (puede ser discreta o continua): son escalas numéricas que se representan en un intervalo (por ejemplo, tiempo, edad, peso corporal, etcétera), en las que se conoce tanto el orden como las cualidades. Un intervalo se caracteriza por un límite inicial, un límite final y un rango, es decir, es una escala de medición cuantitativa que mide la diferencia entre dos variables. Por ejemplo, una escala de intervalos de la temperatura entre 50 y 60 grados tiene un rango de 10 grados.

Variable de razón (puede ser discreta o continua): es el tipo de datos cuantitativos que se caracterizan por un punto de cero absoluto, lo que significa que no hay ningún valor numérico negativo, es decir, existe un valor o es inexistente (cero). Los números se comparan en múltiplos de uno. Por ejemplo, estatura, peso, hijos, baños, etcétera.

Definición entre signo y símbolo en matemáticas

Antes de iniciar el tema se considera conveniente dejar en claro la diferencia entre **signo**, que proviene del latín *signum* y trata de un objeto, fenómeno o acción materia que representa o sustituye a otro; también representa una operación o posición de un valor en una escala numérica que puede ser positivo o negativo. Mientras que un **símbolo** proviene del latín *simbŏlum* y es la representación sensorial de una idea, concepto o tema que guarda un vínculo convencional y arbitrario con un objeto. Ambos conceptos son vistos desde el término matemático con el propósito de referirnos a uno u otro de manera correcta.

Tenemos, entonces, que un signo indica la precisión de la medida de una magnitud; por ejemplo, 738 ± 1 mm quiere decir la longitud comprendida entre 737 y 739 mm (símbolo ±). En física, el signo indica el carácter positivo de una cantidad, como la carga eléctrica (símbolo +).

Un símbolo de la Estadística Descriptiva es un signo abreviado que permite identificar un proceso de esta, sin necesidad de utilizar su nombre completo. Por ejemplo, el símbolo de la sumatoria de todos los datos es: $\sum$; el símbolo de la media es: $\bar{X}$ (una “X” testada); el símbolo de la desviación estándar de una población es: σ (sigma); y el símbolo de la varianza es: σ^2 (sigma al cuadrado).

Signos matemáticos

Son figuras, señales y abreviaturas utilizadas en matemáticas para denotar entidades, relaciones y operaciones.

Por ejemplo, la suma (+) o la resta (–); la suma es el recíproco de la resta y viceversa. La multiplicación [•, x, *, ( )] o la división (/, ÷); la multiplicación es el recíproco de la división y viceversa. La potencia (^, x^y); y el radical ($\sqrt{\quad}$); la potencia es el recíproco de la raíz y viceversa.

Reglas de los signos

Básicamente las reglas de los signos se agrupan de la siguiente manera:

Reglas de los signos para la suma y la resta

1. Si los números tienen el mismo signo, se suman los valores absolutos y al resultado se le coloca el signo común:

$$3 + 5 = +8; (-3) + (-5) = -8.$$

2. Si los números tienen distinto signo, se restan los valores absolutos (al mayor se le resta el menor) y al resultado se le coloca el signo del número con el mayor valor absoluto:

$$-3 + 5 = +2; 3 + (-5) = -2.$$

Reglas de los signos para la multiplicación y la división

1. Al multiplicar números con signos diferentes, se obtienen números con signo negativo:

$$(-)(+) = (-);$$
$$(+)(-) = (-).$$

   Así, $(+2)(-4) = -8$, porque se está multiplicando dos veces el -4. Lo mismo sucederá si se pone primero el número negativo y luego el positivo:

$$(-4)(+2) = -8.$$

2. Al multiplicar números con signos iguales, se obtienen números con signo positivo:

$$(+)(+) = (+);$$
$$(-)(-) = (+).$$

   Esto se explica al recordar que todo número multiplicado por la unidad da el mismo número. Si la unidad fuera negativa, habría que cambiar el signo del número que se multiplica: (-1)(-2) = +2. También, si se multiplica un número positivo por otro positivo, se tendrá otro número positivo:

$$(+1)(+2) = +2.$$

Las leyes de los signos para operaciones matemáticas, se sintetizan en la Figura 2.

Signos para multiplicar	(+)	(–)
(+)	(+)	(–)
(–)	(–)	(+)

(+)(+) = (+)
(+)(–) = (–)
(–)(+) = (–)
(–)(–) = (+)

Figura 2. Leyes de los signos para operaciones matemáticas

Fuente: elaboración propia.

A continuación, se puede observar cómo se aplican las leyes de los signos para la multiplicación (Figura 3).

El producto de signos contrarios da un signo negativo	El producto de signos iguales da un signo positivo
Ejemplos:	*Ejemplos*:
(+3)(-2) = **-6** (-3)(+2) = **-6** (+4)(-1) = **-4** (-12)(+2) = **-24** (-6)(+3) = **-18** (-12)(0) = **0**	(+3)(+2) = **+6** (-3)(-2) = **+6** (+4)(+1) = **+4** (-12)(-2) = **+24** (-6)(-3) = **+18** (-12)(0) = **0**

Figura 3. Ejemplos numéricos de las leyes de los signos para la multiplicación

Fuente: elaboración propia.

Las reglas que se obtuvieron para la multiplicación se utilizan perfectamente en el caso de la división, es decir, la regla de la multiplicación es igual a la de la división, como se observa a continuación (Figura 4).

La división de signos iguales da un signo positivo	La división de signos diferentes da un signo negativo
$\frac{(+)}{(+)} = (+)$ $\quad$ $\frac{(-)}{(-)} = (+)$	$\frac{(+)}{(-)} = (-)$ $\quad$ $\frac{(-)}{(+)} = (-)$
Ejemplos: $\frac{(+2)}{(+1)} = \mathbf{+2}$ $\frac{(-2)}{(-1)} = \mathbf{+2}$	*Ejemplos*: $\frac{(+2)}{(-1)} = \mathbf{-2}$ $\frac{(-2)}{(+1)} = \mathbf{-2}$

Figura 4. Ejemplos numéricos de las leyes de los signos para la división

Fuente: elaboración propia.

Observe que, en la división, al cambiar de lugar al divisor y al dividendo, se modifica el resultado de la división, pero no los signos (Figura 5).

Ejemplos:

$\frac{(+1)}{(+2)} = \mathbf{+0.5}$ $\frac{(+2)}{(+1)} = \mathbf{+2}$	$\frac{(+1)}{(-2)} = \mathbf{-0.5}$ $\frac{(-2)}{(+1)} = \mathbf{-2}$
$\frac{(-1)}{(-2)} = \mathbf{+0.5}$ $\frac{(-2)}{(-1)} = \mathbf{+2}$	$\frac{(-1)}{(+2)} = \mathbf{-0.5}$ $\frac{(+2)}{(-1)} = \mathbf{-2}$

Figura 5. Ejemplos numéricos de las leyes de los signos para la división

Fuente: elaboración propia.

Recuerde que: la división de signos iguales da un signo positivo, y la división de signos diferentes da un signo negativo.

Ejemplos:

$\frac{(-4)}{(-2)} = \mathbf{+2}$	$\frac{(-2)}{(-4)} = +\frac{(2)}{(4)} = \frac{(1)}{(2)} = \mathbf{+0.5}$
$\frac{(-4)}{(+2)} = \mathbf{-2}$	$\frac{(+12)}{(-6)} = \mathbf{-2}$

Nota: es importante señalar que los signo + no se escriben, pues se entiende que, por naturaleza, los números son positivos; en otras palabras, tienen un signo positivo (+).

Ejercicio:

Las operaciones con signos pueden estar combinadas, como se muestra a continuación:

$$(-2) + (-3) - (-2) - (-1) - (-3) = \mathbf{?}$$

Para resolver esta operación, primero se deben quitar todos los paréntesis:

$$-2 -3 + 2 + 1 + 3 = \mathbf{?}$$

Se suman todos los números con el mismo signo, y conservan su signo original.

$$\frac{-2\ -3}{-5} = \mathbf{?} \qquad \frac{(+2\ +1\ +3)}{+6} = \mathbf{?}$$

Se resuelve la operación:

$$- 5 + 6 = \mathbf{?}$$
$$+ 10 - 10 = \mathbf{?}$$

Se suman los números positivos y negativos poniendo su signo. Se ejecuta la operación con el signo del número mayor. Como el "0" no tiene ningún valor **no se pone ningún signo**.

Observe que las operaciones de los signos en este ejemplo, se realizaron en un solo paso de la siguiente manera:

$$-(+\ 4) \cdot (+\ 2) = -8$$

(Menos) por (Más) = Menos, y este (Menos) por (Más) = Menos

$$-(-\ 7) \cdot (+\ 4) = +28$$

(Menos) por (Menos) = (Más), y este (Más) por (Más) = Más

Ejercicio:
Ponga los signos que faltan en las siguientes operaciones:

1. $(-3) + (-3) + \frac{(-6)}{(-2)} - (+3) \bullet (-2) =$ **?**
2. $(+18) - \frac{(-6)}{(-2)} - (+13) =$ **?**
3. $(3) + (3) + (3) - (6) =$ **?**
4. $(-8) \bullet (-3) + (-12) \bullet (-2) + (-2) =$ **?**
5. $(24) + (24) + (2) =$ **?**
6. $2 \bullet 5 =$ **?**
7. $(-2) \bullet (-5) =$ **?**
8. $2 \bullet (-5) =$ **?**
9. $(-2) \bullet 5 =$ **?**
10. $10 \div 5 =$ **?**
11. $(-10) \div (-5) =$ **?**
12. $10 \div (-5) =$ **?**
13. $(-10) \div 5 =$ **?**

Reglas de los signos de una potencia

1. Las potencias de un exponente par son siempre positivas:

$$(+)^{par} = + \qquad 2^6 = 64$$

$$(-)^{par} = + \qquad (-2)^6 = 64$$

2. Las potencias de un exponente impar tienen el mismo signo de la base:

$$(+)^{impar} = + \qquad 2^3 = 8$$

$$(-)^{impar} = - \qquad (-2)^3 = -8$$

Signo de una radicación

La radicación es la operación que consiste en multiplicar el mismo número por sí mismo las veces que resulte otro número determinado.

Por ejemplo: $\sqrt[3]{27} = 3 \longrightarrow 3 \times 3 \times 3 = 27.$

La radicación está compuesta por diferentes términos que en cualquier raíz son conocidos como: radical, subradical, índice y raíz (Figura 6).

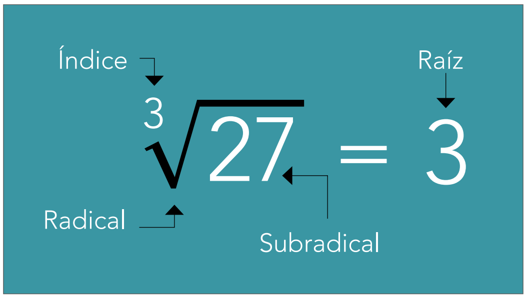


Figura 6. Términos de los componentes de la radicación

Fuente: elaboración propia.

En el caso de la radicación si el índice es impar, el resultado tendrá el mismo signo que la cantidad subradical. Si el índice es par y la cantidad subradical es positiva, el resultado tendrá doble signo: positivo y negativo; si la cantidad subradical es negativa, el resultado será una cantidad imaginaria, es decir, no existe en el campo real.

Reglas de los signos de una radicación

Toda raíz de índice impar de una cantidad subradical positiva, es siempre positiva:

$\sqrt[\text{impar}]{+} = +$

Toda raíz de índice impar de una cantidad subradical negativa, es siempre negativa:

impar $\sqrt{-} = -$

Toda raíz de índice par de una cantidad subradical positiva, siempre tiene doble signo:

par $\sqrt{+} = \pm$

Todas las raíces de índice par y cantidad subradical negativa son cantidades imaginarias, debido a que las raíces no son reales porque toda cantidad real elevada a exponente par es siempre positiva:

par $\sqrt{-}$ = Cantidad imaginaria: $-6^2 = 36$.

Ejercicio:

Resuelva las siguientes operaciones:

1. (2 x 3)(9 x -4) = **?**
2. (-5)(3) = **?**
3. -10 / 2 = **?**
4. 5 + (-4) = **?**
5. -5 + (-4) = **?**
6. (2 x 3 x 9)5 = **?**
7. -20 / -5 = **?**
8. 10 / -2 = **?**
9. (-5)(-3) = **?**
10. (5)(-3) = **?**

Símbolos en matemáticas

El origen y la evolución de los símbolos matemáticos no se conocen bien. El origen del cero es desconocido, aunque hay confirmación de su existencia antes del año 400 d. C. La extensión del sistema de lugares decimales que representan valores inferiores a la unidad, se atribuye al matemático holandés Simon Stevin (conocido también como Simón de Brujas), que llamó a las décimas, centésimas y milésimas, primas, secundas y tercias.

Antes de 1492 se empezó a utilizar un punto para separar la parte decimal de un número. Más tarde se empleó también una raya vertical. En su *Exempelbüchlein* de 1530, el matemático polaco Christoff Rudolff resolvía un problema de interés compuesto haciendo uso de fracciones decimales. El astrónomo alemán Johannes Kepler empezó a utilizar la coma para separar los espacios decimales y el matemático suizo Justus Byrgius usaba fracciones decimales de la forma 3.2.

A pesar de que los antiguos egipcios tenían símbolos para la adición y la igualdad, y los griegos, hindúes y árabes tenían símbolos para la igualdad y las incógnitas, en esos primeros tiempos las operaciones matemáticas solían ser bastante engorrosas, debido a la falta de signos apropiados. Las expresiones de dichas operaciones tenían que ser escritas por completo o mediante abreviaturas de las palabras. Más tarde, los griegos, los hindúes y el matemático alemán Jordanus Nemorarius empezaron a indicar la suma mediante yuxtaposición, mientras que los italianos la denotaban con las letras “P” o “p” atravesadas con una raya, pero estos símbolos no eran uniformes. Ciertos matemáticos utilizaban la “p”, otros la “e” y el italiano Niccolò Fontana, *Tartaglia*, solía expresar esta operación como “Æ”. Los algebristas alemanes e ingleses introdujeron el signo “+”, al que denominaron *signum additorum*, aunque al principio solo se utilizaba para indicar excedentes. El matemático griego Diofanto empleaba el signo “´” para indicar la sustracción. Los hindúes usaban un punto y los algebristas italianos la representaban con una “M” o “m” y con una raya atravesando la letra. Los algebristas alemanes e ingleses fueron los primeros en utilizar el signo actual, al que denominaron *signum subtractorum*. Los signos “+” y “-” fueron usados por primera vez en

Mateo Fabián Itza Ortiz • Angélica María Escárcega Ávila • José María Carrera Chávez
Ernesto Orozco Lucero • Diana Marcela Beristain Ruiz

1489 por el alemán Johann Widmann. El matemático inglés William Oughtred fue el primero en emplear el signo "×" en vez de la palabra "veces". El matemático alemán Gottfried Wilhelm Leibniz utilizaba un punto para indicar la multiplicación y, en 1637, el francés Rene Descartes empezó a usar el signo de igualdad, "=", que creó el matemático galés Robert Recorde. Otro matemático inglés, Thomas Harriot, fue el primero en emplear los símbolos ">" y "<", "mayor que" y "menor que". El matemático francés François Viète introdujo varios signos de agrupación. Los símbolos de diferenciación, "dx", y de integración, o utilizados en el cálculo, son originales de Leibniz, lo mismo que el símbolo de semejanza, "~", empleado en geometría. El matemático suizo Leonhard Euler es el responsable principal de los símbolos "Æ", "f ", "F", usados en la Teoría de Funciones.

A continuación, se presentan los símbolos principales empleados en la Estadística Descriptiva e Inferencial (Cuadro 1).

Cuadro 1. Símbolos principales utilizados en la Estadística Descriptiva e Inferencial

Símbolo	Abreviación	Sonido	Definición
Σ			Sumatoria
Y			Variable dependiente
X			Variable independiente
N			Tamaño de la población
n			Tamaño de la muestra
$\overline{X}$	M		Promedio; media central; media aritmética
M_e	Me Md		Mediana
M_o	Mo		Moda
σ S	D. E. (STD)	Sigma	Desviación estándar de la población o muestra
σ^2 S^2	VAR.	Sigma cuadrada	Varianza de la población o muestra
$S/\sqrt{n}$	E. E. (SE)		Error estándar
$S/\overline{X}$	CV		Coeficiente de Variación
ε		Épsilon	Error experimental
A	A	Alfa	Nivel de significancia
B	B	Beta	Poder de la prueba; probabilidad del error tipo I; Coeficiente Estandarizado de Regresión Múltiple
μ		Mu	Media poblacional

Continúa...

Símbolo	Abreviación	Sonido	Definición
	GL (DF) (*f*)		Grados de libertad
χ^2		Ji Chi	Ji o Chi cuadrada
Sa			Desviación estándar del intercepto (a)
sa^2			Varianza del intercepto (a)
Sb			Desviación estándar de la pendiente (b)
sb^2			Varianza de la pendiente (b)
Sx			Desviación estándar de la variable independiente (x)
sx^2			Varianza de la variable independiente (x)
Sxy			Covarianza
Sy			Desviación estándar de la variable dependiente (y)
sy^2			Varianza de la variable dependiente (y)
t_f			Cuartil de la distribución ***t*** (de Student) para ***f*** grados de libertad
P	*P*		Probabilidad
P	P		Porcentaje
$>$			Mayor que
$\geq$			Mayor igual que
$<$			Menor que
$\leq$			Menor igual que
$\neq$			Desigual
	NS		No significativo
	MS		Media al cuadrado
	MSE		Error de la media al cuadrado
R	*R*		Coeficiente de Correlación Pearson
R	R		Correlación múltiple
r^2	R^2		Coeficiente de Determinación; correlación al cuadrado
r			Regresión lineal
F	F		Frecuencia observada
Fc	*Fc*		Frecuencia calculada o teórica
A	A		Intercepto de la regresión lineal simple
α'	a′		Intercepto de la regresión funcional
B	B		Pendiente de la regresión lineal simple
b'			Pendiente de la regresión funcional
H_0			Hipótesis Nula
H_1			Hipótesis Alternativa
T			Valor calculado de la prueba ***t*** de Student

Fuente: elaboración propia.

Redondeo de cifras

Redondear un número quiere decir reducir el número de cifras manteniendo un valor parecido. El resultado es menos exacto, pero más fácil de usar. Primero hay que saber si se redondea a décimas, centésimas, etcétera. Por ejemplo: 3.1416 redondeado a centésimas es 3.14, porque la cifra siguiente (16) es menor que 50; 1.2635 redondeado a décimas es 1.3, porque la cifra siguiente (63) es 50 o mayor; 1.2635 redondeado a 3 cifras decimales es 1.264, porque la cifra siguiente (50) es 50 o mayor.

Si se quiere redondear a decenas, centenas, etcétera, hay que sustituir las cifras que se quitan por ceros. Por ejemplo: 134.9 redondeado a decenas es 130, porque la cifra siguiente (40) es menor que 50; 12,690 redondeado a millares es 13,000, porque la cifra siguiente (69) es 50 o mayor; 1.239 redondeado a unidades es 1, porque la cifra siguiente (23) es menor que 50.

Para redondear “tantas” cifras significativas, se cuentan de izquierda a derecha las cifras a redondear. No se redondean cifras significativas cuando se inicia con dos ceros (0.006), porque solo se ponen para indicar lo pequeño que es el número. Ejemplo: 1.239 redondeado a 3 cifras significativas es 1.24, porque la cifra siguiente (90) es 50 o mayor; 134.9 redondeado a 1 cifra significativa es 100, porque la cifra siguiente (34) es menor que 50; 0.0165 redondeado a 2 cifras significativas es 0.017, porque la cifra siguiente (50) es 50 o mayor.

El redondeo de cifras sigue tres reglas básicas:

1. Si los primeros dos dígitos a descartarse son menores de 50, el número anterior no cambia. Ejemplo: 3.34489 se redondea a 3.34. A esto se le conoce como redondear abajo.
2. Si los primeros dos dígitos a descartarse son mayores de 50, se le suma 1 al número anterior. Ejemplo: 3.34617 se redondea a 3.35. A esto se le conoce como redondear arriba.
3. Si los primeros dos dígitos a descartarse son 50, se le suma 1 al número anterior si es impar; y si es par no se cambia. Ejemplos: 3.3350 y 3.3450 se redondean a 3.34.

El redondeo inadecuado de cifras se relaciona más a menudo con la precisión que deben tener los promedios. Por ejemplo, el promedio de la suma de 2.4 mm, 2.7 mm y 3.1 mm es 2.733333 mm. ¿Es adecuado redondear esta cifra a varios puntos decimales si solo tenemos tres datos precisos a un punto decimal? La cifra redondeada debe ser 2.7 mm.

El tamaño de la muestra, la amplitud de la variación, la naturaleza del objeto medido y la importancia de la precisión determinan la exactitud óptima de la cifra redondeada. Por ejemplo, si el diámetro promedio de 6 árboles se calcula en 1.8567 m y el rango de variación es 0.54 a 2.59 m, no tiene sentido expresar el promedio con cuatro puntos decimales de precisión (milésimas de milímetro), porque la muestra es pequeña y la variación es grande. Aunque podríamos expresar el promedio con más precisión si la muestra fuera de 100 árboles y la variación fuera menor, hacerlo sería igualmente inútil, porque conocer el diámetro de un árbol grande con la precisión de fracciones de un milímetro es irrelevante. La cifra redondeada debe ser 1.86 m.

El término "aproximadamente" debe usarse con precaución. Por ejemplo, sería impropio decir que el volumen de un estanque es de 33,547 litros, aproximadamente, porque esa cifra es exacta y no aproximada; la cifra aproximada sería 33,500 litros.

Ejercicio:

Redondeo de datos con la precisión indicada:

- **a)** 48.6 (sin decimal) = **?**
- **b)** 2.484 (2 decimales) = **?**
- **c)** 4.556 (1 decimal) = **?**
- **d)** 24.448 (2 decimales) = **?**
- **e)** 136.6 (sin decimal) = **?**
- **f)** 0.04358 (3 decimales) = **?**
- **g)** 143.95 (1 decimal) = **?**
- **h)** 5.56572 (3 decimales) = **?**

Notación científica

En el sistema decimal la notación científica tiene la base 10, es decir, la notación científica significa que un número entre el 1 y el 10 es multiplicado por una potencia de base 10 elevada a la potencia del exponente, y este exponente $(^{y})$ representa el número de veces que se desplaza el punto decimal (a la derecha o a la izquierda).

La notación científica es formada siempre por un número entero (conocido como coeficiente), la base 10 y su exponente, que puede ser positivo o negativo dependiendo de la dirección (a la derecha o a la izquierda) del desplazamiento del punto decimal, como se presenta a continuación (Figura 7).

Figura 7. Componentes de una notación científica

Fuente: elaboración propia.

En cuanto a los decimales en Europa continental se escriben de la forma 1,23; en las islas Británicas, 1·23; y en el continente americano, 1.23. Utilizando la notación científica estándar un número como 0.000000123, se puede escribir 1.23×10^{-7}.

La notación científica es la forma de escribir números muy grandes o muy pequeños en una manera más conveniente y de fácil lectura. Generalmente tiene gran utilidad en matemáticas, física e ingeniería y puede ser usada en microbiología para determinar poblaciones de bacterias. A continuación, se presentan diferentes valores numéricos, sus equivalentes en notación científica y la representación numérica de cada uno (Cuadro 2).

Cuadro 2. Nomenclatura para el valor
de notación científica y su valor numérico

Valor numérico	Representación en notación científica	Representación numérica
Miltrillonésima	10^{-21}	0.000000000000000000001
Trillonésima	10^{-18}	0.000000000000000001
Milbillonésima	10^{-15}	0.000000000000001
Billonésima	10^{-12}	0.000000000001
Milmillonésima	10^{-9}	0.000000001
Millonésima	10^{-6}	0.000001
Milésima	10^{-3}	0.001
Centésima	10^{-2}	0.01
Décima	10^{1}	0.1
Uno	1	1
Diez	10^{1}	10
Cien	10^{2}	100
Mil	10^{3}	1,000
Millón	10^{6}	1,000,000
Mil millones	10^{9}	1,000,000,000
Billón*	10^{12}	1,000,000,000,000
Mil billones	10^{15}	1,000,000,000,000,000
Trillón	10^{18}	1,000,000,000,000,000,000
Mil trillones	10^{21}	1,000,000,000,000,000,000,000

Fuente: elaboración propia.

En nuestro sistema métrico decimal, los valores de miles o millares están representados por una coma (",") y el punto decimal. En el sistema inglés, los millares están representados por un punto (".").

En Estados Unidos de América 10^9 se denomina "billón". Para el resto de los países de habla hispana 10^9 equivale a "mil millones", mientras que el billón se representa como 10^{12}. Igualmente, en las naciones de habla hispana 10^9 recibe también el nombre de "millardo" (palabra proveniente del francés *millard*), además de "mil millones". Por tanto, lo que para los estadounidenses es "one billion dollars or euros" (un billón de dólares o de euros), para los hispanohablantes sería "un millardo de dólares o de euros" o "mil millones de dólares o de euros". Por otra parte, en español 10^4 (10,000) también se denomina "miríada".

1. Método para representar un número entero en notación científica

Cualquier número entero o decimal, independientemente de la cantidad de cifras que posea, se puede reducir empleando la notación científica. Veamos el siguiente ejercicio (Cuadro 3).

Cuadro 3. Ejemplos de valores numéricos (n) expresados en valores de notación científica (e^x también como n^y)

a)	529,745,386	=	5.29×10^{8}
b)	450	=	4.5×10^{2}
c)	590,587,348,584	=	5.9×10^{11}
d)	0.3483	=	3.5×10^{-1}
e)	0.000987	=	9.87×10^{-4}

Fuente: elaboración propia.

Como se puede observar en el Cuadro 3, la notación científica se compone siempre de un solo número entero y el resto pueden ser uno o varios decimales, según la menor o mayor exactitud que requiera una representación numérica determinada. La cantidad de decimales se puede recortar a uno o dos números solamente por medio de la aproximación o redondeo de la cifra, pues el objetivo de emplear la notación científica es, precisamente, acortar las cifras largas, ya sean de números enteros o decimales.

Para convertir en notación científica el número 529,745,386 (inciso a del Cuadro 3) será necesario **contar de derecha a izquierda** los espacios que existen entre el último número de la serie numérica, a partir del "6", hasta llegar al primero ("5" en este caso). Después de contar veremos que hay ocho espacios, por lo que la notación científica de ese número entero la podemos escribir así: 5.29×10^{8}. (El superíndice o exponente 8 representa los espacios que hemos contado desde el "6" hasta el "5").

Si queremos redondear esa cifra para que la notación sea aún más simplificada, podemos escribirla también como: 5.2×10^{8}. Igualmente, se pueden representar más cifras decimales empleando los propios números que forman el número entero, por ejemplo: 5.2975×10^{8}.

Para convertir de nuevo la cifra representada en notación científica en el número entero que le dio origen, realizamos la operación inversa. Por ejemplo: si el número entero 529,745,386 se redondeó originalmente para que su representación decimal en notación científica fuera 5.3×10^8 y queremos restaurar ahora el número original, en este caso será necesario multiplicar $5.3 \times 100{,}000{,}000$ (los ocho ceros se corresponden con el superíndice o exponente 10^8). El resultado de la operación será 530,000,000 en lugar de 529,745,386, que como se podrá comprobar difiere algo del número entero original, debido a la aproximación o redondeo que se realizó anteriormente.

2. Método para representar un número decimal o fraccionario en notación científica

El procedimiento para convertir un número decimal en otro número en notación científica es parecido al anterior. Tomemos por ejemplo el número 0.000987 correspondiente al inciso e) del Cuadro 3. Para realizar la conversión sencillamente **corremos la coma hacia la derecha** de los cuatro espacios que la separan del "9", con lo que obtendremos el siguiente número decimal: 9.87. Por tanto, la notación final quedará de la siguiente forma: 9.87×10^{-4}. Si queremos acortar más la notación podemos redondear y escribirla también como: 9.9×10^{-4}. En el caso de la conversión de decimales a notación científica, el superíndice o exponente del "10" llevará el signo "menos" para indicar que esta notación corresponde a un número fraccionario en lugar de uno entero.

Para convertir de nuevo la notación científica de este ejemplo en decimal, **movemos la coma tantos lugares a la izquierda como números nos indique el superíndice negativo**, agregando los correspondientes ceros para completar la cifra.

Ejercicios:

Expresar los siguientes números usando potencia de 10:

- **a)** $4.83 \times 10^{7} =$ **?**
- **b)** $3.80 \times 10^{-4} =$ **?**
- **c)** $300 \times 10^{8} =$ **?**
- **d)** $8.4 \times 10^{-6} =$ **?**
- **e)** $1.86 \times 10^{6} =$ **?**
- **f)** $70{,}000 \times 10^{-10} =$ **?**

Escribir cada número en notación científica:

a) 24,380,000 = **?**
b) 7,3000,000,000 = **?**
c) 0.000009851 = **?**
d) 0.00018400 = **?**

Obtenga el producto de las siguientes operaciones utilizando notación científica:

a) $(10^3)(10^2)$ = **?**
b) $10^6 / 10^4$ = **?**
c) (4,000,000)(0.0000000002) = **?**
d) $\frac{(0.0009)(160000)}{(0.005)}$ = **?**

Dígitos significativos

Los dígitos o cifras significativas aportan información sobre el resultado de la medición. Ellas representan el uso de una o más escalas de incertidumbre en determinadas aproximaciones. Por ejemplo, se dice que 4.7 tiene dos cifras significativas, mientras que 4.07 tiene tres.

Reglas básicas

1. Cualquier dígito diferente de cero se considera significativo.

   Ejemplo: 27.42 = 4 DS; 348 = 3 DS; 5432 = 4 DS

2. Los ceros entre dígitos diferentes de cero se consideran significativos.

   Ejemplo: 102008 = 6 DS; 403 = 3 DS; 7002 = 4 DS; 10401 = 5 DS

3. Los ceros a la izquierda del primer dígito diferente de cero no se consideran significativos.

   Ejemplo: 0.0035 = 2 DS; 0.004 = 1 DS

4. Para un número mayor que la unidad, todos los ceros que están a la derecha del punto decimal se consideran significativos.

   Ejemplo: 1.020 = 4 DS; 43.20 = 4 DS; 1.01 = 3 DS

5. Para números que no tienen puntos decimales, los ceros que están después del último dígito diferente de cero pueden ser o no dígitos significativos.

   Ejemplo: 7000 = 4 DS; 7×10^3 = 1 DS; 7.00×10^3 = 3 DS

Dígitos significativos en la suma o la resta

En una operación de sumas y restas se considera como dígitos significativos toda la cifra de la parte entera del resultado de la operación, más las cifras de su parte decimal que no superen al menor número de cifras decimales de las cantidades involucradas en la operación.

Ejercicio:

1. Calcule la siguiente operación de dígitos significativos:

$$13.\underbrace{467}_{3\ \text{DS}} + 134.\underbrace{2}_{1\ \text{DS}} + 6.\underbrace{84}_{2\ \text{DS}} = 154.5\cancel{07}$$

Resultado de DS = 153.5

2. Calcule la siguiente operación de dígitos significativos:

$$13.\underbrace{467}_{3\ \text{DS}} + 134.\underbrace{2}_{1\ \text{DS}} - 6.\underbrace{84}_{2\ \text{DS}} = 140.8\cancel{27}$$

Resultado de DS = 140.8

Dígitos significativos en la multiplicación o la división

En una operación de multiplicación o división se considera como dígitos significativos toda la cifra que resulte de la operación (multiplicación o división) que no supere al menor número de dígitos significativos de las cantidades involucradas en la operación.

Ejercicio:

1. Calcule la siguiente operación de dígitos significativos:

$$\underbrace{13.467}_{5\ \mathrm{DS}} \times \underbrace{6.84}_{3\ \mathrm{DS}} = 92.1\cancel{1428}$$

Resultado de DS = 92.1

2. Calcule la siguiente operación de dígitos significativos:

$$\underbrace{13.467}_{5\ \mathrm{DS}} / \underbrace{6.84}_{3\ \mathrm{DS}} = 1.96\cancel{885965}$$

Resultado de DS = 1.968

Funciones o Teoría de Conjuntos

En matemáticas, una función es el término empleado para indicar la relación o correspondencia entre dos o más cantidades. El término **función** fue usado por primera vez en 1637 por el matemático francés Rene Descartes, para designar una potencia x^n de la variable X. En 1694 el matemático alemán Gottfried Wilhelm Leibniz utilizó el término para referirse a varios aspectos de una curva, como su pendiente. Hasta recientemente su uso más generalizado ha sido el definido en 1829 por el matemático alemán J. P. G. Lejeune-Dirichlet (1805-1859), quien escribió:

> Una variable es un símbolo que representa un número dentro de un conjunto de ello. Dos variables X y Y están asociadas de tal forma que al asignar un valor a X entonces, por alguna regla o correspondencia, se asigna automáticamente un valor a Y, se dice que Y es una función (unívoca) de X. La variable X, a la que se asignan libremente valores, se llama variable independiente, mientras que la variable Y, cuyos valores dependen de la X, se llama variables dependientes. Los valores permitidos de X constituyen el dominio de definición de la función y los valores que toma Y constituye su recorrido.

Una función f de A en B es una relación que hace corresponder a cada elemento x ∈ A uno y solo un elemento y ∈ B, llamado imagen de x por f, que se escribe y = f(x).

Definición de función matemática

Iniciaremos definiendo qué es una función matemática. Una función matemática también puede ser llamada solamente función y es la relación que hay entre una magnitud y otra cuando el valor de la primera depende de la segunda (Figura 8). Por ejemplo, si calculamos el peso del huevo decimos que depende del día que lo pesamos, por lo que estaremos estableciendo —sin saberlo— una función entre ambas cosas. Ambas magnitudes son variables que distinguen dos de ellas (variables):

- » *Variable dependiente*: depende del valor de la otra magnitud. En el caso del ejemplo anterior es el peso del huevo.
- » *Variable independiente*: esta define la variable dependiente. En el caso del ejemplo anterior es el día del pesado del huevo.

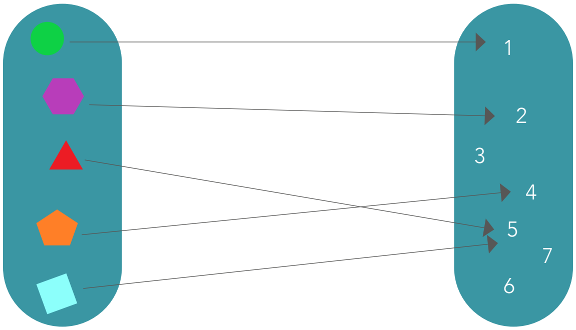


Figura 8. La función es una relación entre dos magnitudes, de tal manera que a cada valor de la primera (conocida como dominio) le corresponde un único valor de la segunda (llamada imagen)

Fuente: elaboración propia.

Aplicaciones de las funciones reales

Generalmente se hace uso de las funciones reales (aun cuando el ser humano no se da cuenta) en el manejo de cifras numéricas en correspondencia con otras, debido a que se están empleando subconjuntos de los números reales. Las funciones son de mucho valor y utilidad para resolver problemas de la vida diaria, de finanzas, de economía, de estadística, de ingeniería, de medicina, de química, de física, de astronomía, de geología y de cualquier área social donde haya que relacionar variables.

Cuando se va al mercado o cualquier centro comercial siempre se relaciona un conjunto de determinados objetos o productos alimenticios con el costo en pesos, para así saber cuánto podemos comprar; si lo llevamos al plano podemos escribir esta correspondencia en una ecuación de función "x" como el precio y la cantidad de producto como "y".

De esta manera, toda función matemática consiste en la relación entre un elemento de un grupo A y otro elemento de un grupo B, siempre que se vinculen de manera única y exclusiva. Por lo tanto, dicha función puede expresarse en términos algebraicos empleando signos de la siguiente manera: en símbolos, $f: A \subseteq B$. Es decir, que para que una relación de un conjunto A en otro B sea función, debe cumplir dos condiciones; a saber:

1. Todo elemento del conjunto de partida A debe tener imagen.
2. La imagen de cada elemento $x \in A$ debe ser única. Es decir, ningún elemento del dominio puede tener más de una imagen.

   * El conjunto formado por todos los elementos de B, que son imagen de algún elemento del dominio, se denomina conjunto imagen o recorrido de f.

Observaciones:

En una función $f: A \subseteq B$ todo elemento $x \in A$ tiene una y solo una imagen $y \in B$.

Un elemento y $\in$ B puede:

- No ser imagen de ningún elemento x $\in$ A;
- Ser imagen de un elemento x $\in$ A;
- Ser imagen de varios elementos x $\in$ A.

La relación inversa f-1 de una función f puede no ser una función. Las formas de expresión de una función mediante el uso de tablas son (Cuadro 4).

Cuadro 4. Cuadro de función f -1 [x, f (x)]

X	Y
-1	1
0	0
½	¼
1	1
2	4

Fuente: elaboración propia.

Gráficamente: cabe aclarar que llamamos gráfica de una función real de variable real al conjunto de puntos del plano que, referidos a un sistema de ejes cartesianos ortogonales, tienen coordenadas [x, f (x)], donde x $\in$ A.

Sistema de coordenadas o plano cartesiano

Las coordenadas o planos cartesianos, también conocidos como coordenadas o planos rectangulares, reciben este nombre porque los creó —y descubrió su importancia— el filósofo y matemático Rene Descartes (1596-1650), quien fue el fundador de la geometría analítica.

El plano cartesiano o sistema de coordenadas cartesianas está determinado por dos rectas reales, una horizontal y otra vertical, llamadas ejes de **coordenadas**, que se cortan entre sí formando cuatro ángulos de 90° cada uno. El plano cartesiano sirve para describir la posición de puntos o espacios geométricos, los cuales se plasman a través de sus coordenadas o pares de coordenadas; en otras palabras, es un sistema de coordenadas cartesianas en un par de rectas graduadas, perpendiculares, que se cortan en un punto $P(0,0)$ llamado origen de coordenadas. Para un punto $P(x, y)$, x e y se llaman coordenadas del punto P.

A la **recta horizontal** se le llama **eje de las abscisas (eje de las x)**, y a su perpendicular o **recta vertical por 0, eje de las ordenadas (eje de las y)**. Este plano está dividido en cuatro cuadrantes (Figura 9).

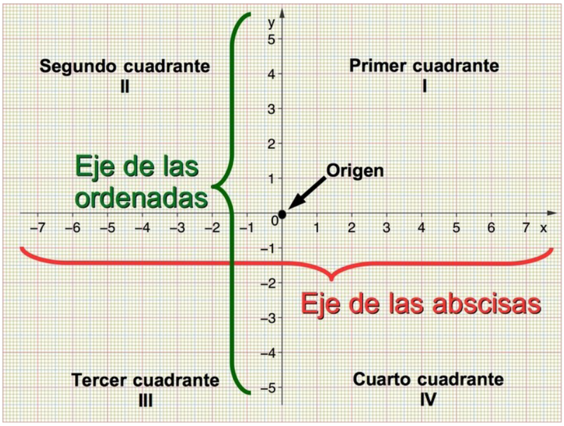


Figura 9. Eje de las abscisas (x) y eje de las ordenadas (y)

Fuente: descargada de internet.

En la Figura 9 se puede observar que hacia la izquierda del origen los valores son negativos y hacia la derecha son positivos sobre el eje horizontal o eje de las abscisas (x). Mientras que hacia abajo del origen los valores son negativos y hacia arriba son positivos sobre el eje vertical o eje de las ordenadas (y). A cada punto P se le asigna una coordenada $P(x, y)$ y en cada cuadrante, se puede encontrar el $P(\pm x, \pm y)$, como se observa en la Figura 10.

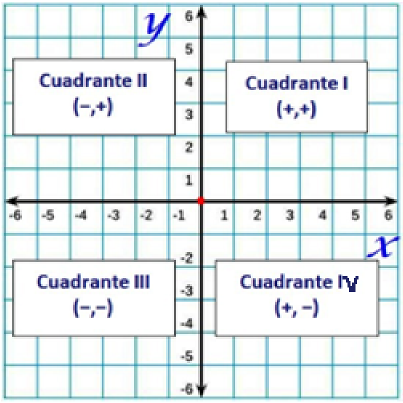


Figura 10. Posición de los cuadrantes en el plano cartesiano y su *P*(*x*, *y*)

Fuente: elaboración propia.

Entonces, el plano cartesiano es un diagrama que permite localizar puntos específicamente dentro de un sistema de coordenadas que se conocen como rectangulares, y localizar cada punto *P*(*x*, *y*) y trazar los puntos *P*(*x*, *y*) a los siguientes *P*(*x*, *y*), y en algunos casos formar figuras geométricas. *En otras palabras,* es una forma de ubicar puntos en el espacio habitualmente en los casos bidimensionales.

En el siguiente ejemplo se muestra cómo fueron colocados los puntos *P*(*x*, *y*) en el espacio (Figura 11).

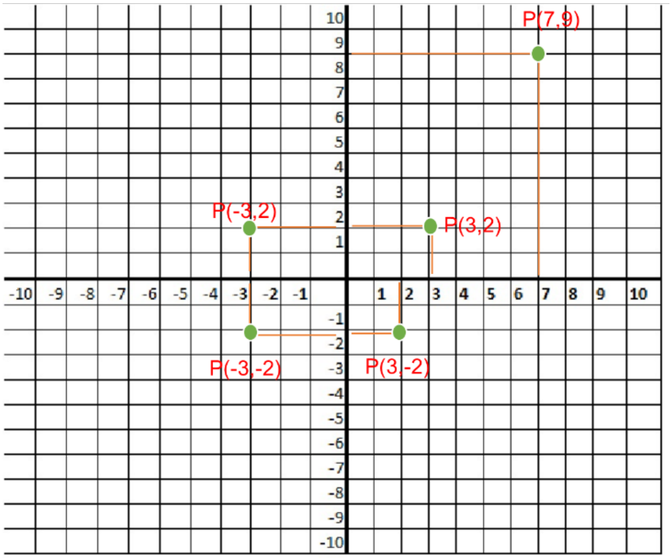
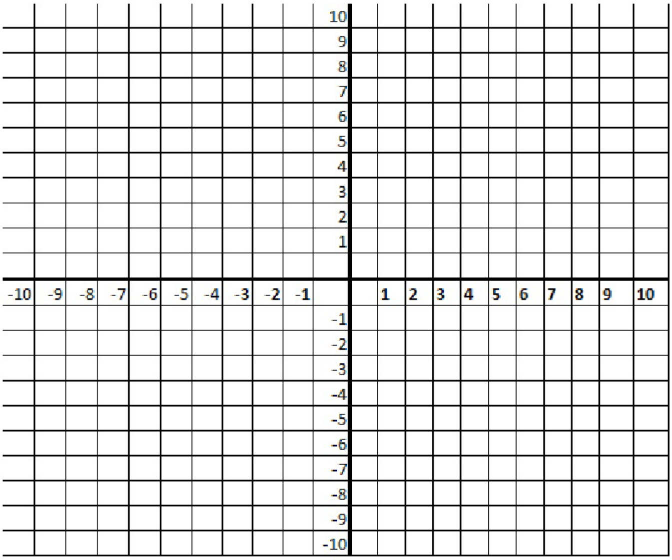


Figura 11. Representación de los puntos *P*(*x*, *y*) en el plano cartesiano

Fuente: elaboración propia.

Ejercicio:

Coloque los siguientes puntos: A(3, 2); B(-3, 2); C(-3, -2); D(3, -2); E(7, 9) en el plano cartesiano y luego una cada uno de los puntos.



Gráficas

La representación de datos de forma gráfica ofrece mensajes más claros donde las conclusiones son fáciles de entender. Una gráfica es una especie de esquema formado por líneas, figuras, mapas, utilizado para representar datos estadísticos (variables cuantitativas) a escala o según una cierta proporción (valores porcentuales), o elementos de un sistema, las etapas de un proceso y las divisiones o subdivisiones de una clasificación. Es importante mencionar que **las gráficas en un texto científico reciben el nombre de Figuras** (primera letra en mayúscula) y deben ser colocadas al pie de la imagen (Figura) y citadas en el texto previo a la aparición de la imagen.

Entre las funciones que cumplen las gráficas, se pueden señalar las siguientes:

- Hacen más visibles los datos, sistemas y procesos.
- Ponen de manifiesto sus variaciones y su evolución histórica o espacial.
- Pueden evidenciar las relaciones entre los diversos elementos de un sistema o de un proceso y representar la correlación entre dos o más variables.
- Sistematizan y sintetizan los datos, sistemas y procesos.
- Aclaran y complementan las tablas y las exposiciones teóricas o cuantitativas.

- El estudio de su disposición y de las relaciones que muestran puede sugerir hipótesis nuevas.

Tipos de gráficas y sus características principales

Para construir una gráfica es importante distinguir el tipo de datos a graficar. Pueden ser datos continuos o datos discretos y la elección de la gráfica dependerá del tipo de datos.

Algunas de las gráficas más importantes son:

1. Gráfica de árbol.
2. Gráfica de áreas o superficies.
3. Gráfica de bandas.
4. Gráfica de barras.
5. Gráfica de bloques.
6. Gráfica circular.
7. Gráfica circular polar.
8. Gráfica de puntos.
9. Gráfica de tallo y hoja diagrama.
10. Histogramas y gráficos de caja.

Para la utilización de las gráficas se requiere un procedimiento específico:

- Decidir la gráfica a emplear.
- Construir gráficas para el control estadístico del proceso.
- Controlar el proceso. Si aparece una anormalidad sobre la gráfica, investigar inmediatamente las causas y tomar acciones apropiadas.

Algunos puntos importantes a considerar previo a la elaboración de la gráfica son:

- Propósito de la gráfica.
- Variable a considerar.
- Tamaño de la muestra.

- Tener un criterio para decidir si conviene investigar causas de variación en el proceso de producción.
- Familiarizar al personal con el uso de la gráfica.

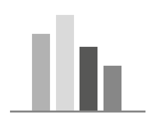
Proceso que se debe seguir para construir una gráfica

La construcción de una gráfica de rangos y promedio resulta de formar una unidad, tanto de la gráfica de promedios como de la de rangos. Consta de dos secciones: la parte superior se dedica a los promedios y la parte inferior, a los rangos. En el eje vertical se establece la escala y a lo largo del eje horizontal se numeran las muestras. En la gráfica se relacionan estos promedios con los intervalos durante los cuales se tomaron las muestras. **En el eje vertical se indican los valores correspondientes a los valores de las muestras. En el eje horizontal se señalan los periodos (de tiempo) en los que se toman las muestras, a semejanza de los de los promedios.**

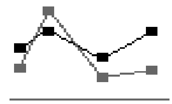
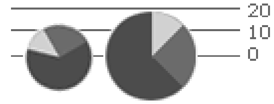
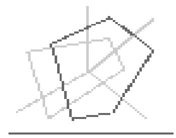
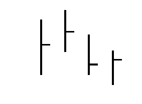
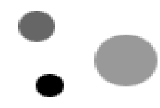
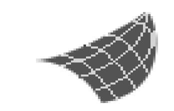
La interpretación de esta gráfica de promedios y rangos sería que, a partir de los datos, podemos determinar el valor central del proceso y su aplicación. Mediante este proceso está bajo control cuando no muestra ninguna tendencia y, además, ningún punto sale de los límites. Se describen los distintos tipos de tendencia, que son patrones de comportamiento anormal de los puntos (inestabilidad o proceso fuera de control estadístico).

En el Cuadro 5 se describen las gráficas principales, así como su descripción y uso principal.

Cuadro 5. Tipos de gráficos (diferencias y usos)

Descripción	Ejemplo
Barras / Columnas Este gráfico sirve para comparar datos entre diferentes segmentos (sectores, empresas, periodos [de tiempo]...).	

Continúa...

Descripción	Ejemplo
Líneas Ayudan a ver la evolución de los datos. Por lo general se emplean para mostrar un mismo tipo de datos y su evolución (valor de la acción y el tiempo; número de ventas y precio).	
Pastel o pie Aquí podemos ver la contribución de cada parte de un total o proporciones. Este gráfico se puede utilizar de forma creativa comparando el tamaño de las "rebanadas" entre sí y el contenido de las mismas.	
Radar En este gráfico podemos ver la superficie creada por varias variables y así poder comparar entidades (dos productos que presentan varias características pueden ser comparados en su totalidad utilizando este gráfico).	
Stocks Aquí se representan datos con cuatro variables: tiempo, máximo, mínimo y cierre.	
Burbujas Aquí el grid (líneas de división del eje) suele ser una variable por sí misma, haciendo que la disposición de las burbujas represente otras variables junto al propio tamaño de las burbujas. Este tipo de gráficos permite concentrar mucha información en poco espacio.	
Superficies Este gráfico se suele emplear para ver la evolución de un dato sujeto a tres variables. Por ejemplo, la dureza de un material dependiendo de la temperatura, densidad y volumen.	

Fuente: elaboración propia.

Las gráficas, llamadas figuras en un texto científico, también suelen elegirse con base en las características de los datos o en función del número de observaciones; en Figuras para representar datos no agrupados o datos agrupados. Por ejemplo, se tiene que la Figura 12 (gráfica) de columnas se puede usar para representar datos no agrupados.

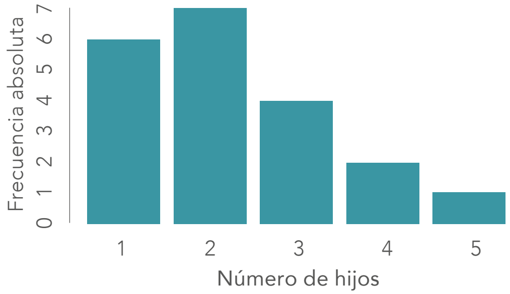


Figura 12. Número de hijos por familia registrados en el municipio de Juárez, Chihuahua, México, en 2022

Fuente: elaboración propia.

También, se pueden representar dos variables en una misma Figura (gráfica) y a este tipo de Figura se le conoce como Figura Combinada (Figura 13), como se observa a continuación:

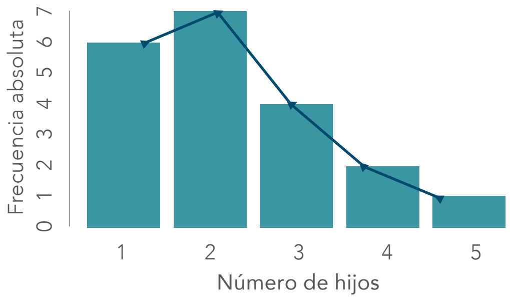


Figura 13. Número de hijos por familia y edad promedio del primogénito registrados en el municipio de Juárez, Chihuahua, México, en 2022

Fuente: elaboración propia.

En datos agrupados, la Figura (gráfica) mayormente empleada es el histograma (Figura 14).

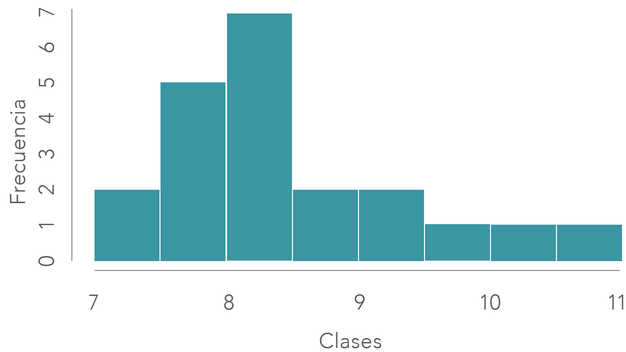


Figura 14. Histograma del paralaje solar en los últimos 11 años

Fuente: elaboración propia.

Otra Figura (gráfica) para datos agrupados, se presenta a continuación (Figura 15):

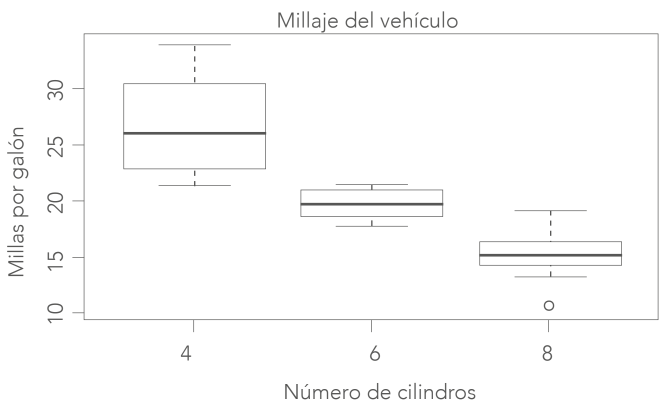


Figura 15. Rendimiento de combustible de un automóvil con base en el cilindraje

Fuente: elaboración propia.

De igual forma, se puede hacer una combinación representando dos variables en una Figura. Recuerde que a los dos ejes conocidos (x, y), se agrega un tercer eje llamado eje de z (ejemplo: Figura 16).

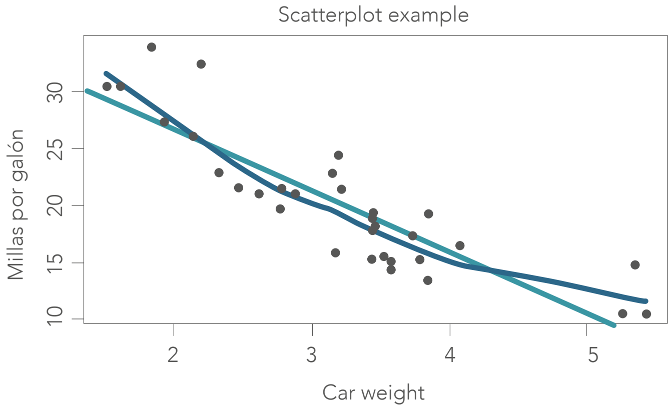


Figura 16. Rendimiento del combustible en relación con el peso del vehículo

Fuente*:* elaboración propia.

Es importante dominar la presentación para mostrar un mensaje fácil de entender. El no ofrecer una conclusión clara hace que la Figura (gráfica) pierda su fuerza en nuestra comunicación. En algunas figuras anteriores, se observa que el peso del automóvil no tiene la unidad de medición, es decir, kilogramos o toneladas, por ejemplo.

Elementos en un gráfico

1. Título principal.
2. Título secundario o subtítulo (opcional).
3. Descripción del gráfico (número de figura al pie de este).
4. Región de datos y símbolos (leyenda).
5. Eje horizontal y escala.
6. Eje vertical y escala.
7. Apuntadores (opcional).
8. Descriptores de señales y marcas.

Recomendaciones generales para elaborar un gráfico (Figura)

1. Haga que sus datos sobresalgan. Evite lo superfluo.
2. Utilice elementos prominentes para mostrar sus datos.
3. Emplee un par de líneas por cada variable. Haga que el interior del rectángulo formado por las líneas de la escala sean la región de sus datos.
4. Coloque marcas afuera de la región de los datos.
5. No apeñusque la región de los datos.
6. No exagere el número de marcas.

Ejemplo: gráfico y sus elementos principales (Figura 17).

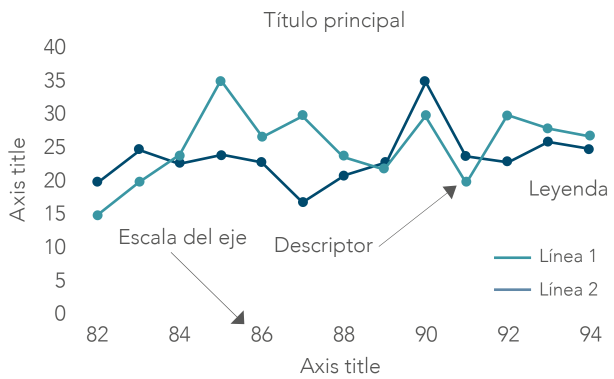


Figura 17. Descripción de los elementos que debe incluir un gráfico

Fuente: elaboración propia.

Es importante mencionar que en un texto científico el título principal de la Figura no aplica, debido a que la descripción de la misma, ubicada al pie de esta, debe contener su propia descripción y ser lo suficientemente clara, para que el lector no tenga que recurrir al texto para poder entenderla.

Distribución de frecuencias

La distribución de frecuencias o la tabla de distribución de frecuencias se puede definir como una ordenación en forma de tabla de los datos estadísticos, asignando a cada dato su frecuencia correspondiente.

Se puede realizar distribución de frecuencias, tanto para variables **cualitativas**, donde observamos y cuantificamos (*n*) las cualidades con las mismas características, como para variables **cuantitativas** buscando los valores **mínimo** y **máximo** conseguidos para obtener una amplitud. En función del número de observaciones decidiremos si se realiza su estudio en forma de **datos no agrupados (individual) o datos agrupados (intervalos, marcas de clases)**.

Las tablas de distribución de frecuencias se utilizan cuando se recolectan datos para representar de manera fácil su análisis. Se pueden elaborar tablas de distribución de frecuencias para datos no agrupados y datos agrupados.

Los datos generalmente tienden a agruparse cuando se tiene mucha información (datos) y dependiendo del objetivo del análisis, por ejemplo, se agrupan en rangos de edades, pesos, entre muchos otros. A este tipo de datos agrupados se les conoce también como intervalos o clases. Entonces, la clase es cada uno de los grupos (intervalos) en los que se dividen los datos.

Definiciones usadas con mayor frecuencia en las tablas de distribución de frecuencias

- » Se define como **frecuencia** al número de veces que se repite un dato, es representada con la letra minúscula f y la frecuencia acumulada con la letra mayúscula F.
- » Se define como **rango** a la diferencia entre el mayor y el menor valor presente en el grupo de datos recopilados.
- » Se define como **marca de clase** al punto medio del intervalo y se calcula promediando el límite inferior y el límite superior.
- » Se define como **límite** a los valores extremos de cada intervalo. Cada intervalo puede tener límites bien definidos conocidos como límite inferior/menor y límite superior/mayor. Estos límites pueden ser límites abiertos, es decir, cuando se da un rango, por ejemplo, de valores mayores o menores de ciertos números.
- » Se define como **intervalo o intervalo de clase** a la forma en la que se agrupan y ordenan los valores observados delimitados por dos valores extremos llamados límites.
- » Se define como **amplitud o ancho del intervalo** a la diferencia de tamaño entre el límite superior y el límite inferior. Generalmente es conocido como ancho del intervalo.
- » Se define como **número de clase** a la posición o número de renglón que ocupa la marca de clase en una columna cuantificada de arriba hacia abajo.

Tabla de distribución de frecuencias para variables cualitativas

Se construye con base en las observaciones y se cuentan las veces que aparece cada una de ellas para construir las frecuencias (f) y de esta, se obtiene el porcentaje (%) de cada una.

Ejemplo 1: Según la Asociación de Lucha contra la Bulimia y la Anorexia, las pautas culturales han determinado que la delgadez sea sinónimo de éxito social. Muchos jóvenes luchan para conseguir el "físico ideal" motivados por modelos, artistas o publicidad comercial.

Durante el mes de marzo del año 2006, en el Colegio "Alcántara" de la ciudad de Talca, después de las vacaciones de verano, se observó con precaución a 27 alumnos con síntomas de anorexia, registrándose los siguientes signos visibles:

Dieta severa	Miedo a engordar	Hiperactividad
Uso de ropa holgada	Dieta severa	Uso de laxantes
Miedo a engordar	Dieta severa	Uso de ropa holgada
Dieta severa	Uso de ropa holgada	Dieta severa
Dieta severa	Dieta severa	Uso de ropa holgada
Hiperactividad	Uso de laxantes	Miedo a engordar
Uso de laxantes	Dieta severa	Uso de ropa holgada
Uso de laxantes	Hiperactividad	Uso de laxantes
Uso de ropa holgada	Hiperactividad	Dieta severa

Resuma la información anterior en el Cuadro 6 de distribución de frecuencias:

Cuadro 6. Distribución de los signos visibles de 27 alumnos con síntomas de anorexia, en el Colegio "Alcántara" de la ciudad de Talca, durante el mes de marzo del año 2006

Signo visible	Número de alumnos	Porcentaje de alumnos
Dieta severa	9	33.2
Miedo a engordar	3	11.1
Hiperactividad	4	14.8
Uso de laxantes	5	18.5
Uso de ropa holgada	6	22.2
Total	27	100.0

Fuente: elaboración propia.

Realice un gráfico para sintetizar la información obtenida del ejemplo anterior (Figura 18).

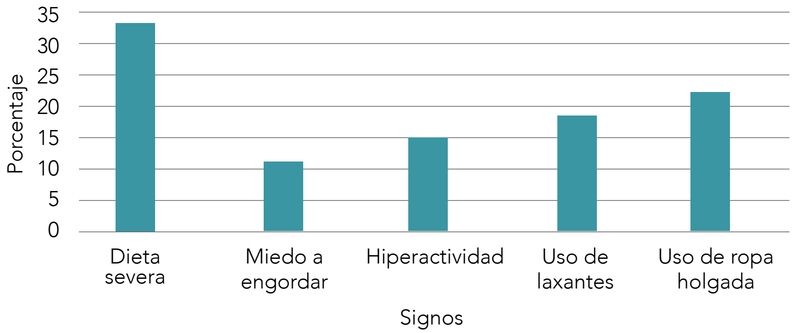


Figura 18. Distribución de los signos visibles de 27 alumnos con síntomas de anorexia, en el Colegio "Alcántara" de la ciudad de Talca, durante el mes de marzo del año 2006

Fuente: elaboración propia.

Tabla de distribución de frecuencias para variables cuantitativas

Un intervalo puede definirse como la franja en la que se divide el recorrido de un conjunto de datos. Un rango puede definirse como la diferencia entre el mayor y el menor valor.

Construcción de intervalos

Teniendo en cuenta la amplitud total de las observaciones (valor máximo menos valor mínimo observados), tomaremos una decisión sobre el número total de intervalos, o bien, sobre la amplitud o tamaño de estos.

Teóricamente se establece que el número ideal de intervalos debe ser la raíz cuadrada del número de observaciones disponibles. Para N observaciones: Criterio de Kaiser **n.° de intervalos** $\approx \sqrt{N}$.

Al establecer dos intervalos consecutivos, por ejemplo, de 10 a 20 y de 20 a 30, hemos de decidir si el valor 20 (final de uno e inicio del siguiente) pertenece al primer intervalo o al segundo. Para ello, empleamos los símbolos "[" y "(".

[o] = El valor situado junto a él pertenece al intervalo;
(o) = El valor situado junto a él no pertenece al intervalo.

Frecuencia absoluta (n)

Para datos no agrupados en intervalos es el número de veces que se presenta cada valor de la variable, es decir, para un cierto valor de la variable, la frecuencia absoluta nos da el número de observaciones menores o iguales que dicho valor.

Se representa por f_i:

$$\sum_{i-1}^{i-n} fi = \mathrm{n}$$

donde la suma de las frecuencias absolutas es igual al número total de datos, que se representa como N.

$$f_1 + f_2 + f_3 + \ldots + f_n = \mathrm{N}.$$

Si los datos se agrupan en intervalos es el número de observaciones que pertenecen a dicho intervalo.

Frecuencia relativa (r)

Es el cociente entre la frecuencia absoluta (f_i) y el número total de observaciones (N). Se puede expresar en tantos por ciento y se representa por n_i:

$$n_i = \frac{fi}{N}.$$

Frecuencia absoluta acumulada (N)

Para un cierto valor de la variable, la frecuencia absoluta acumulada nos da el número de observaciones menores o iguales que dicho valor. Se representa por F_i:

$$\sum_{i-1}^{i-n} Fi = \mathrm{N}.$$

La suma de las frecuencias absolutas acumuladas es igual al número total de datos, que se representa por N:

$$f_1 + f_2 + f_3 + \ldots + f_n = \mathrm{N}.$$

Proporción o Porcentaje (p):

Es la frecuencia relativa multiplicada por 100 (es la expresión de las frecuencias en %).

De igual modo que se definió para las frecuencias absolutas, se define para las FRECUENCIAS RELATIVAS ACUMULADAS (R) y los PORCENTAJES ACUMULADOS (P).

Tabla completa de frecuencias. Ecuaciones para el cálculo de sus elementos

Intervalo de clase	Marca de clase x	n	r	p	N	R	P
a – b	x_1	n_1	$r_1 = n_1 / N$	$p_1 = r_1 \bullet 100$	n_1	r_1	p_1
c – ...	x_2	n_2	$r_2 = n_2 / N$	$p_2 = r_2 \bullet 100$	$n_1 + n_2$	$r_1 + r_2$	$p_1 + p_2$
	…	…	…	…	…	…	…
	x_i	n_i	$r_1 = n_1 / N$	$p_i = r_i \bullet 100$	$n_1 + n_2 + \ldots + n_i$	$r_1 + r_2 + \ldots + r_i$	$p_1 + p_2 + \ldots + p_i$
	…	…	…	…	…	…	…
		$\boldsymbol{\Sigma\, n_i = N}$	$\boldsymbol{\Sigma\, r_i = 1}$	$\boldsymbol{\Sigma\, p_i = 100}$			

Fuente: elaboración propia.

Ejemplo de una tabla completada

x	n	r	p	N	R	P
2	5	0.125	12.5	5	0.125	12.5
3	10	0.250	25	15	0.375	37.5
4	16	0.400	40	31	0.775	77.5
5	6	0.150	15	37	0.925	92.5
6	3	0.075	7.5	**40**	**1.000**	**100**
	40	**1**	**100**			

Ejemplo de tabla de distribución de frecuencias

Supuesto: Valor máximo = 87, Valor mínimo = 11. Luego: AMPLITUD = 87 – 11 = 76.

Si decidimos construir 8 intervalos, la amplitud de cada uno será de 10 unidades (valor aproximado de (76 / 8). El primer intervalo no tiene por qué iniciarse en 11 (mínimo); es más, se aconseja tomar siempre valores "visualmente agradables" (5, 10, 15...).

Con esto los intervalos serían: [10, 20) [20, 30) [30, 40) [40, 50) [50, 60) [60, 70) [70, 80) [80, 90].

Si partimos de la decisión de que los intervalos tengan 15 unidades de amplitud simplemente iniciaremos su construcción hasta llegar a un intervalo que contenga al valor máximo observado.

[10, 25) [25, 40) [40, 55) [55, 70) [70, 85) [85, 90].

Intervalo de clase	Marca de clase x	Recuento	n	N
$(e_1 \bullet e_2)$	x_1	///	n_1	n_1
$(e_2 \bullet e_3)$	x_2	///// ///// /	n_2	$n_1 + n_2$
…	…	…	…	…
$(e_i \bullet e_i + 1)$	x_i	///// ///	n_i	$n_1 + n_2 + \ldots + n_i$
…	…	…	…	…
			$\sum n_i = N$	

Fuente: elaboración propia.

Medidas de centralización o de posición

Las medidas de tendencia central son parámetros estadísticos que informan sobre el centro de la distribución de la muestra o población estadística. Las medidas de tendencia central son usadas para ubicar la media de un grupo de datos y estos pueden ser considerados como datos no agrupados o datos agrupados.

Supongamos que queremos saber cuál es el peso corporal que más frecuencia absoluta tiene y de este último podremos utilizar la moda. También estas medidas de tendencia central sirven para comparar —así como interpretar— los resultados obtenidos en relación con los distintos valores observados y para comparar los resultados con otros grupos.

Medidas de localización (posición)

Las medidas de posición son promedios y pueden ser de tendencia central o no. Son coeficientes de tipo promedio que tratan de representar una determinada distribución. Pueden ser de dos tipos:

1. Centrales:
a) Media:

    i. Aritmética
    ii. Ponderada
    iii. Geométrica
    iv. Armónica

b) Mediana
c) Moda

2. No centrales:

a) Cuantiles:

    i. Cuartiles
    ii. Deciles
    iii. Centiles o percentiles

Medidas de dispersión

Son complementarias de la posición en el sentido de que señalan la dispersión en conjunto de todos los datos de la distribución respecto de la medida o medidas de localización adoptadas.

1. *Medidas de dispersión absoluta*: recorrido.
2. *Medidas de dispersión relativa*: recorrido intercuartílico; desviación típica o estándar; varianza; error estándar.
3. Coeficiente de Variación de Pearson.
4. Diagrama de caja.

Medidas de forma

Estudian la asimetría-simetría y deformación (apuntamiento, aplastamiento) respecto de una distribución modelo denominada distribución NORMAL.

1. Coeficiente de Asimetría y Coeficiente de Curtosis.

Medidas de concentración

Estudian la concentración de una distribución frente a la uniformidad.

1. Índice de Concentración de Gini.
2. Curva de Lorenz.

La agrupación de los datos de las medidas de tendencia central se divide en:

1. **Datos no agrupados:** se caracterizan por ser ordenados, lineales o consecutivos.
2. **Datos agrupados:** se caracterizan por estar en un rango o intervalo de datos.

Medidas de tendencia central de datos no agrupados

Son aquellas representadas como un conjunto de indicadores estadísticos que van a mostrar hacia qué valores se agrupan los datos numéricos; en otras palabras, son medidas estadísticas que buscan resumir en un solo valor un conjunto de valores.

Las medidas de tendencia central son, entonces, medidas que permiten conocer cómo se comportan los datos, en una distribución normal, con las características de un conjunto de datos que más se repiten o que estén más hacia el centro del conjunto de datos. Si los representamos sobre una gráfica (Figura), se observarían como se presenta en la Figura 19.

Las tres medidas de tendencia central usadas con mayor frecuencia son (Cuadro 7):

1. **Media aritmética:** se suman todos los datos y se dividen entre el número de ellos.
2. **Mediana:** se acomodan todos los datos en orden ascendente y se encuentra el dato central, que está en medio de ellos.
3. **Moda:** es el dato que más se repite.

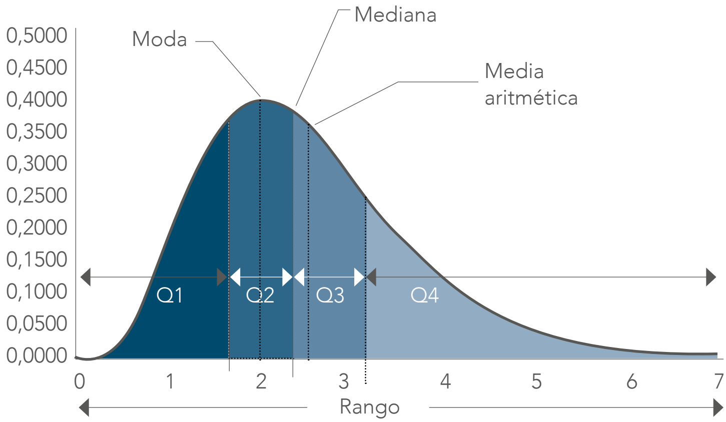


Figura 19. Representación de la frecuencia de datos que se distribuyen normalmente

Fuente: descargada de internet.

Media aritmética o promedio

Iniciaremos por "tratar" de diferenciar los dos términos empleados en la estadística descriptiva: "media" y "promedio". El primero (media) puede ser sinónimo de "promedio" o referirse a un tipo de cálculo totalmente distinto (media geométrica, ponderada o armónica), mientras que el segundo (promedio) se refiere a una operación específica del conjunto de dos o más valores de datos. La "media" suele definirse como media o media aritmética; en otras palabras, al referirnos como media implícitamente aceptamos que hablamos de una media aritmética. Entonces, la media es simplemente un método para describir el promedio de la muestra, que es calculada de la misma forma (misma ecuación) que el promedio; para todos los propósitos son lo mismo.

Es el presente libro abordaremos que la media es el valor promedio de un conjunto de datos numéricos, calculada como la suma del conjunto de valores dividida entre el número total de valores. En la media aritmética todos los valores son iguales de importantes.

A continuación, se muestra la ecuación de la media aritmética:

$$Media\ aritmética = \frac{\sum_{i=1}^{n} x_i}{N}.$$

La media aritmética es un tipo de media también llamada promedio, que otorga la misma ponderación a todos los valores; todos son ponderados como 1. La media aritmética se simboliza o se escribe como $\overline{X}$ ("x" testada) y es la suma de todos los valores de una variable dividida entre el número total de datos de los que se dispone, por ejemplo: sean $x1$, $x2$, $x3$... x_n los datos de un estudio estadístico, que es igual a la suma de todos los valores dividida por n datos:

Ejemplo: consideremos 10 pacientes con edades de 21, 32, 15, 59, 60, 61, 64, 60, 71 y 80 años; calcule la media de edad de los pacientes.

La media de edad de estos sujetos será: $\frac{\sum_{i=1}^{n} x_i}{N}$

Donde:

- » $\sum$ Sumatoria = Suma de los valores desde el primero hasta el *n*-ésimo valor. Existen 10 valores;
- » N = Número total de observaciones: edad de 10 pacientes;
- » *X* = Variable *x*, que es sobre la que se calcula la media aritmética. En este caso serían las edades de los pacientes; e
- » *i* = Posición de cada observación. En este ***ejemplo*** es la posición de cada edad: la edad 21 (1), la edad 32 (2), la edad 15 (3), la edad 59 (4) … edad *x* (*i*).

Cálculo de la media aritmética del ***ejemplo*** planteado:

$$\bar{X} = \frac{21 + 32 + 15 + 59 + 60 + 61 + 64 + 60 + 71 + 80}{10} = 52.30 \text{ años}$$

Propiedades de la media aritmética

- » **Propiedad 1.** La suma de las desviaciones de los valores de la variable respecto a la media es cero (0).
- » **Propiedad 2.** La media aritmética de los cuadrados de las desviaciones de los valores de la variable respecto a una constante cualquiera, se hace mínima cuando dicha constante coincide con la media aritmética.

» **Propiedad 3.** Si a todos los valores de la variable se les suma una misma cantidad, la media queda aumentada en dicha cantidad.
» **Propiedad 4.** Si todos los valores de la variable se multiplican por una misma constante, la media queda multiplicada por dicha constante.
» **Propiedad 5.** Si de un conjunto de valores se pueden obtener dos o más subconjuntos disjuntos, la media aritmética del conjunto se relaciona con la media aritmética de cada uno de los subconjuntos disjuntos.

Observaciones sobre la media aritmética

1. La media se puede hallar solo para variables cuantitativas.
2. La media es independiente de las amplitudes de los intervalos.
3. La media es muy sensible a las puntuaciones extremas.

Si tenemos una distribución con los siguientes pesos en kilogramos:

65, 69, 65, 72, 66, 75, 70 y 110 kg,

la media es igual a 74 kg, que es una medida de centralización poco representativa de la distribución.

Corolario: Si una variable es transformación lineal de otra variable (suma de un número y multiplicación por otro), la media aritmética de la primera variable sigue la misma transformación lineal respecto a la media aritmética de la segunda variable, siendo $y = ax_i + b$, donde a y b son números reales.

Media ponderada

Otorga diferentes pesos a los distintos valores sobre los que se calcula. Se diferencia de la media aritmética en que la ponderación no les da la misma importancia a los valores, es decir, cada dato podrá tener un valor diferente al anterior o el siguiente. Por ejemplo, al tomar el ejemplo de las edades de los 10 pacientes, en la media aritmética, los valores (edad) tienen la misma importancia y se representaría como: 21 (1), 32 (1), 15 (1), 59 (1), 60 (1), 61 (1), 64 (1), 60 (1), 71 (1) y 80 (1) años, y la edad se multiplica por el mismo valor (1).

Para la media ponderada se pondrá de ejemplo el cálculo de las calificaciones de una materia: la calificación para Patología tendría una ponderación del 20%, mientras que la calificación de Anatomía, una ponderación del 35%; de esta forma, cada asignatura tiene un valor diferente con base en su dificultad, especialización e importancia.

Ejemplo: examen de Patología, 8.4 (0.20); examen de Anatomía, 9.1 (0.35). Es importante mencionar que, si bien cada dato tiene una importancia diferente, puede tener una o varias iguales, pero no en su totalidad, como en el caso de la media aritmética.

Para calcular la media ponderada se deben hacer los siguientes pasos:

1. Multiplicar cada dato estadístico por su peso correspondiente: $\sum x_i * P_i$.
2. Sumar todos los productos calculados en el paso anterior.
3. Sumar todas las ponderaciones: $\sum p_i$.
4. Dividir la suma de todos los productos pesos por ponderación entre la suma de las ponderaciones: $\frac{\sum x_i * P_i}{\sum p_i}$.
5. El resultado obtenido es la media ponderada de la muestra estadística: $\overline{x_p}$.
6. La ecuación es: $\bar{X}_P = \frac{\sum x_i * P_i}{\sum p_i}$.

Media geométrica

Responde a la expresión $\sqrt[n]{\sum x_i}$ (radical) y se define como la raíz *n*-ésima del producto de todos los valores de la variable meramente positivos. La media geométrica se calcula como un producto conjunto; en otras palabras, todos los valores se multiplican entre sí. De modo que, si uno de ellos es cero, el producto total sería cero; otro factor que se debe tener en cuenta es que necesita números **únicamente positivos**. El logaritmo de la media geométrica es la media aritmética de los logaritmos de los valores de la variable. Uno de sus principales usos es para calcular medias sobre porcentajes e índice. La media geométrica tiene la ventaja de ser menos sensible a valores extremos (elevados o menores) que podrían alterar la media de una muestra estadística. Por el contrario, su desventaja es que no puede utilizar números negativos o ceros.

Para calcular la media geométrica se deben hacer los siguientes pasos:

1. Multiplicar cada dato estadístico correspondiente: $X_1 * X_2 * X_3 \dots X_n$.
2. Obtener el radical de n: $\sqrt[N]{x_1 \cdot x_2 \cdot \dots \cdot x_N}$.
3. El resultado obtenido es la media ponderada de la muestra estadística: $\bar{X}_G$.
4. La ecuación es: $\bar{X}_G = \sqrt[N]{x_1 \cdot x_2 \cdot \dots \cdot x_N}$.

Media armónica

La media armónica (H) es la inversa de la media aritmética, es decir, la media armónica es una medida estadística recíproca a la media aritmética, que es la suma de un conjunto de valores entre el número de observaciones. Se utiliza para promediar velocidades, tiempos, rendimientos o en el área de la electrónica. Sin embargo, su uso no está muy extendido en otras disciplinas. Su problema es cuando algún valor de la variable es cero o próximo a cero y no se puede calcular.

Cuadro 7. Diferentes medias empleadas en la Estadística Descriptiva

Media ponderada	Media geométrica	Media armónica
Es aplicable cuando a cada valor (x_i), se le asigna un peso (p_i): $\bar{X}_P = \frac{\sum x_i * P_i}{\sum p_i}$	$\bar{X}_G = \sqrt[N]{x_1 \cdot x_2 \cdot \dots \cdot x_N}$ Con frecuencia f_i para cada x_i: $(\mathrm{N} = \sum f_i)$	$\bar{X}_A = \frac{N}{\sum \left[\frac{1}{x_1}\right]}$ Con frecuencia f_i para cada x_i: $(\mathrm{N} = \sum f_i)$

Fuente: elaboración propia.

Mediana o media posicional

Es un estadístico de posición central que parte la distribución de los datos en dos, es decir, deja la misma cantidad de datos de un lado que de otro. Sin embargo, para calcular la mediana es importante que los datos estén ordenados de mayor a menor o viceversa: de menor a mayor. Por el contrario, la media aritmética puede estar desplazada hacia un lado u otro, según la distribución, pero la mediana siempre se sitúa en el centro de esta.

La ecuación no calcula el valor de la mediana, sino que nos da la posición en la que se encuentra el valor dentro del conjunto de datos. Con base en esto, se debe tener en cuenta si el número total de datos (n) es par o impar. En otras palabras, si los valores (x) se ordenan según sus valores numéricos y si n es un número impar, la mediana es la x que ocupa la posición central; y si n es un número par, la mediana es la media aritmética de las dos x centrales. Las ecuaciones para la mediana serán las siguientes:

1. Cuando el número de observaciones es impar:

Mediana = $(n + 1) / 2$ → Valor medio de la observación
Ejemplo: 5, 6, 6, 7, 8, 9, 10 → La mediana es 7.

2. Cuando el número de observaciones es par:

Mediana = $(n1 + n2) / 2$ → Media aritmética de las observaciones
Ejemplo: 5, 6, 6, 7, 8, 9, 10, 12 → La mediana es → 7 + 8 = 15 / 2 = 7.5.

Moda

Es el valor x que aparece con mayor frecuencia. Si dos o más x aparecen con igual máxima frecuencia es **bimodal**, siendo la moda las dos x que aparecen con más frecuencia; o es **trimodal** con las tres x más frecuentes; y si tiene más de tres modas, se le conoce como **multimodal**. Es importante mencionar que cuando las x solo tienen una frecuencia, se dice que el conjunto de las x no tiene moda.

Una característica de la moda para datos agrupados es que puede ser ubicada tanto en variables cualitativas (*variables categóricas*: ordinales y nominales) como en variables cuantitativas (*variables numéricas*: discretas y continuas), debido a que la moda es, en otras palabras, la frecuencia con la que una variable o "respuesta" aparece en un conjunto de datos.

Medidas de dispersión de datos no agrupados

Normalmente la estadística también se ocupa de la dispersión de la distribución, es decir, si los datos aparecen sobre todo alrededor de la media o si están distribuidos por todo el rango. En otras palabras, las medidas de dispersión son un conjunto de variables utilizadas para calcular el grado de la distribución de los datos y su relación entre variables en función de un valor de referencias, por ejemplo, la media aritmética. En otras palabras, expresan cómo se distribuyen los datos en torno a alguna de las medidas de centralización e indican si una variable se mueve mucho, poco, más o menos que otra. La razón de las medidas de dispersión es conocer de manera resumida una característica de la variable estudiada y, en este sentido, deben acompañar a las medidas de tendencia central. Ambas ofrecen información que podrá ser empleada para comparar y tomar decisiones.

Una medida de dispersión es la diferencia entre dos percentiles, por lo general entre el 25 y el 75. El percentil p es un número tal que un p por ciento de los datos son menores o iguales que p. En particular, los percentiles 25 y 75 se denominan cuartiles inferior y superior, respectivamente.

La desviación estándar es otra medida de dispersión, pero más útil que los percentiles, pues está definida en términos aritméticos, como se explica a continuación.

Las medidas de dispersión tratan de medir el grado de dispersión que una variable estadística tiene en torno a una medida de posición o tendencia central, indicándonos lo representativa que es la medida de posición. A mayor dispersión habrá menor representatividad de la medida de posición y viceversa.

Tipos y categorías de medidas de dispersión

Las medidas de dispersión son de dos tipos:

1. **Medidas de dispersión absoluta:** rango, desviación media, varianza y desviación típica, que se usan en los análisis estadísticos generales.
2. **Medidas de dispersión relativa:** determinan la dispersión de la distribución estadística independientemente de las unidades en las que se exprese la variable. Se trata de parámetros más técnicos y utilizados en estudios específicos, y entre ellas se encuentran los Coeficientes de Apertura, el rango relativo, el Coeficiente de Variación (Índice de Dispersión de Pearson) y el Índice de Dispersión Mediana.

Las medidas de dispersión absoluta generalmente se clasifican en cuatro categorías (Figura 20) y pueden variar según las necesidades particulares del investigador; son las siguientes:

1. *Rango de variación*: se trata de un número que indica la distancia (diferencia) entre el valor máximo y el menor de todos ellos. Dicho valor se toma de una población estadística determinada. Hay dos maneras de expresar esta medida:

a) La diferencia entre los valores mayor y menor.
b) Los valores mayor y menor del grupo.

Rango = X mayor — X menor.

2. *Varianza*: esta medida representa la variación que puede sufrir un conjunto de datos respecto a la media. Se define como el cociente entre la suma de los cuadrados de las desviaciones de los valores de la variable y el número de datos del estudio. Se refiere a la media aritmética de los cuadrados de las desviaciones de cada valor respecto de la media aritmética de los datos, es decir, es el promedio de las desviaciones de la media elevadas al cuadrado. La varianza se representa con el símbolo σ^2 (sigma cuadrada) para el universo o población y con el símbolo s^2 (s cuadrado) cuando se trata de la muestra.

Varianza para una población

$$\boxed{\sigma^2 = \frac{\sum_1^n (x_I - \bar{x})^2}{N}}$$

Varianza para una muestra

$$\boxed{s^2 = \frac{\sum_1^n (x_I - \bar{x})^2}{n-1}}$$

Donde:

σ^2 o s^2 = Variable sobre la que se pretende calcular la varianza;
X_i = Observación número *i* de la variable *X*; *i* puede tomar valores entre 1 y n;
N o n = Número de observaciones;
$\bar{x}$ = Media de la variable *X*.

¿Por qué se elevan al cuadrado las diferencias X_i - media?
La razón por la que las diferencias se elevan al cuadrado es que la suma de las diferencias X_i – media sería cero, la cual es una propiedad de la varianza. Para evitarlo, tal como ocurre con la desviación típica, se elevan al cuadrado. El resultado es la unidad de medida en la que se miden los datos, pero elevada al cuadrado.

3. *Desviación estándar o típica*: corresponde a una desviación que es "habitual" entre el valor y la media. Se define sencillamente como la raíz cuadrada de la varianza. Se trata de un evento más probable y, por lo tanto, se emplea como tal en el cálculo de dispersión. Se representa como la raíz cuadrada de la varianza, se simboliza por σ (sigma) cuando pertenece al universo o población y por s cuando pertenece a la muestra.

Desviación estándar para una población

$$\sigma = \sqrt{\frac{\sum_1^n (x_I - \bar{x})^2}{N}}$$

o

$$\sigma = \sqrt{\sigma^2}$$

Desviación estándar para una muestra

$$s = \sqrt{\frac{\sum_1^n (x_I - \bar{x})^2}{n-1}}$$

$$s = \sqrt{s^2}$$

Donde:

σ o s = Variable sobre la que se pretende calcular la desviación estándar;

X_i = Observación número *i* de la variable *X*; *i* puede tomar valores entre 1 y n;

N o n = Número de observaciones;

$\bar{x}$ = Media de la variable *X*.

4. *Coeficiente de Variación*: aunque está considerado como una dispersión relativa es muy utilizado en la estadística, así que el Coeficiente de Variación de Spearman, también conocido como Coeficiente de Variación de Pearson, es una medida estadística que ofrece información respecto a la relación de la media aritmética y la desviación estándar de un conjunto de datos. En otras palabras, es la variación de un conjunto de datos respecto de su media aritmética y, al mismo tiempo, las dispersiones que tienen los datos dispersos entre sí mismos. Esta medida de dispersión se expresa como un porcentaje, que es el resultado de dividir la desviación estándar sobre la media aritmética; sus siglas son CV.

$$\mathrm{CV} = \frac{\sigma}{\bar{X}} \, x \, 100$$

Donde:
CV = Coeficiente de Variación;
σ = Desviación estándar de la población;
S = Desviación estándar de la muestra;
$\bar{x}$ = Media aritmética.

Propiedades y aplicaciones del Coeficiente de Variación

1. El Coeficiente de Variación no tiene unidades.
2. El Coeficiente de Variación se expresa en porcentaje; sin embargo, solo en algunas ocasiones puede expresarse su valor numérico, es decir, 0 a 1 o incluso mayor que 1.
3. El Coeficiente de Variación depende de la desviación estándar y la media aritmética.
4. El Coeficiente de Variación es muy común en estadística y la probabilidad aplicada, empleándose para comparar las variaciones en distintos conjuntos de datos o distintas poblaciones.
5. *Error estándar de la media*: es otra medida de dispersión de gran utilidad en la estadística, pero inferencial, debido a que mide la variabilidad de la distribución muestral, es decir,

mide la variación o precisión que tiene la media muestral en comparación con la media de la población utilizando la desviación estándar. En general, cuanto mayor sea el tamaño de la muestra menor será el error estándar de la media.

Error estándar de la muestra para una población

$$\boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}}$$

Error estándar de la muestra para una muestra

$$\boxed{S_{\bar{x}} = \frac{s}{\sqrt{n}}}$$

Donde:

$\sigma_{\bar{x}}$ o $S_{\bar{x}}$ = Variable sobre la que se pretende calcular el error estándar de la muestra;

σ o s = Sigma con el que se pretende calcular la desviación estándar;

N o n = Número de observaciones.

Características de las medidas de dispersión

Son seis principales:

1. Las medidas de dispersión **indican qué tan diseminados se encuentran los datos de una distribución**.
2. Permiten **conocer qué tan cerca o lejos de la media, se encuentran los datos**.
3. Las medidas de variabilidad te dan la posibilidad de **saber la homogeneidad o heterogeneidad de las distribuciones de los datos**.
4. Su aplicación es **fácil y rápida**.
5. Sus valores de dispersión **siempre son positivos o cero**, en el caso de que estos sean iguales.
6. El uso de las medidas de dispersión **se puede aplicar en diversos ámbitos**, como el sector salud, industrial, económico, empresarial, etcétera.

MEDIDAS DE DISPERSIÓN

VARIANZA	DESVIACIÓN ESTÁNDAR
$\sigma^2 = \dfrac{\sum_1^N (x_i - \bar{X})^2}{N}$	$\sigma = \sqrt{\dfrac{\sum_1^N (x_i - \bar{X})^2}{N}}$

- **X** → Variable sobre la que se pretenden calcular la varianza.
- **x_i** → Observación número i de la variable X. i puede tomará valores entre 1 y n.
- **N** → Número de observaciones.
- **$\bar{x}$** → Es la media de la variable X.

RANGO ESTADÍSTICO	COEFICIENTE DE VARIACIÓN
$R = Máx_x - Mín_x$	$CV = \dfrac{\sigma_x}{\lvert \bar{X} \rvert}$
•**R** → Es el rango. •**Máx** → Es el valor máximo de la muestra o población. •**Mín** → Es el valor mínimo de la muestra o población estadística. •**x** → Es la variable sobre la que se pretende calcular esta medida.	•**X** → Variable sobre la que se pretenden calcular la varianza. •**σ_x** → Desviación típica de la variable X. •**$\lvert \bar{x} \rvert$** → Es la media de la variable X en valor absoluto con $\bar{x} \neq 0$.

Figura 20. Categorías de las medidas de dispersión de datos no agrupados

Fuente: elaboración propia.

Medidas de posición no central de datos no agrupados

Las medidas de posición no central son los cuantiles, que son aquellas que pueden ser divididas en dos partes iguales en un conjunto de datos, que al ser ordenados de forma ascendente o descendente ubican la medida de posición llamada mediana, que parte la distribución en dos (50%), es decir, deja la misma cantidad de valores de un lado que de otro. La mediana se puede dividir en cuatro partes iguales y a esto se le denomina cuartiles; cuando se divide en diez partes iguales, se le denomina deciles; y cuando se divide en cien partes iguales, se le denomina percentiles. De esta forma, reflejan los valores superiores, medios e inferiores.

Cuartiles

Son usados en la estadística descriptiva y se refieren a las medidas de posición no central que permiten reconocer otros puntos característicos de la distribución, los cuales no son centrales. En otras palabras, los cuartiles son los tres valores de la variable que dividen a un conjunto de datos ordenados en cuatro partes iguales (de 25% cada parte). De manera que para resolver un problema sobre cuartiles solamente tenemos que hallar: Q1; Q2 y Q3.

Así que tenemos que los datos menores a Q1 representan el 25% de los datos, los que están debajo de Q2 son el 50%, mientras que aquellos menores a Q3 son el 75%.

El concepto de cuartil es propio de la estadística descriptiva y es de gran utilidad para el análisis de datos.

La ecuación para calcular la posición de un cuartil es:

$$Q1 = \frac{(n+1)}{4}; \quad Q2 = \frac{2(n+1)}{4}; \quad Q3 = \frac{3(n+1)}{4}$$

El cuartil 2 equivale al 50% y es igual a la mediana.

Donde:
Q1; Q2; Q3 = Cuartil;
n = Total de datos.

Una de las características importantes del cuartil es que, al ordenar los datos en forma ascendente, el Q2 es la mediana, que puede ser calculada de la misma forma, donde si la cantidad de datos es impar, el valor que está en el centro será la mediana, es decir, Q2. Sin embargo, si la cantidad de datos es par, se suman los 2 datos centrales y se divide entre 2, que será igual a la mediana, Q2. Solamente habrá que calcular Q1 y Q3.

Deciles

Son los nueve valores de la variable que dividen un conjunto de datos ordenados en 10 partes iguales correspondientes al 10%, 20%... y hasta el 90% (de 10% cada parte). De manera que para resolver un problema sobre deciles solamente tenemos que hallar: D1; D2; D3; D4; ... D9.

El quinto decil coincide con la mediana: D5 = Me y también coincide con el segundo cuartil: D5 = Q2 Me.

La ecuación para calcular la posición de un decil es:

$$D1 = \frac{(n+1)}{10}; \quad D5 = \frac{5(n+1)}{10}; \quad D9 = \frac{9(n+1)}{10}$$

El decil 5 equivale al 50% y es igual a la mediana.

Donde:
D1 … D9 = Decil;
n = Total de datos.

Características del decil

Los deciles tienen una serie de características y estas son:

- Como estadísticos de posición son útiles para conocer qué lugar ocupan los datos en una distribución. Así, el decil 8 es el límite superior de los datos que representan el 80% del total.
- Permiten conocer cuáles de ellos se sitúan en los niveles más altos (> 90%) y en los más bajos (< 10%).
- Son muy frecuentes en comparaciones económicas, como niveles de renta, salarios o ingresos.
- Además, junto a los cuartiles, quintiles o percentiles, son los cuantiles más utilizados en la estadística descriptiva.

Percentiles

Son una medida estadística de posición, que divide la distribución ordenada de los datos en 100 partes iguales (de 1% cada parte). De manera que para resolver un problema sobre percentiles solamente tenemos que hallar: P1; P2; P3; P4; ... P99. Esta medida de posición no central aporta información sobre el porcentaje de observaciones de una variable, ordenados de menor a mayor, que se sitúan por debajo del valor de este.

La ecuación para calcular la posición de un percentil es:

$$P1 = \frac{(n+1)}{100}; \quad P50 = \frac{50(n+1)}{100}; \quad P99 = \frac{99(n+1)}{100}$$

El percentil 50 equivale al 50% y es igual a la mediana.

Donde:
P1 … P99 = Percentil;
n = Total de datos.

Características del percentil

Las características más relevantes del percentil son:

- Es similar a otras medidas de posición no central. Por tanto, nos informa sobre la posición de un dato respecto a otros.
- Aporta información más detallada que otros. Por ejemplo, algunos índices de impacto de revistas científicas lo utilizan en lugar del cuartil.
- Es útil para agrupar una gran cantidad de datos. Cuando se trabaja con muchos casos, los otros cuantiles pueden arrojar grupos demasiado numerosos y difíciles de interpretar.
- *Inconveniente relacionado con el punto anterior*: no es útil para muestras con pocos casos, ya que los grupos serían demasiado pequeños. Por eso en estas circunstancias se recomienda emplear otros, como el cuartil o el decil.

Medidas de tendencia central de datos agrupados

Las medidas de tendencia central para un grupo de datos se utilizan para describir, ubicar e identificar el comportamiento del punto (valor) alrededor del cual se centran los datos recabados del grupo. En otras palabras, los datos agrupados son los que se clasifican en categorías (variables cualitativas) o clases (variables cuantitativas), que es el número de subconjuntos, tomando como criterio su frecuencia. Se trata de hacer un análisis al simplificar —y hacer fácil— el manejo de grandes cantidades de datos y tener un mayor orden de estos, para poder así establecer cuáles son sus tendencias.

Los datos agrupados son organizados en una tabla de frecuencia, como ya se ha visto en otra sección del presente libro. Una vez que los datos han sido reunidos y tabulados comienza el análisis, con el objeto de calcular un número único que represente o resuma todos los datos. Dado que, por lo general, la frecuencia de los intervalos centrales es mayor que el resto, este número se suele denominar valor o medida de *tendencia central*. Se representan como la frecuencia, es decir, el número de veces que se repite un dato, que es indispensable para identificar la moda. Y el rango medio, $\frac{Valor\ más\ alto + Valor\ más\ bajo}{2}$, indica el centro de los datos desde otra perspectiva, como el promedio de los valores mayor y menor, que también se conoce como medio rango.

Media aritmética para datos agrupados

La media es la más utilizada para caracterizar valores numéricos. **En datos agrupados, la ecuación ser representa como:**

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i . f_i}{N}$$

Donde:

$\overline{X}$ =	Media aritmética de datos agrupados;
f_i =	Frecuencias de la clase o frecuencias absolutas;
$\sum_{i=1}^{n} x_i \cdot f_i$ =	Sumatoria del producto punto medio o marca de clase por su frecuencia;
N =	Suma total de las frecuencias.

A continuación, se presenta un ***ejemplo*** que corresponde a los datos de las edades de individuos del ejercicio encontrados en la sección de Datos no agrupados:

Se agruparon las edades de 42 individuos (en años) en la siguiente tabla de frecuencias. Calcule la media aritmética de estos.

Intervalo de la media

Número de clase	Intervalo	Clase	Frecuencia absoluta	Producto
	Li Ls	x_i	f_i	$x_i \cdot f_i$
1	[10, 20)	15	1	15
2	[20, 30)	25	8	200
3	[30, 40)	35	10	350
4	**[40, 50)**	**45**	**9**	**405**
5	[50, 60)	55	8	440
6	[60, 70)	65	4	260
7	[70, 80)	75	2	150
		N =	**42**	**1,820**

Ecuación: $\bar{X} = \frac{\sum_{i=1}^{n} x_i \cdot f_i}{N} = \frac{1,820}{42} = 43.33.$ El resultado se ubica en el intervalo que contenga la $\overline{X}$ = 43, siendo este el 4.° número de clase de la tabla de frecuencias, es decir, el intervalo entre 40 a 50 años.

Media ponderada para datos agrupados

La media ponderada es una medida de posición central de la estadística descriptiva, que se usa cuando en un conjunto de datos todos o algunos de los datos tienen un peso distinto. Para calcular la media aritmética ponderada para datos agrupados, se siguen los siguientes pasos:

1. Se determina el punto medio de cada clase.
2. Se multiplica el punto medio de cada clase con la frecuencia absoluta o relativa.
3. Se suman los productos anteriores y se divide el resultado entre la suma total de frecuencias absolutas o relativas, es decir, el total de los datos.
4. Se divide el total obtenido entre la suma de los pesos.

Ecuación de media ponderada para datos agrupados

$$\overline{Xp} = \frac{\sum_{i=1}^{n} x_i \cdot w_i \cdot f_i}{\sum_{i=1}^{n} w_i \cdot f_i}.$$

Ejemplo de media ponderada para datos agrupados

Un evaluador desea ponderar una calificación en función de tres exámenes; sin embargo, estos poseen distintos pesos (w_i), dado que tienen diferentes niveles de importancia. Esta importancia, peso w_i, que asignamos a cada valor, es independiente de la frecuencia absoluta que se tenga.

Los resultados del examen obtenidos por el primer estudiante fueron: 9, 8, 7. Los exámenes tuvieron una duración de 1, 2 y 3 horas, respectivamente. Tomándose en cuenta esa ponderación calcular la calificación media ponderada.

Resultado	Frecuencia	Duración ponderación			
x_i	f_i	w_i	$x_i \cdot f_i$	$w_i \cdot f_i$	$x_i \cdot f_i \cdot w_i$
9	1	1	9	1	9
8	1	2	8	2	16
7	1	3	7	3	21
N =	**3**	**6**	**24**	**6**	**46**

Media aritmética de datos agrupados

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i \cdot f_i}{N} = \frac{24}{3} = \mathbf{8.}$$

Media ponderada de datos agrupados

$$\overline{Xp} = \frac{\sum_{i=1}^{n} x_i \cdot w_i \cdot f_i}{\sum_{i=1}^{n} w_i \cdot f_i} = \frac{46}{6} = \mathbf{7.6.}$$

Mediana para datos agrupados

La mediana se encuentra en el intervalo donde la frecuencia acumulada llega hasta la mitad de la suma de las frecuencias absolutas. Cuando se tienen datos agrupados no se pueden ordenar los números individualmente, sino que estos se representan en intervalos con su respectiva frecuencia; por ejemplo, el intervalo [10 – 20) tiene una frecuencia de 4, es decir, se ubican 4 individuos entre el 10 y el 20, y no se conoce con exactitud qué valor es o cuáles se repiten. Entonces, la mediana en datos agrupados determina un dato estimado de en dónde puede estar la mediana y, por lo tanto, para encontrar la mediana de datos agrupados se tiene que utilizar otra ecuación:

$$Me = L_i + \frac{\frac{N}{2} - F_{i-1}}{f_i} * a_i$$

Donde:

Me = Mediana de datos agrupados;

L_i = Límite inferior del intervalo donde se encuentra la frecuencia acumulada mayor a $\frac{N}{2}$;

$\frac{N}{2}$ = Suma de todas las frecuencias clase o frecuencias absolutas dividida entre 2;

F_{i-1} = Frecuencia acumulada del intervalo anterior al intervalo que contiene la mediana;

f_i = Frecuencia del intervalo en el intervalo de la mediana;

a_i = Amplitud que tiene el intervalo de la mediana, que se obtiene restando el límite superior del límite inferior del intervalo que contiene la mediana.

En otras palabras, se puede precisar la mediana de un grupo de datos agrupados como el valor que se encuentra en el intervalo donde la frecuencia acumulada llega hasta la mitad de la suma de las frecuencias del intervalo o frecuencias absolutas.

A continuación, se presenta un ***ejemplo*** y los pasos para obtener la mediana:

Se agruparon las edades de 42 individuos (en años) en la siguiente tabla de frecuencias. Calcule la mediana de estos.

Número de clase	Intervalo de clase	Clase	Frecuencia absoluta	F_i acumulado
	Li *Ls*	x_i	f_i	F_i
1	[10, 20)	15	1	1
2	[20, 30)	25	8	9
3	[30, 40)	35	10	19
4	**[40, 50)**	**45**	**9**	**28**
5	[50, 60)	55	8	36
6	[60, 70)	65	4	40
7	[70, 80)	75	2	42
		N =	**42**	

Intervalo de la mediana

Como se observa en la tabla anterior, se colocan las frecuencias del intervalo (frecuencia absoluta) y en la siguiente columna, los valores acumulados de dichas frecuencias del intervalo. A continuación, se divide N entre 2 para ubicar el intervalo de clase donde se encuentra la mediana, debido a que la mediana es el valor central.

Lo siguiente es ubicar en la columna de frecuencias acumuladas (F_i) el valor obtenido, es decir, 21 o el valor inmediato superior al valor obtenido; el intervalo de clase de la mediana en este caso es: [40, 50). Ahora se usará la **ecuación de la mediana para datos agrupados:**

$$Me = L_i + \frac{\frac{N}{2} - F_{i-1}}{f_i} * a_i$$

Sustituyendo las variables por los valores siguientes:

$$Me = 40 + \frac{21 - 19}{9} * 10 = 40 + \frac{2}{9} * 10 = 40 + 0.222 * 10 = 42.222.$$

La mediana de los datos agrupados, entonces, se ubica en el valor 42.222 del intervalo de clase de 40 a 50 años.

Moda para datos agrupados

La moda —como la conocemos— en un conjunto de datos o datos no agrupados cambia la forma de encontrarla cuando se organizan datos agrupados, porque no hay forma de contar cuántas veces aparece un número entre los resultados organizados en intervalos de clase. Hay que precisar que la ubicación de la moda en datos agrupados no es más que el número o variable que más veces se repite en un conjunto de datos; sin embargo, al no tener un número específico, sino que se ubica en un intervalo, este no se puede contar uno a uno y, por lo tanto, es muy probable que la moda que se obtenga no sea un número que esté en los datos recopilados, ya que el número que se calcula en la moda para datos agrupados simplemente es una estimación de un número que se aproxima al valor exacto de la moda, empero puede ser que en algunos casos la moda sí coincida con los datos, pero esto no sucederá siempre.

Otro punto a considerar de la moda para datos agrupados es que se necesita un proceso con un número de por medio; por lo tanto, para encontrar la moda de datos agrupados se tiene que usar otra **ecuación:**

$$Mo = L_i + \frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} * a_i$$

Donde:

Mo = Moda de datos agrupados;

L_i = Límite inferior del intervalo donde se encuentra el intervalo o la clase modal;

f_i = Frecuencia absoluta del intervalo o clase modal;

f_{i-1} = Frecuencia absoluta del intervalo anterior al intervalo o clase modal;

f_{i+1} = Frecuencia absoluta del intervalo posterior al intervalo o clase modal;

a_i = Amplitud que tiene el intervalo, que se obtiene restando el límite superior del límite inferior del intervalo que contiene el intervalo o clase modal.

También, se utiliza **otra fórmula de la moda** que proporciona un valor aproximado de esta. Es importante mencionar que si bien esta ecuación está disponible no genera más que una aproximación de la moda:

$$Mo = L_i + \frac{f_{i+1}}{f_{i-1} + f_{i+1}} * a_i$$

Donde:

Mo = Moda de datos agrupados;

L_i = Límite inferior del intervalo donde se encuentra el intervalo o clase modal;

f_{i-1} = Frecuencia absoluta del intervalo anterior al intervalo o clase modal;

f_{i+1} = Frecuencia absoluta del intervalo posterior al intervalo o clase modal;

a_i = Amplitud que tiene el intervalo, que se obtiene restando el límite superior del límite inferior del intervalo que contiene el intervalo o clase modal.

A continuación, se presenta un ***ejemplo*** y los pasos para obtener la moda:

Tomando el mismo ***ejemplo***, se agruparon las edades de 42 individuos (en años) en la siguiente tabla de frecuencias. Calcule la moda de estos.

Intervalo modal

Número de clase	Intervalo de clase	Clase	Frecuencia absoluta	F_i acumulado
	Li *Ls*	x_i	f_i	F_i
1	[10, 20)	15	1	1
2	[20, 30)	25	8 ←	9
3	**[30, 40)**	**35**	→ **10**	**19**
4	[40, 50)	45	9 ←	28
5	[50, 60)	55	8	36
6	[60, 70)	65	4	40
7	[70, 80)	75	2	42
		N =	**42**	

Como se observa en la tabla anterior, se colocan las frecuencias del intervalo (frecuencia absoluta) y en la siguiente columna, los valores acumulados de dichas frecuencias del intervalo. A continuación, se busca en la frecuencia absoluta (f_i) el mayor número y esta corresponde a la clase modal, y es la f_i = 10 la que posee la mayor frecuencia (véase: flecha color negro). Teniendo identificado el intervalo o clase modal usaremos la **ecuación**:

$$Mo = L_i + \frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} * a_i = 30 + \frac{10-8}{(10-8)+(10-9)} * 10 = 36.66 \text{ años}$$

Donde:
Mo = Moda de datos agrupados;
L_i = 30;
f_i = 11;
f_{i-1} = 8;
f_{i+1} = 9;
a_i = 30 – 40 = 10.

La moda será, entonces, 36 años que se encontrará en el número de clase 3, donde se observa en el intervalo de clase de 30 a 40 años. Como ya se ha mencionado, la ubicación de la moda en datos agrupados no tiene un número específico, sino que se ubica en un intervalo; por lo tanto, es muy probable que la moda que se obtenga no sea un número que esté en los datos recopilados, ya que el número que se calcula en la moda para datos agrupados simplemente es una estimación de un número que se aproxima al valor exacto de la moda, empero puede ser que en algunos casos la moda sí coincida con los datos, pero esto no sucederá siempre.

Propiedades de la moda para datos agrupados

Sus propiedades principales son:

- Cálculo sencillo.
- Interpretación muy clara.
- Al depender solo de las frecuencias, puede calcularse para variables cualitativas. Es por ello el parámetro más utilizado para resumir una población donde no es posible realizar otros cálculos; por ejemplo, cuando se enumeran en medios periodísticos las características más frecuentes de determinado sector social. Esto se conoce informalmente como "retrato robot".

Sus inconvenientes principales son:

- Su valor es independiente de la mayor parte de los datos, lo que la hace muy sensible a variaciones muestrales. Por otra parte, en variables agrupadas en intervalos, su valor depende excesivamente del número de intervalos y de su amplitud.
- Emplea muy pocas observaciones, de tal modo que grandes variaciones en los datos fuera de la moda no afectan en modo alguno a su valor.
- No siempre se sitúa hacia el centro de la distribución.
- Puede haber más de una moda en el caso en el que dos o más valores de la variable presenten la misma frecuencia (distribuciones bimodales o multimodales).

Medidas de dispersión de datos agrupados

Normalmente la estadística también se ocupa de la *dispersión* de la distribución, es decir, si los datos aparecen sobre todo alrededor de la media o si están distribuidos por todo el rango. Una medida de dispersión es la diferencia entre dos **percentiles**, por lo general entre el 25 y el 75. El percentil *p* es un número tal que un *p* por ciento de los datos son menores o iguales que *p*. En particular, los percentiles 25 y 75 se denominan **cuartiles** inferior y superior, respectivamente. La **desviación estándar** es otra medida de dispersión, pero más útil que los percentiles, pues está definida en términos aritméticos, como se explica a continuación. La **desviación** de un elemento del conjunto es su diferencia respecto a la media; por ejemplo, en la sucesión $x_1, x_2, \ldots, x_n$ la desviación de x_1 es: $x_1 - \sigma$ y el cuadrado de la desviación es: $(x_i - \sigma)^2$. La **varianza** es la media del cuadrado de las desviaciones. Por último, la desviación estándar, representada por la letra griega sigma (σ o S), es la raíz cuadrada de la varianza y se calcula de la siguiente manera: si la desviación estándar es pequeña, los datos están agrupados cerca de la media; si es grande, están muy dispersos.

Medidas de dispersión de datos agrupados

Las medidas de dispersión tratan de medir el grado de dispersión que una variable estadística tiene en torno a una medida de posición o tendencia central, indicándonos lo representativa que es la medida de posición. A mayor dispersión habrá menor representatividad de la medida de posición y viceversa.

Como ya se ha visto en las medidas de dispersión para datos no agrupados, las medidas de dispersión de datos agrupados tienen una variante en su ecuación, que permite calcular el valor de la dispersión en un intervalo de clase; en otras palabras, no posicionan un valor de la dispersión, sino un intervalo donde se ubica la dispersión calculada.

Varianza para datos agrupados

La varianza es la media aritmética del cuadrado de las desviaciones respecto a la media de una distribución estadística.

Las **ecuaciones para calcular la varianza de una población y de una muestra para datos agrupados** son las siguientes:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2 * f_i}{N} \qquad S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2 * f_i}{n-1}$$

Donde:

σ^2 o S^2 = Varianza de la población o de la muestra;
X_i = Marca de clase;
$\bar{X}$ = Media aritmética;
f_i = Frecuencia absoluta;
n = Tamaño de la muestra;
N = Tamaño de la población.

Es importante mencionar que para calcular la varianza de una población (σ^2), se divide entre N.

Propiedades de la varianza para datos agrupados

- » Siempre es un valor no negativo, que puede ser igual o distinto de cero.
- » La varianza es la media de la dispersión cuadrática óptima por ser la menor de todas.
- » Si a todos los valores de la variable se le suma una constante, la varianza no se modifica.
- » Si todos los valores de la varianza se multiplican por una constante, la varianza queda multiplicada por el cuadrado de dicha constante.

A continuación, se presenta un ***ejemplo*** y los pasos para obtener la varianza.

Se agruparon las edades de 42 individuos (en años) en la siguiente tabla de frecuencias. Calcule la varianza de estos.

Número de clase	Intervalo	Clase	Frecuencia	Producto	Diferencia		
	Li Ls	x_i	f_i	$x_i \cdot f_i$	$x_i - \bar{X}$	$(x_i - \bar{X})^2$	$(x_i - \bar{X})^2 \cdot f_i$
1	[10, 20)	15	1	15	-28.33	802.59	802.58
2	[20, 30)	25	8	200	-18.33	335.99	2,687.91
3	[30, 40)	35	10	350	-8.33	69.39	693.88
4	[40, 50)	45	9	405	1.67	2.79	25.10
5	[50, 60)	55	8	440	11.67	136.19	1,089.51
6	[60, 70)	65	4	260	21.67	469.59	1,878.35
7	[70, 80)	75	2	150	31.67	1,002.99	2,005.97
	∑ =		**42**	**1,820**	**11.69**	**2,819.52**	**9,183.33**

Como se observa en la tabla anterior, se colocan las frecuencias del intervalo (frecuencia absoluta f_i) y marca de clase (x_i), que se obtiene como valor promedio del intervalo de clase [(*Li* + *Ls*) / 2]. La **ecuación de la varianza**, $S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2 * f_i}{n-1}$, nos dice que debemos obtener primero la media aritmética para datos agrupados, que es:

$\bar{X} = \frac{\sum_{i=1}^{n} x_i \cdot f_i}{N}$. Se agrega la columna de la sumatoria del producto $x_i \cdot f_i$, que es $= \frac{1,820}{42} = 43.33$. Entonces, se tiene que $S^2 = \frac{9,183.33}{42-1} = 223.98$ años.

Desviación estándar para datos agrupados

La desviación estándar o desviación típica (σ) es una medida de centralización o dispersión para variables de razón (ratio o cociente) y de intervalo, de gran utilidad en la Estadística Descriptiva. Se define como la raíz cuadrada de la varianza. Junto con este valor, la desviación estándar es una medida (cuadrática) que informa de la media de distancias que tienen los datos respecto de su media aritmética, expresada en las mismas unidades que la variable.

La **ecuación para calcular la desviación estándar de una muestra para datos agrupados**, es la siguiente:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{X})^2 \cdot f_i}{N}}$$

Donde:

σ o S =	Desviación estándar de la población o de la muestra;
x_i =	Marca de clase;
$\bar{X}$ =	Media aritmética;
f_i =	Frecuencia absoluta;
n =	Tamaño de la muestra;
N =	Tamaño de la población.

Es importante mencionar que para calcular la desviación estándar de una población (σ), se divide entre N.

Para continuar con el cálculo de la desviación estándar de una muestra, se tomará el ejemplo anterior donde se obtuvo que la varianza, siendo esta 223.98, al obtener la raíz cuadrada, $S^2 = \sqrt{223.98}$, nos genera la desviación estándar. Entonces, la desviación estándar de la muestra (S) = 14.97 años.

Coeficiente de Variación

El Coeficiente de Correlación de Spearman es exactamente el mismo que el Coeficiente de Correlación de Pearson. Ambos realizan cálculos sobre las observaciones respecto a la desviación estándar entre la media aritmética y se expresa en porcentaje. Sin embargo, el Coeficiente de Correlación de Pearson es mayormente conocido en la correlación estimada entre X e Y para hallar el Coeficiente de Correlación de Pearson para un conjunto de rangos apareados. Entonces, el Coeficiente de Variación, que se denomina por sus siglas CV, es una medida estadística que ofrece información respecto de la dispersión relativa de un conjunto de datos relacionando la media aritmética y la desviación estándar de un conjunto de datos. En otras palabras, el Coeficiente de Variación mide la dispersión entre datos respecto de su media aritmética.

No hay una ecuación del CV para datos agrupados, pues se usa la misma que se ha presentado en datos no agrupados.

El CV es empleado cuando se quieren comparar conjuntos de datos con medias aritméticas diferentes, debido a que la desviación estándar resulta insuficiente para realizar un análisis de los datos dispersos de una población o muestra con diferentes enfoques de una situación determinada.

El CV es útil, por ejemplo, para conocer la variación del peso corporal de una parvada ubicada en diferentes casetas o granjas, o por estación del año, entre otras variables. El CV ayuda a tomar decisiones acerca de sucesos que se quieren comparar y tiene la ventaja de que se puede representar en porcentaje y sin unidades.

Error estándar de la media

Esta medida de dispersión, de gran utilidad en la estadística pero inferencial, es el error estándar de la media para datos agrupados, que se calcula de la misma forma que se presentó para datos no agrupados y tiene las mismas propiedades.

Coeficiente de Curtosis

Recibe también el nombre de Coeficiente de Concentración Central. Mide el grado de aplastamiento o apuntamiento de la gráfica de la distribución de la variable estadística. Una mayor concentración de datos en torno al promedio hará que la forma sea alargada, siendo un tanto más plana (o aplastada) cuanto mayor sea la dispersión de estos (Figura 21).

Determina la forma de la distribución en relación con su grado de aplastamiento:

$$K = \frac{\frac{\sum n_i \cdot (x_i - \bar{x})^4}{N}}{\sigma^4} - 3$$

Interpretación

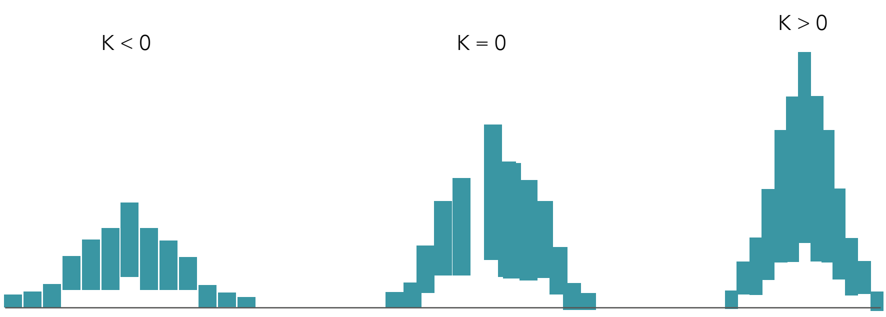


Figura 21. Gráficas de curtosis en relación con su valor K

Fuente: elaboración propia.

Curva de Lorenz

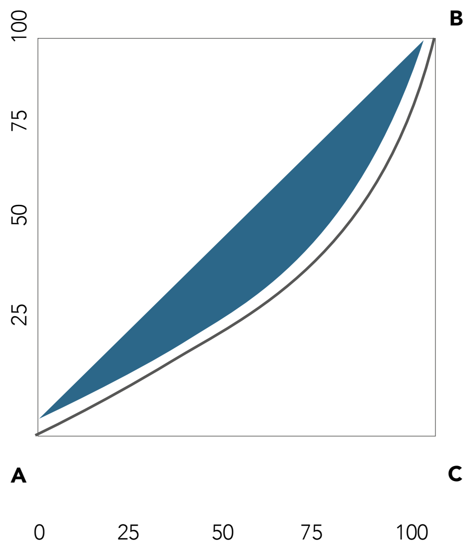


Figura 22. Curva de Lorenz

Fuente: elaboración propia.

Sobre un rectángulo de 100 unidades de lado, se dibuja la poligonal que resulta de unir los puntos (Pi, Qi).

Esta poligonal (curva de Lorenz) determina con la diagonal AB un recinto (sombreado en la figura) que mide el grado de concentración. Cuando el área sombreada es muy pequeña (la curva de Lorenz se aproxima a la diagonal AB), se presenta una baja concentración o, lo que es lo mismo, indica uniformidad en el reparto de los valores de la variable.

La mayor concentración se producirá cuando la zona sombreada coincida con el triángulo ABC.

Índice de Concentración de Gini

Haciendo uso de la tabla de cálculos anterior, necesaria para la obtención de la curva de Lorenz, definiremos el presente estadístico. Otros, como el Índice de Dalton, el de Paridad, etcétera, pueden ser empleados con idéntica interpretación a la que tratamos con el Índice de Gini, si bien omitimos su estudio.

$G = \frac{\sum_{i=1}^{k-1}(P_i - Q_i}{\sum_{i=1}^{k-1} P_i} = \frac{TD}{TP - 100}$ El Índice de Gini (expresión de la izquierda) coincide geométricamente con el cociente entre el área sombreada (definida por la curva de Lorenz) y la del triángulo ABC.

- Concentración mínima: G = 0
- Concentración máxima: G = 1.

Medidas de posición no central de datos agrupados

Las medidas de posición no central son los cuantiles, los cuales dividen de forma igual la distribución ordenada (de menor a mayor) de un conjunto de datos, reflejando los valores superiores, medios e inferiores, que permiten conocer otros puntos característicos de la distribución que no son los valores centrales.

Estas medidas de posicionamiento son representadas en una tabla de frecuencias y sus valores son calculados de forma diferente a los ya mencionados en datos no agrupados.

Cuartiles para datos agrupados

La **ecuación para calcular el cuartil de datos agrupados** es la siguiente:

$$Q_k = L_i + \frac{\frac{k * N}{4} - F_{i-1}}{f_i} * a_i$$

Donde:

Q_k = Cuartil para datos agrupados; k = 1, 2, 3;

L_i = Límite inferior del intervalo donde se encuentra el cuartil;

F_{i-1} = Frecuencia acumulada del intervalo anterior al intervalo o clase del cuartil;

f_i = Frecuencia absoluta del intervalo o clase del cuartil;

N = Suma de las frecuencias absolutas;

a_i = Amplitud que tiene el intervalo, que se obtiene restando el límite superior del límite inferior del intervalo que contiene el intervalo o clase del cuartil.

Se continuará con el mismo problema anterior de las edades. Se tiene, entonces, que hallar el cuartil (25% de las edades). Primero se debe encontrar la mayor frecuencia absoluta donde se halla el cuartil de interés (ejemplo: Q_1) usando la siguiente ecuación para el cálculo del Q_1, que es: $\frac{k * N}{4}$. Al sustituir los valores para calcular el primer cuartil es: $\frac{1 * 42}{4} = 10.5$; se ubica esta frecuencia acumulada y se encuentra en el número de clase 3 con una frecuencia absoluta de 10. Cuando se quiera hallar el Q_2 o Q_3, entonces $k = 2$ o 3, respectivamente, para ubicar la frecuencia acumulada correspondiente y continuar con el procedimiento.

Número de clase	Intervalo de clase	Clase	Frecuencia absoluta	F_i acumulado
	Li *Ls*	x_i	f_i	F_i
1	[10, 20)	15	1	1
2	[20, 30)	25	8	9
3	**[30, 40)**	**35**	**10**	**19**
4	[40, 50)	45	9	28
5	[50, 60)	55	8	36
6	[60, 70)	65	4	40
7	[70, 80)	75	2	42
		N =	**42**	

F_i

Entonces, se tiene que la ecuación para calcular el cuartil es: $Q_k = L_i + \frac{\frac{k * N}{4} - F_{i-1}}{f_i} * a_i$. Al sustituir los valores entonces queda: $Q_1 = 30 + \frac{\frac{1 * 42}{4} - 9}{10} * 10 = 31.5$

La respuesta es 31.5, donde se ubica el Q_1, es decir, el 25% de los individuos tiene 31.5 años.

Deciles para datos agrupados

Para hallar los deciles (D) para datos agrupados, se aplica la siguiente ecuación:

$$D_k = L_i + \frac{\frac{k * N}{10} - F_{i-1}}{f_i} * a_i$$

Donde:

D_k = Decil para datos agrupados; k = 1, 2, 3, …, 9;
L_i = Límite inferior del intervalo donde se encuentra el decil;
F_{i-1} = Frecuencia acumulada del intervalo anterior al intervalo o clase del decil;
f_i = Frecuencia absoluta del intervalo o clase del decil;
N = Suma de las frecuencias absolutas;
a_i = Amplitud que tiene el intervalo, que se obtiene restando el límite superior del límite inferior del intervalo que contiene el intervalo o clase del decil.

Se continuará con el mismo problema anterior de las edades. Se tiene, entonces, que hallar el decil (50% de las edades). Primero se debe encontrar la mayor frecuencia absoluta donde se halla el decil de interés usando la siguiente ecuación para el cálculo del D_1, que es: $\frac{k * N}{10}$. Al sustituir los valores para calcular el quinto decil es: $\frac{5 * 42}{10} = 21$; se ubica esta frecuencia acumulada y se encuentra en el número de clase 4 con una frecuencia absoluta de 9. Cuando se quiera hallar el D_2, D_3 … D_k, entonces k = 2, 3 … k, respectivamente, para ubicar la frecuencia acumulada correspondiente y continuar con el procedimiento.

Número de clase	Intervalo de clase	Clase	Frecuencia absoluta	F_i acumulado
	Li *Ls*	x_i	f_i	F_i
1	[10, 20)	15	1	1
2	[20, 30)	25	8	9
3	[30, 40)	35	10	19
4	**[40, 50)**	**45**	**9**	**28**
5	[50, 60)	55	8	36
6	[60, 70)	65	4	40
7	[70, 80	75	2	42
		N =	**42**	

F_i

Entonces, se tiene que la ecuación para calcular el decil es: $D_k = L_i + \frac{\frac{k*N}{10} - F_{i-1}}{f_i} * a_i$. Al sustituir los valores entonces queda: $D_1 = 30 + \frac{\frac{5*42}{10} - 19}{9} * 10 = 32.22$.

La respuesta es 32.22, donde se ubica el D_5, es decir, el 50% de los individuos tiene 32.22 años.

Percentiles para datos agrupados

Para hallar los percentiles (P) para datos agrupados, se aplica la siguiente ecuación:

$$P_k = L_i + \frac{\frac{k*N}{100} - F_{i-1}}{f_i} * a_i$$

Donde:

P_k = Decil para datos agrupados; $k = 1, 2, 3, \ldots, 99$;

L_i = Límite inferior del intervalo donde se encuentra el percentil;

F_{i-1} = Frecuencia acumulada del intervalo anterior al intervalo o clase del percentil;

f_i = Frecuencia absoluta del intervalo o clase del percentil;

N = Suma de las frecuencias absolutas;

a_i = Amplitud que tiene el intervalo, que se obtiene restando el límite superior del límite inferior del intervalo que contiene el intervalo o clase del percentil.

Se continuará con el mismo problema anterior de las edades. Se tiene, entonces, que hallar el percentil (75% de las edades). Primero se debe encontrar la mayor frecuencia absoluta donde se halla el percentil de interés usando la siguiente ecuación para el cálculo del P_1, que es: $\frac{k * N}{100}$. Al sustituir los valores para calcular el 75.º percentil es: $\frac{75 * 42}{100} =$ 31.5; se ubica esta frecuencia acumulada y se encuentra en el número de clase 5 con una frecuencia absoluta de 8. Cuando se quiera hallar el P_2, $P_3 \ldots P_k$, entonces $k = 2, 3\ldots k$, respectivamente, para ubicar la frecuencia acumulada correspondiente y continuar con el procedimiento.

Número de clase	Intervalo de clase	Clase	Frecuencia absoluta	F_i acumulado
	Li *Ls*	x_i	f_i	F_i
1	[10, 20)	15	1	1
2	[20, 30)	25	8	9
3	[30, 40)	35	10	19
4	[40, 50)	45	9	28
5	**[50, 60)**	**55**	**8**	**36**
6	[60, 70)	65	4	40
7	[70, 80)	75	2	42
		N =	**42**	

F_i

Entonces, se tiene que la ecuación para calcular el percentil es: $P_k = L_i + \frac{\frac{k * N}{100} - F_{i-1}}{f_i} * a_i$. Al sustituir los valores entonces queda: $P_1 = 30 + \frac{\frac{75 * 42}{100} - 28}{8} * 10 = 34.37$.

La respuesta es 34.37, donde se ubica el P_{75}, es decir, el 75% de los individuos tiene 34.37 años.

Distribución binomial (*n*, *p*)

Es una distribución de probabilidad discreta que describe el número de éxitos al realizar *n* experimentos independientes entre sí acerca de una variable aleatoria; en otras palabras, la distribución binomial se entiende como una serie de pruebas o ensayos en la que solo podemos tener dos resultados (éxito o fracaso), siendo el éxito nuestra variable aleatoria.

Entonces, se tiene que una distribución binomial es una distribución de probabilidad ampliamente utilizada de una variable aleatoria discreta, que toma valores naturales. Esta describe varios procesos de interés para los administradores, porque ayuda a aplicar comprensión de la incertidumbre y probabilidad en la toma de decisiones.

Distribuciones discretas

Se dividen en el cálculo de probabilidades en:

1. Uniforme discreta.
2. Binomial.
3. Hipergeométrica.
4. Geométrica.
5. Binomial negativa.
6. Poisson.

La distribución binomial describe datos discretos resultantes de un experimento. Esta distribución, también denominada función de la distribución de Bernoulli en honor del matemático suizo Jacob Bernoulli, quien vivió en el siglo XVII, suele representarse por B(*n*, *p*), donde *n* es el número de pruebas de que consta el experimento, *p* es la probabilidad de éxito (A) y la probabilidad de fracaso (A^-), se presenta como *q*, y $q = 1 - p$. La distribución binomial es, entonces, una distribución discreta que, en un experimento, se puede considerar de distribución binomial si cumple con lo siguiente:

- » Que solo hay dos posibles sucesos resultantes del experimento: A o A^- (éxito y fracaso).
- » Las probabilidades de cada suceso (A, $\bar{A}$) son las mismas en cualquier realización del experimento (*p* y $q = 1 - p$, respectivamente). Es decir, si se tira una moneda varias veces, no cambia la probabilidad de obtener cara.
- » Toda ejecución del experimento es independiente del resto.

Para hallar la distribución binomial B(*n*, *p*) se aplica la siguiente ecuación:

$$P_{(x)} = \begin{bmatrix} n \\ x \end{bmatrix} p^x \, q^{n-x}$$

Donde:
n = Número de ensayos/experimentos;
x = Número de éxitos;
p = Probabilidad de éxito;
q = Probabilidad de fracaso (1 – *p*).

Es importante señalar que la expresión entre corchetes no es una expresión matricial, pues representa un cálculo de una combinatoria sin repetición, que se obtiene usando la siguiente ecuación:

$$C_{n,\ x} = \begin{bmatrix} n \\ x \end{bmatrix} = \frac{n!}{x!\,(n-x)!}$$

Donde:
n! = *n* factorial.

El signo de exclamación en la expresión anterior representa el símbolo de factorial. De esta forma, la ecuación completa para calcular B(*n*, *p*) es la siguiente:

$$P_{(x)} = \frac{n!}{x!\,(\,n - x\,)!}\, p^{x}\, q^{n - x}$$

Donde:

$n =$	Número de ensayos/experimentos;
$n! =$	n factorial;
$x =$	Número de éxitos;
$p =$	Probabilidad de éxito;
$q =$	Probabilidad de fracaso ($1 - p$).

Propiedades de la distribución binomial

Para que una variable aleatoria se considere que sigue una distribución binomial, tiene que cumplir las siguientes propiedades:

1. En cada ensayo, experimento o prueba solo son posibles dos resultados (**éxito o fracaso**).
2. La **probabilidad de éxito** ha de ser constante. Esta se representa mediante la **letra *p***. La probabilidad de que salga cara al lanzar una moneda es 0.5 y esta es constante, dado que la moneda no cambia en cada experimento y las probabilidades de sacar cara son constantes.
3. La **probabilidad de fracaso** ha de ser también constante. Esta se representa mediante la **letra $q = 1 - p$**. Es importante fijarse que, mediante esta ecuación, sabiendo p o sabiendo q, podemos obtener la que nos falte.
4. El resultado obtenido en cada experimento es independiente del anterior. Por lo tanto, lo que ocurra en cada experimento no afecta a los siguientes.
5. Los sucesos son mutuamente excluyentes, es decir, no pueden ocurrir los dos al mismo tiempo. No se puede tener ÉXITO y FRACASO al mismo tiempo o que, al lanzar una moneda, salga cara y cruz al mismo tiempo.
6. Los sucesos son colectivamente exhaustivos, es decir, al menos uno de los dos ha de ocurrir. Si no se tiene ÉXITO, se

tiene FRACASO y, si se lanza una moneda, si no sale cara ha de salir cruz.

7. La variable aleatoria que sigue una distribución binomial, se suele representar como X ~ (*n*, *p*), donde ***n* representa el número de ensayos o experimentos** y ***p*, la probabilidad de éxito**.

A continuación se presenta un ***ejemplo*** de B(*n*, *p*):

En una granja de gallinas se sospecha que el 80% de las parvadas tiene la presencia de ectoparásitos y de estos, el más común es el *Menopon gallinae*. En un congreso nacional se reúnen 4 colegas y discuten acerca de la presencia de ectoparásitos en sus granjas. ¿Cuál es la probabilidad de que 3 de ellos hayan identificado un *Menopon gallinae*?

En primer lugar, se deben identificar las definiciones de las variables del experimento:

n = 4 (es el total de la muestra);

x = Número de éxitos, que en este ejemplo son 3. Buscamos la probabilidad de que 3 de los 4 colegas hayan identificado al mismo ectoparásito;

p = Probabilidad de éxito (0.8);

q = Probabilidad de fracaso (0.2), que es el resultado obtenido de $q = 1 - p$ ($0.2 = 1 - 0.8$).

La ecuación es : $P_{(x)} = \frac{n!}{x!(n-x)!} p^x q^{n-x}$; entonces debemos sustituir las variables en la ecuación: $P_{(3)} = \frac{4!}{3!(4-3)!} 0.8^3 \, 0.2^{4-3}$.

Para calcular el numerador del factorial de 4 (4!), se multiplica *n* hasta su mínima expresión, es decir, 1. En este ejemplo se tiene que 4 se multiplica por el inmediato menor (3) y su producto por 2, y este por 1. Así tenemos que 4! = 4 * 3 * 2 * 1 = 24. Y el denominador factorial de 3! sería 3 * 2 *1 = 6, multiplicado por el factorial de (4 – 3)! = 1!; entonces 6(1) = 6. El resultado de $\frac{24}{6} = 4$.

Fuera de la operación de la división tenemos dos números. El primero es 0.8^3 = 0.512 y el segundo, 0.2 (dado que 4 – 3 = 1 y cualquier número elevado a 1 es el mismo). Por tanto, el resultado final sería: 4 *

(0.512 * 0.2) = 0.4096. Al pasar el resultado a su expresión porcentual, se tiene que la probabilidad es de 40.96% de que 3 de los 4 colegas hayan identificado la misma especie de ectoparásito en su granja.

Los datos del número de éxito pueden graficarse (Figura 23). Para ello, usando el mismo ejemplo, se calculará el valor de $x = 0$, $x = 1$, $x = 2$ y el que se ha calculado previamente: $x = 3$, siendo los valores numéricos obtenidos 0.00960, 0.05120, 0.15360 y el obtenido previamente: 0.4096, respectivamente; se obtiene la siguiente gráfica:

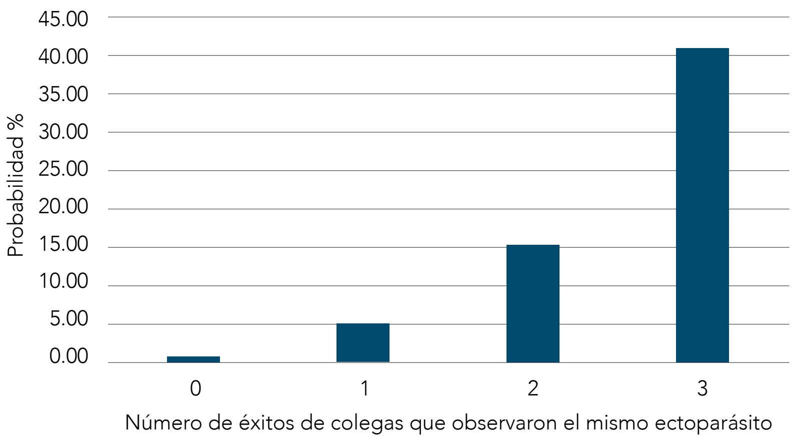


Figura 23. Probabilidad de encontrar el mismo ectoparásito en granjas de postura comercial

Fuente: elaboración propia.

Distribución normal (μ, σ)

La distribución normal es la más importante de todas las distribuciones de probabilidad. También conocida como distribución gaussiana o Campana de Gauss, debe su nombre a las valiosas aportaciones del matemático Carl Friedrich Gauss (1700 a 1800) al estudiar la distribución de los errores en las observaciones astronómicas.

La distribución normal es una distribución variable continua que puede tomar cualquier número real. Queda especificada por dos parámetros de los que depende su función de densidad y que resulta ser la media aritmética y la desviación estándar de la distribución. Por ello también se le representa como: $X \sim N\,(\mu, \sigma)$.

Se presenta como la función de densidad de probabilidad de una variable aleatoria que sigue una distribución normal (Figura 24).

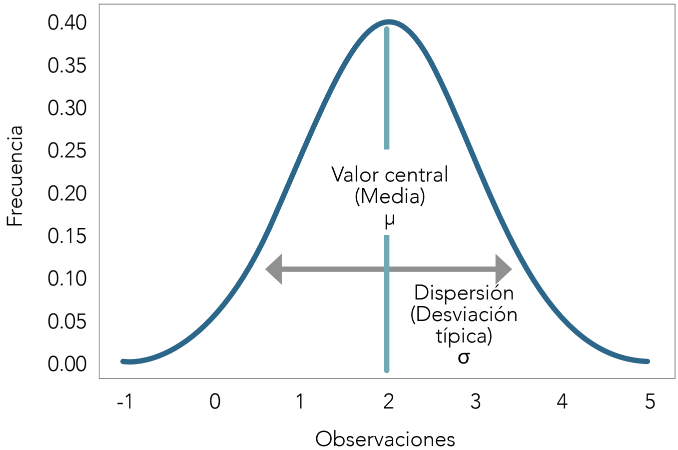


Figura 24. Función de densidad de una distribución normal

Fuente: elaboración propia.

Distribuciones continuas

Son las siguientes:

1. Uniforme.
2. Normal.
3. Lognormal.
4. Logística.
5. Beta.
6. Gamma.
7. Exponencial.
8. Chi cuadrada.
9. *t* de Student.
10. F de Snedecor.

La distribución normal es un modelo teórico capaz de aproximar satisfactoriamente el valor de una variable aleatoria a una situación ideal, es decir, se adapta una variable aleatoria a una función que depende de la media y la desviación estándar.

Este modelo de distribución se utiliza para explicar las distribuciones de frecuencias de un determinado fenómeno y se aplica en series que siguen una distribución normal, aproximadamente, es decir, que las frecuencias parciales menores se ubican en los extremos y las frecuencias más altas se ubican al centro. Es una curva asintótica (nunca toca el eje de la abscisa). Para efectos prácticos, el área dentro de la campana representa la totalidad de una serie comprendida entre dos líneas verticales situadas a ± 3 desviaciones estándar de cada lado de la media.

Para determinar el valor de las desviaciones estándar, se utilizan las constantes estadísticas de Z: *t de Student* (el número de veces que se suman o se restan las desviaciones estándar): se usa la constante de *t de Student* cuando se manejan muestras pequeñas en variables cuantitativas; estas constantes manejan grados de libertad ($n - 1$), que denotan el número de valores muestrales que no pueden calcularse partiendo de otros valores.

Puede observarse que, en cada mitad, la curva es primero cóncava hacia arriba y luego cóncava hacia abajo, habiéndose dado el nombre de "punto de inflexión" a aquel en el cual la curva cambia de dirección. Hay, por consiguiente, dos puntos de inflexión: uno izquierdo y otro derecho. La distancia que separa a cada punto de inflexión de la línea central que representa el promedio constituye una desviación estándar. Aunque teóricamente la curva nunca toca la línea horizontal, para propósitos prácticos puede considerarse que la totalidad de su área se encuentra comprendida entre dos líneas verticales situadas a tres desviaciones estándar del promedio.

Los matemáticos han demostrado que un 68% de toda el área de la curva, aproximadamente, se encuentra comprendida entre las dos líneas verticales que pasan por los puntos de inflexión, lo cual equivale a decir que el 68% del área se encuentra entre el promedio más una desviación estándar y en el promedio menos una desviación estándar, se encuentra el 95% del área de la curva, aproximadamente, y prácticamente el 100% del área se encuentra entre el promedio ± 3 desviaciones estándar. Usando una notación matemática: $X \pm 1$ D. E. incluye el 68.26% del área de la curva; ± 2 D. E., el 95.45%; y ± 3 D. E., el 99.73%.

Lo anterior es importante por dos razones principales: en primer lugar, porque ya señalamos que los resultados dados por el azar siguen una curva normal y en segundo lugar, porque se ha visto que casi todas las constantes fisiológicas de los individuos (peso, estatura, tensión arterial, etcétera) y, en general, las diferentes características de toda población se distribuyen formando una curva normal.

Esto quiere decir que las propiedades de la curva pueden aplicarse a cualquier característica que tenga una distribución normal. Las constantes estadísticas (Z, t) son iguales al número de veces que se suman o restan las desviaciones estándar. Para determinar los valores de desviaciones estándar, se utiliza la tabla de áreas bajo la curva normal tipificada o de 0 a Z. Se ubica el valor más cercano al estimado, obteniendo el valor de las desviaciones estándar correspondientes. Por ejemplo: 95% = 0.95 / 2 = 0.4750. Tenemos que calcular el valor que está debajo de la mitad de la curva; entonces tenemos que: 0.5 – 0.4750 = 0.0250. Buscamos en la tabla de Z el valor más cercano a 0.0250 y es -1.96; entonces si tenemos debajo de toda la curva que: 1 – 0.0250 = 0.9750, buscamos en la tabla de Z el valor más cercano a 0.9750 y es 1.96. Entonces el valor Z de 95% = 1.96 desviaciones estándar (σ).

Siguiendo con el mismo ***ejemplo***, pero con 99% de confianza o seguridad, tenemos que 0.99 / 2 = 0.4950, por lo que hay que calcular el valor que está debajo de la mitad de la curva; entonces tenemos que: 0.5 – 0.4950 = 0.0050. Buscamos en la tabla de Z el valor más cercano a 0.0050, que es -2.57 (0.0051); entonces si tenemos debajo de toda la curva que: 1 – 0.0051 = 0.9949, buscamos en la tabla de Z el valor más cercano a 0.9949 y es 2.57. Entonces el valor z de 99% = 2.57 desviaciones estándar (σ).

Propiedades de la distribución normal

1. *Distribución unimodal*: los valores que son más frecuentes o que tienen más probabilidad de aparecer están alrededor de la media. En otras palabras, cuando nos alejamos de la media, la probabilidad de aparición de los valores y su frecuencia descienden.

2. La distribución normal estándar es una distribución normal con media de 0 y desviación estándar de 1.
3. También es llamada distribución *Z*.
4. Un valor *z* es la distancia entre un valor seleccionado llamado *x*.
5. Un puntaje *Z* lo que hace es decirnos a cuántas unidades de desviación estándar del promedio está un puntaje determinado, o sea, no contamos en cantidad de puntos, sino en cantidades de desviaciones estándar. Para utilizar el puntaje *Z* requerimos que la distribución sea normal y conocer la media y la desviación estándar de los puntajes.
6. La media de la población (μ) dividida entre la desviación estándar (σ). La ecuación es: $Z = (\overline{X} - \mu) / \sigma$.
7. La forma de la campana de Gauss depende de los parámetros μ y σ.
8. La media indica la posición de la campana, de modo que para diferentes valores de μ la gráfica es desplazada a lo largo del eje horizontal.
9. Por otra parte, la desviación estándar determina el grado de apuntamiento de la curva. Cuanto mayor sea el valor de la desviación estándar más se dispersarán los datos en torno a la media, y la curva será más plana.
10. Un valor pequeño de este parámetro indica, por tanto, una gran probabilidad de obtener datos cercanos al valor medio de la distribución.

Características de una distribución probabilística normal (Figura 25)

1. La distribución de probabilidad normal es simétrica respecto a su media.
2. La curva normal decrece uniformemente en ambas direcciones, a partir del valor central.
3. Es asintótica. Esto significa que la curva se acerca cada vez más al eje de las *x*, pero en realidad nunca llega a tocarlo. Esto es, los puntos extremos de la curva se extienden indefinidamente en ambas direcciones.
4. Es una distribución simétrica: el valor de la media, la mediana y la moda coinciden.

Tiene forma de campana. Es asintótica al eje de las abscisas (para $x = \pm\infty$).
Es simétrica respecto a la media (μ), donde coinciden la mediana (Mn) y la moda (Mo).
Los puntos de inflexión tienen como abcisas los valores $\mu \pm \sigma$.

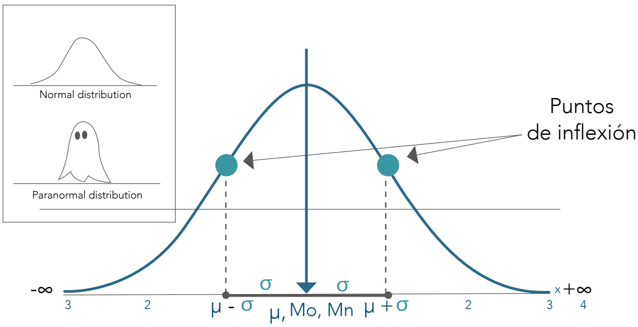


Figura 25. Características de la distribución normal

Fuente: descargada de internet.

¿Qué necesitamos para representar una distribución normal?

1. Una variable aleatoria.
2. Calcular la media.
3. Calcular la desviación típica.
4. Decidir la función que queremos representar: función de densidad de probabilidad o función de distribución.

En resumen, la importancia de la distribución normal se debe principalmente a que existen muchas variables asociadas a fenómenos naturales que siguen el modelo de la distribución normal.

» *Caracteres morfológicos* de individuos (personas, animales, plantas...) de una especie; por ejemplo: tallas, pesos, envergaduras, diámetros, perímetros...
» *Caracteres fisiológicos*; por ejemplo: efecto de una misma dosis de un fármaco o de una misma cantidad de abono.
» *Caracteres sociológicos*; por ejemplo: consumo de cierto producto por un mismo grupo de individuos; puntuaciones de un examen.

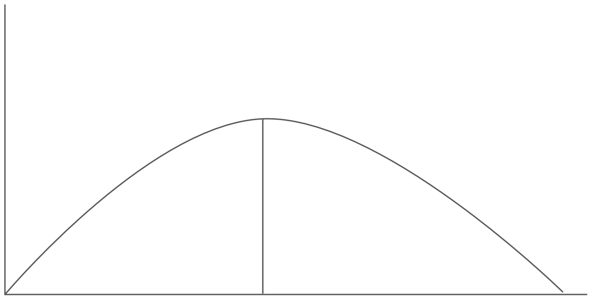
- *Caracteres psicológicos;* por ejemplo: cociente intelectual; grado de adaptación a un medio...
- *Errores* cometidos al medir ciertas magnitudes.
- *Valores estadísticos* muestrales; por ejemplo: la media.
- *Otras distribuciones*, como la binomial o la de Poisson, son aproximaciones normales.

Y, en general, cualquier característica que se obtenga como suma de muchos factores.

Distribución normal reducida (tipificada)

La probabilidad es una herramienta matemática que nos permite hacer inferencias y tomar decisiones. Entonces, se puede inferir acerca de una población normal necesariamente de una u otra manera al calcular probabilidades asociadas a ella, pero si nos fijamos con cuidado, para cada variable aleatoria, tenemos un valor de M (media) y de σ (desviación estándar); por lo tanto, para cada una de esas variables aleatorias, se tendrá que realizar un proceso de integración, proceso que complica el procedimiento. En otras palabras, la distribución normal queda definida por dos parámetros: su media aritmética y su desviación estándar, y se expresa como: $N(\mu, \sigma)$, que representa la distribución normal.

Para facilitar el proceso todas las variables aleatorias normales deben ser transformadas a la variable aleatoria denominada normal estándar, que se caracteriza por tener una media igual a 0 (cero) y una desviación estándar igual a 1 (uno), que se indica como una Z. De aquí su definición: Z es una variable aleatoria normal llamada normal estándar, con media cero y desviación estándar uno.



Esta variable aleatoria permite por medio de valores tabulados (tabla de *Z*) calcular las probabilidades correspondientes debajo de la curva.

Para ello, se utiliza la siguiente ecuación para convertir una variable aleatoria normal estándar. Se calcula el valor de *Z* por medio de la siguiente expresión:

Si tenemos que la variable *X* es $N(\mu, \sigma)$, entonces la variable tipificada de *X* es: $Z = \frac{x - \mu}{S}$, donde S = Desviación estándar de la muestra, que sigue una distribución normal de: $\mu = 0$ y $\sigma = 1$.

Si el valor que se obtiene es un número negativo ($x < \mu$), nos señala que *X* está a la izquierda de su media, y si el valor es positivo ($x > \mu$), nos indica que *X* está a la derecha de su media cuando el valor es cero ($x = 0$); entonces el valor de $Z = 0$ (Figura 26).

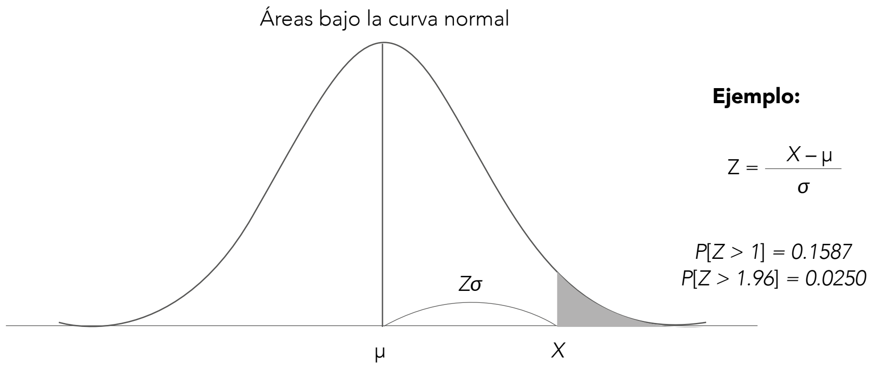


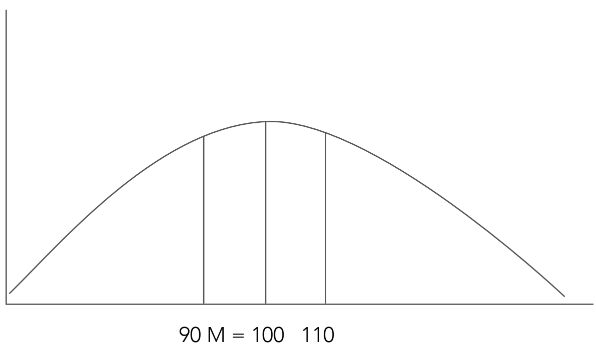
Figura 26. Área (gris) desde la cola derecha (positivo) hacia *X* correspondiendo a *P*(*Z* > *X*)

Fuente: elaboración propia.

Por ejemplo, si tenemos que la media del peso de los cerdos al finalizar la engorda es de 100 kg y que la desviación estándar es de 10 kg, encontrar el valor estándar para 90 y 110 kg.

$$Z = \frac{x-\mu}{S} \quad \textit{Se tiene que}: \; Z = \frac{90-100}{10} = -\frac{10}{10} = -1.$$

$$Z = \frac{x-\mu}{S} \quad \textit{Se tiene que}: \; Z = \frac{110-100}{10} = \frac{10}{10} = 1.$$



Para calcular una probabilidad usando la tabla es necesario considerar lo siguiente (Figura 27):

1. El área bajo toda la curva es 1.
2. La curva es simétrica, pues tiene como centro la media. Esto significa que el área bajo la curva es 0.500 de la media hacia la derecha o -0.500 de la media hacia la izquierda.
3. La tabla solo nos da la probabilidad entre la media y el valor de nuestro interés.

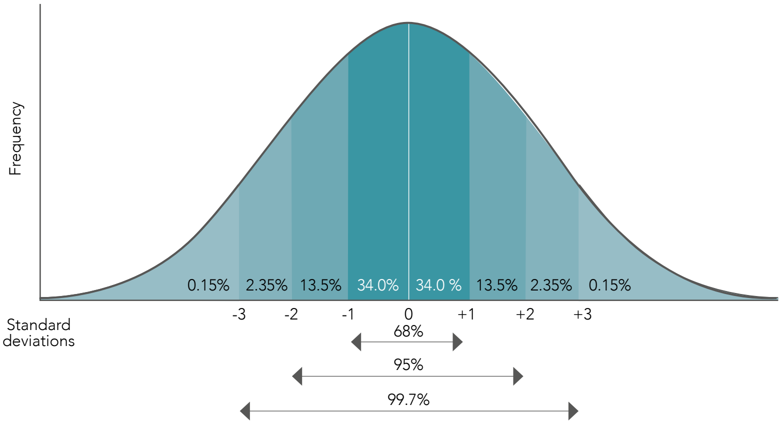


Figura 27. Área abajo de la curva de la distribución normal

Fuente: descargada de internet.

A continuación, se presentan ***ejemplos*** de la distribución normal:

En una población de estudiantes de la carrera de Medicina, se obtiene el peso corporal de 450 alumnos. Se sabe que el peso se distribuye normalmente y se obtiene un peso promedio de 80 kg y una desviación estándar de 14.0 kg:

Ecuación:

$$Z = \frac{x - \mu}{S}$$

Media = 80 kg
D. E. = 14 kg

a) Calcule la probabilidad de tener un peso localizado entre 75.0 y 90.0 kg.

$Z = \frac{90-80}{14} = 0.7142$ (buscar en la tabla de *Z* el valor 0.71);
$Z = \frac{75-80}{14} = -0.3571$ (buscar en la tabla de *Z* el valor -0.36);
p(75 < = *x* < = 90) = 0.7611 – 0.3594 = 0.4017 = 40.17%.

b) Calcule la probabilidad de un valor de 75.0 o menor.

$Z = \frac{75-80}{14} = -0.3571$ (buscar en la tabla de ***Z*** el valor -0.36);
p(*x* < = 75) = 0.3594 = 35.94%.

c) Calcule la probabilidad de un peso localizado entre 55.0 y 70.0: *p*(55 < = *x* < = 70).

$Z = \frac{70-80}{14} = -0.7142$ (buscar en la tabla de *Z* el valor -0.71);
p(55 < = ***x*** < = 70) = 0.2389;
$Z = \frac{55-80}{14} = -1.7857$ (buscar en la tabla de *Z* el valor -1.79);
p(55 < = *x* < = 70) = 0.0367;
p(55 < = *x* < = 70) = 0.2389 – 0.0367 = 0.2022 = 20.22%.

En otro ejemplo podemos calcular *n* individuos. A continuación, se presenta un ***ejemplo***:

Varias pruebas de inteligencia dieron una puntuación que sigue una ley normal con una media de 100 y una desviación típica de 15. En una población de 2,500 individuos, ¿cuántos individuos se espera que tengan un coeficiente superior a 125?

p(*x* > 125) = Z = 125 – 100 / 15 = 1.6666
(buscar en la tabla de *Z* el valor 1.67)
Z = 1.67 = 0.9525.

Como está del lado derecho, entonces tenemos que el procedimiento es 1 – 0.9525 = 0.0475 y este resultado se multiplica por 2,500 = 118.75 = 119 individuos que estarán por arriba del coeficiente de 125.

Referencias

Definición de: Función matemática (https://definicion.de/funcion-matematica/).

Devore, J. L. (2008). *Probabilidad y estadística para ingeniería y ciencias* (7.ª ed.). CENGAGE Learning Inc. Corporativo Santa Fe. 742 págs.

Distribución binomial. Consultado: 10 de octubre de 2022. https://economipedia.com/definiciones/distribucion-binomial.html

Distribución binomial o de Bernoulli. Consultado: 25 de octubre de 2022. https://www.matematicasonline.es/BachilleratoCCSS/primero/apuntes/binomial.pdf

Distribución normal. Consultado: 10 de diciembre de 2022. https://www.uv.es/ceaces/pdf/normal.pdf

Equipo Editorial Etecé. Función matemática. De: Argentina. Para: Concepto.de. https://concepto.de/funcion-matematica/. Última edición: 16 de julio de 2021. Consultado: 18 de agosto de 2022.

Gráficos para representar datos. Consultado: 10 de enero de 2023. http://www.desarrolloweb.com/articulos/875.php

Huntsberger, D. V. y Billingsley, P. (1983). *Elementos de estadística inferencial*. Compañía Editorial Continental, 404 págs.

Itza-Ortiz, M. y Pérez Casio, F. (2014). *Estadística descriptiva: Manual de estudio*. UACJ. 60 págs. ISBN: 978-607-520-093-4.

Johnson, R. (1996). *Estadística elemental* (3.ª reimpr.). Ed. Trillas.

Milton, J. S. (2007). *Estadística para biología y ciencias de la salud* (3.ª ed.). McGraw-Hill.

Moda para datos agrupados. Consultado: 27 de agosto de 2022. https://www.fhybea.com/moda-para-datosagrupados.html#:~:text=La%20moda%20en%20un%20conjunto,se%20trabaja%20con%20datos%20agrupados

Ruiz Muñoz, D. (2004). *Manual de estadística*. Editado por eumed·net. 91 págs. ISBN: 84-688-6153-7.

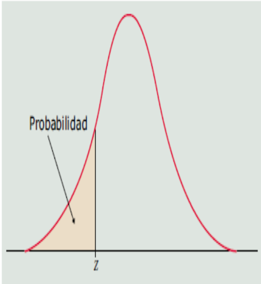
Rumsey, D. (2005). *Statistics Workbook for Dummies*. Wiley Pub.

Todo sobre la mediana. Cálculo de la mediana de datos agrupados. Consultado: 27 de agosto de 2022. https://www.superprof.es/apuntes/escolar/matematicas/estadistica/descriptiva/mediana.html#:~:text=dos%20puntuaciones%20centrales.-,F%C3%B3rmula%20y%20c%C3%A1lculo%20de%20la%20mediana%20para%20datos%20agrupados,en%20el%20que%20se%20encuentre

Variables estadísticas y sus características. Consultado: 20 de noviembre de 2022. http://www.vitutor.com/estadistica/descriptiva/a_2.html

Wayne, D. (1991). *Bioestadística. Base para el análisis de las ciencias de la salud*. Limusa Bibliografía.

Anexo

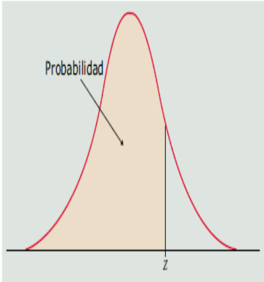


El valor de la tabla para z es el área bajo la curva de la normal estándar a la izquierda de z

TABLA A: Probabilidades de la normal estándar

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Fuente: Devore (2008).



El valor de la tabla para z es el área bajo la curva de la normal estándar a la izquierda de z

TABLA A: Probabilidades de la normal estándar (cont.)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998