Universidad Autónoma
del Estado de México

E-BOOK
PENSAMIENTO MATEMÁTICO I
1ER. SEMESTRE

Vasti Galicia Aldama
Víctor Eric Velásquez Núñez
María Isabel Arzate Ocampo
Ana Esperanza Carreño Contreras
Maximiliano Adán Evangelista Esquivel



ADMINISTRACIÓN UNIVERSITARIA 2021 - 2025

SD

Doctor en Ciencias e Ingeniería Ambientales
Carlos Eduardo Barrera Díaz
Rector

Doctor en Ciencias Computacionales
José Raymundo Marcial Romero
Secretario de Docencia

Doctora en Ciencias Sociales
Martha Patricia Zarza Delgado
Secretaria de Investigación y Estudios Avanzados

Doctor en Ciencias de la Educación
Marco Aurelio Cienfuegos Terrón
Secretario de Rectoría

Doctora en Humanidades
María de las Mercedes Portilla Luja
Secretaria de Difusión Cultural

Doctor en Ciencias del Agua
Francisco Zepeda Mondragón
Secretario de Extensión y Vinculación

Doctora en Ciencias Económico Administrativas
Eréndira Fierro Moreno
Secretaria de Administración

Doctor en Educación
Octavio Crisóforo Bernal Ramos
Secretario de Finanzas

Doctora en Ciencias Administrativas
María Esther Aurora Contreras Lara Vega
Secretaria de Planeación y Desarrollo Institucional

Doctora en Derecho
Luz María Consuelo Jaimes Legorreta
Oficina de la Abogacía General

Licenciada en Comunicación
Ginarely Valencia Alcántara
Directora General de Comunicación Universitaria

Maestra en Salud Animal
Trinidad Beltrán León
Secretaria Técnica de la Rectoría

Doctor en Ciencias Sociales
Luis Raúl Ortiz Ramírez
Director de Centros Universitarios
y Unidades Académicas Profesionales
Región A y Encargado del Despacho Región B

Doctora en Investigación Educativa
Eva Lilia García Escobar
Directora de Estudios de Nivel Medio Superior

PENSAMIENTO MATEMÁTICO I

DIRECCIÓN DE PUBLICACIONES UNIVERSITARIAS
Editorial de la Universidad Autónoma del Estado de México

Doctor en Ciencias e Ingeniería Ambientales
Carlos Eduardo Barrera Díaz
Rector

Doctora en Humanidades
María de las Mercedes Portilla Luja
Secretaria de Difusión Cultural

Doctor en Administración
Jorge Eduardo Robles Alvarez
Director de Publicaciones Universitarias

Doctora en Investigación Educativa
Eva Lilia García Escobar
Directora de Estudios de Nivel Medio Superior

Maestro en Tecnología Educativa
Alberto Torres Gutiérrez
Director de Educación Continua y Digital

PENSAMIENTO MATEMÁTICO I

Vasti Galicia Aldama
Víctor Eric Velásquez Núñez
María Isabel Arzate Ocampo
Ana Esperanza Carreño Contreras
Maximiliano Adán Evangelista Esquivel

Universidad Autónoma
del Estado de México

"2024, Conmemoración del 60 Aniversario de la Inauguración de Ciudad Universitaria"

Primera edición, Agosto 2024

PENSAMIENTO MATEMÁTICO I
Libro de texto digital

Autores
Ana Esperanza Carreño Contreras
Plantel "*Dr. Ángel Ma. Garibay Kintana*" de la Escuela Preparatoria de la UAEMéx
Maria Isabel Arzate Ocampo
Plantel "*Dr. Ángel Ma. Garibay Kintana*" de la Escuela Preparatoria de la UAEMéx
Maximiliano Adán Evangelista Esquivel
Plantel "*Dr. Ángel Ma. Garibay Kintana*" de la Escuela Preparatoria de la UAEMéx
Vasti Galicia Aldama
Plantel "*Cuauhtémoc*" de la Escuela Preparatoria de la UAEMéx
Víctor Erick Velásquez Núñez
Plantel "*Dr. Pablo González Casanova*" de la Escuela Preparatoria de la UAEMéx

Universidad Autónoma del Estado de México
Av. Instituto Literario 100 Ote. Col. Centro
Toluca, Estado de México
C. P. 50000
Tel: (52) 722 277 38 35 y 36
http://www.uaemex.mx

El contenido de este material es responsabilidad de los autores. Queda prohibida la reproducción o transmisión total o parcial del contenido de la presente obra, sin el consentimiento previo y por escrito del editor. La Universidad Autónoma del Estado de México se deslinda de cualquier acción legal derivada de este material.

® Derechos reservados, 2024

ISBN: En trámite

Hecho en México
Made in Mexico

Mensaje

La **Universidad Autónoma del Estado de México,** en la búsqueda permanente de garantizar una educación inclusiva, equitativa y de excelencia, se mantiene a la vanguardia en los avances tecnológicos, desarrollando libros electrónicos (e-book), con el propósito de implicarte interactivamente con los contenidos de las asignaturas que integran el plan de estudios del Bachillerato Universitario.

Nuestro objetivo es contribuir a la formación integral de estudiantes con conocimientos rigurosos, valores humanistas y habilidades digitales, utilizando metodologías activas, que desarrollen el pensamiento crítico y analítico, así como la capacidad de afrontar los retos de la era digital.

Los cambios sociales, ambientales y culturales del mundo contemporáneo presentan serios desafíos al proceso educativo; por ello, nuestra Escuela Preparatoria promueve formas de aprendizaje innovadoras, en las que se desarrollan la comunicación asertiva, la creatividad y el trabajo colaborativo. De este modo buscamos garantizar la impartición de una educación capaz de estimular la curiosidad intelectual de los preparatorianos y la exploración de todos los campos del conocimiento, de modo que cada estudiante sea capaz de ampliar sus horizontes culturales, reconociendo sus habilidades, gustos y aptitudes.

El material educativo que tienes en tus manos ha sido elaborado por docentes de las áreas de conocimiento correspondientes. La creatividad, el pensamiento crítico y el espíritu científico de los colectivos docentes se reúnen en un e-book para facilitar la comprensión de los contenidos y comprometerte con tu aprendizaje.

El libro digital que ahora puedes consultar tiene una estructura lógica de fácil manejo, que te permitirá navegar por los contenidos y recursos académicos que responden a tus necesidades; podrás acceder a ellos desde una computadora, tableta o un teléfono celular.

Te invito a aprovechar al máximo esta nueva experiencia de aprendizaje, disfruta esta herramienta que potenciará tus habilidades digitales, tus conocimientos y tu creatividad.

SOMOS UAEMéx

Patria, Ciencia y Trabajo

Doctor en Ciencias e Ingeniería Ambientales
Carlos Eduardo Barrera Díaz
Rector



Módulo I

Introducción a la estadística y probabilidad

Propósito de Módulo

Analiza e interpreta datos estadísticos y probabilísticos la representación de los conjuntos y subconjuntos de números, para poder identificarlos y relacionarlos en su entorno.

Introducción

En este módulo se abarcan **contenidos** qué te permitirán analizar e interpretar datos e informaciones relevantes sobre la **probabilidad y estadística,** realizarás como primera actividad una evaluación diagnóstica para conocer que tanto sabes sobre estos rubros. En los primeros temas encontrarás los símbolos matemáticos, los números reales y sus operaciones; los cuales serán estudiados por la importancia que tienen para analizar datos, calcular estadísticas y tomar decisiones en diferentes contextos.

Durante el módulo te **encontrarás varios juegos** que te permitirán practicar lo aprendido. Al igual, se te presentan ejemplos sobre la recolección de datos y su organización para valorar el pensamiento estadístico y probabilístico, se analizan los instrumentos de recolección de información, el diseño de encuestas, la conceptualización y clasificación de la estadística, los datos, las variables, clasificación, población y la muestra para realizar estos estudios. Además, se analiza la representación de los conjuntos y subconjuntos de números con la finalidad de identificarlos y relacionarlos en el entorno como sucesos que faciliten el estudio de la probabilidad.

Revisarás elementos aritmético-algebraicos mediante la elaboración de tablas de distribución de frecuencias, seleccionando el modelo matemático correcto de acuerdo con la naturaleza de sus variables de manera que permita explicar una situación o fenómeno. También analizarás técnicas de conteo que promueven una mayor conciencia en la toma de decisiones al conocer la probabilidad de un caso favorable en un espacio muestral de enumeración no sencilla.

Se aborda el Teorema de Bayes para determinar probabilidades condicionales mediante ejemplos para favorecer la argumentación en situaciones o fenómenos matemáticos en un contexto inmediato. Se proponen problemas de aplicación para socializar lo aprendido en la solución de un problema ya sea teórico o práctico.

Al final del módulo se tiene una actividad auto evaluable con el propósito de medir tu aprendizaje, así como un juego de gamificación que rete y te motive a dar lo mejor de ti en tu formación académica.

Actividad 1.



Evaluación Diagnóstica
(conocimientos previos)

Antes de comenzar con la revisión de los contenidos del módulo 1 de Pensamiento Matemático I, te invitamos a responder la evaluación diagnóstica

Eso te ayudará a descubrir los conocimientos previos que tienes sobre los temas a abordar y, al mismo tiempo, le permitirá a tu docente conocer tus fortalezas y debilidades.

Para comenzar con el ejercicio, revisa en la plataforma **SEDUCA** la sección Actividades (Estudiante) y responde a la actividad denominada: M1 Examen Diagnóstico..

1.1 Símbolos matemáticos

Definición

Los **símbolos matemáticos** son elementos gráficos que representan ideas, operaciones, relaciones o conceptos de tipo matemático; pueden ser letras, figuras, números, entre otros elementos pictóricos, incluso combinaciones de ellos.

Son representaciones visuales cuyo significado determina una operación matemática.

Y son parte del lenguaje matemático que hacen que en ocasiones las consideremos indescifrables, y más allá de establecer una interacción, orden, significado entre ellos, permite que sea un lenguaje universal independientemente del idioma de los países.

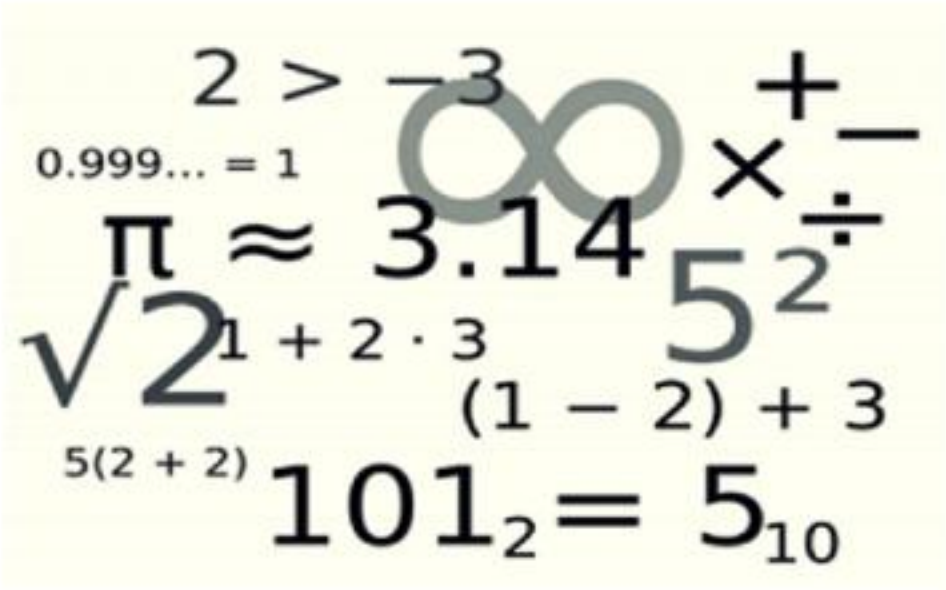
Sirven para indicar operaciones matemáticas en nuestra vida diaria expresando sumas, restas, multiplicaciones, divisiones y en forma menos frecuente raíces, potenciación, etc.

$+ - \% \sqrt{} \sqrt[3]{}$

¿Has considerado la importancia de los símbolos matemáticos?



En nuestra vida diaria hacemos uso de la simbología matemática de forma tan frecuente que nos pasa inadvertida su importancia y frecuencia de aplicación. Pensemos la hora en que nos levantamos y los tiempos invertidos en cada actividad para irnos a nuestro trabajo o a la escuela, el dinero necesario para transportarnos, el número de vasos con agua que necesita si un vaso es de 250 ml y se recomienda tomar 2l, los incisos que aprobaremos en un examen con mínimo con 6.0 de calificación, etc.





Actividad 2.



Símbolo matemático correcto

Instrucciones:
Utilizando las operaciones de suma, resta, multiplicación, división, potenciación y radicación, coloca los símbolos matemáticos y de agrupación que consideres deben estar en las siguientes operaciones para que proporcione el resultado indicado.
Da clic en la **imagen.**

En seguida se muestran los **símbolos matemáticos** más significativos o de uso más frecuente.

Tabla 1. Símbolos matemáticos comunes

Símbolos matemáticos comunes			
Símbolo	Significado	Símbolo	Significado
: =	Se define como	$\sqrt[3]{\square}$	Raíz cúbica
+ ∞	Más infinito	■	Fin de demostración
- ∞	Menos infinito	$\|x\|$	Valor absoluto de x
%	Por ciento	$\sqrt[2]{\square}$	Raíz cuadrada
Operadores binarios básicos			
+	Suma	±	Más menos
−	Resta	∩	Intersección
/	División	∪	Unión
÷	División	∨	Ó
×	Multiplicación	∧	Y

Símbolos matemáticos comunes			
Operadores racionales comunes			
=	Igual a	≅	aproximadamente igual con
≈	Aproximadamente igual a	∈	Pertenece a
≡	Idénticamente igual a	⊂	Contenido
<	Menor que	⊆	Contenido igual
>	Mayor que	⊃	Superconjunto de
≤	Menor o igual	⊇	Superconjunto igual de
≥	Mayor o igual	∝	Proporcional a
⊥	Perpendicular a	||	Paralelo a
Σ	Suma	Δ	Incremento
≫	Mucho mayor que		
Símbolos lógicos			
∀	para todo	∃	Existe
∃!	existe un único	|	tal que
⇒	implica	⇔	sí y solo si
∵	porque	∴	por lo tanto
Símbolos estadísticos			
α	Nivel de significación (error tipo I)	$\hat{X}$	Moda
$\bar{X}$	Media de una muestra	$\tilde{X}$	Mediana
f	Frecuencia	σ	Desviación estándar
s	Desviación estándar de una muestra	r	coeficiente de correlación de una muestra
Q_i	i-ésimo cuartil	σ^2	Varianza
P(A)	Probabilidad del evento A	**P(B|A)**	Probabilidad de que ocurra el evento B, dado que ocurre el evento A
n!	Factorial		

Nota. Fuente: Elaboración propia

1.2 Números reales y sus operaciones

El uso de símbolos matemáticos te lleva en muchas ocasiones a la realización de operaciones, entre las básicas tenemos a la suma, resta, multiplicación y división donde el resultado obtenido puede ser **un** número de contar los cuales reciben el nombre de números naturales o bien el resultado de la operación puede ser un cero o éstos mismos números de contar pero negativos que son llamados números enteros, también puede ser que el resultado de la operación represente una parte de un número entero y en ese caso estamos hablando de los números que tienen a las fracciones o los números con parte decimal finita o infinita periódica, llamado conjunto de números racionales o bien de los números irracionales que tienen parte decimal infinita y no periódica.

Tabla 2. Conjuntos numéricos

Conjuntos numéricos	Notación
Números naturales	$\mathbb{N} = \{1,2,3,4,\ldots\}$,
Números enteros	$\mathbb{Z} = \{\ldots,-3,-2,-1,0,1,2,3\ldots\}$
Números racionales	$\mathbb{Q} = \{x \mid x = \frac{a}{b} \text{ donde } a \wedge b \in \mathbb{Z} \text{ con } b \neq 0\}$
Números irracionales	$\mathbb{I} = \{x \mid x \text{ tiene parte decimal infinita y no periódica}\}$
Números reales	$\mathbb{R} = \mathbb{Q} \cup \mathbb{I}$

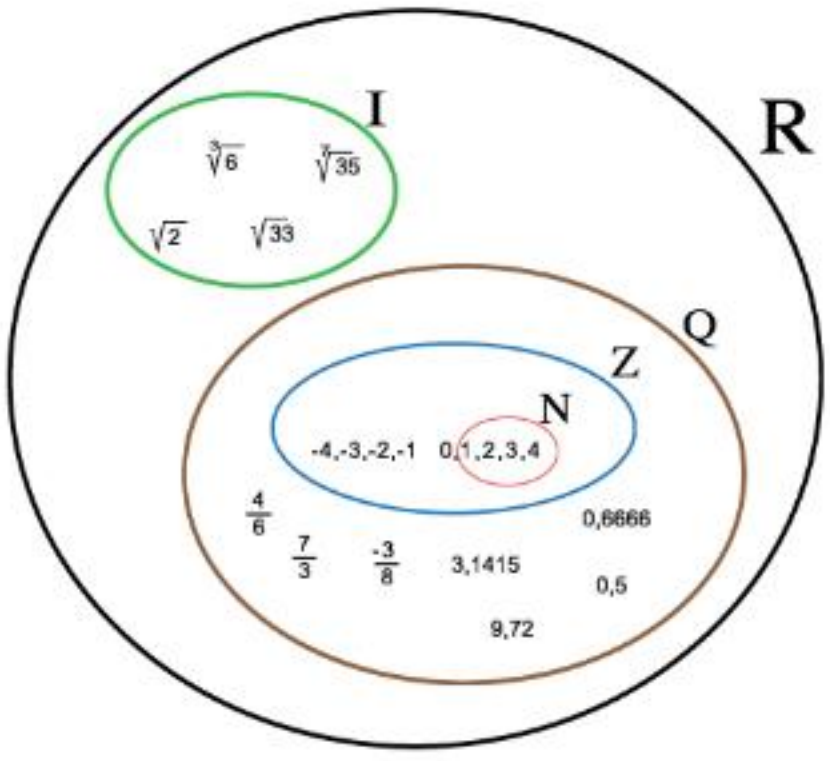
Nota. Fuente: Elaboración propia

Para su mayor comprensión ver el **Figura 1.**

El conjunto de los números reales, denotado por "$\mathbb{R}$" está conformado por la unión del conjunto de los números racionales con el conjunto de los números irracionales. Esto es:

$$\mathbb{R} = \mathbb{Q} \cup \mathbb{I}$$

Figura 1. Conjuntos numéricos.

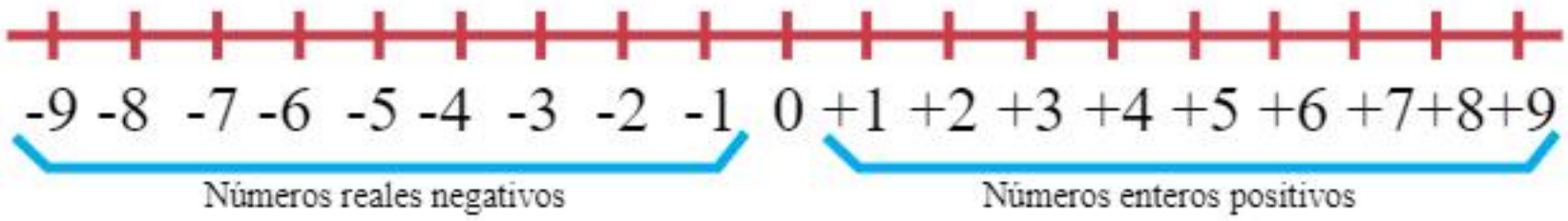


N = Números naturales (enteros positivos)
Z = Números enteros (positivos y negativos)
Q = Números racionales (fracciones)
I = Irracionales

Nota. Fuente: Elaboración propia

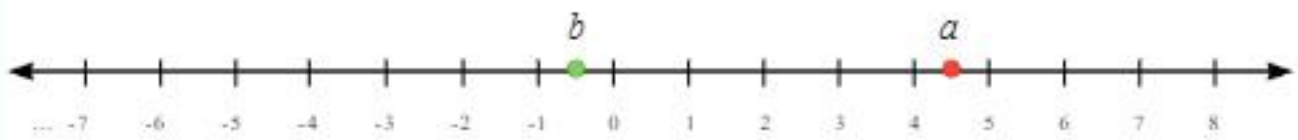
1.2.1 Relación de orden

Los números reales se pueden representar geométricamente en la recta numérica, la cual tiene un punto fijo **O** que se llama origen y que corresponde al número real cero. Considerando un segmento unitario como unidad de medida, a la derecha de O se indican los puntos que corresponden a los números reales positivos ($\mathbb{R}^+$)y a la izquierda de O los puntos que corresponden a los números reales negativos ($\mathbb{R}^-$). De esta manera, a cada número real le corresponde un único punto de la recta, y a cada punto de la recta, un único número real.

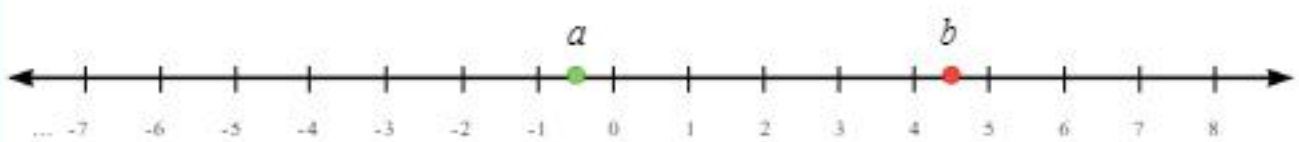


Para los números reales se puede definir una relación de orden respecto a la recta numérica, por ejemplo:

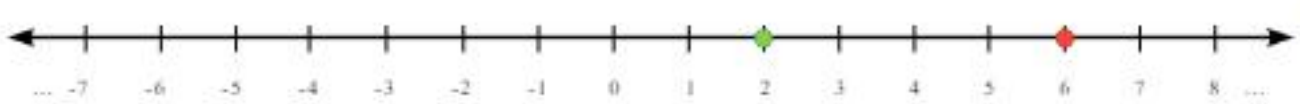
Se tiene que un número *a* es mayor que otro número *b* si el punto asociado al número *a* se ubica a la derecha del punto asociado al número *b*.



Por otro lado, un número *a* es menor que otro número *b* si el punto asociado al número *a* se localiza a la izquierda del punto asociado al número *b*.



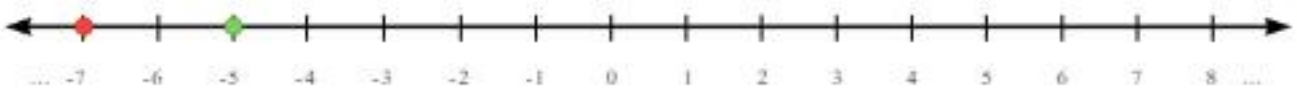
¿Cómo simbolizamos si un número es mayor que otro?



En la figura anterior, podemos ver que el número 6 está ubicado a la derecha del número 2, por lo que, decimos que el número 6 es mayor que el número 2.

Utilizamos el símbolo >, para indicar que un número es mayor que otro. Por ejemplo, el número 6 es mayor que el número 2, y lo representamos de la siguiente forma: 6 > 2

¿Cómo simbolizamos si un número es menor que otro?



Si observamos la posición de los puntos asociados a estos números en la recta anterior, podemos notar que el del número -7 está ubicado a la izquierda del punto correspondiente al número -5, por lo que, decimos que el número -7 es menor que el número -5.

Utilizamos el símbolo <, para indicar que un número es menor que otro. Por ejemplo, el número -7 es menor que el número -5, y lo representamos de la siguiente forma: -7 < -5

De lo anterior, podemos decir que sólo una de las expresiones siguientes es verdadera:

$$a > b, \quad a < b \quad o\,bien \quad a = b$$

Esta propiedad recibe el nombre de ***ley de tricotomía.***

Cabe aclarar que $a < b$ equivale a $b > a$,
esto es, que, si a **es menor que** b, equivale a decir que b **es mayor que** a



Actividad 3.

Relación de orden

Instrucciones:
Utilizando los símbolos operadores racionales comunes mayor que (>), menor que (<) o igual (=) **arrastra** de la tabla inferior la respuesta correcta y relaciona los siguientes pares de números. Da clic en la **imagen.**

1.2.2 Jerarquía de operaciones

En el conjunto de los números reales se pueden definir las siguientes operaciones:

Adición	Multiplicación	Potenciación
Es la operación aritmética la cual resulta de agregar una o más cantidades para obtener un total. Se expresa de la siguiente manera: $a+b=c$ Los términos de la operación, a y b se consideran los sumandos y el resultado, c, adición (conocida como suma).	La multiplicación en los números reales implica el sumar uno de los valores con él mismo tantas veces como indica el otro valor. Se expresa de la siguiente manera: $\underbrace{a+a+\cdots+a}_{n\ veces}=n\times a=c$ Los términos de la operación, n y a se llaman los factores y el resultado c se llama producto.	La potenciación consiste en multiplicar un número llamado base por sí mismo, las veces que exprese el exponente. Esto es: $\underbrace{a*a*\cdots*a}_{n\ veces}=a^n$ Donde a es la base y n el exponente.

Se sabe por las propiedades de las operaciones básicas con números reales de la adición y la multiplicación que sus respectivas operaciones opuestas son: la sustracción (existencia del inverso aditivo) y la división (existencia del inverso multiplicativo). Revisa las siguientes propiedades:

Propiedad	Sean $a, b \wedge c \in \mathbb{R}$	Ejemplo de la propiedad
Cerradura	$a+b \in \mathbb{R}$	$\frac{3\pi}{2}+\frac{4}{5}=\frac{15\pi+8}{10}$
Asociativa	$a+(b+c)=(a+b)+c$	$\frac{5}{2}\left(\frac{3}{2}+\frac{4}{2}\right)=\left(\frac{5}{2}+\frac{3}{2}\right)+\frac{4}{2}$ $\frac{12}{2}=\frac{12}{2}$ $6=6$
Conmutativa	$a+b=b+a$	$\frac{4}{5}+2.6=2.6+\frac{4}{5}$ $3.4=3.4$
Neutro Aditivo	Existe $0 \in \mathbb{R}$ tal que $a+0=a$	$4.56+0=4.56$
Inverso Aditivo	Para $a \in \mathbb{R}$, existe $-a \in \mathbb{R}$ tal que $a+(-a)=a-a=0$	Si $3 \in \mathbb{R}$, existe $-3 \in \mathbb{R}$ tal que $3+(-3)=3-3=0$

Tabla 3. Propiedades para la adición y la sustracción. (Elaboración propia)

Propiedad	Sean $a, b \wedge c \in \mathbb{R}$	Ejemplo de la propiedad
Cerradura	$a \cdot b \in \mathbb{R}$	$(-2)\left(\frac{\pi}{5}\right)=-\frac{2\pi}{5}$
Asociativa	$a\,(b \cdot c)=(a \cdot b)\,c$	$3(2.5 \cdot 7.1)=(3 \cdot 2.5)7.1$ $3(17.75)=(7.5)7.1$ $53.25=53.25$
Conmutativa	$a \cdot b = b \cdot a$	$(2.5 * 5)=(5 * 2.5)$ $12.5=12.5$
Neutro Multiplicativo	Para $a \in \mathbb{R}$, existe $1 \in \mathbb{R}$ tal que $1 \cdot a = a$	$1 \cdot \sqrt{3}=\sqrt{3}$
Inverso Multiplicativo	Para $a \in \mathbb{R}$, $a \neq 0$ $a\left(\frac{1}{a}\right)=\frac{a}{a}=1$	$6\left(\frac{1}{6}\right)=\frac{6}{6}$ $1=1$
Distributiva de la multiplicación con respecto a la adición	$a(b+c)=ab+ac$	$8\left(\frac{2}{5}+0.5\right)=\left(8*\frac{2}{5}\right)+(8*0.5)$ $8(0.9)=\frac{16}{5}+4$ $7.2=7.2$

Tabla 4. Propiedades para el producto y el cociente. (Elaboración propia)

Ahora bien, para resolver operaciones combinadas es necesario tener en cuenta que se aplica la llamada **jerarquía de operaciones** que consiste en lo siguiente:

1. **Se efectúan las operaciones dentro de los signos de agrupación**
2. **Se calculan las potencias y las raíces**
3. **Se efectúan los productos y los cocientes**
4. **Se realizan las sumas y restas**

Si respetamos esta jerarquía y no se tienen errores de tipo aritmético, el resultado será único. Si dos personas tienen resultados diferentes al realizar un determinado ejercicio, implica que alguien no respetó la jerarquía o bien, hubo algún error de tipo aritmético en el cálculo.

Ejemplo 1: Realiza la siguiente operación de acuerdo con la jerarquía establecida

$$4-6\div(2+1)+2\times4^2-\sqrt{10+6}\times3=$$

Se efectúan las operaciones dentro de los signos de agrupación. El paréntesis y la raíz están fungiendo como signos de agrupación	$4-6\div(2+1)+2\times4^2-\sqrt{10+6}\times3$
Se calculan las potencias y las raíces	$4-6\div(3)+2\times4^2-\sqrt{16}\times3$
Se efectúan los productos y los cocientes	$4-6\div(3)+2\times16-4\times3$
En este punto únicamente se tienen sumas y restas y se resuelven	$4-2+32-12$
Resultado	22

Ejemplo 2: Realiza la siguiente operación de acuerdo con la jerarquía establecida

$$12+5\,(2-3)^2+\{[3+1-4\,(7-5)+2(6-8)]\}=$$

Se efectúan las operaciones dentro de los signos de agrupación	$12+5\,(2-3)^2+\{2[3+1-4\,(7-5)+2(6-8)]\}=$
Las operaciones en el interior de cada paréntesis se han resuelto, sin embargo, esos mismos paréntesis aún continúan porque cada uno involucra otra operación, por lo que debemos resolverlos	$12+5\,(-1)^2+\{2[3+1-4\,(2)+2(-2)]\}=$
Nuevamente resultan operaciones dentro de un signo de agrupación por lo que se resuelven	$12+5\,(1)+\{2[3+1-8-4]\}$

En este paso tenemos paréntesis y corchete, sin embargo, ambos están representando multiplicación, por lo que se resuelven	$12+5\,(1)+\{2[-8]\}$
Una vez más, la llave está representando multiplicación, por lo que se resuelve	$12+5+\{-16\}$
En este punto únicamente se tienen sumas y restas y se resuelven	$12+5-16$
Resultado	1

Ejemplo 3: Realiza la siguiente operación de acuerdo con la jerarquía establecida

$$\left[5-\left((2)^3-\frac{10}{2}\right)\right]\left[5+((3)(2)-4)\right]-3+(8-(2)(3))$$

Buscamos signos de agrupación (azul) que en su interior tengan operaciones que se pueden resolver (rosa):	$\left[5-\left((2)^3-\frac{10}{2}\right)\right]\left[5+((3)(2)-4)\right]-3+(8-(2)(3))=$
Esos mismos paréntesis (azul) aún continúan porque cada uno involucra otra operación, por lo que debemos resolverlas	$[5-(8-5)]\,[5+(6-4)]-3+(8-6)=$
Nuevamente resultan operaciones dentro de un signo de agrupación por lo que se resuelven	$[5-3]\,[5+2]-3+2=$
Los corchetes están representando multiplicación por lo que se resuelve	$[2]\,[7]-3+2=$
En este punto únicamente se tienen sumas y restas y se resuelven	$14-3+2=$
Resultado	13

Las operaciones aritméticas son importantes no solo en las matemáticas, se utilizan en la vida cotidiana para calcular el cambio en una compra, determinar cuánta comida se necesita para una receta. Son esenciales en las ciencias como la física, la ingeniería, la informática y la economía para realizar cálculos, modelar fenómenos y resolver problemas. También se utilizan para analizar datos, calcular estadísticas y tomar decisiones informadas en muchos contextos sociales hasta la investigación científica.





Actividad 4.



Jerarquía de operaciones

Instrucciones:

Descarga el archivo adjunto, dando clic en la imagen, imprímelo y realiza el ejercicio.

1.3 Instrumentos de recolección de información

La investigación es una de las actividades que se desarrolla en todas las ciencias sociales y exactas, de manera que la implementación de técnicas e instrumentos de recolección de información forma parte de las metodologías de la investigación para alcanzar los objetivos del estudio en cuestión. Una buena técnica determina en gran medida la calidad de la información recabada.

La recolección de datos es una forma de obtener conocimiento científico.

Los **métodos o técnicas de recolección de datos o información** son el medio, los procedimientos o actividades orientados al desarrollo de instrumentos de evaluación, medición a través de los cuales se establece la relación entre el investigador y el consultado.

Algunas técnicas de recolección de datos son:

Encuesta	• Técnica en la cual se plantea un listado de preguntas para obtener datos precisos.
Observación participante	• Técnica que consiste justamente en observar el desarrollo de un fenómeno social, ambiental, experimental que se desea analizar.
Entrevista	• Técnica que consiste en una conversación generalmente planificada, donde el investigador plantea una serie de preguntas previamente definidas sobre algún tema a una o varias personas, con el fin de obtener información específica.
Grupos focales	• Técnica que consiste en reunir un grupo de personas con características relacionadas con la investigación y realizar una entrevista grupal orientada hacia la información que se desea obtener.
Documentos o registros	• Ténica que consiste en examinar los datos presentes en documentos ya existentes como bases de datos, actas, informes, registros de asistencias, etc.
Delphi	• Técnica que consiste en indagar con una serie de expertos un tema determinado para orientar la toma de decisiones.

Figura 2. Técnicas de recolección de datos (Elaboración propia

Las técnicas de recolección de datos necesitan del desarrollo de instrumentos específicos relativos el fenómeno o tema que se desea investigar. Se define un **instrumento de recolección de datos** como la herramienta con la que se cuenta para documentar o registrar la información recabada de la realidad la cual pude ser cualitativa o cuantitativa

Algunos instrumentos de recolección de datos son:

1. Formularios o cuestionarios:

Los primeros se caracterizan por crear preguntas que no tienen ningún tipo de puntuación o corrección. Mientras que los cuestionarios caracterizan por estar orientados, pueden ser de tipo test o pueden ser preguntas de texto con una puntación.

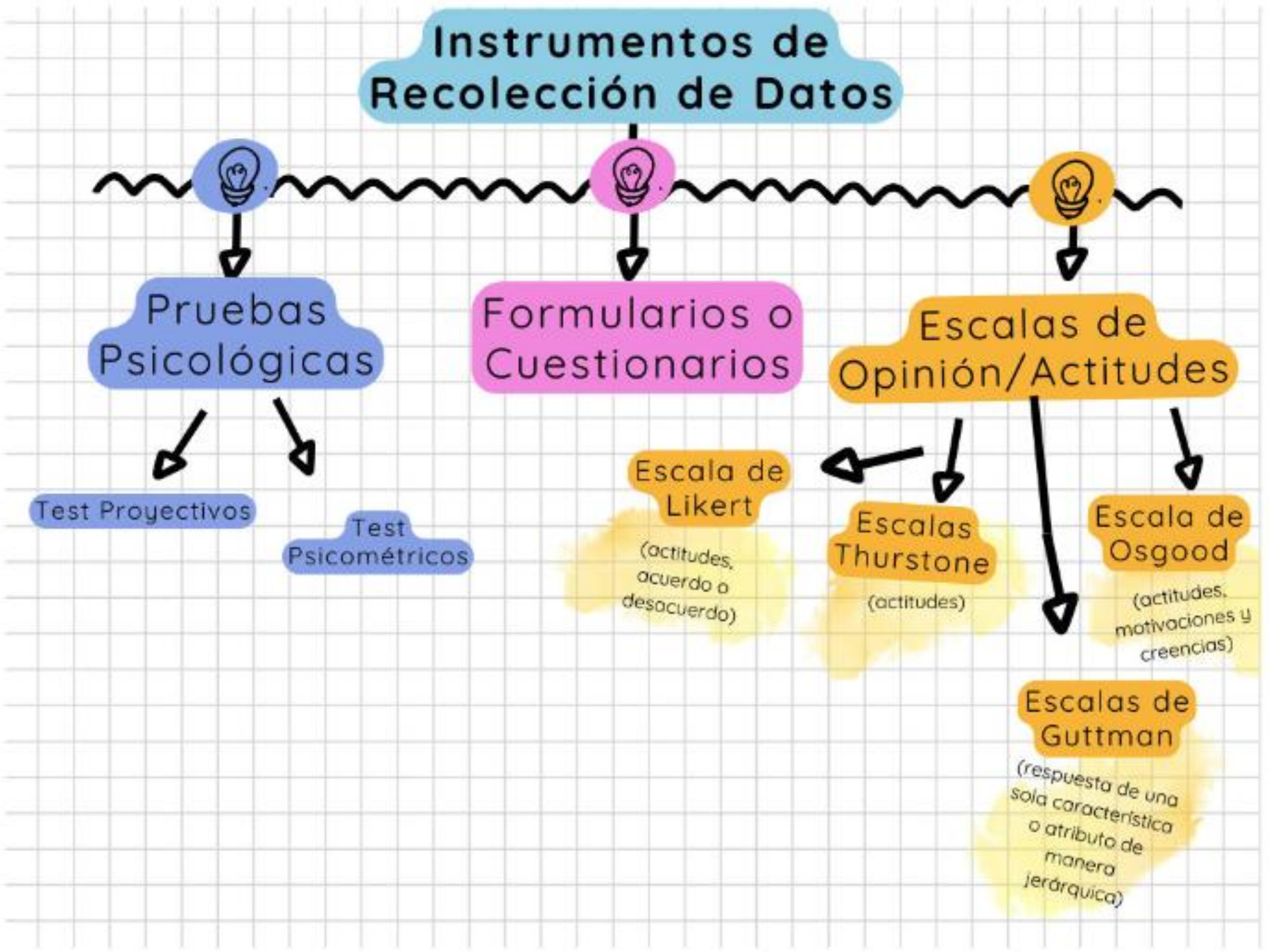
2. Pruebas psicológicas:

Son instrumentos estandarizados utilizados para medir una o más características de las personas como personalidad, inteligencia, atención de manera rápida, etc.

3. Escalas de Opinión, actitudes e intereses:

Son instrumentos utilizados para la medición de aspectos relacionados a la intensidad de un sentimiento o la actitud de un grupo de personas respecto a un tema en particular.

Figura 3. Instrumentos de recolección de datos



Nota. Fuente: Elaboración propia

1.4 Diseño de encuestas

De los métodos y técnicas de investigación para la recolección de datos una de las más importantes es la **encuesta** que es una investigación o búsqueda sistemática y ordenada de obtener información, datos y comentarios por medio de una serie de preguntas específicas, con la finalidad de recopilar información de un grupo o población de interés, cuyas respuestas servirán para analizarlas, interpretarlas y tener un panorama que ayude a tomar decisiones o a generar alguna estrategia o acción específica de la investigación que se está realizando.

En una encuesta hay dos actores el **encuestado** (persona o personas que diseñan o aplican la encuesta) y el o los **encuestados** (quienes dan respuesta a los cuestionarios).

Para diseñar una encuesta es importante:

1. Tener claro el **objetivo,** éste establece una guía para adaptar las preguntas con la finalidad de obtener las respuestas necesarias.

2. Considerar los **métodos de aplicación o recopilación** de datos los cuáles son:

- **Online:** método para realizar un cuestionario usando internet como medio de distribución y hoy es uno de los tipos de encuestas más populares. Son fáciles de diseñar y de desplegar, se puede aplicar desde cualquier lugar en cualquier momento.
- **Correo:** método que en enviar un cuestionario a través del servicio postal para que después de ser respondido, sea regresado al remitente. La encuesta por correo electrónico es más usual que la tradicional.
- **Telefónico:** método para realizar el cuestionario mediante una llamada telefónica en la que el encuestador debe capturar las respuestas del encuestado.
- **Físico o presencial:** método que consiste en realizar la encuesta cara a cara, por lo que debe existir una interacción entre encuestador y encuestado. Normalmente el encuestador toma nota de las respuestas, aunque en ocasiones solo entrega el cuestionario y es el encuestado quien lo llena.

3. Definir el **tipo de encuesta:** según el tipo de pregunta y respuesta, las cuáles pueden ser abiertas o cerradas.

- **Respuesta abierta:** permiten al encuestado tener la libertad de responder libremente cada pregunta, lo que permite obtener respuestas más profundas y también obtener respuestas en las que no había pensado. En este tipo de encuestas suele ser difícil de cuantificar, por lo que se requieren herramientas de análisis de texto para interpretar las respuestas.

- **Respuesta cerrada:** los participantes elegirán alguna opción, cuya principal ventaja es que son más fáciles de cuantificar.



Figura 4. Método de aplicación o recolección de datos

La investigación y todo el trabajo que esto implica para recolección de datos no tiene sentido hasta que esta se registra, analiza y se interpreta. ¿Quién se encarga de esto?

1.5 Conceptualización y clasificación de la estadística

La estadística la podemos apreciar como cifras en porcentajes, números y gráficos que reflejan el resultado de un análisis a cierta actividad o información de interés y la presentan en diversos medios de comunicación tanto visuales como auditivos en áreas a nivel científico, comercial, educativo, social, etc., siendo la estadística una gran herramienta de análisis y comprensión de información a situaciones problema y hechos reales o ficticios.



Definición

La **estadística** es una rama de las matemáticas que permite mediante diferentes métodos aritméticos y sobre un conjunto de datos recogerlos, organízalos, resumirlos y arrojar un análisis e interpretación de estos.

La estadística se utiliza prácticamente en cualquier área de conocimiento, tiene una gran flexibilidad al utilizar diferentes tipos de datos, por ejemplo:

- Identificar algún tipo de carne más vendida en una carnicería.
- En el pronóstico del clima.
- Al llenar la quiniela de algún deporte considerando el desempeño de los equipos.
- En encuestas de salida en elecciones políticas e ir informando el porcentaje de votos a favor de los participantes.
- Permitir a las escuelas tener un panorama de su matrícula.

Figura 5. Ejemplos de estadísticas



Nota. Fuente: Shutterstock.com

Clasificación de la estadística

Estadística Descriptiva	Estadística Inferencial
Consiste en la presentación de datos en forma de tablas y gráficas. Esta comprende cualquier actividad para resumir o describir los mismos factores pertinentes adicionales, esto se refiere a no intentar nada que vaya más allá de los datos. • Su objetivo es recopilar, ordenar y representar la información proporcionada por los datos, ofreciendo interpretación cuantitativa de variables como estatura de una población, ingresos o egresos en un negocio entre otros. • Se hará uso de fórmulas para que con la información recabada se proporcione un análisis de ésta al obtener medidas de tendencia central (media, mediana y moda), de variabilidad (rango, desviación estándar, varianza y rango intercuartílico). • Utiliza tablas y gráficos, fundamental para mostrar la información de manera visual y concreta.	Determina y deduce las características de una población a partir del análisis de una muestra de ésta. Ello por la imposibilidad de analizar cada elemento de la población en casi todas las situaciones y extrapolando para generalizar a toda la población. Los tipos de estadística inferencial son: **• Estadística paramétrica** Rama de la estadística inferencial que considera que los datos se pueden trabajar mediante una distribución de probabilidad y hace uso de pruebas estadísticas que se adecúan a distribuciones de probabilidad conocidas. Fundamentalmente sirve estimar un parámetro, ya sea mediante estimaciones puntuales o por intervalos. **• Estadística no paramétrica** Rama de la estadística inferencial que estudia las variables que no encajan en una distribución de probabilidad o donde los parámetros de la distribución son indeterminados. En resumen, la estadística no paramétrica se utiliza cuando las variables no pueden definirse con modelos teóricos.
Comprende tres categorías principales: 	

Tabla 5. Clasificación de la estadística (Elaboración propia)



Actividad 5.



Introducción a la estadística descriptiva

Instrucciones:
Te invitamos a revisar el siguiente material referente a la estadística descriptiva. Da clic en la **imagen.**

En resumen

La estadística inferencial nos sirve cuando deseamos conocer las cantidades reales empacadas o envasadas de un producto al público, funciona entorno a lo que marcan las cantidades en los empaques o envases. Este proceso en la mayoría de las ocasiones suele ser de forma automática y siempre habrá una posibilidad de recibir más o menos producto. Una incertidumbre bastante frecuente es cuando al cargar gasolina vemos e interpretamos si fue o no despachada de forma correcta; sólo con el hecho de que la aguja indicadora del auto levante o por el rendimiento en kilómetros que nos proporciona el auto; lo podemos intuir.

Durante el desarrollo de módulo, la aplicaremos para hacer deducciones sobre los parámetros estadísticos de una población, donde solamente analizaremos una porción de ella.

La estadística descriptiva, por su parte, la utilizaremos para mostrar las frecuencias de una muestra de datos en alguno de los diferentes gráficos. Nos ayudará a determinar la muestra de un estudio estadístico, a través de calcular su media aritmética, mediana, moda y desviación estándar.

1.5.1 Dato estadístico, variable y su clasificación

En el acontecer diario, vemos recopilada información a través de encuestas, experimentos sociales, sondeos, entre otros. Dicha información puede representarse mediante diversas formas, como tablas, gráficos, diagramas, etc. A continuación, presentaremos el concepto de dato estadístico, variable y cómo clasificarlos.

Dato estadístico
Los datos estadísticos, son los valores que se obtienen al llevar a cabo un estudio de tipo estadístico. Serán el resultado de la observación del fenómeno que queremos analizar.

Un dato estadístico de forma individual no tiene relevancia, consideremos que el procesamiento de estos es el que arrojará la información deseada.

Obtener y procesar los datos estadísticos es de gran relevancia para considerarlo y/o aplicarlo en diversos ámbitos.

Se dividen en:

Datos discretos

Los datos discretos son aquellos que se pueden contar o medir, por ejemplo: el número de peces que contiene el acuario de Veracruz o como la estatura de los niños de un jardín de niños.

Datos continuos

Los datos categóricos se subdividen en:
- Datos nominales que identifican una categoría (Tipos de razas de perros)

- Datos ordinales que se utilizan para clasificar u ordenar (Ranking semanal de canciones de moda).

Variable estadística

Característica medible u observable que se puede registrar en un conjunto de datos.

Tipos:

Variable Cuantitativa:

Son variables indicadas numéricamente. Las hay de 2 tipos:

Variable continua: Toman un valor infinito de valores entre un intervalo de datos (puede ser número decimal). Nivel de triglicéridos de una persona adulta (menos de 150 mg/dl).

Variable discreta: Toman un valor finito de valores enteros entre un intervalo de datos. Número de tortas vendidas en un receso escolar de alguna institución educativa

Variable Cualitativa:

Son variables que se expresan por normativa, en palabras.

Variable ordinal: Expresa diferentes niveles y orden. (Tabla de clasificación de la liga MX en México)

Variable nominal: Expresa un nombre claramente diferenciado. Por ejemplo, tipos de osos que pueden ser oso malayo, oso labiado, oso de anteojos u oso pardo.

Existen también más subtipos de variables: económicas, categóricas, dicotómicas, dependientes, cardinales.

Es importante mencionar que cualquier variable cualitativa puede ser utilizada para un modelo matemático asignando números a la misma para generar variables cuantitativas, por ejemplo, para el oso malayo asignar un 1, para el oso labiado un 2, para el oso de anteojos un 3 y para el oso pardo un 4. Las variables dicotómicas que indica SI asignar un 1 y NO asignar un 0.

Para el diseño de una investigación, la estadística ayuda a decidir el tipo de dato que se estudiará, sin embargo, es importante determinar el sector (ciudad, pueblo, alumnos, manada, etc) y la proporción de este a la que estudiará u observará para garantizar la validez y la fiabilidad de los resultados de la investigación. El sector y la porción de este suelen llamarse población y muestra.

1.6 Población y muestra

Al ver nuestro entorno, nos es fácil identificar que pertenecemos a un núcleo familiar, éste a su vez se encuentra en una colonia, misma que pertenece a una ciudad, la cual es parte de un municipio que pertenece a un estado y que finalmente junto con otros 30 estados más conforman nuestro país; considerando lo anterior tendremos infinidad de situaciones que sean de nuestro interés para obtener datos estadísticos y que involucren alguna de las poblaciones anteriores, considerar a todos los individuos que conforman las poblaciones en ocasiones resulta complicado, caro o imposible de realizar, ante esas situaciones lo que se hace es tomar una muestra de la misma.

Hay más tipos de poblaciones y en todas ellas tendremos las mismas limitantes, por ende, son muy importantes las muestras que representen las características o información de la población total.

Definición

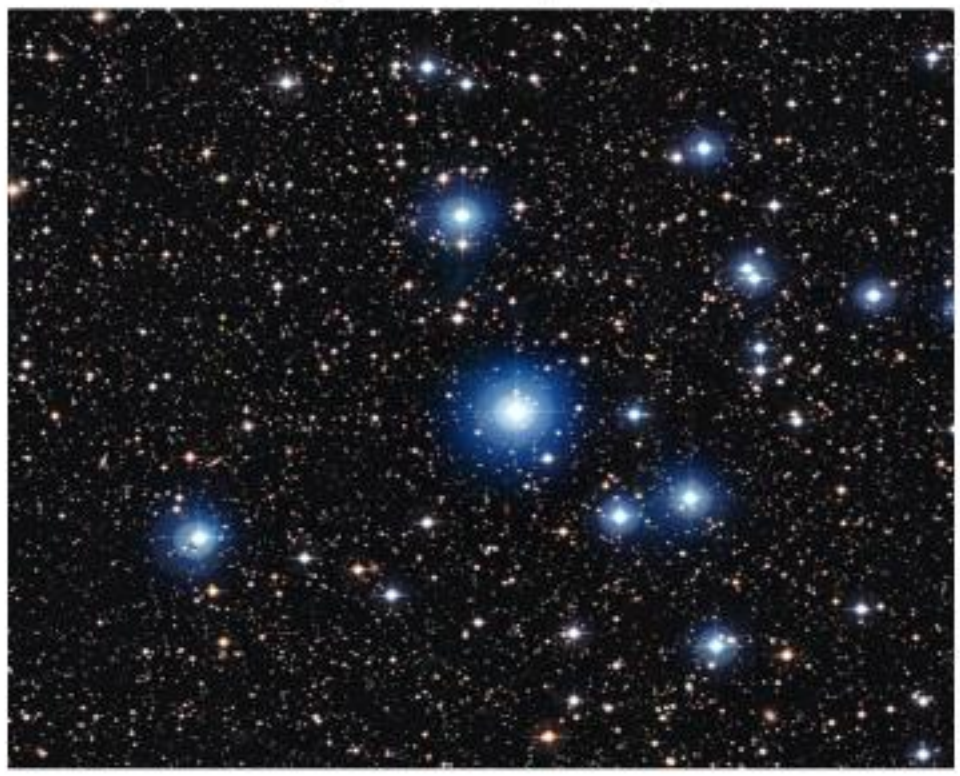
Población o población estadística
En estadística, una población es un conjunto de elementos, cosas o individuos que comparten características similares con las que se pretende hacer un estudio o análisis estadístico.
Existen dos tipos de población estadística:
• Población estadística finita: población cuyo número de elementos es contable o finito.
• Población estadística infinita: población cuyo número de elementos es indeterminado o infinito.

Figura 6. Perros callejeros (ejemplos)



Por ejemplo, el número de perros callejeros en una comunidad es una población finita ya que, aunque fueran muchos, es contable.

Figura 7. El universo (estrellas)



Por el contrario, el número de estrellas en el universo es una población infinita porque aún no se sabe con certeza la cantidad existente y, por lo tanto, no se puede dar un número exacto.

Definición

Muestra
La muestra estadística también se puede considerar un subconjunto de la población la cual será analizada y que permitirá obtener información confiable de la población ahorrando tiempo y dinero

Para relacionar ambos conceptos se muestran los siguientes ejemplos:

Situación de interés	Población para analizar	Muestra
Se desea saber en una farmacia lo que en promedio gastó un cliente durante el día.	Juntar las notas de las ventas del día.	Elegir aleatoriamente 50 notas de venta.
Sueldo mensual que gana un docente en nivel medio superior en la UAEMéx.	Talones de pago de los docentes.	Elegir aleatoriamente 100 talones de pago.
La edad promedio de los alumnos adscritos en un plantel universitario de la UAEMéx.	Los alumnos del plantel universitario.	Elegir los que su apellido paterno inicia con B.

Siendo la muestra la que nos permitirá trabajar y proporcionar la información deseada, es importante saber el número que permita hacer un diagnóstico de forma confiable, para ello haremos uso de la siguiente fórmula:

$$n = \frac{N}{1 + N(e)^2}$$

donde:
n = tamaño de la muestra
N = tamaño total de la población
e = error máximo (expresado en decimal)

Ejemplo 1:
Tenancingo, Estado de México contaba hasta el censo del año 2020 con 104,677 habitantes, se desea conocer la edad promedio de dicha población, determinar el tamaño de la muestra que permitirá conocer de forma confiable el dato estadístico. Considerar un error del 3%.

Utilizando la fórmula propuesta

$$n = \frac{N}{1 + N(e)^2}$$

sustituiremos:
n = tamaño de la muestra
N = 104,677
e = 0.03

$$n = \frac{104\,677}{1 + 104\,677(0.03)^2}$$

$$n = 1100 \ habitantes.$$

Para una población de 104,677 habitantes considerando un error del 3% es suficiente encuestar a una muestra de 1100 habitantes.

Ejemplo 2:
La matrícula aceptada por la UAEMéx a nivel bachillerato en el periodo 2022-2023 en sus tres modalidades fue de 8073 alumnos, determinar el tamaño de la muestra para determinar el promedio académico con el cual ingresaron. Considerar un error del 5%

Utilizando la fórmula propuesta

$$n = \frac{N}{1 + N(e)^2}$$

sustituiremos:

n = tamaño de la muestra
N = 8073 alumnos
e = 0.05

$$n = \frac{8073}{1 + 8073(0.05)^2}$$

$$n = 381 \ alumnos$$

Para una población de 8073 alumnos considerando un error del 5% es suficiente encuestar a una muestra de 381 alumnos.

Considerando los tamaños de las muestras obtenidas, se aprecia una disminución considerable de los elementos a encuestar que se reflejará en ahorro de tiempo y dinero.

1.7 Conjuntos

Los seres humanos desde pequeños somos dados a ordenar, agrupar y clasificar todo tipo de cosas u objetos, por ejemplo, los carritos de juguetes, las pelotas, y con más edad hemos aprendo a clasificar los contenidos de películas en películas de comedia, terror, románticas, etc. Los estilos de ropa en casual, formal, urbana, deportiva, fiesta, etc. Tipos de alimentos en el plato del buen comer. En matemáticas no nos quedamos atrás, podemos clasificar números en números enteros, números positivos, números negativos entre otros.

Definición

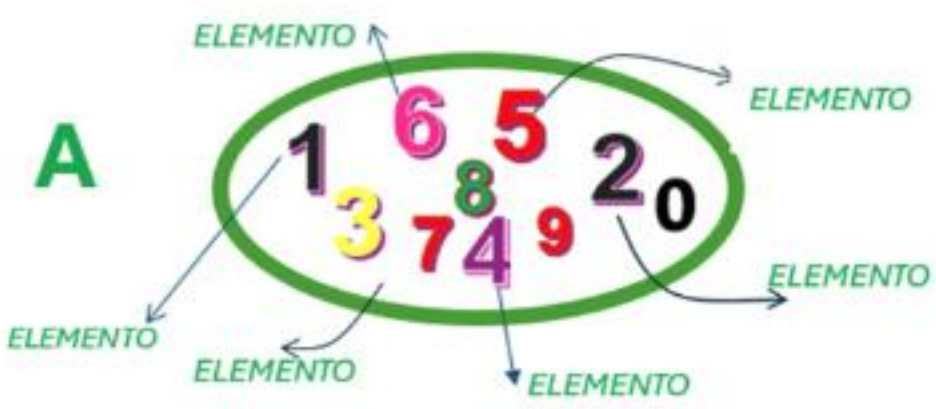
Un **conjunto** es una colección o familia de objetos con alguna característica en común. Generalmente se denotan por letras mayúsculas A, B, C, … a cada objeto del conjunto se llama elemento.

Los conjuntos se pueden representar de manera gráfica colocando los elementos del conjunto en círculos (cuadrados, rectángulos, o curva cerrada) que permiten agruparlos. A este tipo de gráfico se le conoce como diagrama de Venn.

Si a es un elemento del conjunto A, dicho elemento pertenece al conjunto y escribimos $a \in A$. En caso contrario, si a no es un elemento de A se simboliza $a \in A$.

El siguiente diagrama de Venn representa el conjunto A, donde A es el conjunto de los dígitos, es decir, $0,1,2,3,4,5,6,7,8,9 \in A$.

Figura 8. Diagrama de Venn del conjunto de dígitos



Por otro lado, un conjunto se puede expresar mediante símbolos matemáticos en dos formas por **extensión** y por **comprensión.**

Figura 9. Representación del conjunto (Elaboración propia)



Ejemplo 1:
Representación en su forma por extensión y comprensión del conjunto de dígitos

- Por extensión $A = \{0,1,2,3,4,5,6,7,8,9\}$ Se lee "A es el conjunto de los números $0,1,2,3,4,5,6,7,8,9$,"
- Por comprensión $\{x|x \text{ es un dígito}\}$ Se lee "A es el conjunto de los x, tal que x es un dígito"

Ejemplo 2:
Representación en su forma por extensión y comprensión del conjunto de días de la semana

- Por extensión $B = \{lunes, martes, miércoles, jueves, viernes, sábado, domingo\}$

Se lee "B es el conjunto de los días de la semana: lunes, martes, miércoles, jueves, viernes, sábado, domingo"

• Por comprensión $\{x \mid x \text{ es un día de la semana}\}$

Se lee "B es el conjunto de los *x*, tal que *x* es un día de la semana

Ejemplo 3:
Representación en su forma por extensión y comprensión de los siguientes sucesos: Al lanzar un dado al azar y obtener un seis

• Por extensión C = {6} Se lee "C es el conjunto del suceso de lanzar un dado al azar y obtener un seis"

• Por comprensión $\{x|x$ *es obtener un seis al lanzar un dado al azar*$\}$

Se lee "C es el conjunto de los *x*, tal que *x* es un seis al lanzar un dado al azar"

Ejemplo 4:
Representación en su forma por extensión y comprensión de los siguientes sucesos: Al lanzar un dado al azar y obtener un seis

• Por extensión D = {2,4,6} Se lee "D es el conjunto del suceso de lanzar un dado al azar y obtener un número par"

• Por comprensión $\{x|x$ *es obtener un número par al lanzar un dado al azar*$\}$

Se lee "D es el conjunto de los *x*, tal que *x* es un número par al lanzar un dado al azar"



Actividad 6.



Conjuntos por extensión y por comprensión

Instrucciones:
Te invitamos a resolver el siguiente juego sobre los conjuntos por extensión y por comprensión.

Da clic en la **imagen.**

1.7.1 Operaciones con conjuntos

Figura 10. Conjuntos (ejemplo)



Nota. Fuente: Shutterstock.com

¿Quiénes de tus compañeros estudian inglés o realizan algún deporte fuera del horario de clases?

¿Ahora en preparatoria, sabes cuántos compañeros de tu salón de clases son egresados de una misma secundaria?

Estas son algunas situaciones – problema de la vida cotidiana que pueden ser resueltas por medio de las operaciones del álgebra de conjuntos como lo es la Unión y la Intersección, veamos a detalle estas operaciones:

Unión de conjuntos

Cuando te arreglas para asistir a la escuela, unes varios elementos de tu armario, como puede ser la ropa con algunos accesorios, unes estos elementos y formas un solo conjunto que es tu atuendo, en matemáticas la unión de conjuntos tiene una definición formal que es la siguiente:

Definición

La unión de dos conjuntos A y B es un conjunto formado por todos los elementos de dichos conjuntos, es decir, es el conjunto de elementos que pertenecen al conjunto A o al conjunto B. Considera que la Unión de conjuntos puede ser entre dos o más conjuntos.

Definición

Y la forma de escribir esta operación entre conjuntos es:

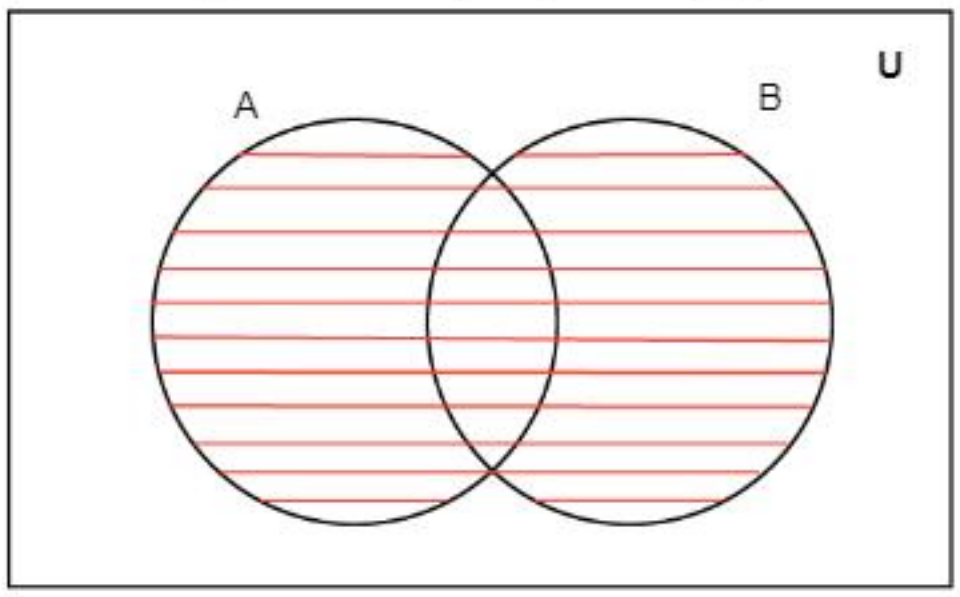
$$\textbf{Notación: } A \cup B = \{x | x \in A \vee x \in B\}$$

Se lee "A unión B es el conjunto de las *x*, tal que *x* pertenece al conjunto A ó *x* pertenece al conjunto B"

Nota: Si se repiten los elementos, solamente se escriben una vez y se sugiere ordenar los elementos con algún criterio.

Su representación gráfica utilizando un Diagrama de Venn-Euler es:

Figura 11. Unión de conjuntos (Elaboración propia)



Observa que se colorean ambos conjuntos.

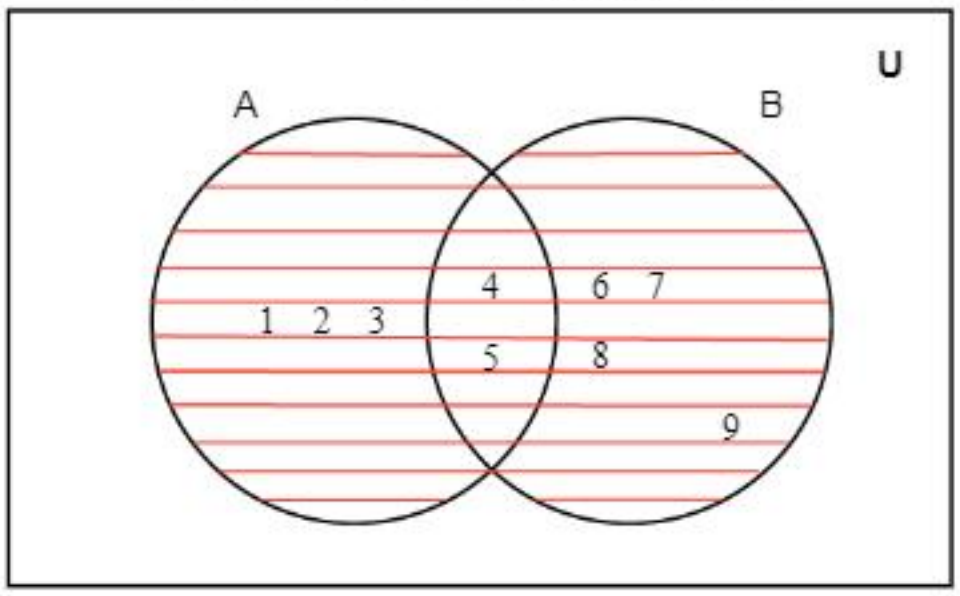
Ejemplo 1:
Sean los siguientes conjuntos: $A=\{1,2,3,4,5\}$ y $B=\{4,5,6,7,8,9\}$. Determine $A \cup B$

Se unen todos los elementos de ambos conjuntos y se obtiene: $A \cup B = \{1,2,3,4,5,6,7,8,9\}$

Nota: Los elementos 4 y 5 pertenecen a ambos conjuntos, sin embargo, sólo se escriben una vez y observen que los elementos siguen un orden, en este caso están escritos de forma ascendente.

Su representación gráfica utilizando un Diagrama de Venn-Euler es:

Figura 12. $A \cup B = \{1,2,3,4,5,6,7,8,9\}$(Elaboración propia)



Los elementos que se repiten se escriben en la parte central, donde se tiene el área común entre el conjunto A y el conjunto B.

Ejemplo 2:
Sean los siguientes conjuntos: $C = \{x | x$ *es obtener un seis al lanzar un dado al azar*$\}$ y $D = \{x | x$ es obtener un número par al lanzar un dado al azar$\}$. Determine $C \cup D$

Los conjuntos son: $C = \{6\}$ y $D = \{2,4,6\}$

Se unen todos los elementos de ambos conjuntos y se obtiene: $C \cup D = \{2,4,6\}$

Nota: El elemento 6 pertenece a ambos conjuntos, sin embargo, sólo se escribe una vez y observen que los elementos siguen un orden, en este caso están escritos de forma ascendente.

Definición

La intersección de dos conjuntos A y B es un conjunto formado por los elementos que son comunes al conjunto A y al conjunto B, es decir, es el conjunto de elementos que pertenecen al conjunto A y al conjunto B. Considera que la Intersección de conjuntos puede ser entre dos o más conjuntos.

Definición

Y la forma de escribir esta operación entre conjuntos es:

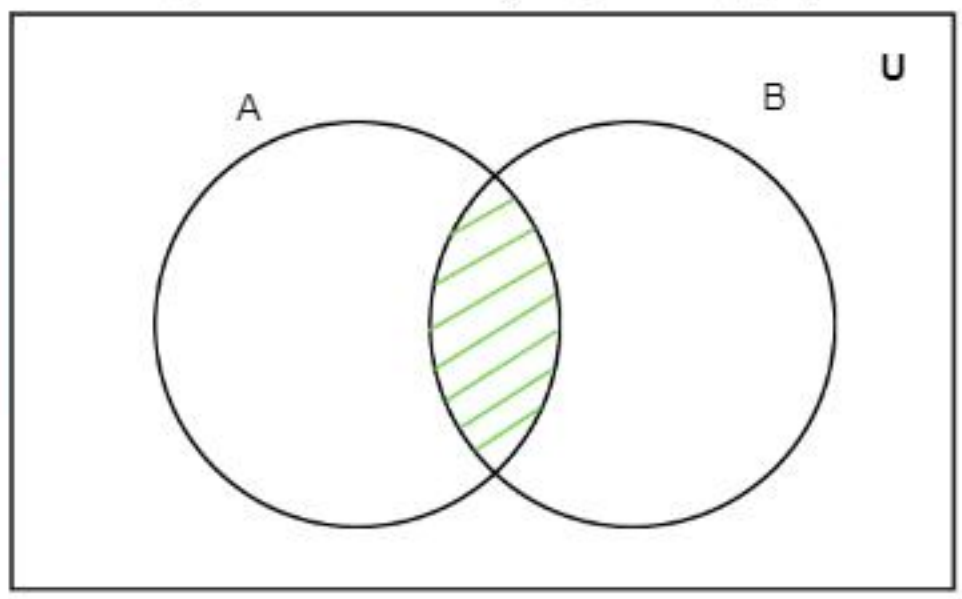
Notación: $A \cap B = \{x | x \in A \wedge x \in B\}$

Se lee "A intersección B es el conjunto de las x, tal que x pertenece al conjunto A y x pertenece al conjunto B"

Nota importante: Se sugiere ordenar los elementos con algún criterio.

Su representación gráfica utilizando un Diagrama de Venn-Euler es:

Figura 13. Intersección de conjuntos (Elaboración propia)



Observa que se colorea sólo la parte en común.

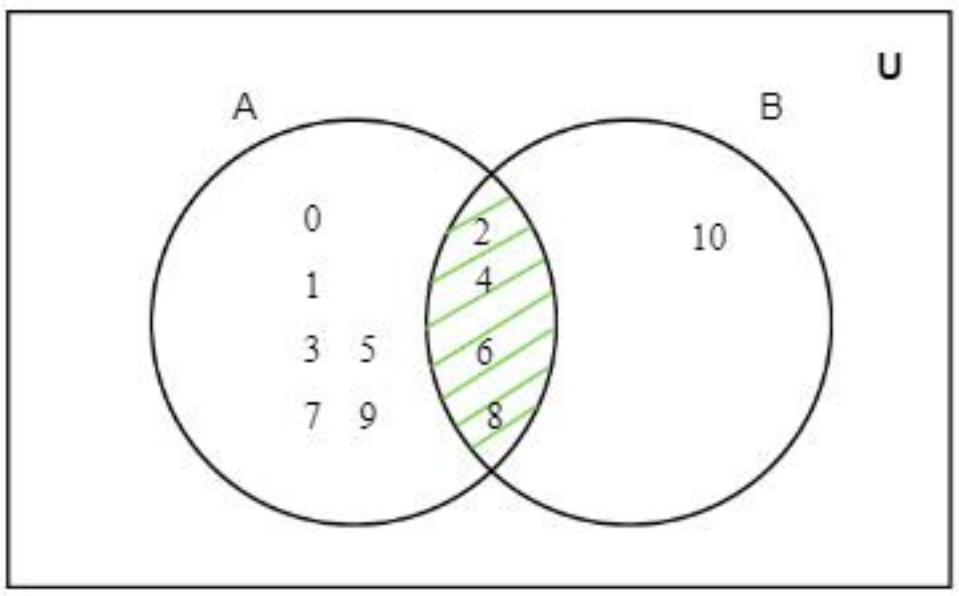
Ejemplo 3:
Sean los siguientes conjuntos $A = \{0,1,2,3,4,5,6,7,8,9\}$ y $B = \{2,4,6,8,10\}$ Determina $A \cap B$

Se identifican cuáles son los elementos que se repiten en ambos conjuntos y se obtiene:

$$A \cap B = \{2,4,6,8\}$$

Nota: *Los elementos 2,4,6,8 son elementos comunes a ambos conjuntos y observen que los elementos están escritos en orden ascendente.*

Su representación gráfica utilizando un Diagrama de Venn-Euler es:



Los elementos que se repiten se escriben en la parte central, donde se tiene el área común entre el conjunto A y el conjunto B.

Ejemplo 4:
Sean los siguientes conjuntos:
C={x|x *es obtener un número menor o igual a tres al lanzar un dado al azar*}
y D={x|x *es obtener un número impar al lanzar un dado al azar*}. Determine $C \cap D$

Los conjuntos son: C ={1,2,3} y D ={1,3,5}

Se identifican cuáles son los elementos que se repiten en ambos conjuntos y se obtiene:

$$A \cap B = \{1,3\}$$

Un conjunto puede ser también el resultado de un experimento probabilístico.



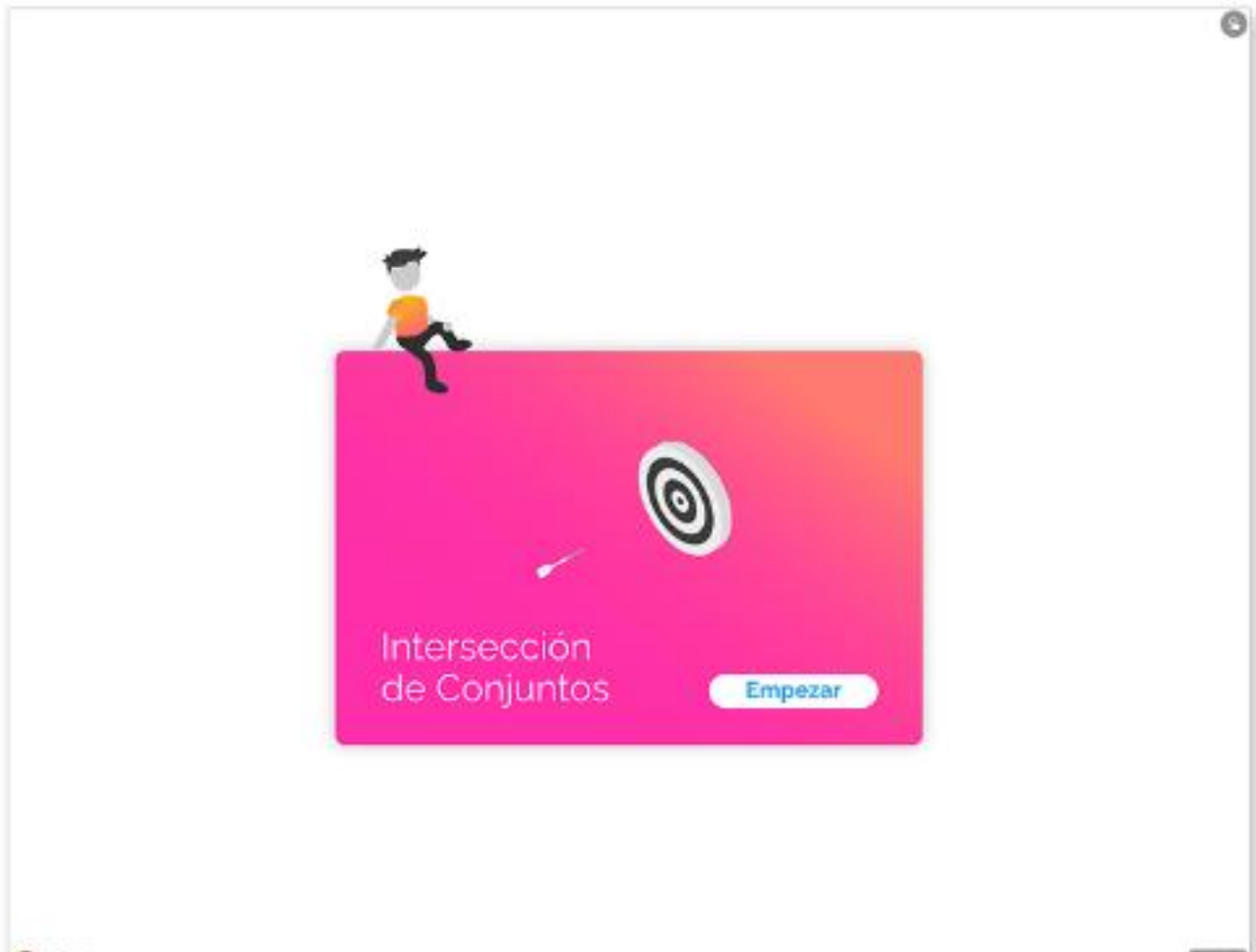
Actividad 7.



Unión e intersección de conjuntos

Instrucciones:
Ejercicio 1. Te invitamos a jugar la ruleta y contestar correctamente las siguientes preguntas sobre unión de conjuntos.

Da clic en la **imagen.**



Actividad 7.



Unión e intersección de conjuntos

Instrucciones:

Ejercicio 2. Te invitamos a jugar tiro de dardos y contestar correctamente las siguientes preguntas sobre intersección de conjuntos.

Da clic en la **imagen.**

1.8 El papel de la probabilidad

Alguna vez te has preguntado **¿qué tan probable es que me gane la lotería al comprar solamente 1 billete?**



"El Sorteo Mayor se celebra todos los martes y participan 60 mil números (billetes) que van del 00001 al 60,000".

Con esta información se tiene 1 probabilidad en 60,000 de ganar la lotería; esto se puede representar con la siguiente expresión matemática:

$$\text{Probabilidad de ganar} = \frac{1}{60000} = 0.001666$$

Multiplicando la cantidad anterior por 100 se obtiene el resultado en forma de porcentaje:

$$(0.00001666) = \mathbf{0.001666}\ \%$$

Por lo tanto, tu probabilidad de ganar la lotería con un solo billete sería de apenas del 0.001666 %.

La información obtenida en un evento probabilístico nos permite tomar decisiones, considerando que, a mayor probabilidad del evento, solemos tomar con mayor facilidad esa opción y viceversa.

En el caso del billete de lotería el desánimo por la baja probabilidad puede ser causa de que ni siquiera nos tomemos la molestia de comprar el billete; no obstante, si la probabilidad fuera mayor, quizá no nos lo pensaríamos dos veces e iríamos directo a su compra.

Observa que si deseas aumentar tus probabilidades de ganar la Lotería Nacional existen dos opciones:

1) Disminuir el número de billetes emitidos, lo cual es poco posible porque eso ya está definido.
2) Aumentar el número de boletos que compras.

De forma similar, intuitivamente empleamos la probabilidad para tomar decisiones en la vida cotidiana, por ejemplo:

- Cuando hemos visto el cielo nublado en un tono grisáceo, con una sensación de humedad en el ambiente, la mayoría de las veces ha llovido; por lo tanto, en un día que observemos esas características, es muy probable que tomemos la decisión de salir bien abrigados y de ser posible con un paraguas.
- Cuando realizan una compra en línea, algunas personas prefieren revisar primero las calificaciones y comentarios sobre el vendedor, si observan que la mayoría son elevadas y positivas respectivamente, existe una mayor probabilidad de que las personas finalicen sus compras.
- En un servicio de telefonía, internet o televisión; cuando nos percatamos que de forma continua tenemos fallas, es muy probable que tomemos la decisión de cambiar de proveedor.

Definición

Probabilidad
Para calcular la probabilidad de un evento probabilístico se puede emplear la siguiente expresión:

$$P(A)=\frac{n}{N}$$

Donde:
n = Total de casos favorables para el suceso A
N = Espacio muestral, total de casos posibles del experimento
P(A) = Probabilidad de que ocurra el suceso A

El resultado de la probabilidad de que ocurra el suceso A se suele escribir en porcentaje en algunos casos, para expresarlo de esa forma solo se debe multiplicar por 100:

$$\%P(A)=100\ P(A)$$

NOTA: Esta fórmula solo es válida cuando los eventos que ocurren en un espacio muestral son EQUIPROBABLES, es decir, todos y cada uno tienen la misma posibilidad de ocurrir, sin mostrar preferencia hacia uno u otro en particular, esto se conoce como la ley de Laplace.

Fuente: Economipedia

1.8.1 Evento y espacio muestral

Retomemos el ejemplo de la Lotería Nacional, recordemos que solo se emiten 60,000 billetes, estos, representan al denominado **"ESPACIO MUESTRAL" (casos posibles), el cual representa la cantidad total de posibilidades que pueden ocurrir en una situación definida (experimento).** A continuación, se comparten algunos ejemplos de experimentos y su espacio muestral:

Tabla 4. Ejemplo de experimentos

Experimento	Espacio muestral	
Jugar a la Lotería Nacional	Billetes emitidos	= 60000
Lanzar una moneda	Águila o sol	= 2
Lanzar un dado	Números del 1 al 6	= 6
Lanzar dos dados	Combinaciones con ambos dados*	= 36
Extraer una canica de una bolsa que tiene 3 canicas verdes y 7 azules	Total de canicas	= 10

* Aquí se ilustra el espacio muestral del lanzamiento de dos dados:

Tabla 7. Espacio muestral que se obtiene al lanzar dos dados al mismo tiempo

Dado 1	Dado 2	Dado 1	Dado 2	Dado 1	Dado 2	Dado 1	Dado 2	Dado 1	Dado 2	Dado 1	Dado 2
1	1	2	1	3	1	4	1	5	1	6	1
1	2	2	2	3	2	4	2	5	2	6	2
1	3	2	3	3	3	4	3	5	3	6	3
1	4	2	4	3	4	4	4	5	4	6	4
1	5	2	5	3	5	4	5	5	5	6	5
1	6	2	6	3	6	4	6	5	6	6	6

Ahora bien, a la cantidad de boletos que compremos se denomina **"EVENTO" o "SUCESO" (casos favorables), el cual representa el total de opciones o selecciones que se consideran del espacio muestral para un caso particular.** A continuación, se muestran algunos ejemplos:

Tabla 8. Ejemplo de casos favorables

Evento	Casos favorables	
Comprar 10 billetes de la Lotería Nacional	Se compraron 10 billetes	= 10
Caer águila al lanzar una moneda	Solo 1 cara es águila	= 1
Lanzar un 3 con un dado	El dado solo tiene un 3	= 1
Obtener un número par al lanzar un dado	El dado tiene 3 números pares	= 3
Extraer una canica azul de una bolsa que tiene 3 canicas verdes y 7 azules	La bolsa tiene 7 canicas azules	= 7

A continuación, se muestras algunos ejemplos de cómo calcular la probabilidad de que ocurra un evento en un experimento:

Ejemplo 1:
Experimento 1: Obtener un número 5 al lanzar un dado.

Espacio muestral	Casos favorables	$P(A)=\frac{n}{N}$
El dado tiene 6 caras $N=6$	Solo 1 cara es cinco $n=1$	$P(A)=\frac{1}{6}=0.166\overline{6}$

Conclusión: La probabilidad de obtener un 5 al lanzar un dado es del 16.67%

Ejemplo 2:
Experimento 2: Obtener un número par al lanzar un dado.

Espacio muestral	Casos favorables	$P(A)=\frac{n}{N}$
El dado tiene 6 caras $N=6$	De las seis caras, 3 son pares $n=3$	$P(A)=\frac{3}{6}=0.5$

Conclusión: La probabilidad de obtener un número par al lanzar un dado es del 50%.

Ejemplo 3:
Experimento 3: Ganar un premio en una rifa al comprar 5 boletos, sabiendo que se han emitido 100 boletos en total.

Espacio muestral	Casos favorables	$P(A)=\frac{n}{N}$
En total se tienen 100 boletos $N=100$	Se han comprado 5 boletos $n=5$	$P(A)=\frac{5}{100}=0.05$

Conclusión: La probabilidad de ganar el premio es del 5%.

Ejemplo 4:
Experimento 4: Extraer 1 canica roja de una bolsa que tiene cuatro canicas rojas y seis negras.

Espacio muestral	Casos favorables	$P(A)=\frac{n}{N}$
La bolsa tiene 10 canicas $N=10$	De las diez canicas, 4 son rojas $n=4$	$P(A)=\frac{4}{10}=0.4$

Conclusión: La probabilidad de extraer una canica roja es del 40%.

Ejemplo 5:
Experimento 5: En una colonia de 800 habitantes, 250 están a favor de la construcción de un monumento. ¿Cuál es la probabilidad de que al preguntar a 1 habitante esté a favor de la construcción del monumento?

Espacio muestral	Casos favorables	$P(A)=\frac{n}{N}$
La colonia tiene 800 habitantes $N=800$	250 habitantes están a favor $n=250$	$P(A)=\frac{250}{800}=0.3125$

Conclusión: La probabilidad de que al preguntar a 1 habitante esté a favor de la construcción del monumento es del 31.25%.

Ejemplo 6:
Experimento 6: Se tienen dos dados. ¿Cuál es la probabilidad de que al lanzarlos al mismo tiempo la suma de puntos sea 8?

El espacio muestral es más complejo, por lo que se representa en la siguiente tabla:

Tabla 9. Ejemplo de espacio muestral

Dado 1	Dado 2	SUMA
1	1	2
1	2	3
1	3	4
1	4	5
1	5	6
1	6	7
2	1	3
2	2	4
2	3	5
2	4	6
2	5	7
2	6	8
3	1	4
3	2	5
3	3	6
3	4	7
3	5	8
3	6	9

Dado 1	Dado 2	SUMA
4	1	5
4	2	6
4	3	7
4	4	8
4	5	9
4	6	10
5	1	6
5	2	7
5	3	8
5	4	9
5	5	10
5	6	11
6	1	7
6	2	8
6	3	9
6	4	10
6	5	11
6	6	12

Espacio muestral	Casos favorables	$P(A)=\frac{n}{N}$
Los dados pueden caer de 36 formas diferentes $N=36$	Solo 5 opciones suman ocho $n=5$	$P(A)=\frac{5}{36}=0.1389$

Conclusión: La probabilidad de que al lanzar dos dados al mismo tiempo la suma de puntos sea 8 es del 13.89%.

Ejemplo 7:
Experimento 7: Se tiene un dado y dos monedas. ¿Cuál es la probabilidad de que al lanzarlos al mismo tiempo se obtenga un número par y dos águilas?

El espacio muestral es más complejo, por lo que se representa en la siguiente tabla:

Tabla 10. Ejemplo de espacio muestral

Dado	Moneda 1	Moneda 2
1		
1		
1		
1		
2		
2		
2		
2		
3		
3		
3		
3		

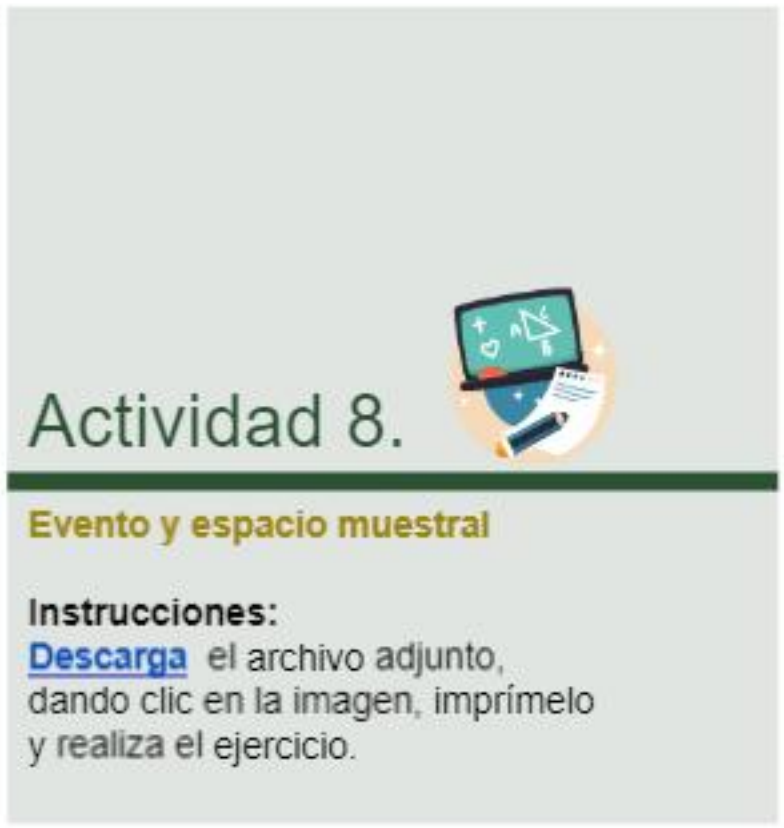
Dado	Moneda 1	Moneda 2
4		
4		
4		
4		
5		
5		
5		
5		
6		
6		
6		
6		

Espacio muestral	Casos favorables	$P(A)=\frac{n}{N}$
Se tienen 24 opciones $N=24$	Solo 3 cumplen lo solicitado $n=3$	$P(A)=\frac{3}{24}=\frac{1}{8}=0.125$

Conclusión: La probabilidad de que al lanzar un dado y dos monedas al mismo tiempo se obtenga un número par y dos águilas es del 12.5%.







1.9 Tabla de distribución de frecuencia

Definición

Una tabla de distribución de frecuencias es una forma organizada de registrar datos estadísticos. La forma de organizar los datos depende de su naturaleza, pueden ser cualitativos (nominales u ordinales) o cuantitativos (discretos o continuos).

Datos cualitativos

Como aprendiste en el tema 1.3. Instrumentos de recolección de información, los datos estadísticos que en adelante solo denominaremos "datos", se pueden obtener mediante encuestas, cuestionarios, formularios, u otros medios, obteniendo información denominada "datos NO agrupados" como la que muestra en los siguientes ejemplos:

Ejemplo 1:
Tabla de frecuencias para datos cualitativos nominales (variables no medibles)

De los siguientes sabores, ¿Cuál es tu favorito en un helado?
- Chocolate (C)
- Fresa (F)
- Limón (L)
- Nuez (N)
- Pistache (P)

Con el sabor preferido por cada persona de una muestra de 40 se obtuvieron los siguientes datos:

C	C	F	F	C	C	F	N
P	N	C	C	L	L	C	C
F	L	P	N	L	F	C	L
C	C	F	C	C	C	F	N
P	P	F	C	N	L	C	C

Cada sabor se denomina categoría o clase, la cantidad de respuestas obtenidas para un mismo sabor será su frecuencia, por lo que debemos contar cuántas veces fue elegido cada sabor, el resultado ordenado en una tabla se denomina "tabla de distribución de frecuencias":

De los siguientes sabores, ¿Cuál es tu favorito en un helado?

Tabla 11. Frecuencias (1)

Categoría	Frecuencia
Chocolate	17
Fresa	8
Limón	6
Nuez	5
Pistache	4
TOTAL	40

NOTA: El sabor de un helado no es medible, se mide el número de personas que eligen ese sabor.

Ejemplo 2:
Tabla de frecuencias para datos cualitativos ordinales (implica una variable no medible, respetando un orden o jerarquía)

¿Qué tan satisfecho te encuentras con el servicio de cable ofrecido por nuestra empresa?

a) Nada satisfecho (0)
b) Poco satisfecho (1)
c) Neutral (2)
d) Satisfecho (3)
e) Muy satisfecho (4)

NOTA: Observa que a las opciones se les ha asignado un número (puede ser ascendente o descendente) que permite identificar un orden o jerarquía entre ellas.

Con la percepción del servicio por cada persona de una muestra de 40 se obtuvieron los siguientes datos:

1	0	2	0	1	0	2	0
0	0	1	2	2	3	2	3
2	3	0	2	0	1	0	0
2	0	2	1	2	2	0	4
0	4	4	3	3	0	4	3

En este caso cada nivel de satisfacción corresponde a la categoría, de igual forma se debe contar cada respuesta obtenida para cada nivel con la finalidad de obtener su frecuencia:

¿Qué tan satisfecho te encuentras con el servicio de cable ofrecido por nuestra empresa?

Tabla 12. Frecuencias (2)

Categoría	Frecuencia
Nada satisfecho	14
Poco satisfecho	5
Neutral	11
Satisfecho	6
Muy satisfecho	4
TOTAL	40

NOTA: El grado de satisfacción no es medible como tal, depende de la percepción de las personas a quienes se les pregunta, por lo tanto, se mide cuántos eligen cada nivel de satisfacción.

Datos cuantitativos

Es común que los datos cuantitativos (que sí son contables) se organicen en tablas de frecuencias con intervalos definidos los cuales se denominan "intervalos de clase". Estos intervalos se emplean para representar a la cantidad de datos que han sido seleccionados entre dos valores denominados límites de clase; el valor mínimo que se denomina "límite inferior" y el valor máximo que se denomina "límite superior".

Intervalo de clase		
1	--	4
5	--	8
9	--	12
13	--	16
...	-	...

Límite inferior del intervalo

Límite superior del intervalo

Cuando se realiza una encuesta, los datos suelen ser recolectados en desorden, esto se conoce como "Datos NO agrupados"

Ejemplo 3:
Diseño de una tabla de distribución de frecuencias para el siguiente caso: un circo registró las edades de las 40 personas (al total de datos se denomina n) que asistieron a una función, obteniendo los siguientes resultados:

Tabla 13. Frecuencias (5)

5	4	26	8	18
8	15	21	35	7
40	35	38	11	11
17	18	15	32	10
16	32	32	29	12
15	42	41	30	8
36	10	12	11	18
9	12	48	8	7

Para organizar los datos de la tabla anterior se puede emplear uno de los siguientes criterios:

A. Definir el número de intervalos y calcular el tamaño del intervalo:

$$tamaño\ del\ intervalo = \frac{Rango + Variación}{Número\ de\ intervalos}$$

B. Definir el tamaño del intervalo y calcular el número de intervalos:

$$número\ de\ intervalos = \frac{Rango + Variación}{Número\ del\ intervalo}$$

En cualquier caso:
Rango = Dato mayor – Dato menor
Variación = Diferencia mínima entre los datos estadísticos

Como caso particular se desea elaborar una tabla de frecuencias con 5 intervalos, entonces:

5	4	26	8
8	15	21	35
40	35	38	11
17	18	15	32
16	32	32	29
15	42	41	30
36	10	12	11
9	12	48	8
12	8	18	7
18	7	11	10

Dato menor = 4
Dato mayor = 48
Rango = 48 – 4 = 44
Variación = 1

$$Tamaño\ del\ intervalo = \frac{44+1}{5} = \frac{45}{5} = 9$$

El primer intervalo iniciará con el dato menor 4 (recuerda que el valor mínimo se denomina límite inferior del intervalo) y como el tamaño del intervalo resultó ser un número entero, el último intervalo terminará con el dato mayor 48.

El límite superior del mismo intervalo se obtendrá así:

$$Límite\ superior_1 = Límite\ inferior_1 + Tamaño\ del\ intervalo - Variación$$
$$Límite\ superior_1 = 4+9-1=12$$

El límite inferior del siguiente intervalo será el límite superior del intervalo anterior más la variación:

$$Límite\ inferior_2 = Límite\ superior_1 + variación$$
$$Límite\ inferior_2 = 12+1=13$$

Tabla 15. Intervalos (1)

Número de intervalo	Intervalo de clase	
	Límite inferior	Límite superior
1	4	12
2	13	21
3	22	30
4	31	39
5	40	48

$12 + 1 = 13$
$21 + 1 = 22$
$30 + 1 = 31$
$39 + 1 = 40$

Después se cuenta cuántos datos corresponden a cada intervalo, el número obtenido se llama FRECUENCIA:

5	4	26	8
8	15	21	35
40	35	38	11
17	18	15	32
16	32	32	29
15	42	41	30
36	10	12	11
9	12	48	8
12	8	18	7
18	7	11	10

Tabla 16. Intervalos (2)

Número de intervalo	Intervalo de clase		Frecuencia
	Límite inferior	Límite superior	
1	4	12	17
2	13	21	9
3	22	30	3
4	31	39	7
5	40	48	4

$n = 40$

A esta tabla de frecuencias se le conoce como tabla de "Datos agrupados".

Ejemplo 4:
La siguiente tabla muestra los kilómetros que deben recorrer 80 estudiantes desde su casa hasta la escuela. Con esa información se diseñó una tabla de distribución de frecuencias con un tamaño de cada intervalo de 3 km.

Tabla 17. Intervalos (3)

13.8	5.5	4.8	7.0	2.7	17.7	13.1	12.1
8.5	11.9	4.5	16.6	7.1	3.2	15.6	14.7
7.9	16.8	9.7	16.9	16.7	13.0	10.1	9.6
6.0	11.1	14.2	17.4	2.3	16.5	16.3	2.0
4.4	12.2	16.1	13.6	8.7	5.3	8.5	11.8
7.3	7.5	14.3	9.1	2.1	11.3	13.7	10.3
3.7	2.1	4.4	2.5	15.0	13.9	17.0	5.0
8.3	3.1	4.6	16.0	18.0	15.1	5.3	2.5
16.4	10.9	5.8	6.5	10.4	12.9	14.7	6.3
2.4	6.1	17.8	9.9	11.3	14.7	12.3	9.1

Dato menor = 2.0
Dato mayor = 18.0
Rango = 18.0 – 2.0 = 16.0
Variación = 0.1

$$\textit{Número de intervalos} = \frac{16 + 0.1}{3} = 5.36$$

Se emplearán 6 intervalos

NOTA: Si el número de intervalos no resulta un valor entero, se debe redondear siempre hacia arriba, un número de intervalos de 5.36 significa que debemos emplear 6 intervalos.

La tabla de frecuencias es de la siguiente forma:

Tabla 18. Intervalos (4)

Número de intervalo	Intervalo de clase		Frecuencia
	Límite inferior	Límite superior	
1	2	4.9	16
2	5	7.9	14
3	8	10.9	13
4	11	13.9	15
5	14	16.9	17
6	17	19.9	5

$n = 80$

$$\textit{Lím sup}_1 = \textit{Lím inf}_1 + \textit{Tamaño del intervalo} - \textit{Variación}$$

$$\textit{Lím sup}_1 = 2+3-0.1=4.9$$

$$\textit{Lím inf}_2 = \textit{Lím sup}_1 + \textit{variación}$$

$$\textit{Límite inferior}_2 = 4.9 + 0.1 = 5$$

NOTA: El dato mayor NO coincide con el límite superior del último intervalo porque se redondeó el número de intervalos.

Ejemplo 5:
Tabla de frecuencias para datos cuantitativos discretos (variables solo pueden tomar valores definidos y separados)

Una empresa registró el número de paquetes enviados cada día de las 12:00 a las 15:00.

Tabla 19. Frecuencias (3)

Intervalo de clase			Frecuencia
1	--	4	9
5	--	8	9
9	--	12	3
13	--	16	12
17	--	20	7
		Total	50

NOTA: El número de paquetes es discreto, esto porque no se puede enviar medio paquete, ni tres cuartos de paquete, se cuentan los paquetes de forma entera, por lo tanto, la separación entre ellos es entera.

Ejemplo 6:
Tabla de frecuencias para datos cuantitativos continuos (variables representan mediciones, por lo que pueden tomar cualquier valor)

Un hospital registró la estatura de 100 personas que asistieron a consulta un día determinado.

Tabla 20. Frecuencias (4)

Intervalo de clase			Frecuencia
1.50	--	1.55	15
1.56	--	1.60	21
1.61	--	1.65	45
1.66	--	1.70	12
1.71	--	1.75	6
1.76	--	1.80	1

NOTA: La estatura es continua, ya que puede tomar cualquier valor (entero o decimal) en un rango definido.





Actividad 9.

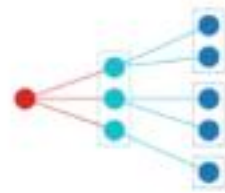


Tabla de distribución de frecuencias

Instrucciones:
Descarga el archivo adjunto, dando clic en la imagen, imprímelo y realiza el ejercicio.

1.10 Problemas de conteo

Cuando el número de elementos del espacio muestral de un experimento es muy grande y su enumeración no es sencilla o se convierte en una imposibilidad práctica, se utilizan algunas herramientas como el **Análisis Combinatorio.**



Principio Fundamental del Conteo o del Análisis Combinatorio

Figura 15. Conteo o Análisis Combinatorio



Ejemplo 1:

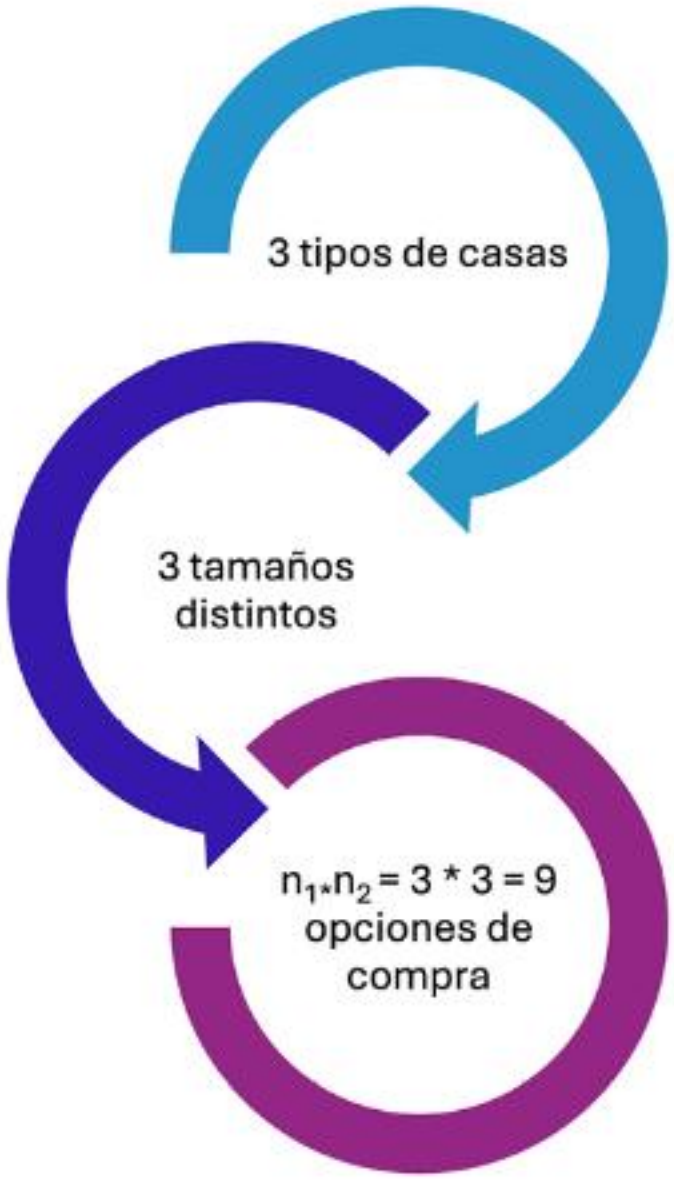
Un constructor ofrece a sus posibles compradores de casas varios estilos:



Figura 16. Ejemplo del principio del conteo (1)

De acuerdo con el Principio Fundamental del Conteo el número de maneras diferentes que un comprador puede adquirir una de estas casas es:

Figura 17. Ejemplo del principio del conteo (2)



Nota. Fuente: Elaboración propia

1.10.1 Factorial, Combinaciones y Permutaciones

El principio fundamental del conteo se puede generalizar para más de dos sucesos, así para k sucesos que pueden ocurrir de $n_1, n_2, n_3, \ldots n_k$ maneras distintas, el número de maneras distintas en que los k sucesos pueden ocurrir está dado por:

Definición

Factorial
Por definición $0! = 1$
Es una función matemática en la que se multiplican todos los números enteros positivos que hay desde el número de la fórmula y el 1.

El factorial de n se denota por $n!$, con n entero positivo

Ejemplo 1: El factorial de 5

$$5! = 5 * 4 * 3 * 2 * 1 * = 120$$

Ejemplo 2: El producto del factorial de 4 y 3

$$4!3! = (4 * 3 * 2 * 1) * (3 * 2 * 1) = 144$$

Ejemplo 3: El cociente del factorial de 6 y 2

$$\frac{6!}{2!} = \frac{6 * 5 * 4 * 3 * 2 * 1}{2 * 1} = \frac{720}{2} = 360$$

Definición

Combinaciones
La combinación de tamaño r de un conjunto de tamaño n son las posibles elecciones de r elementos diferentes que pueden tomarse de un conjunto de n elementos y se le denomina combinación de n en r. (Possani y Barreiro, 2008).

El número de combinaciones de n elementos tomados de r en r se representa por:

nCr C(n,r) ó $C_{n,r}$ = $\frac{n!}{r!(n-r)!}$

Las combinaciones se diferencian entre sí por la naturaleza del elemento ya que en estas no importa el orden en que estén colocados.

Ejemplo 4: El número de combinaciones de los números 1,2,3 tomadas de 2 en 2 es:

$$C(3,2) = \frac{3!}{2!(3-2)!} = \frac{3*2*1}{2*1(1)} = \frac{6}{2} = 3$$

Las combinaciones obtenidas son:

1,2 1,3 2,3

Nótese que 1,2 es la misma combinación que 2,1; porque cuando se selecciona no importa el orden de los elementos.

Definición

Permutaciones
La permutación simple corresponde a un conjunto de n elementos, en los que es posible un acomodo de los elementos en una fila no respetando el orden de los elementos, pero sí el orden de aparición (Possani y Barreiro, 2008).

En este libro hablaremos de tres tipos de permutaciones:

1. Permutaciones sin repetición de n elementos tomados todos a la vez
2. Permutaciones sin repetición de n elementos tomados r de cada vez, donde $r \leq n$
3. Permutaciones de n elementos de los cuales p_1 son de un tipo, p_2 de otro tipo, …, p_k de otro tipo, donde $p_1 + p_2 + \ldots + p_k = n$

Ejemplo 5: ¿Cuántos números de 4 cifras diferentes se pueden formar con los dígitos 1,2,3,4?

$$P4 = 4!$$

$4! = 4 * 3 * 2 * 1 = 24$ Se pueden formar 24 cifras diferentes y son:

1234	1243	1342	1324	1423	1432
2134	2143	2341	2314	2413	2431
3124	3142	3214	3241	3412	3421
4123	4132	4231	4213	4312	4321

1. Para las permutaciones sin repetición de n elementos tomados todos a la vez su fórmula es $P_n = n!$

2. Para las permutaciones sin repetición de n elementos tomados r de cada vez, donde $r \leq n$, su fórmula es:

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

Ejemplo 6: El número de permutaciones de los números 1,2,3 tomados de 2 en 2 es:

$$P(3,2) = \frac{3!}{(3-2)!} = \frac{3!}{1!} = \frac{3*2*1}{1} = 6$$

Y son: 12, 13, 21, 23, 31, 32
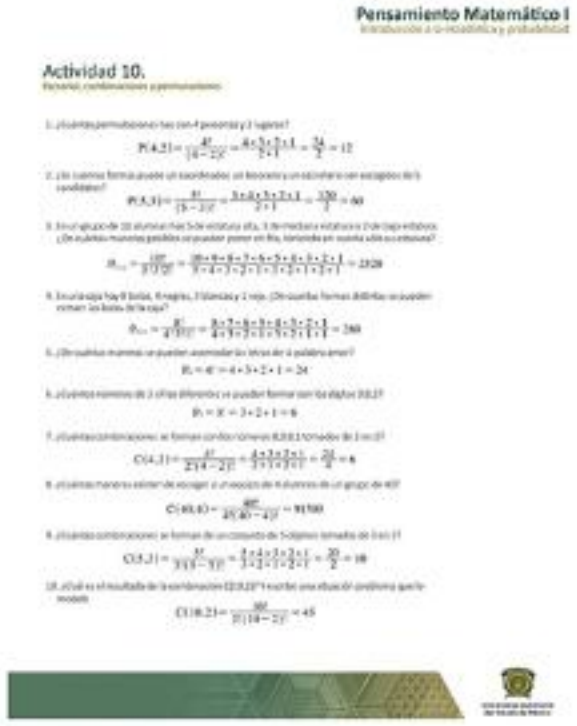
Ejemplo 7: ¿De cuántas maneras se pueden acomodar las letras de la palabra infinito?

Hay tres letras i, hay dos letras n, hay una letra f, hay una letra t y hay una letra o, entonces:

$$P_{P_1,P_2,\ldots P_k} = \frac{8!}{3!2!1!1!1!} = 3360$$

Las letras de la palabra infinito se pueden acomodar de 3360 maneras.





Actividad 10.

Factorial, combinaciones y permutaciones

Instrucciones:

Descarga el archivo adjunto, dando clic en la imagen, imprímelo y realiza el ejercicio.

1.11 Fuentes de información

Según Hernández Sampieri, Fernández Collado y Baptista Lucio (2020, p. 56), "Las fuentes de información son recursos que proporcionan datos, conocimientos o evidencia sobre un tema específico, facilitando la comprensión, análisis y estudio del mismo".

Definición

Las fuentes de información deben contener información verídica y comprobable, pueden ser:

- Directas: Aquellas donde se produce el hecho
- Indirectas: Aquellas donde el hecho se manifiesta indirectamente

De acuerdo con García (2019, p.57) las fuentes de información pueden clasificarse en:

Primarias

• Aquellas que nos brindan información nueva u original y que no ha sido filtrada o modificada.

Secundarias

• Su objetivo no es ofrecer información, sino indicar la fuente que nos proporciona la información y puede hacer algún comentario o análisis.

Terciarias

• Hace referencia a una fuente primaria o secundaria, sin hacer ningún comentario.

Figura 18. Fuentes de información (1)

Ejemplo 1:

Primarias

Libros
Monografías
Periódicos
Leyes
Videos/películas
Entrevistas
Fotografías
Discursos
Diarios personales

Secundarias

Artículos periodísticos
Biografías
Documentales
Enciclopedias
Estadísticas
Bases de datos
Antologías

Terciarias

Bibliografías
Diccionarios
Índices
Catálogos
Listas de lectura
Directorios
Artículos de encuestas

Figura 19. Fuentes de información (2)

En la actualidad existen diversas bases de datos donde podemos encontrar información relevante de algunas variables socioeconómicas, en México unas de las más relevantes son:

Banco mundial
http://databank.worldbank.org/data/databases.aspx
Instituto Nacional de Estadística Geografía e Informática:
https://www.inegi.org.mx/
Data México (Datos de Comercio):
https://www.economia.gob.mx/datamexico/es/profile/geo/mexico



Actividad 11.



Fuentes de información

Instrucciones:
Te invitamos a resolver
el siguiente reto.
Da clic en la **imagen.**

1.12 Teorema de Bayes

Definición

El Teorema de Bayes tiene múltiples aplicaciones en las ciencias, principalmente en la Biología, tecnología, negocios y en la medicina, ya que nos permite tener una certeza de algún suceso dada una información de antemano, como en la medicina que nos permite calcular las probabilidades que intervienen en la detección de diferentes enfermedades.

$$P(A \mid B) = \frac{P(A) * P(B \mid A)}{P(B)}, \text{donde } P(B) \neq 0$$

La fórmula anterior se lee como: la probabilidad de que ocurra el evento A dado que ha ocurrido el evento B, es igual a la probabilidad de que ocurra el evento A multiplicado por la probabilidad de que ocurra el evento B dado que ha ocurrido el evento A, todo esto dividido entre la probabilidad de que ocurra el evento B.

Ejemplo 1:
En diciembre el 30% de los pacientes llegan al hospital con fiebre. Además, el 20% de los pacientes tienen COVID. La probabilidad de que un paciente tenga COVID dado que presenta fiebre es del 50% ¿Cuál es la probabilidad de que una persona tenga fiebre, dado que tiene COVID?

Evento A: Pacientes con fiebre $P(F) = 0.3$
Evento B: Pacientes con COVID $P(C) = 0.2$
Probabilidad de que un paciente tenga COVID dado que presenta fiebre: $P(C|F) = 0.5$
Probabilidad de que un paciente tenga fiebre dado que tiene COVID: $P(F \mid C) = ?$

$$P\left(\text{  } \,\middle|\, \text{}\right) = \frac{P\left(\text{  }\right) * P\left(\text{  } \,\middle|\, \text{  }\right)}{P\left(\text{  }\right)}$$

$$P(F|C) = \frac{P(F) * P(C|F)}{P(C)}$$

$$P(F|C) = \frac{0.3 * 0.5}{0.2}$$

$$P(F|C) = 0.75$$

Respuesta: La probabilidad de que una persona tenga fiebre dado que tiene COVID es del 75%

Ejemplo 1:
En una granja se tienen dos razas de puercos, los negros (pelón mexicano) y los rosas (cerdo blanco). El 40% de los puercos son negros y el 60% son rosas, de los negros el 30% son machos, mientras que de los rosas el 40% son hembras. Si se selecciona una hembra, ¿Cuál es la probabilidad de que sea de color negro?

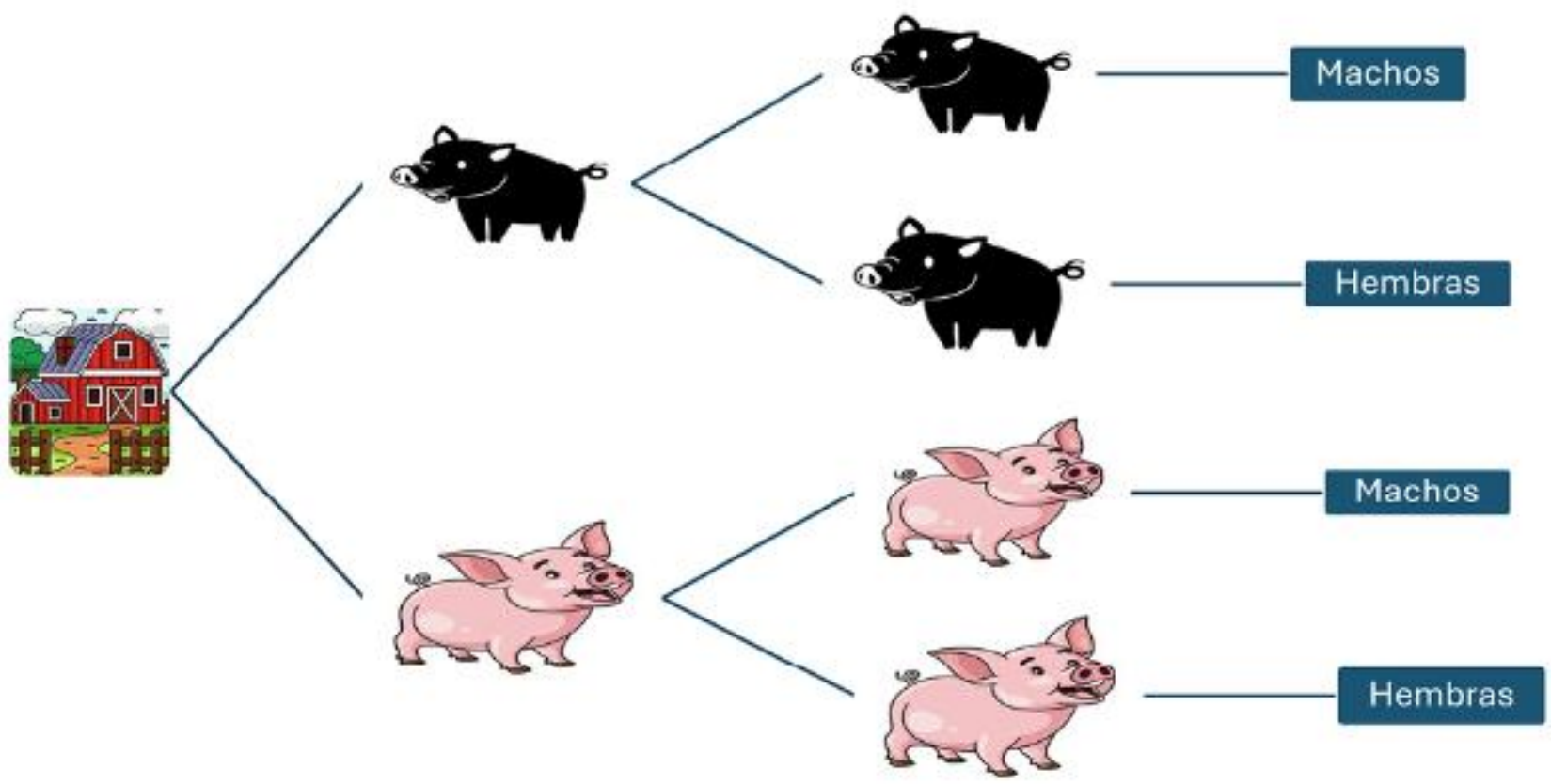


Figura 20. Probabilidad (ejemplo)

$$P(N|H) = \frac{P(N) * P(H|N)}{P(H)}$$

$$P(N|H) = \frac{(0.4)(0.7)}{(0.4)(0.7) + (0.6)(0.4)}$$

$$P(N|H) = \frac{(0.28)}{(0.28) + (0.24)}$$

$$P(N|H) = \frac{(0.28)}{(0.52)}$$

$$P(N|H) = 0.5385$$

Respuesta: La probabilidad de que el puerco sea de color negro si se selecciona una hembra es del 53.85%

1.13 Problemas de aplicación

Ejemplo 1:
Ejercicio 1: En el Plantel "Dr. Pablo González Casanova" de la Escuela Preparatoria y de acuerdo con la agenda estadística de la UAEMéx del año 2022, 628 alumnos egresaron en el ciclo escolar 2021-2022, ¿De qué tamaño debió de ser la muestra para conocer la carrera de su preferencia a nivel superior? Considere un error del 3%

Solución:
Utilizando la fórmula

$$n = \frac{N}{1 + N(e)^2}$$

sustituimos:
n = tamaño de la muestra
N = 628
e = 0.03

$$n = \frac{628}{1 + 628(0.03)^2}$$

$$n = 401 \text{ alumnos.}$$

Para una población de 628 alumnos considerando un error del 3% es suficiente encuestar a una muestra de 401 alumnos.

Ejemplo 2:
En una feria un juego consiste en lanzar dos dados y dos monedas. Para ganar, una persona debe obtener en las dos monedas la misma cara, en un dado un número impar y en el otro dado un múltiplo de tres. ¿Cuál es la probabilidad de ganar en este juego?

Solución:

M1	M2	D1	D2
A	A	1	1
A	A	1	2
A	A	1	3
A	A	1	4
A	A	1	5
A	A	1	6
A	A	2	1
A	A	2	2
A	A	2	3
A	A	2	4
A	A	2	5
A	A	2	6

M1	M2	D1	D2
A	S	1	1
A	S	1	2
A	S	1	3
A	S	1	4
A	S	1	5
A	S	1	6
A	S	2	1
A	S	2	2
A	S	2	3
A	S	2	4
A	S	2	5
A	S	2	6

M1	M2	D1	D2
S	A	1	1
S	A	1	2
S	A	1	3
S	A	1	4
S	A	1	5
S	A	1	6
S	A	2	1
S	A	2	2
S	A	2	3
S	A	2	4
S	A	2	5
S	A	2	6

M1	M2	D1	D2
S	S	1	1
S	S	1	2
S	S	1	3
S	S	1	4
S	S	1	5
S	S	1	6
S	S	2	1
S	S	2	2
S	S	2	3
S	S	2	4
S	S	2	5
S	S	2	6

A	A	3	1
A	A	3	2
A	A	3	3
A	A	3	4
A	A	3	5
A	A	3	6
A	A	4	1
A	A	4	2
A	A	4	3
A	A	4	4
A	A	4	5
A	A	4	6
A	A	5	1
A	A	5	2
A	A	5	3
A	A	5	4
A	A	5	5
A	A	5	6
A	A	6	1
A	A	6	2
A	A	6	3
A	A	6	4
A	A	6	5
A	A	6	6

A	S	3	1
A	S	3	2
A	S	3	3
A	S	3	4
A	S	3	5
A	S	3	6
A	S	4	1
A	S	4	2
A	S	4	3
A	S	4	4
A	S	4	5
A	S	4	6
A	S	5	1
A	S	5	2
A	S	5	3
A	S	5	4
A	S	5	5
A	S	5	6
A	S	6	1
A	S	6	2
A	S	6	3
A	S	6	4
A	S	6	5
A	S	6	6

S	A	3	1
S	A	3	2
S	A	3	3
S	A	3	4
S	A	3	5
S	A	3	6
S	A	4	1
S	A	4	2
S	A	4	3
S	A	4	4
S	A	4	5
S	A	4	6
S	A	5	1
S	A	5	2
S	A	5	3
S	A	5	4
S	A	5	5
S	A	5	6
S	A	6	1
S	A	6	2
S	A	6	3
S	A	6	4
S	A	6	5
S	A	6	6

S	S	3	1
S	S	3	2
S	S	3	3
S	S	3	4
S	S	3	5
S	S	3	6
S	S	4	1
S	S	4	2
S	S	4	3
S	S	4	4
S	S	4	5
S	S	4	6
S	S	5	1
S	S	5	2
S	S	5	3
S	S	5	4
S	S	5	5
S	S	5	6
S	S	6	1
S	S	6	2
S	S	6	3
S	S	6	4
S	S	6	5
S	S	6	6

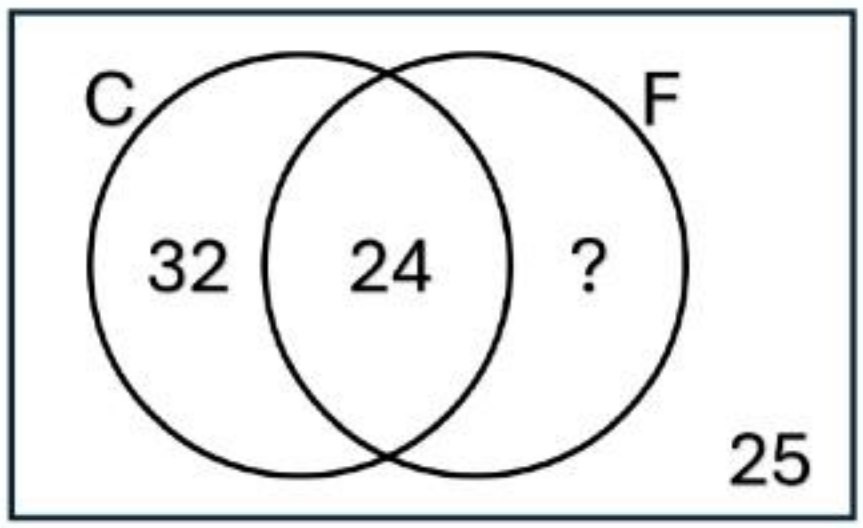
Nomenclatura: **M1** = moneda 1, **M2** = moneda 2, **D1** = dado 1, **D2** = dado 2, **A** = cara águila, **S** = cara sol.

En la tabla se observa que el espacio muestral con dos monedas y dos dados es de 144, mientras que únicamente 22 de esas opciones cumplen los requisitos, de donde la probabilidad de ganar es de:

[redacted]

La probabilidad de ganar en el juego de feria es del $15.2\overline{7}\%$

Ejemplo 3:
Una heladería registró las últimas 100 ventas de las personas que compraron helado de fresa, helado de chocolate, ambos o sabores diferentes; conforme al siguiente diagrama, ¿Cuál es la probabilidad de que una persona elija únicamente helado de fresa?



Solución:

El espacio muestral corresponde al total de ventas registradas que es $n = 100$.

Después se debe averiguar cuántas personas eligieron únicamente helado de fresa, lo que corresponde al evento de interés N, esto se puede resolver porque se conoce el total de personas y se restan aquellas que eligieron algún sabor conforme al diagrama:

Total de personas	=	Eligieron solo chocolate	+	Eligieron ambos	+	Eligieron solo fresa	+	Eligieron otro sabor
100	=	32	+	24	+	?	+	25

De donde:

Eligieron solo fresa	=	Total de personas	-	Eligieron solo chocolate	-	Eligieron ambos	-	Eligieron otro sabor
?	=	100	-	32	-	24	-	25

Eligieron solo fresa $= N =$ 19

De donde la probabilidad solicitada es:

$$P(A) = \frac{n}{N}$$

$$P(A) = \frac{19}{100}$$

$$P(A) = 0.19$$

$$P(A) = 19\ \%$$

La probabilidad de que una persona elija únicamente helado de fresa es del 19 %.

Ejemplo 4:
Un cine registró las edades de 80 asistentes a una función obteniendo los siguientes datos:

37	9	9	12	37	31	15	18
6	10	29	35	9	12	25	23
40	21	20	20	14	9	35	35
21	9	22	22	24	28	34	36
22	15	31	13	20	11	34	14
12	25	38	35	16	6	11	13
30	28	34	28	21	11	31	26
14	34	31	36	20	22	21	26
22	22	31	7	10	25	6	20
20	39	27	31	25	30	13	31

Con esta información determina:

a. ¿Cuál es el dato mayor?
b. ¿Cuál es el dato menor?
c. ¿Cuál es el rango?
d. ¿Cuál es la variación?
e. Considerando 7 intervalos ¿Cuál es el tamaño de cada intervalo?
f. Realiza una tabla de distribución de frecuencias con 7 intervalos
g. ¿Cuál es la probabilidad de que asista al cine una persona entre los 16 y 20 años?
h. ¿Cuál es el intervalo de edad con mayor probabilidad de asistencia? ¿Cuál es esa probabilidad?
i. ¿Cuál es el intervalo de edad con menor probabilidad de asistencia? ¿Cuál es esa probabilidad?
j. ¿Cuál intervalo de edad tiene una probabilidad de asistencia del 13.75 %?

Solución:

37	9	9	12	37	31	15	18
6	10	29	35	9	12	25	23
40	21	20	20	14	9	35	35
21	9	22	22	24	28	34	36
22	15	31	13	20	11	34	14
12	25	38	35	16	6	11	13
30	28	34	28	21	11	31	26
14	34	31	36	20	22	21	26
22	22	31	7	10	25	6	20
20	39	27	31	25	30	13	31

a. ¿Cuál es el dato mayor?
40

b. ¿Cuál es el dato menor?
6

c. ¿Cuál es el rango?
= 40 – 6 = 34

d. ¿Cuál es la variación?
Los datos difieren en una unidad = 1

e. Considerando 7 intervalos ¿Cuál es el tamaño de cada intervalo?

$$Tamaño\ del\ intervalo = \frac{Rango + Variación}{Número\ de\ intervalos}$$

$$Tamaño\ del\ intervalo = \frac{34 + 1}{7}$$

$$Tamaño\ del\ intervalo = \frac{35}{7}$$

$$Tamaño\ del\ intervalo = 5$$

f. Realiza una tabla de distribución de frecuencias con 7 intervalos

Primero definimos los límites de cada intervalo:

Límite inferior del primer intervalo = $Límite\ inferior_1$ = Dato menor = 6

$$Límite\ superior_1 = Límite\ inferior_1 + Tamaño\ del\ intervalo - Variación$$

$$Límite\ superior_1 = 6 + 5 - 1 = 10$$

El límite inferior del siguiente intervalo será el límite superior del intervalo anterior más la variación:

$$Límite\ inferior_2 = Límite\ superior_1 + variación$$

$$Límite\ inferior_2 = 10 + 1 = 11$$

Continuando de forma sucesiva hasta el intervalo 7, los límites de los intervalos quedan de la siguiente forma:

Intervalo	Lim inferior	Lim superior
1	6	10
2	11	15
3	16	20
4	21	25
5	26	30
6	31	35
7	36	40

Observemos que como el tamaño del intervalo es exacto, el límite superior del último intervalo coincide con el dato mayor.

Por último, de la tabla de datos contamos cuántos de ellos pertenecen a cada intervalo:

37	9	9	12	37	31	15	18
6	10	29	35	9	12	25	23
40	21	20	20	14	9	35	35
21	9	22	22	24	28	34	36
22	15	31	13	20	11	34	14
12	25	38	35	16	6	11	13
30	28	34	28	21	11	31	26
14	34	31	36	20	22	21	26
22	22	31	7	10	25	6	20
20	39	27	31	25	30	13	31

Intervalo	Lim inferior	Lim superior	Frecuencia
1	6	10	11
2	11	15	14
3	16	20	8
4	21	25	16
5	26	30	9
6	31	35	15
7	36	40	7

Para contestar las siguientes preguntas, se calculará el porcentaje de probabilidad de cada intervalo recordando que el espacio muestral es el total de asistentes, mientras que el número de eventos favorables corresponde a la frecuencia de cada intervalo:

Intervalo	Lim inferior	Lim superior	Frecuencia	Probabilidad	%
1	6	10	11	$= \frac{11}{80} = 0.1375$	13.75 %
2	11	15	14	$= \frac{14}{80} = 0.1750$	17.50 %
3	16	20	8	$= \frac{8}{80} = 0.1000$	10 %

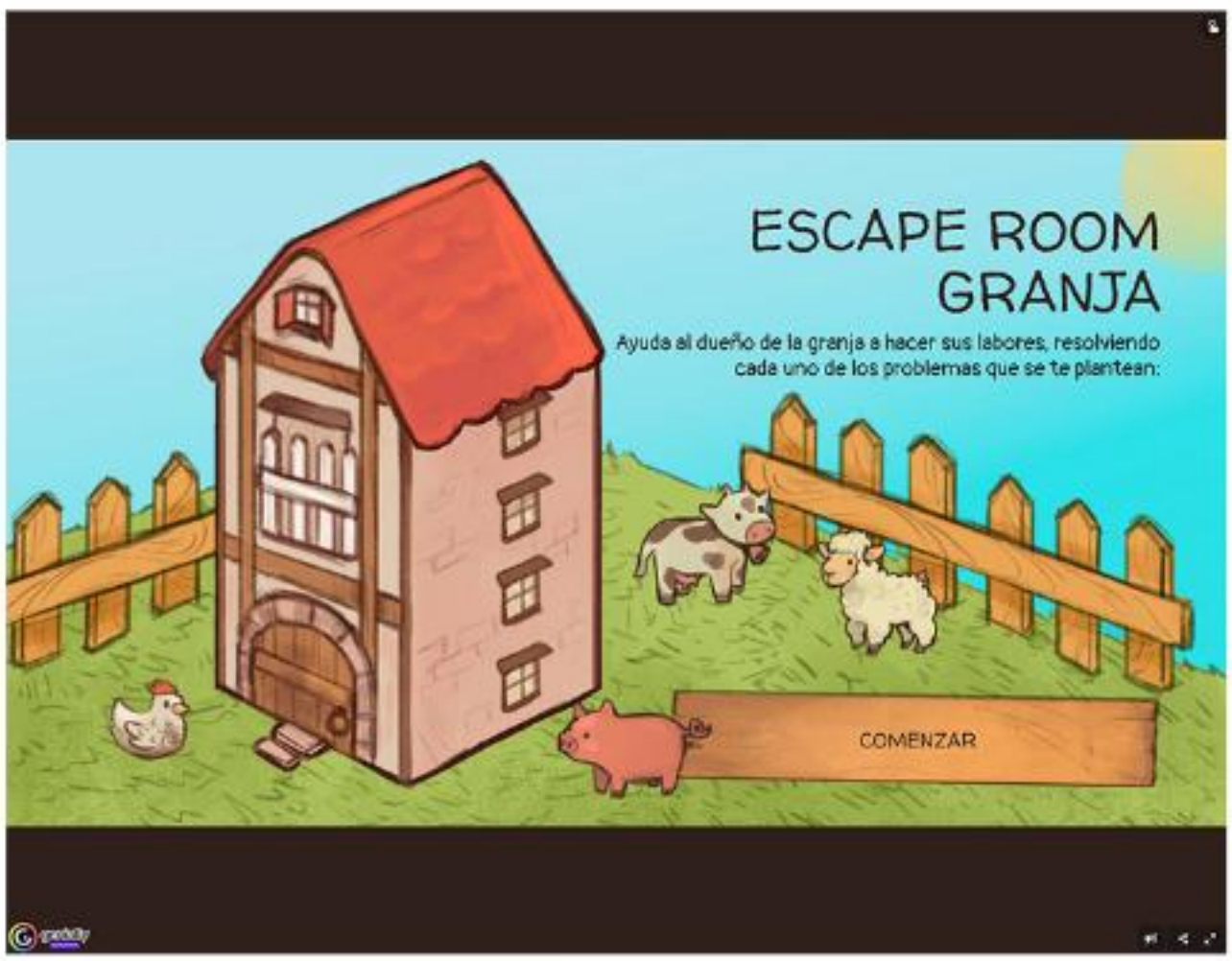
4	21	25	16	$= \frac{16}{80} = 0.2000$	20 %
5	26	30	9	$= \frac{9}{80} = 0.1125$	11.25 %
6	31	35	15	$= \frac{15}{80} = 0.1875$	18.75 %
7	36	40	7	$= \frac{7}{80} = 0.0875$	8.75 %
		TOTAL =	80	TOTAL =	100 %

g. ¿Cuál es la probabilidad de que asista al cine una persona entre los 16 y 20 años?
La probabilidad es del 10 %

h. ¿Cuál es el intervalo de edad con mayor probabilidad de asistencia? ¿Cuál es esa probabilidad?
El que tiene la mayor frecuencia, el intervalo 4 con 16 asistentes de los 21 a los 25 años.
La probabilidad es del 20 %

i. ¿Cuál es el intervalo de edad con menor probabilidad de asistencia? ¿Cuál es esa probabilidad?
El que tiene la menor frecuencia, el intervalo 7 con 7 asistentes de los 36 a los 40 años.
La probabilidad es del 8.75 %

j. ¿Cuál intervalo de edad tiene una probabilidad de asistencia del 13.75 %?
El intervalo 1 con 11 asistentes de los 6 a los 10 años



Actividad 12.

Ejercicios de Aplicación

Instrucciones:
Te invitamos a resolver el siguiente reto.
Da clic en la **imagen.**

Resumen del Módulo I

Los temas vistos en este módulo te han permitido **conocer el papel de la probabilidad** en diferentes situaciones o fenómenos al analizar e interpretar datos para obtener información relevante que te permita una acertada y oportuna toma de decisiones; el mundo de las matemáticas comienza en el conocimiento de los **números y los símbolos** ahora has aprendido diferentes operaciones que te permiten comprender mejor la estadística y su importancia en la ciencia y la investigación.

Se **fortaleció** tu pensamiento estadístico y probabilístico al conocer la recolección de datos y su organización, analizar los instrumentos de recolección de información y conceptualizar los elementos que intervienen en estos estudios, como: variables, muestra, población, entre otros. Al igual, te fuiste familiarizado con el lenguaje y la simbología de esta disciplina.

En la vida cotidiana está presente la toma de decisiones de acuerdo con los diferentes sucesos que vives; en este módulo has aprendido sobre la teoría de conjuntos y como los sucesos que satisfacen ciertas condiciones se relacionan con el estudio de la probabilidad.

También aprendiste a realizar **tablas de distribución** de frecuencias para organizar datos, así como diferentes técnicas de conteo como estrategias matemáticas de la probabilidad, sin olvidar el aprendizaje adquirido sobre el cálculo de probabilidades condicionales con el famoso Teorema de Bayes.

El **aprendizaje es continuo** y se verá fortalecido con los siguientes temas sobre la representación gráfica de los datos y su variabilidad.



Actividad 13.



Ejercicio de gamificación

Instrucciones:
Hemos llegado al final del módulo I, donde has aprendido a analizar e interpretar datos estadísticos y probabilísticos, así como la representación de los conjuntos y subconjuntos de números, mide tu aprendizaje, te invitamos a resolver el siguiente reto:

Da clic en la **imagen.**



Actividad 14.

Cuestionario Autoevaluable

Instrucciones:
Como actividad final, se te pide realizar un cuestionario; para comenzar, revisa en la plataforma **SEDUCA** la sección Actividades (Estudiante) y responde a la actividad denominada: M1 Cuestionario Final

Glosario

Combinaciones. La combinación de tamaño r de un conjunto de tamaño n son las posibles elecciones de r elementos diferentes que pueden tomarse de un conjunto de n elementos y se le denomina combinación de n en r. (Possani y Barreiro, 2008).

Conjunto. Es una colección o familia de objetos con alguna característica en común.

Cuestionarios. Se caracterizan por estar orientados, pueden ser de tipo test o pueden ser preguntas de texto con una puntación.

Datos discretos. Son aquellos que se pueden contar o medir.

Datos continuos. Son aquellos que dentro de un intervalo pueden tomar cualquier valor.

Dato estadístico. Son los valores que se obtienen al llevar a cabo un estudio de tipo estadístico. Serán el resultado de la observación del fenómeno que queremos analizar.

Delphi. Técnica que consiste en indagar con una serie de expertos un tema determinado para orientar la toma de decisiones.

Documentos o Registros. Técnica que consiste en examinar los datos presentes en documentos ya existentes como bases de datos, actas, informes, registros de asistencias, etc.

Encuesta. Técnica en la cual se plantea un listado de preguntas para obtener datos precisos.

Entrevista. Técnica que consiste en una conversación generalmente planificada, donde el investigador plantea una serie de preguntas previamente definidas sobre algún tema a una o varias personas, con el fin de obtener información específica.

Espacio muestral. Representa la cantidad total de posibilidades que pueden ocurrir en una situación definida.

Estadística. Rama de las matemáticas que permite mediante diferentes métodos aritméticos y sobre un conjunto de datos recogerlos, organízalos, resumirlos y arrojar un análisis e interpretación de estos.

Estadística descriptiva. Consiste en la presentación de datos en forma de tablas y gráficas.

Estadística inferencial. Determina y deduce las características de una población a partir del análisis de una muestra de ésta.

Evento o Suceso. (Casos favorables), Representa el total de opciones o selecciones que se consideran del espacio muestral para un caso particular.

Formularios. Se caracterizan por crear preguntas que no tienen ningún tipo de puntuación o corrección.

Escalas de opinión, actitudes e intereses. Son instrumentos utilizados para la medición de aspectos relacionados a la intensidad de un sentimiento o la actitud de un grupo de personas respecto a un tema en particular.

Factorial. Es una función matemática en la que se multiplican todos los números enteros positivos que hay entre el número de la fórmula y el 1

Fuente de información primaria. Aquellas que nos brindan información nueva u original y que no ha sido filtrada o modificada.

Fuente de información secundaria. Su objetivo no es ofrecer información, sino indicar la fuente que nos proporciona la información y puede hacer algún comentario o análisis.

Fuente de información terciaria. Hace referencia a una fuente primaria o secundaria, sin hacer ningún comentario.

Grupos focales. Técnica que consiste en reunir un grupo de personas con características relacionadas con la investigación y realizar una entrevista grupal orientada hacia la información que se desea obtener.

Instrumento de recolección de datos. Como la herramienta con la que se cuenta para documentar o registrar la información recabada de la realidad la cual pude ser cualitativa o cuantitativa.

Jerarquía de operaciones. Realizar operaciones siguiendo este orden: Se efectúan las operaciones dentro de los signos de agrupación, se calculan las potencias y las raíces, posteriormente se efectúan los productos y los cocientes y al final se realizan las sumas y restas.

Métodos o técnicas de colección de datos o información. Son el medio, los procedimientos o

actividades orientados al desarrollo de instrumentos de evaluación, medición a través de los cuales se establece la relación entre el investigador y el consultado.

Muestra. Subconjunto de la población la cual será analizada y que permitirá obtener información confiable de la población ahorrando tiempo y dinero.

Números enteros. Incluyen a los números naturales, al cero y a los negativos que no tienen parte decimal, son $\mathbb{Z} = \{\ldots,-3,-2,-1,0,1,2,3\ldots\}$.

Números irracionales. Son los números que tienen parte decimal infinita y no periódica, son $\mathbb{I} = \{\sqrt{2} = 1.4142\ldots,\sqrt{3} = 1.7320508\ldots,\pi = 3.1415\ldots,\ldots\}$.

Números naturales. Se utilizan para contar y comienzan desde el 1, son $\mathbb{N} = \{1,2,3,4,\ldots\}$.

Números racionales. Son los números que tienen a las fracciones o los números con parte decimal finita o infinita periódica, son $\mathbb{Q} = \left\{x \mid x = \frac{a}{b} donde\ a \wedge b\ !\ \mathbb{Z}\ con\ b \neq 0\right\}$.

Números reales. Son los números que resultan de la unión del conjunto de los números racionales con el conjunto de los números irracionales. Esto es: $\mathbb{R} = \mathbb{Q} \cup \mathbb{I}$

Observación participante. Técnica que consiste justamente en observar el desarrollo de un fenómeno social, ambiental, experimental que se desea analizar.

Permutaciones. La permutación simple corresponde a un conjunto de n elementos, en los que es posible un acomodo de los elementos en una fila no respetando el orden de los elementos, pero sí el orden de aparición (Possani y Barreiro, 2008).

Población estadística. Es un conjunto de elementos, cosas o individuos que comparten características similares con las que se pretende hacer un estudio o análisis estadístico.

Pruebas psicológicas. Son instrumentos estandarizados utilizados para medir una o más características de las personas como personalidad, inteligencia, atención de manera rápida, etc.

Respuesta abierta. Permiten al encuestado tener la libertad de responder libremente cada pregunta, lo que permite obtener respuestas más profundas y también obtener respuestas en las que no había pensado.

Respuesta cerrada. Los participantes elegirán alguna opción.

Símbolos matemáticos. Son representaciones en forma de caracteres, indican que se debe hacer una consideración de lógica o de aritmética.

Tabla de distribución de frecuencias. Es una forma organizada de registrar datos estadísticos. La forma de organizar los datos depende de su naturaleza, pueden ser cualitativos (nominales u ordinales) o cuantitativos (discretos o continuos).

Variable cualitativa ordinal. Se expresan por normativa, en palabras y manifiesta diferentes niveles y orden.

Variable cualitativa nominal. Expresa un nombre claramente diferenciado.

Variable cuantitativa continua. Es una variable indicada numéricamente que toma un valor infinito de valores entre un intervalo de datos (puede ser número decimal).

Variable cuantitativa discreta. Es una variable indicada numéricamente que toma un valor finito de valores enteros entre un intervalo de datos.

Variable estadística. Característica medible u observable que se puede registrar en un conjunto de datos.

Referencia Bibliográfica

- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2020). Metodología de la Investigación. McGraw-Hill Interamericana.
- García M.A.C. (2019). Fuentes de información. Boletín Científico de las Ciencias Económico Administrativas del ICEA, 8 (15). https://doi.org/10.29057/icea.v8i15.4864
- Possani, E. y Barreiro, L. (2008). Estadística y probabilidad. México: Santillana;. (15). https://doi.org/10.29057/icea.v8i15.4864



Módulo II

Variabilidad y representación gráfica

Propósito de Módulo

Analiza las relaciones existentes entre dos variables y su representación gráfica para identificarlas y relacionarlas con su entorno.

Introducción

Los datos estadísticos son la base para tomar decisiones informadas. Imagina que eres el gerente de una tienda y necesitas decidir cuánto inventario comprar para la próxima temporada. Aquí es donde los datos sobre las ventas anteriores, las tendencias del mercado y las preferencias de los clientes entran en juego. Sin datos, estaríamos a ciegas. Científicos, economistas y sociólogos validan sus hipótesis con datos estadísticos. Por ejemplo, si alguien afirma que el cambio climático está afectando la biodiversidad, necesitamos datos que respalden o refuten esta afirmación.

Los datos también nos permiten construir modelos matemáticos que predicen eventos futuros. Desde pronósticos del tiempo hasta predicciones económicas, los datos son la base de estos modelos. En campos como la medicina y las finanzas, los datos estadísticos nos ayudan a evaluar riesgos. ¿Cuál es la probabilidad de que un medicamento cause efectos secundarios?, ¿cuál es el riesgo de invertir en una determinada acción?

En este módulo **aprenderás** que los **datos estadísticos** se clasifican como variables cualitativas y cuantitativas, con diversos ejemplos se te guiará para que puedas identificar cuándo un fenómeno o experimento involucra a cada tipo de variables.

Es importante resaltar que los datos estadísticos son la base en la recopilación de información para la toma de decisiones, así que para su fácil comprensión y análisis los solemos representar en gráficas.

Si bien es cierto que se podría emplear cualquier tipo de gráfica para cualquier tipo de datos, algunas resultan más cómodas que otras para ciertos casos. En este módulo aprenderás sobre los diferentes tipos de gráficas y los usos que les puedes dar. Cada gráfica posee características adecuadas para los datos, por ejemplo, algunas son más convenientes para datos cualitativos y otras para datos cuantitativos; también hay ejemplos propios de los datos bivariados.

Es importante resaltar que no basta con crear gráficas; también debemos interpretarlas correctamente. ¿Qué nos dicen los datos?, ¿hay alguna tendencia?, ¿existe alguna relación? La interpretación precisa es clave para tomar decisiones basadas en evidencia. En el módulo se te brindarán herramientas como coeficiente de correlación y la recta de regresión, con ellas, podrás identificar que tan relacionados están los datos y si son útiles o no para la toma de decisiones futuras.

Al igual que el módulo anterior, este módulo te retará a poner en práctica tus conocimientos y habilidades mediante actividades interactivas. Recuerda realizar la actividad auto evaluable y mide tu progreso.

Actividad 1.



Evaluación Diagnóstica
(conocimientos previos)

Antes de comenzar con la revisión de los contenidos del Módulo 2 de Pensamiento Matemático I, te invitamos a responder la evaluación diagnóstica.

Eso te ayudará a descubrir los conocimientos previos que tienes sobre los temas a abordar y, al mismo tiempo, le permitirá a tu docente conocer tus fortalezas y debilidades.

Para comenzar con el ejercicio, revisa en la plataforma **SEDUCA** la sección Actividades (Estudiante) y responde a la actividad denominada: M2 Examen Diagnóstico.

2.1 Variables cuantitativas y cualitativas

En el primer módulo nos familiarizamos con diferentes términos como dato, variable, población, muestra, entre ellos identificamos las variables estadísticas: cualitativas y cuantitativas; revisemos estas últimas más a fondo.

Te recordamos que una variable es una característica medible u observable que se puede registrar en un conjunto de datos.

Figura 1. Variables



Las variables estadísticas se dividen en cualitativas y cuantitativas

2.1.1 Variables cualitativas

Es un tipo de variable estadística que describe cualidades, circunstancias o características de un objeto o persona que permiten expresar una característica, atributo, cualidad o categoría no numérica.

a. Características de las variables cualitativas

- Imposible medirlas numéricamente.
- No proporciona datos específicos y muchas veces tampoco un orden.
- Especifica una cualidad, característica o condición.
- Cuando la variable puede tomar sólo dos posibles valores, se llama dicotómica.
- Cuando la variable puede tomar tres o más valores, se conocerá como politómica.

b. Tipos de variables cualitativas

Figura 2. Variables



La variable cualitativa puede ser nominal, ordinaria o binaria.

Nominal
Variable que se representa sin orden ni números, y matemáticamente es menos precisa.

Por ejemplo, los colores: verde, blanco, rojo, etc.

Ordinaria
Conocida también como variable cuasi cuantitativa, es representada por una modalidad que no requiere números, indican atributos o cualidades, se expresan con palabras, y las categorías se ordenan jerárquicamente.

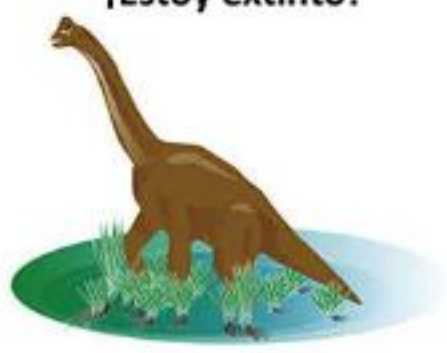
Por ejemplo, la calidad de atención recibida por un cliente en una zapatería dependiendo de la opinión de los clientes. Los posibles valores son mala, regular, buena y excelente.

Binaria
Trabaja con valores específicos de tipo binario.
Por ejemplo, el foco puede estar encendido o apagado.

Ejemplo 1. Variables cualitativas

a. Estado civil: soltero, casado.
b. Profesión: docente, enfermero, secretaria, azafata, etc.
c. El tamaño de una bolsa: chica, mediana, grande, extragrande o jumbo.

d. Criterio al haber recibido un servicio público: pésimo, malo, regular, bueno o excelente.
e. Color de ojos: cafés, verdes, azules.

Variable Nominal Descripciones Desordenadas	**Variable Ordinal** Descripciones Ordenadas	**Variable Binaria** Solo 2 resultados mutuamente excluyentes
 **Soy una tortuga**	 **Estoy feliz**	**¡Estoy extinto!** 
 **Soy un caracol**	 **Estoy bien**	**Soy hembra** 
 **Soy una mariposa**	 **Estoy increíble**	**¡Encendido!** 

2.1.2 Variables cuantitativas

Definición

Son aquellas que pueden expresarse mediante números, es decir solo puede tomar un valor numérico.

Por ejemplo:

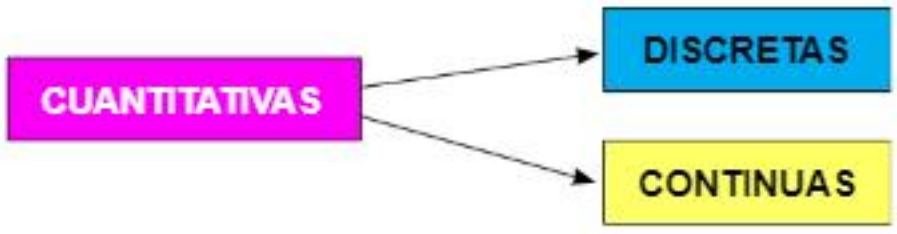
- La cantidad de peces dentro de una pecera es una variable cuantitativa porque solo puede ser un número (12 peces, 23 peces, 2 peces, etc.)
- El número de alumnos en un aula de clases (30 alumnos, 38 alumnos, 42 alumnos, etc.)
- Cantidad de docentes en un plantel educativo (5 docentes, 59 docentes, 68 docentes, etc.)
- Cantidad de tela para elaborar una prenda (2.5m, 3m, 1.8m, etc)

a. Características de las variables cuantitativas:

- Expresan sus valores con números.
- Son utilizadas generalmente en encuestas o entrevistas.
- Utilizan gráficos llamados diagramas integrales y diagramas diferenciales para mostrar la frecuencia relativa de las variables.
- También pueden servirse de diagramas de barra para otorgar cifras.

b. Tipos de variables cuantitativas

Figura 3. Tipos de variables cuantitativas



Las variables cuantitativas pueden ser discretas o continuas

Discreta
Son aquellas que pueden adquirir un valor de un conjunto numérico dado, es decir no existen valores intermedios entre ellos, dicho de otra manera, es un valor entero.

Por ejemplo:
- Una persona puede tener 0, 1, 2, 3 o más hermanos, no dos hermanos y medio.
- La cantidad de perros que hay en un refugio canino puede ser de 32 perros, pero no 32.3 perros.
- La cantidad de enfermeras de un hospital puede ser de 15 pero no 14.7 enfermeras.

Continua
Esta variable puede asignar un valor cualquiera dentro de un intervalo o medición, es decir, siempre habrá opción de que haya más valores entre dos valores conocidos. Normalmente se representan por valores decimales para ser más específico.

- Por ejemplo, la estatura de nacimiento de un bebé puede ser de 51 centímetros.
- El contenido de agua en un vaso puede ser de 132.4 mililitros.
- El peso de una tortilla de maíz puede ser de 28 gramos.

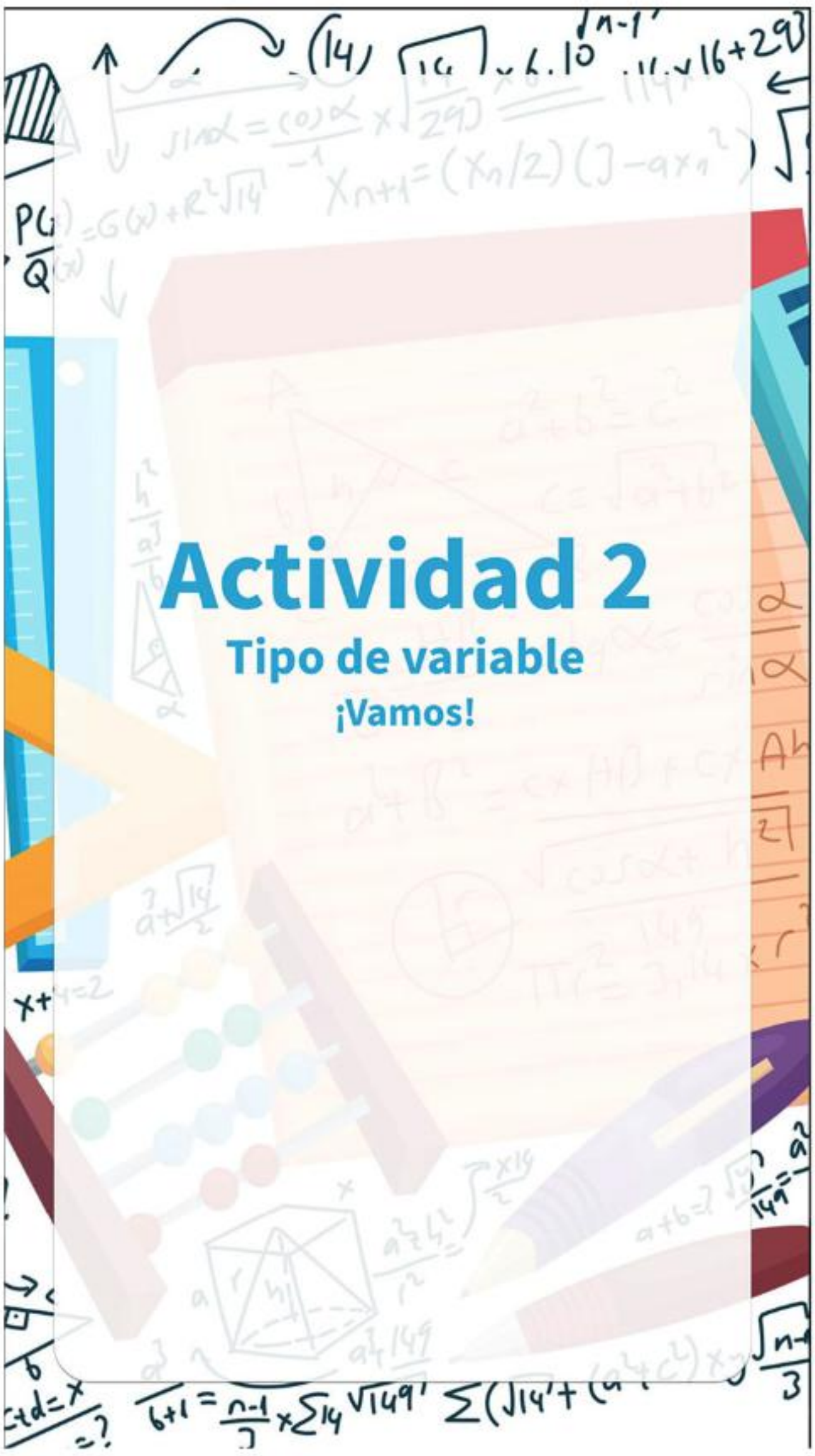
Ejemplo 2. Variables cuantitativas

- Cantidad de personas que se suben a un juego mecánico: 10, 15, 24, etc.
- Cantidad de integrantes de una familia: 2, 3, 4, 5, etc.
- Velocidad de un automóvil de Toluca a México: 45.4 kl/h, 66.8 km/h, 100 km/h, etc.
- Precio de una fritura: $12, $20, $35, etc.
- El peso de una bolsa de frituras: 35g, 29g, 40.5gr, etc.
- Factor de protección solar de una crema: 10, 20, 30, 50, etc.
- Cantidad de asistentes a un partido de futbol: 1520, 3480, 47906, etc.
- Cantidad de clientes atendidos en un negocio: 4, 7, 28, 76, etc.

Figuras 4. Variable cuantitativa discreta y variable cuantitativa



En resumen, las variables cualitativas y cuantitativas las encontramos en todas las ramas del conocimiento y en múltiples situaciones de nuestra vida diaria para poder utilizarlas como mejor lo necesitemos.



Actividad 2.



Variables cualitativas y cuantitativas

Instrucciones:
Te invitamos a resolver el siguiente ejercicio sobre variables cualitativas y cuantitativas, para iniciar da clic en la **imagen.**

2.2 Representación gráfica de datos

La representación gráfica nos permite visualizar los datos cuantitativos o cualitativos de forma más expresiva, con solo ver las representaciones gráficas podremos entender la naturaleza de los datos, sus características y tomar decisiones.

Las representaciones gráficas para datos cuantitativos más usuales son:

- Gráfica de barras
- Histograma
- Polígono de frecuencias
- Ojiva (mayor que y menor que)
- Circulograma

Figura 5. Gráfica de Barras sobre las estaturas de los estudiantes



2.2.1 Barras

Definición

a. Gráfica de barras

Estos diagramas se conforman por barras rectangulares, cuya longitud es proporcional a los valores o cantidades representadas, son muy útiles si se requiere presentar a una variable en dos situaciones distintas.

Su representación en el plano cartesiano se ubica de la siguiente manera:

Eje x: intervalos o categorías de datos cuantitativos
Eje y: frecuencia (Fi)

Ejemplo 1. Gráfica de barras para datos cuantitativos

En un grupo de 50 estudiantes se recopilaron sus estaturas, siendo la mayor de 178 cm y la menor de 145 cm. Obteniendo los siguientes resultados:

Tabla 2. Resultados de estaturas obtenidas

Núm.	Intervalo (Altura)	Frecuencia Fi
1	145-151	5
2	152-158	9
3	159-165	20
4	166-172	13
5	173-179	3

Definición

Representación para datos cualitativos

Gráfica de barras

Eje x: categoría o características
Eje y: frecuencia (Fi)

Ejemplo 2. Gráfica de barras para datos cualitativos

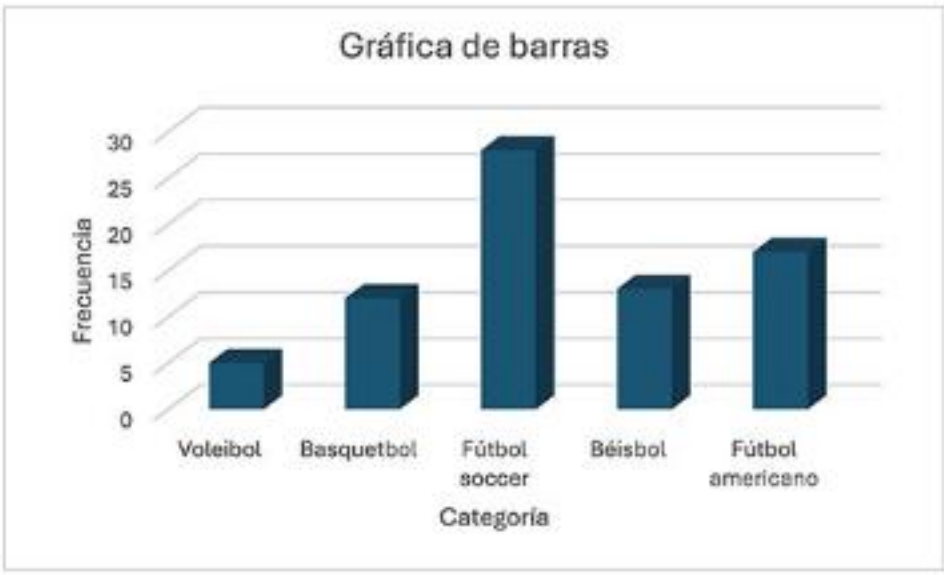
Un periodista realiza un reportaje sobre el deporte favorito de las personas y de una muestra representativa de 75 personas se obtuvieron los siguientes resultados:

Tabla 3. Resultados del deporte favorito elegido

Categoría	Frecuencia
Voleibol	5
Basquetbol	12
Fútbol soccer	28
Béisbol	13
Fútbol americano	17
Total	75

Con la información obtenida se elabora la gráfica de barras con datos cualitativos.

Figura 6. Gráfica de barras sobre el deporte favorito de las personas



2.2.2 Histogramas

Definición

a. Histograma

Es una representación gráfica en forma de barras, que a diferencia con la gráfica de barras en el histograma las barras están unidas, ya que indican continuidad.

Eje x: Límites reales de clase*
Eje y: Frecuencia (Fi)

* En el módulo I aprendiste como los datos estadísticos cuantitativos se organizan en tablas de frecuencias con intervalos denominados "intervalos de clase", recordarás que estos intervalos se emplean para representar a la cantidad de datos que han sido seleccionados entre dos valores denominados límites de clase; el valor mínimo que se denomina "límite inferior" y el valor máximo que se denomina "límite superior".

Ahora bien, para obtener los límites reales de clase utilizaremos el concepto de variación también visto en el módulo I que es la diferencia mínima entre los datos estadísticos, de manera que:

- El **límite real inferior** se calcula restando la mitad de la variación al límite inferior del intervalo.
- El **límite real superior** se calcula sumando la mitad de la variación al límite superior del intervalo.

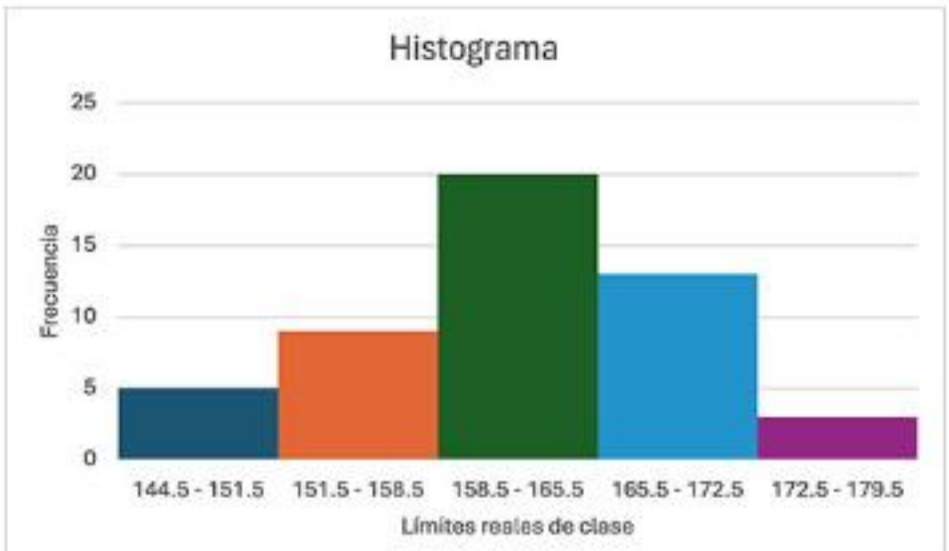
Ejemplo 3. Histograma

Con la información anterior, del ejemplo 1 considerando que la variación es uno obtenemos los siguientes resultados:

Tabla 4. Resultados de las estaturas de los alumnos, incluye límites reales de clase

Núm.	Intervalo (Altura)	Frecuencia *Fi*	Lím. real de clase
1	145-151	5	144.5 - 151.5
2	152-158	9	151.5 - 158.5
3	159-165	20	158.5 - 165.5
4	166-172	13	165.5 - 172.5
5	173-179	3	172.5 - 179.5

Figura 7. Histograma sobre las estaturas de los alumnos



2.2.3 Polígonos de frecuencias

Definición

a. Polígono de frecuencias

Es una representación gráfica de líneas que se construye trazando una línea que una los puntos medios de los rectángulos o barras de un histograma. Se traza uniendo las marcas de clase, para realizarla es necesario anexar un intervalo anterior y otro al final con frecuencia cero.

Eje x: marca de clase*
Eje y: frecuencia (Fi)

*__Marca de clase:__ Son los puntos medios de los intervalos de clase y se calculan con la siguiente fórmula:

$$M_c = \frac{límite\ sup erior + límite\ inf erior}{2}$$

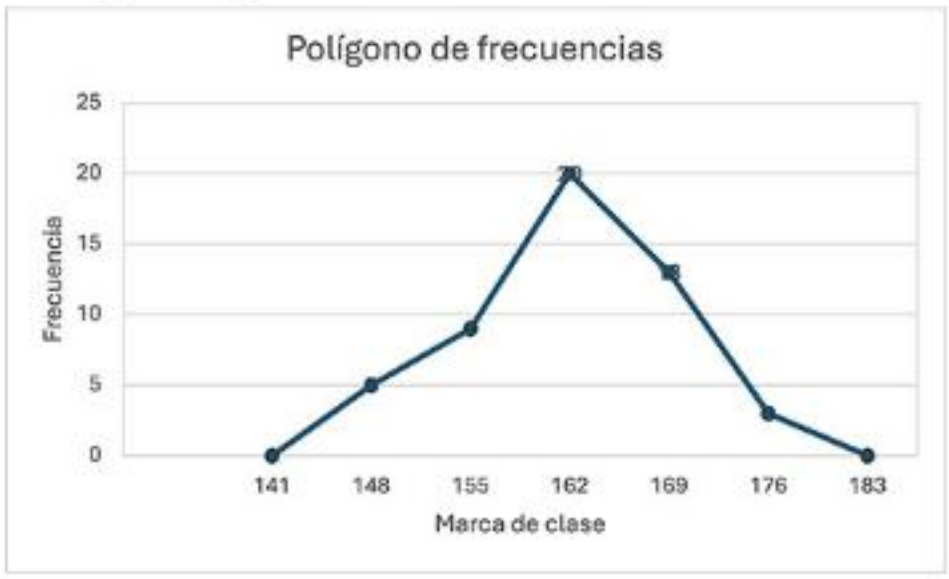
Ejemplo 4. Polígono de frecuencia

Con la información anterior, del ejemplo 1 obtenemos los siguientes resultados:

Tabla 5. Resultados de las estaturas de los alumnos, incluye límites reales de clase y marcas de clase

Núm.	Intervalo (Altura)	Frecuencia *Fi*	Lím. real de clase	Marca de clase
		0		141
1	145-151	5	144.5 - 151.5	148
2	152-158	9	151.5 - 158.5	155
3	159-165	20	158.5 - 165.5	162
4	166-172	13	165.5 - 172.5	169
5	173-179	3	172.5 - 179.5	176
		0		183

Figura 8. Polígono de frecuencias sobre las estaturas de los alumnos



2.2.4 Ojiva

Definición

Es la representación gráfica de líneas que muestra relaciones entre factores, puede representarse de dos formas por **la ojiva mayor que** o **la ojiva menor que;** revisemos un poco de cada una de ellas.

a. Ojiva mayor que

Eje x: Límite real inferior
Eje y: Frecuencia acumulada

Para obtener la frecuencia acumulada se resta al número total de datos, la frecuencia de cada intervalo y se anexa un intervalo al final con frecuencia cero.

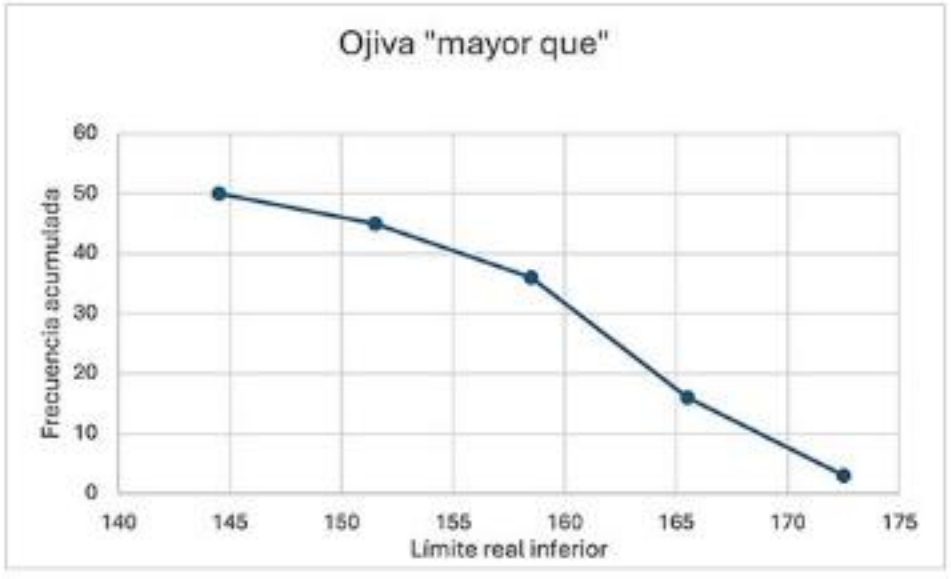
Ejemplo 5. Ojiva mayor que

Del ejemplo 1 tenemos que:

Tabla 6. Resultados de las estaturas de los alumnos, incluye límite real inferior y frecuencia acumulada

Núm.	Intervalo (Altura)	Frecuencia *Fi*	Lím. real inferior	Frecuencia acumulada para ojiva mayor que
1	145-151	5	144.5	50
2	152-158	9	151.5	45
3	159-165	20	158.5	36
4	166-172	13	165.5	16
5	173-179	3	172.5	3
	180-186	0		0

Figura 9. Ojiva "mayor que" sobre las estaturas de los alumnos



b. Ojiva "menor que"

Eje x: Límite real superior
Eje y: Frecuencia acumulada*

***Frecuencia Acumulada**
"La frecuencia acumulada de un intervalo es la suma de la frecuencia absoluta para ese intervalo con la frecuencia del intervalo anterior." (Vargas, 2021, p.10)

Ejemplo 6. Ojiva menor que

Del ejemplo 1 tenemos que:

Tabla 7. Resultados de las estaturas de los alumnos, incluye frecuencia, límite real superior y frecuencia acumulada

Núm.	Intervalo (Altura)	Frecuencia *Fi*	Límite real superior	Frecuencia acumulada
				0
1	145-151	5	151.5	5
2	152-158	9	158.5	14
3	159-165	20	165.5	34
4	166-172	13	172.5	47
5	173-179	3	179.5	50

Figura 10. Ojiva "menor que" sobre las estaturas de los alumnos



2.2.5 Circulograma

Definición

Es la representación gráfica en forma circular, también conocida como gráfica de pastel, que muestra el valor de la frecuencia como un sector circular dentro de un círculo completo. Son útiles para hacer comparaciones entre datos, sobre todo si se expresa en forma de porcentaje.

Para realizarlo se debe calcular el ángulo del círculo y el registro del porcentaje de la frecuencia.

Su cálculo se da en grados sexagesimales con la siguiente fórmula:

$$ángulo = \frac{360^{\circ}(\,frecuencia\ absoluta\,)}{total\ de\ datos}$$

$$porcentaje = \frac{Frecuencia\ absoluta(\,100\,)}{total\ de\ datos}$$

Eje x: Ángulo

Eje y: Porcentaje

Ejemplo 7. Circulograma para datos cuantitativos

Del ejemplo 1 tenemos que:

Tabla 8. Resultados de las estaturas de los alumnos, incluye frecuencia, ángulo y porcentaje

Núm.	Intervalo (Altura)	Frecuencia *Fi*	Ángulo	Porcentaje
1	145-151	5	36°	10%
2	152-158	9	64.8°	18%
3	159-165	20	144°	40%
4	166-172	13	93.6°	26%
5	173-179	3	21.6°	6%
	Suma	50	360°	100%

Para obtener los ángulos se realizó lo siguiente:

$$ángulo\ del\ primer\ intervalo = \frac{360^{\circ}(5)}{50} = 36^{\circ}$$

$$ángulo\ del\ segundo\ intervalo = \frac{360^{\circ}(9)}{50} = 64.8^{\circ}$$

$$ángulo\ del\ tercer\ intervalo = \frac{360^{\circ}(20)}{50} = 144^{\circ}$$

$$ángulo\ del\ cuarto\ intervalo = \frac{360^{\circ}(13)}{50} = 93.6^{\circ}$$

$$ángulo\ del\ quinto\ intervalo = \frac{360^{\circ}(3)}{50} = 21.6^{\circ}$$

Para obtener los porcentajes se realizó lo siguiente:

$$porcentaje\ del\ primer\ intervalo = \frac{5(100)}{50} = 10\%$$

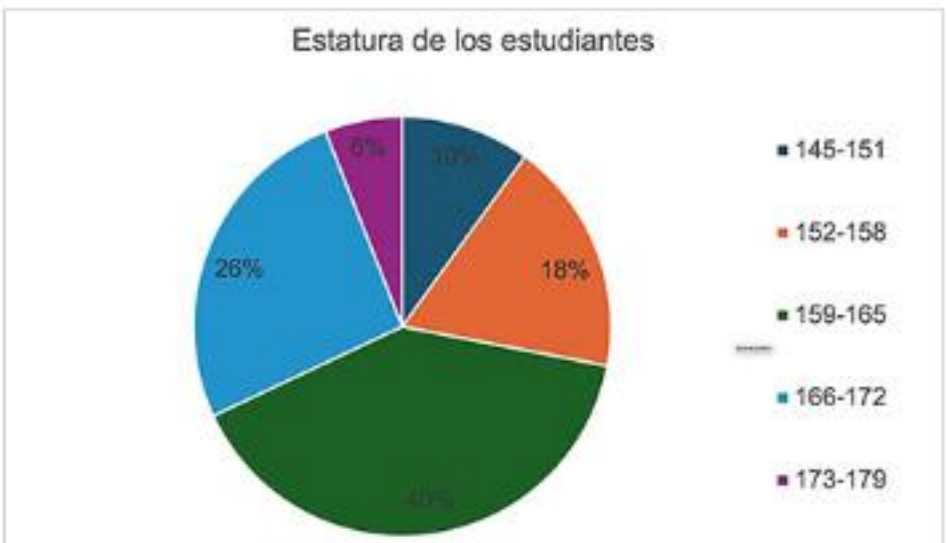
$$porcentaje\ del\ primer\ intervalo = \frac{9(100)}{50} = 18\%$$

$$porcentaje\ del\ primer\ intervalo = \frac{20(100)}{50} = 40\%$$

$$porcentaje\ del\ primer\ intervalo = \frac{13(100)}{50} = 26\%$$

$$porcentaje\ del\ primer\ intervalo = \frac{3(100)}{50} = 6\%$$

Figura 11. Circulograma sobre las estaturas de los alumnos

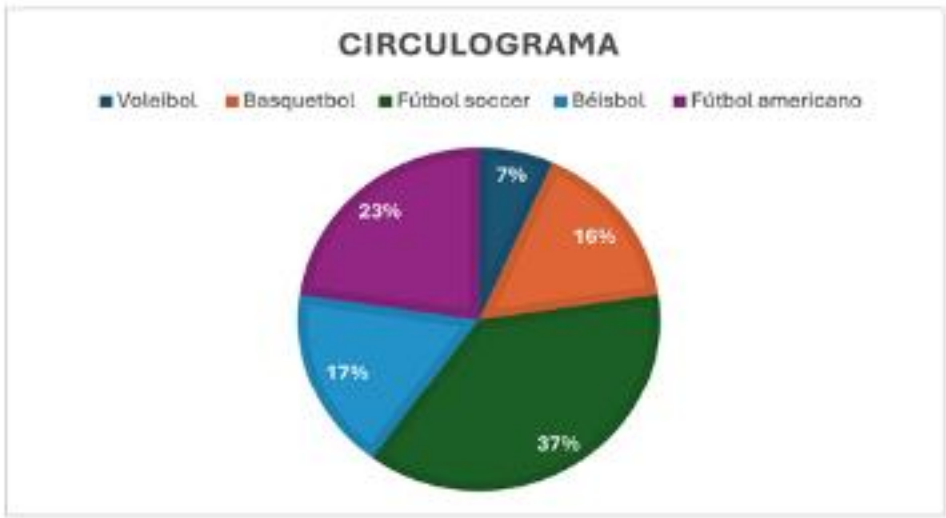


Ejemplo 8. Circulograma para datos cualitativos

Tabla 9. Resultados sobre el deporte favorito, incluye frecuencia, ángulo y porcentaje

Categoría	Frecuencia	Ángulo	Porcentaje
Voleibol	5	24°	7%
Basquetbol	12	57.6°	16%
Fútbol soccer	28	134.4°	37%
Béisbol	13	62.4°	17%
Fútbol americano	17	81.6°	23%
Total	75	360°	100%

Figura 12. Circulograma sobre el deporte favorito





Actividad 3.



Representaciones gráficas (cuantitativos)

Instrucciones:
Te invitamos a resolver el siguiente ejercicio sobre las representaciones gráficas que puedes utilizar para datos cuantitativos, para iniciar da clic en la **imagen.**

2.3 Datos bivariados

"Dos variables se relacionan entre sí cuando el cambio de una categoría a otra (para variables cualitativas) o de un valor a otro (para variables cuantitativas), provoca una modificación en las categorías o valores de la otra variable" (Sagaró, 2019).

En otras palabras, los datos bivariados serán aquellos en cuyo estudio se analizan simultáneamente 2 variables, una que será independiente y la otra, dependiente de esa última. La dependencia obedece a la naturaleza de los datos estudiados y del interés del experimentador, por ejemplo, cuando se desea saber el comportamiento de las ventas de una tienda, se puede elegir como variable independiente al número de clientes, ya que su asistencia a las compras no puede ser controlada ni determinada por la tienda, mientras que el número de ventas sí, apoyándose de descuentos, promociones, publicidad u otras alternativas. En los casos de fenómenos temporales, a veces también llamados series de tiempo, la variable independiente es el lapso en el que se desarrolla el fenómeno.

De lo anterior, en ocasiones en un mismo experimento se da la oportunidad de analizar el comportamiento de dos variables, Porras (s/f) nos define esta relación de la siguiente manera:

Definición

"La descripción bivariada trata de estudiar la relación entre pares de atributos medidos simultáneamente en la misma muestra y comprende un conjunto de herramientas que se enfoca en el análisis de dos variables con el fin de determinar las relaciones empíricas entre ellas."

El objetivo del análisis de los datos bivariados es tratar de predecir el comportamiento de las variables involucradas, con la finalidad de tomar decisiones respecto a una o ambas variables.

Ejemplo 1. Venta de autos

Una empresa que se dedica a la venta de autos ha realizado una tabla que muestra las ventas en seis años consecutivos de su vehículo estrella, obteniendo los siguientes datos:

Año	Número de autos vendidos
2013	52,372
2014	84,263
2015	94,386
2016	97,681
2017	105,360
2018	132,157

Tabla 10. Registro de venta de autos

En la tabla 10 podemos observar que el número de autos vendidos por la empresa aumenta conforme pasan los años, por tal motivo, se podría tomar la decisión de contratar más personal, mejorar los procesos de producción o incluso abrir una nueva fábrica del auto en cuestión.

Ejemplo 2. Venta de tortas

Una tortería registró el promedio de ventas a lo largo del día durante un mes, obteniendo los siguientes datos:

Hora	Número de tortas vendidas
10:00	58
11:00	97
12:00	136
13:00	254
14:00	421
15:00	316
16:00	285
17:00	137

Tabla 11. Registro de venta de tortas

De la tabla 11 se tiene que el número de tortas vendidas aumenta desde las 10:00 hasta las 14:00, después disminuye el número de tortas vendidas. Con esta información se podrían tomar las siguientes decisiones:

1. Definir el horario de entrada y salida del personal, manteniendo poco personal en las primeras y últimas horas y, en mayor número en horas pico.
2. Definir horarios y volumen para la cocción de ingredientes.
3. Definir horarios y volumen para la extracción de ingredientes refrigerados.

Ejemplo 3. Cantidad de tierra que deben procesar para obtener plata de un yacimiento

Una empresa minera ha medido, a lo largo de una semana, la cantidad de tierra que deben procesar para obtener plata de un yacimiento, obteniendo los siguientes datos:

Toneladas de tierra procesada	Kilogramos de plata obtenidos
50	234
36	542
45	531
29	284
48	487
57	332
42	197

Tabla 12. Registro de cantidad de tierra que deben procesar para obtener plata de un yacimiento

De la tabla 12 podemos concluir que en esta ocasión los datos no muestran una relación que permita realizar una toma de decisiones. Podrían o no, ocurrir las siguientes situaciones que generan una menor o mayor extracción de plata:

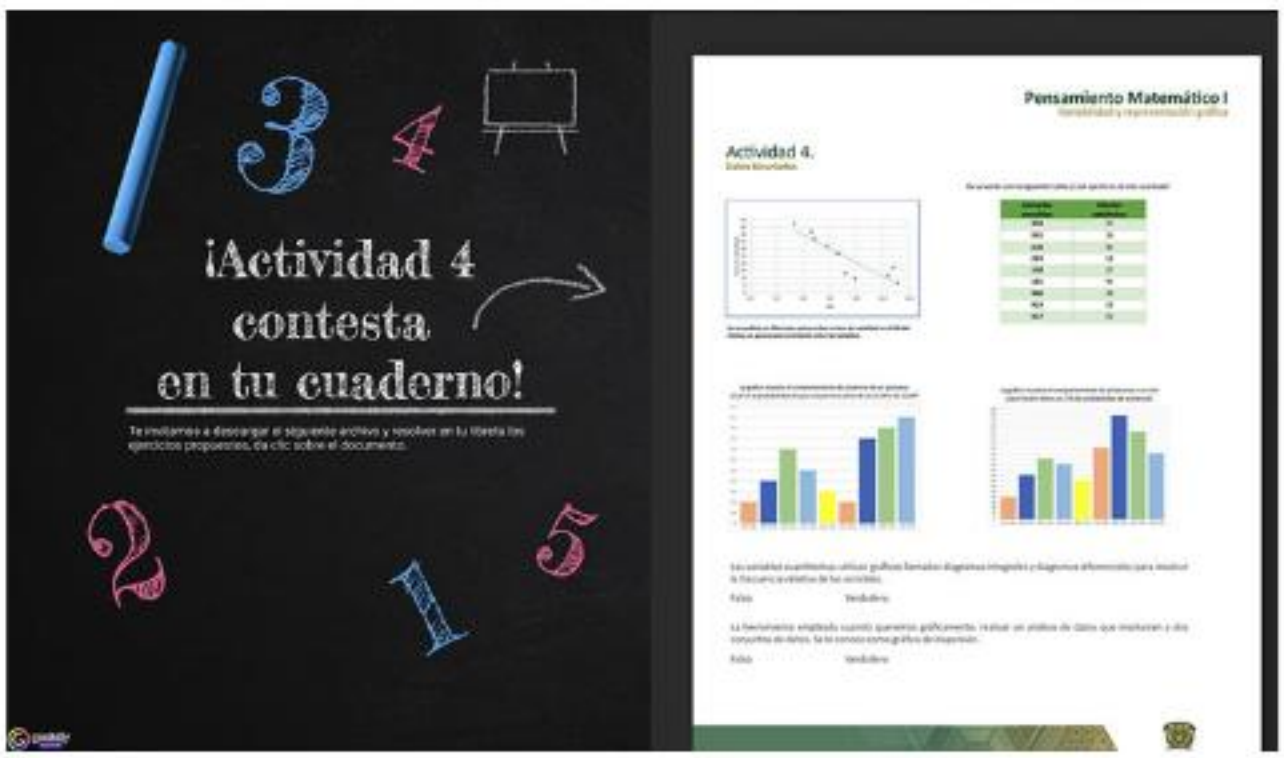
1. La extracción en las zonas de excavación se ha dificultado por la presencia de roca dura.
2. Las condiciones climáticas lluviosas generan lodo y zonas fangosas, en consecuencia, la maniobrabilidad del suelo se ve interrumpida.
3. Existen zonas con betas ricas en el metal de interés, por lo que ese día la extracción fue mayor.
4. Problemas y dificultes técnicas con el equipo de extracción.

Como podemos observar en este último ejemplo, no siempre se puede predecir el comportamiento de las variables, ya que en el contexto en el que ocurre el experimento se pueden presentar condiciones que no se habían contemplado en el mismo.

Ante esta situación, en algunos casos se pueden modificar los parámetros del experimento y repetirlo una vez más, con la finalidad de observar si la relación entre las variables involucradas aumenta o incluso disminuye.

Del mismo modo, existen fenómenos en los que no podemos controlar a las variables, y simplemente presentan un comportamiento aleatorio y poco predecible; por ejemplo, fenómenos meteorológicos, movimientos sociales, comportamiento de la bolsa de valores, entre otros.

Además, las conclusiones obtenidas en los ejemplos citados obedecen simplemente a la intuición, al análisis cualitativo del comportamiento de las variables involucradas. Un análisis más profundo y cuantitativo de la relación entre datos bivariados se abordará en los temas siguientes, esto permitirá una toma de decisiones mejor fundamentada.



Actividad 4.

Datos bivariados

Instrucciones:

Descarga el archivo adjunto, dando clic en la imagen, imprímelo y realiza el ejercicio.

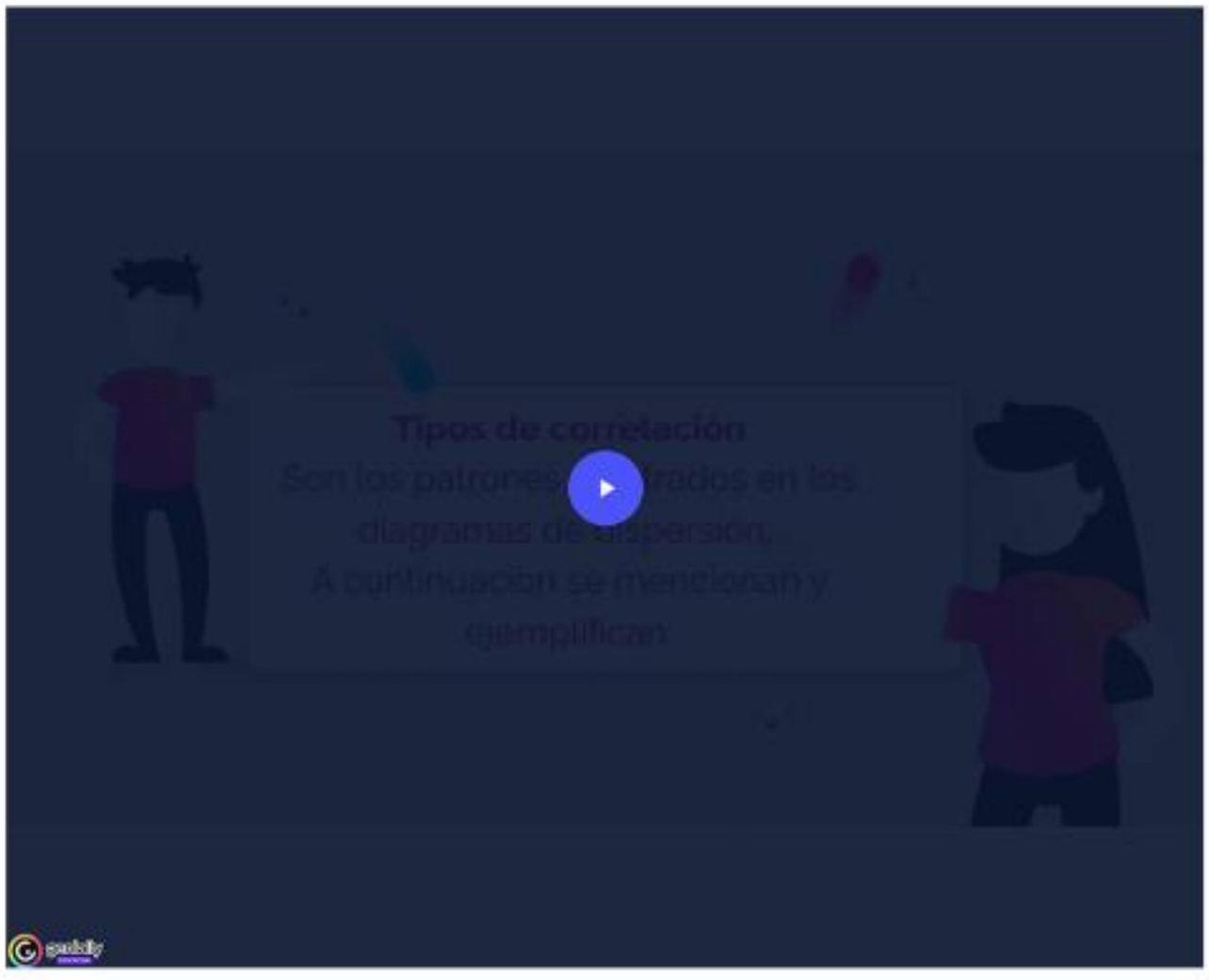
2.4 Diagramas de dispersión

Definición

El diagrama de dispersión es una herramienta empleada cuando queremos gráficamente llevar a cabo un análisis de datos que involucran a dos conjuntos de datos. También se conoce como gráfico de dispersión, gráfico de puntos, diagrama de XY, diagrama de Scattergram. Analizando el diagrama de dispersión podemos determinar la relación o correlación existente entre una variable y otra.

Se llama diagrama de dispersión, porque al colocar en el plano cartesiano todos los pares ordenados generados al comparar dos variables, estos puntos se van dispersando y generando un diagrama de acuerdo con el grado de relación existente entre ambas.

Se pueden interpretar varios tipos de correlación a través de los patrones mostrados en los diagramas de dispersión, a continuación, se muestran en la siguiente presentación:



Actividad 5.



Diagramas de dispersión

Instrucciones:
Da clic en la **imagen.**

El diagrama de dispersión nos permitirá tener un control sobre algún proceso, y haciendo un análisis de las variables involucradas, permitirá mejorarlo o cambiarlo de ser necesario dependiendo del comportamiento entre ellas. Es una herramienta utilizada cuando se desea realizar un análisis gráfico de datos bivariados, es decir, los que se refieren a dos conjuntos de datos.

Suele ser común que los datos bivariados sean representados en un plano cartesiano para un análisis más práctico, esta gráfica se conoce como **Diagrama de dispersión.**

2.5 Coeficiente de correlación

En muchas áreas de estudio es importante conocer la relación existente entre las diversas cantidades, por ejemplo: la que hay entre el precio de un producto y su consumo o entre el interés y el capital. Para analizar la posible relación entre las cantidades, se pueden construir coordenadas y a cada una de ellas se le asigna una variable.

La relación que hay entre dos variables puede ser: lineal, cuadrática, cúbica, exponencial, entre otras. Es importante considerar en caso de que dos variables tengan alguna relación, se dice que existe una correlación entre ellas.

Definición

El coeficiente de correlación es una cantidad que permite conocer que tan bien o con qué intensidad se ajusta un conjunto de pares ordenados a una relación lineal (recta).

Se dice que dos cantidades diferentes están correlacionadas linealmente si los puntos a una dispersión se pueden ubicar en una recta (con mayor o menor grado de presión).

El coeficiente de correlación se representa por el símbolo ***r***.

Para calcular el coeficiente de correlación se usa la fórmula:

$$r = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\sqrt{\Sigma(X-\overline{X})^2\Sigma(Y-\overline{Y})^2}}$$

Donde X es la variable a la que se le asignó la abscisa, es la coordenada que se le asignó el eje de las coordenadas. $\overline{X}$ y $\overline{Y}$ son las medias aritméticas* respectivas.

Para interpretar el coeficiente de correlación se tiene que:

i. Sí $r = 0$, entonces se dice que no existe correlación lineal entre las variables y los puntos en el gráfico no mostrarán patrón alguno.

ii. Sí $r = 1$ o -1, entonces se dice que si existe una correlación lineal, todos los puntos en el gráfico se alinearán en una línea con pendiente positiva o negativa.

iii. Sí $-1 < r < 1$, entonces la correlación no es exacta, es decir, entonces el resultado diferirá con valores alrededor de una línea recta. En problemas de la vida real, son el resultado que predomina.

*Media aritmética o promedio, es el valor que se obtiene al sumar todos los datos y dividirlos entre el total de ellos.

¿Qué limitaciones hay que tener en cuenta?

Sólo aporta información sobre las dos variables involucradas y no refiere en lo absoluto a posibles injerencias fuera de los datos bivariados o a valores atípicos; ni detectan relaciones curvilíneas de forma correcta.

¿Cómo se utiliza el coeficiente de correlación?

Para las variables en análisis, el resultado compara la distancia de cada dato respecto a la media de la variable, dicha comparación indica hasta qué punto la relación se ajusta a una línea imaginaria ubicada entre los datos. Esto hace referencia a que la correlación examina las relaciones lineales.

En la siguiente figura se puede observar lo mencionado:

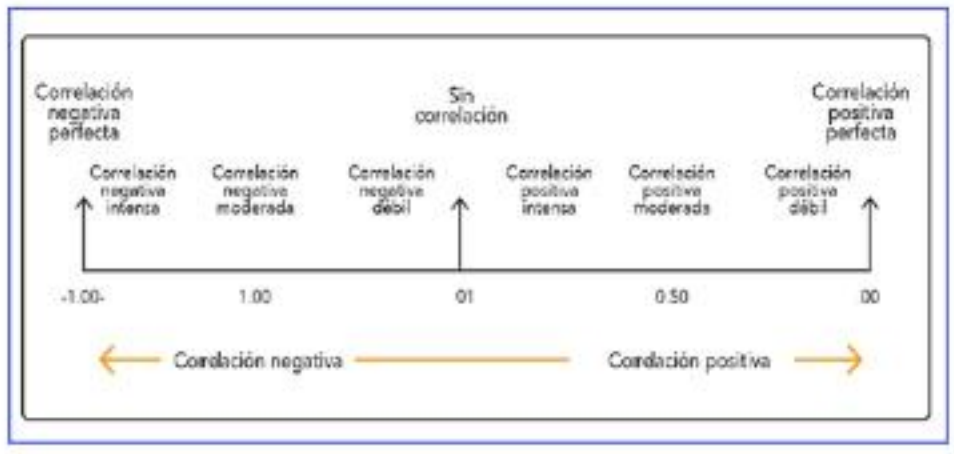


Figura 13. Coeficiente de correlación
Tomado para fines educativos de:
http://www.seduca2.uaemex.mx/ckfinder/uploads/files/u3tema_1_coeficiente.pdf

Ahora bien veamos algunos ejemplos de cómo se obtiene el coeficiente de correlación:

Ejemplo 1. Coeficiente de correlación en venta de autos.

Del tema 2.3 obtuvimos de la venta de autos, la tabla 13 que muestra las ventas en seis años consecutivos de su vehículo estrella.

Tabla 13. Registro de número de autos vendidos en seis años consecutivos

Año	Número de autos vendidos
2013	52,372
2014	84,263
2015	94,386
2016	97,681
2017	105,360
2018	132,157

Primero se calcularán las cantidades $\overline{X}$ y $\overline{Y}$

Tabla 14. Datos para el cálculo del coeficiente de correlación (1)

Años	Número de autos vendidos	$X-\overline{X}$	$Y-\overline{Y}$	$(X-\overline{X})(Y-\overline{Y})$	$(X-\overline{X})^2$	$(Y-\overline{Y})^2$
2013	52,372	-2.5	-41997.83	104994.58	6.25	1763817724.71
2014	84,263	-1.5	-10106.83	15160.25	2.25	102148012.65
2015	94,386	-0.5	16.17	-8.08	0.25	261.47
2016	97,681	0.5	3311.17	1655.59	0.25	10963846.77
2017	105,360	1.5	10990.17	16485.26	2.25	120783836.63
2018	132,157	2.5	37787.17	94467.93	6.25	1427870216.61
$\sum$= 12093	$\sum$= 566219			$\sum$= 232755.50	$\sum$= 17.5	$\sum$= 3425583898.83
$\overline{X}$ = 2015.5	$\overline{Y}$ = 94369.83					

Obtenemos el coeficiente de correlación:

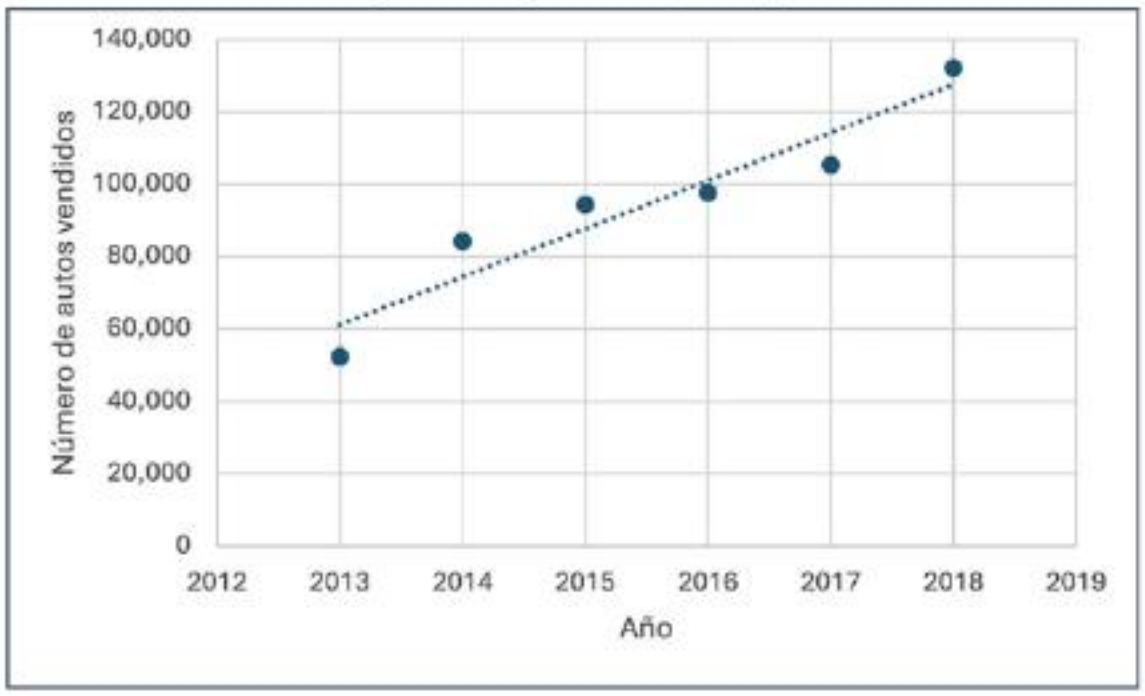
$$r=\frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\sqrt{\Sigma(X-\overline{X})^2\Sigma(Y-\overline{Y})^2}} \qquad r=\frac{232755.50}{\sqrt{(17.5)(3425583898.83)}}$$

El coefiente de correlación es:

$$r=0.950634614$$

Obtenemos el diagrama de dispersión de los datos y el coeficiente de correlación de $r = 0.950634614$ indica una relación lineal positiva, mostrándonos que los datos son directamente proporcionales.

Figura 14. Diagrama de dispersión (1)



Ejemplo 2. Coeficiente de correlación en cantidad de tierra que deben procesar para obtener plata de un yacimiento..

Del tema 2.3 obtuvimos la Tabla 15 del registro de cantidad de tierra que deben procesar para obtener plata de un yacimiento, los resultados fueron lo siguientes:

Tabla 15. Registro de cantidad de tierra.

Toneladas de tierra procesada	Kilogramos de plata obtenidos
50	234
36	542
45	531
29	284
48	487
57	332
42	197

Primero se calcularán las cantidades $\overline{X}$ y $\overline{Y}$

Tabla 16. Datos para el cálculo del coeficiente de correlación (2)

Toneladas de tierra procesada	Kilogramos de plata obtenidos	$X-\overline{X}$	$Y-\overline{Y}$	$(X-\overline{X})(Y-\overline{Y})$	$(X-\overline{X})^2$	$(Y-\overline{Y})^2$
50	234	6.14	-138.43	-849.9602	37.6996	19162.8649
36	542	-7.86	169.57	-1332.8202	61.7796	28753.9849

Toneladas de tierra procesada	Kilogramos de plata obtenidos	$X-\overline{X}$	$Y-\overline{Y}$	$(X-\overline{X})(Y-\overline{Y})$	$(X-\overline{X})^2$	$(Y-\overline{Y})^2$
45	531	1.14	158.57	180.7698	1.2996	25144.4449
29	284	-14.86	-88.43	1314.0698	220.8196	7819.8649
48	487	4.14	114.57	474.3198	17.1396	13126.2849
57	332	13.14	-40.43	-531.2502	172.6596	1634.5849
42	197	-1.86	-175.43	326.2998	3.4596	30775.6849
$\overline{X}$ = 43.86	$\overline{Y}$ = 372.43			∑= -418.57	∑= 514.86	∑= 126417.71

Obtenemos el coeficiente de correlación:

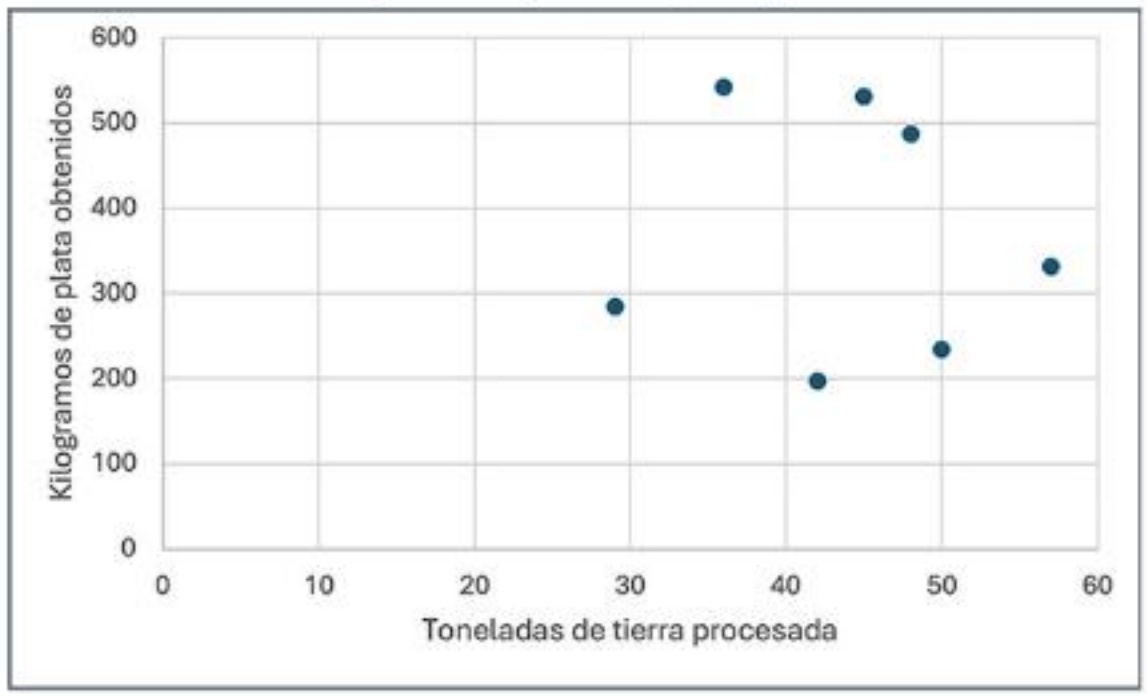
$$r = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\sqrt{\Sigma(X-\overline{X})^2\Sigma(Y-\overline{Y})^2}} \qquad r = \frac{-418.57}{\sqrt{(514.86)(126417.71)}}$$

El coefiente de correlación es:

$$r = -0.051882661$$

Obtenemos el diagrama de dispersión de los datos y el coeficiente de correlación de $r = -0.051882661$ indica una relación nula entre las toneladas de tierra procesada y los kilogramos de plata obtenidos.

Figura 15. Diagrama de dispersión (2)



Ejemplo 3. Coeficiente de correlación para la tasa de natalidad

Se ha considerado que en el mundo la tasa de natalidad por cada mil habitantes es inversamente proporcional al producto interno bruto de cada país, podría considerarse que un país desarrollado tiene una menor tasa de natalidad y que en contraparte un país en desarrollo tiene una tasa elevada. Aleatoriamente elegimos 10 países con su crecimiento del PIB (%) según el **Banco Mundial** y su correspondiente tasa de natalidad (nacimientos/1000 habitantes) de la página **Index Mundi** y tenemos la siguiente tabla:

Tabla 17. Registro de PIB y tasa de natalidad

País	PIB	Tasa de natalidad
Georgia	10.4	11.56
Egipto	6.6	27.22
Islandia	7.2	13.28
Emiratos Árabes	7.9	9.5
Nigeria	3.3	47.51
Mónaco	11.1	6.37
Uganda	4.6	42.27
Panamá	10.8	17.06
Liberia	4.8	37.29
Togo	5.8	31.97

Primero se calcularán las cantidades $\overline{X}$ y $\overline{Y}$

Tabla 18. Datos para el cálculo del coeficiente de correlación (3)

PIB	Tasa de natalidad	$X-\overline{X}$	$Y-\overline{Y}$	$(X-\overline{X})(Y-\overline{Y})$	$(X-\overline{X})^2$	$(Y-\overline{Y})^2$
10.4	12	3.15	-12.84	-40.446	9.9225	164.8656
6.6	27	-0.65	2.82	-1.833	0.4225	7.9524
7.2	13	-0.05	-11.12	0.556	0.0025	123.6544
7.9	10	0.65	-14.9	-9.685	0.4225	222.01
3.3	48	-3.95	23.11	-91.2845	15.6025	534.0721
11.1	6	3.85	-18.03	-69.4155	14.8225	325.0809
4.6	42.27	-2.65	17.87	-47.3555	7.0225	319.3369
10.8	17.06	3.55	-7.34	-26.057	12.6025	53.8756
4.8	37.29	-2.45	12.89	-31.5805	6.0025	166.1521
5.8	31.97	-1.45	7.57	-10.9765	2.1025	57.3049
$\overline{X}=7.25$	$\overline{Y}=24.4$			$\sum=-328.08$	$\sum=68.93$	$\sum=1974.3$

Obtenemos el coeficiente de correlación:

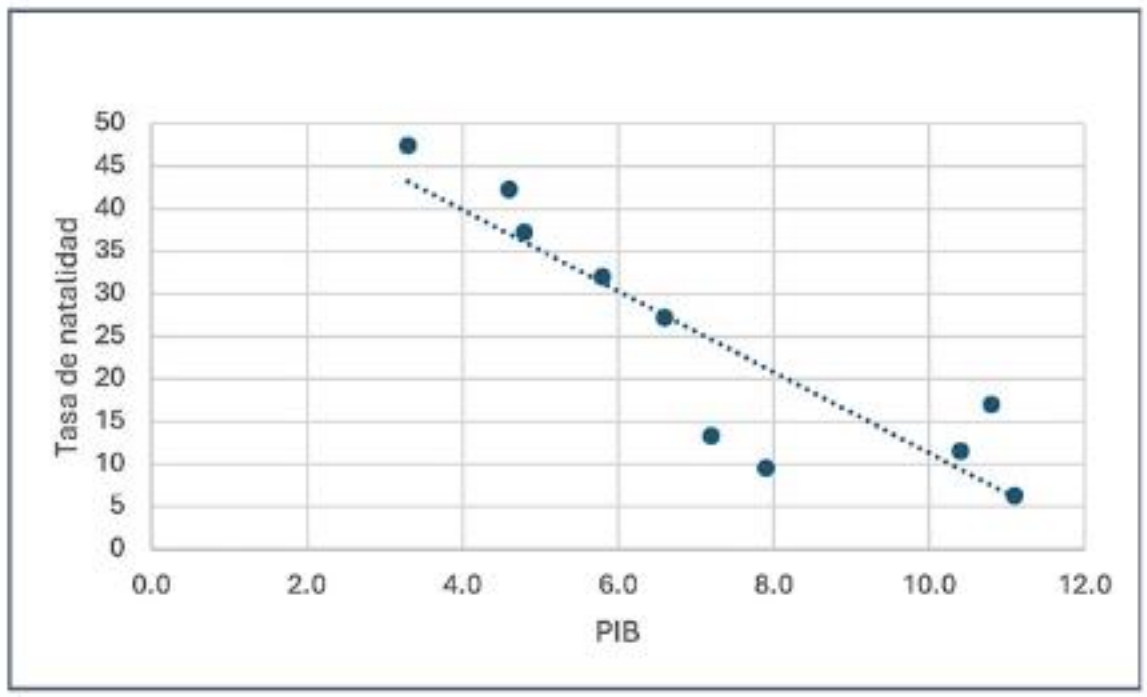
$$r=\frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\sqrt{\Sigma(X-\overline{X})^2\Sigma(Y-\overline{Y})^2}} \qquad r=\frac{-328.08}{\sqrt{(68.93)(1974.3)}}$$

El coefiente de correlación es:

$$r=-0.889366891$$

Obtenemos el diagrama de dispersión de los datos y el coeficiente de correlación de $r = -0.889366891$ que indica una relación lineal negativa, mostrándonos que los datos son inversamente proporcionales y que efectivamente, entre mejor PIB tenga el país, menor tasa de natalidad habrá en él y un PIB bajo conllevará a un elevado índice de natalidad.

Figura 16. Diagrama de dispersión (3)



Ejemplo 4. Coeficiente de correlación para una colonia de bacterias

Una colonia de bacterias inicia con 1000 bacterias, al transcurrir una hora se divide generando una nueva población de 2000 bacterias, así sucesivamente durante 10 horas. Tabulando tenemos:

Tiempo (horas)	Núm. De bacterias
0	1000
1	2000
2	4000
3	8000
4	16000
5	32000
6	64000
7	128000
8	256000
9	512000
10	1024000

Primero se calcularán las cantidades $\overline{X}$ y $\overline{Y}$

Tabla 20. Datos para el cálculo del coeficiente de correlación (4)

Tiempo (horas)	Núm. De bacterias	$X-\overline{X}$	$Y-\overline{Y}$	$(X-\overline{X})(Y-\overline{Y})$	$(X-\overline{X})^2$	$(Y-\overline{Y})^2$
0	1,000	-5	-185090.91	925454.55	25	3.4259E+10
1	2,000	-4	-184090.91	736363.64	16	3.3889E+10
2	4,000	-3	-182090.91	546272.73	9	3.3157E+10
3	8,000	-2	-178090.91	356181.82	4	3.1716E+10
4	16,000	-1	-170090.91	170090.91	1	2.8931E+10
5	32,000	0	-154090.91	0	0	2.3744E+10
6	64000	1	-122090.91	-122090.91	1	1.4906E+10
7	128000	2	-58090.91	-116181.82	4	3374553825
8	256000	3	69909.09	209727.27	9	4887280865
9	512000	4	325909.09	1303636.36	16	1.0622E+11
10	1024000	5	837909.09	4189545.45	25	7.0209E+11
$\overline{X}=5$	$\overline{Y}=186090.91$			$\sum=8199000$	$\sum=110$	$\sum=1.0172E+12$

Obtenemos el coeficiente de correlación:

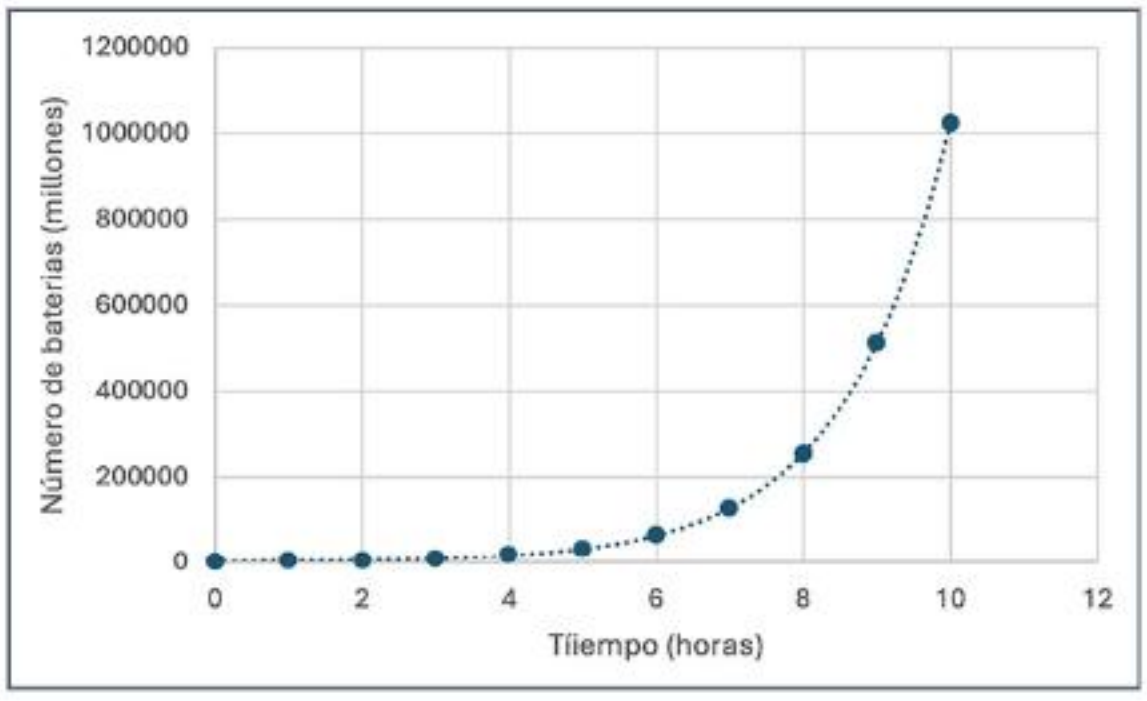
$$r=\frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\sqrt{\Sigma(X-\overline{X})^2\Sigma(Y-\overline{Y})^2}} \qquad r=\frac{8199000}{\sqrt{(100)(1.0172E+12)}}$$

El coefiente de correlación es:

$$r=-0.775116803$$

Obtenemos el diagrama de dispersión de los datos y el coeficiente de correlación de $r=0.775116803$ que indica una relación exponencial, mostrándonos que la variación es cada vez mayor que en el periodo anterior.

Figura 17. Diagrama de dispersión (4)



Ejemplo 5. Coeficiente de correlación para el número de bacterias en un cultivo

En la materia de Biología, un grupo de estudiantes compararon el número de bacterías presentes en un cultivo contra el tiempo transcurrido para evaluar el funcionamiento de un antibiótico. Los resultados fueron los siguientes:

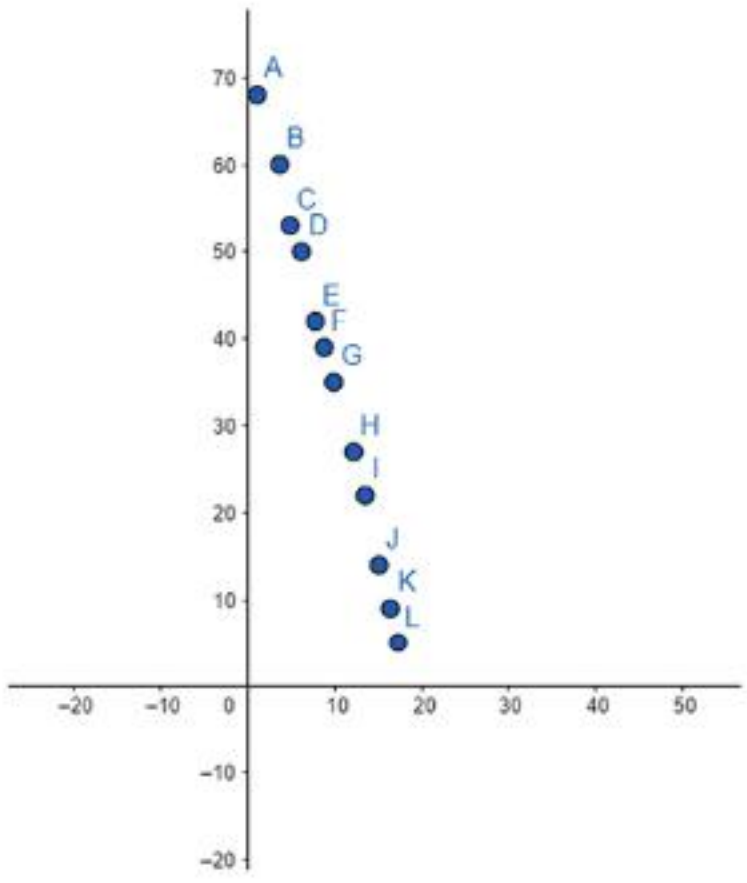
Tabla 21. Número de bacterias respecto al tiempo transcurrido

Tiempo (minutos)	Número de bacterias
1.1	68
3.7	60
4.9	53
6.2	50
7.8	42
8.8	39
9.9	35
12.2	27
13.5	22
15.1	14
16.4	9
17.3	5

De los datos anteriores obtener:

a. Diagrama de dispersión

Figura 18. Diagrama de dispersión (5)



b. El coeficiente de correlación

Tabla 22. Datos para el cálculo del coeficiente de correlación

Tiempo (minutos)	Número de bacterias	$X-\overline{X}$	$Y-\overline{Y}$	$(X-\overline{X})(Y-\overline{Y})$	$(X-\overline{X})^2$	$(Y-\overline{Y})^2$
1.1	68	-8.64	32.67	-282.27	74.65	1067.33
3.7	60	-6.04	24.67	-149.01	36.48	608.61
4.9	53	-4.84	17.67	-85.52	23.43	312.23
6.2	50	-3.54	14.67	-51.93	12.53	215.21
7.8	42	-1.94	6.67	-12.94	3.76	44.49
8.8	39	-0.94	3.67	-3.45	0.88	13.47
9.9	35	0.16	-0.33	-0.05	0.03	0.11
12.2	27	2.46	-8.33	-20.49	6.05	69.39
13.5	22	3.76	-13.33	-50.12	14.14	177.69
15.1	14	5.36	-21.33	-114.33	28.73	454.97
16.4	9	6.66	-26.33	-175.36	44.36	693.27
17.3	5	7.56	-30.33	-229.29	57.15	919.91
$\overline{X}=9.74$	$\overline{Y}=35.33$			$\sum$=-1174.77	$\sum$=302.19	$\sum$=4576.67

Obtenemos el coeficiente de correlación:

$$r=\frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\sqrt{\Sigma(X-\overline{X})^2\Sigma(Y-\overline{Y})^2}}$$

$$r=\frac{-1174.77}{\sqrt{(302.19)(4576.67)}}$$

El coefiente de correlación es:

$$r=-.999=-1$$

Tiene un coefiente de correlación negativo perfecto, lo que indica una relación lineal negativa, mostrándonos que los datos son inversamente proporcionales y que efectivamente para evaluar el funcionamiento de un antibiótico, entre más tiempo transcurra el número de bacterías presentes en un cultivo disminuye.

La correlación no implica forzosamente causa y efecto. Una correlación de 1 o -1 entre dos variables pudiera implicar una tercera variable no medida. Es decir, el hecho de que dos cosas estén relacionadas no significa que necesariamente una sea causa de la otra.

Aun así, los diagramas de dispersión y las correlaciones son herramientas útiles para ahondar en los datos, nos ayudan a detectar tendencias y relaciones que sólo ellas nos pueden mostrar y con su interpretación entender y mejorar el mundo que nos rodea.

2.6 Ecuación de la recta de regresión y aplicación a serie de tiempos

En la investigación es fundamental comprobar la hipótesis, una variable depende de otra u otras. Por lo que analizar una estimación de la magnitud de impacto en una variable sobre otra, ayuda a predecir los valores futuros. Sin embargo, la relación que nos interesa es cuando los puntos de una dispersión tienen una correlación lineal, por lo que se pueden ajustar a una recta; llamada recta de regresión.

Definición

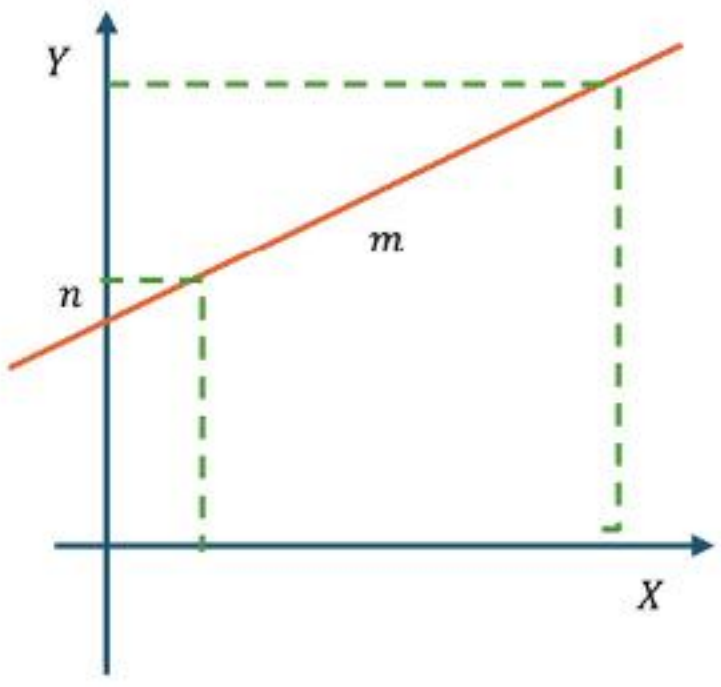
La recta de regresión es la representación gráfica de la regresión lineal que se define como un modelo estadístico que permite observar la tendencia lineal de un conjunto de puntos obtenida de los posibles resultados y tiene la forma:

$$y = mx + b$$

Donde:
y es la variable dependiente
x es la variable independiente
m es la pendiente de la recta
b es la ordenada al origen

Su representación gráfica es:

Figura 19. Recta de regresión (1)



De la gráfica se deduce:

- Si el signo de m es positivo indica que la recta crece.
- Si el signo de m es negativo entonces la recta decrece.

En este caso la pendiente está dada por:

$$m = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\Sigma(X-\overline{X})^2}$$

Una característica de la recta de regresión es que pasa por un punto llamado centro de gravedad con coordenadas ($\overline{X}$, $\overline{Y}$), es decir, las medias aritméticas de los valores de X y Y.

Por lo que la ecuación de la regresión lineal se puede escribir en su forma ordenada al origen como

$$y = mx + b$$

Donde

$$m = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\Sigma(X-\overline{X})^2}$$

$$b = \overline{Y} - m\overline{X}$$

Por otro lado, si la correlación es nula; entonces las rectas de regresión son perpendiculares.

Ejemplo 1. Ecuación de la recta de regresión

Traza la gráfica de la recta indicando la pendiente, si se sabe que la ordena al origen es $b = 2$; tiene pendiente $m = 1$.

Solución:

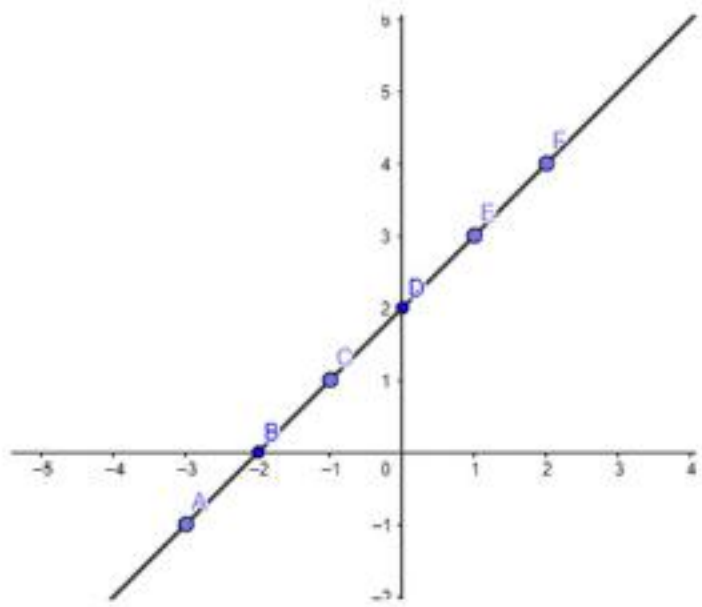
$$y = (1)x + 2$$

Para trazar la gráfica tabulamos valores $x \in [-3,3]$

Tabla 23. Tabulación

x	$y = x + 2$	(x, y)
−3	$-3 + 2 = -1$	(−3,−1)
−2	$-2 + 2 = 0$	(−2, 0)
−1	$-1 + 2 = 1$	(−1,1)
0	$0 + 2 = 2$	(0, 2)
1	$1 + 2 = 3$	(1, 3)
2	$2 + 2 = 4$	(2, 4)
3	$3 + 2 = 5$	(3, 5)

Figura 20. Recta de regresión (2)



Ejemplo 2. Ecuación de la recta de regresión del ejemplo 5 del tema 2.5

En la materia de Biología, un grupo de estudiantes compararon el número de bacterías presentes en un cultivo contra el tiempo transcurrido para evaluar el funcionamiento de un antibiótico. Los resultados fueron los siguientes:

Tabla 24. Número de bacterias respecto al tiempo transcurrido

Tiempo (minutos)	Número de bacterias
1.1	68
3.7	60
4.9	53
6.2	50
7.8	42
8.8	39
9.9	35
12.2	27
13.5	22
15.1	14
16.4	9
17.3	5

Donde se obtuvieron los siguientes resultados:

Tabla 25. Datos para el cálculo del coeficiente de correlación y la recta de regresión (1)

Tiempo (minutos)	Número de bacterias	$X-\overline{X}$	$Y-\overline{Y}$	$(X-\overline{X})(Y-\overline{Y})$	$(X-\overline{X})^2$
1.1	68	-8.64	32.67	-282.27	74.65
3.7	60	-6.04	24.67	-149.01	36.48
4.9	53	-4.84	17.67	-85.52	23.43
6.2	50	-3.54	14.67	-51.93	12.53
7.8	42	-1.94	6.67	-12.94	3.76

Tiempo (minutos)	Número de bacterias	$X-\overline{X}$	$Y-\overline{Y}$	$(X-\overline{X})(Y-\overline{Y})$	$(X-\overline{X})^2$
8.8	39	-0.94	3.67	-3.45	0.88
9.9	35	0.16	-0.33	-0.05	0.03
12.2	27	2.46	-8.33	-20.49	6.05
13.5	22	3.76	-13.33	-50.12	14.14
15.1	14	5.36	-21.33	-114.33	28.73
16.4	9	6.66	-26.33	-175.36	44.36
17.3	5	7.56	-30.33	-229.29	57.15
$\overline{X}=9.74$	$\overline{Y}=35.33$			$\sum$= -1174.77	$\sum$=302.19

El centro de gravedad tiene coordenadas $(\overline{X},\overline{Y})=(9.74, 35.33)$
La pendiente es:

$$m=\frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\Sigma(X-\overline{X})^2}$$

$$m=\frac{-1174.77}{302.19}$$

$$m=-3.89$$

La ordenada al origen es:

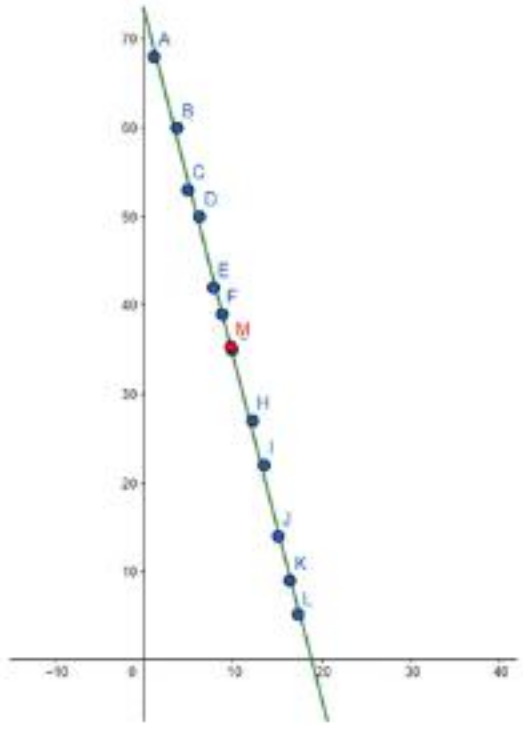
$$b=\overline{Y}-m\overline{X}$$

$$b=35.33-(-3.89)(9.74)$$

$$b=73.22$$

Entonces **la ecuación de la recta de regresión es:** $y=-3.89x+73.22$

Figura 21. Recta de regresión (3)



La recta describe el comportamiento lineal del número de bacterías presentes en un cultivo en relación con el tiempo transcurrido para evaluar el funcionamiento de un antibiótico.

Definición

Series de Tiempo

Una **serie de tiempo** es un conjunto o una colección de valores o fenómenos observados durante una serie de periodos temporales secuencialmente ordenada. Los periodos pueden ser de horas, por día, semanales, mensuales, trimestrales, o anuales.

El análisis de series de tiempo hace referencia al proceso de analizar los datos disponibles para identificar el patrón o la tendencia en los datos.

Lo que permite extraer y modelar las relaciones entre datos a lo largo del tiempo, sea extrapolando (hacia futuro) o interpolando (hacia el pasado) el comportamiento de datos no observados.

Estos pueden ser representados mediante un graficado secuencial, donde cada x_t fue observado en el instante t

Ejemplo 3. Series de tiempo para temperaturas

Supongamos que la temperatura máxima en Toluca ha sido observada diariamente del 1 al 10 de abril del 2024. Obteniendo los datos:

x_1=29°, x_2=29°, x_3=27°, x_4=26°, x_5=27°, x_6=28°, x_7=28°, x_8=27°, x_9=27°, x_{10}=27°

Ejemplo 4. Series de tiempo para venta de autos

Del ejemplo 1 del Tema 2.3

Una empresa que se dedica a la venta de autos ha realizado una tabla que muestra las ventas en seis años consecutivos de su vehículo estrella obteniendo los siguientes datos:

Tabla 26. Registro de número de autos vendidos

Año	Número de autos vendidos
2013	52,372
2014	84,263
2015	94,386
2016	97,681
2017	105,360
2018	132,157

Obtener la recta de ajuste, la gráfica y las ventas que se esperan para el año 2024 y 2025.

Primero se calcularán las cantidades $\overline{X}$ y $\overline{Y}$

Tabla 27. Datos para el cálculo del coeficiente de correlación y la recta de regresión (2)

Años	Número de autos vendidos	$X-\overline{X}$	$Y-\overline{Y}$	$(X-\overline{X})(Y-\overline{Y})$	$(X-\overline{X})^2$	$(Y-\overline{Y})^2$
2013	52,372	-2.5	-41997.83	104994.58	6.25	1763817724.71
2014	84,263	-1.5	-10106.83	15160.25	2.25	102148012.65
2015	94,386	-0.5	16.17	-8.08	0.25	261.47
2016	97,681	0.5	3311.17	1655.59	0.25	10963846.77
2017	105,360	1.5	10990.17	16485.26	2.25	120783836.63
2018	132,157	2.5	37787.17	94467.93	6.25	1427870216.61
$\sum$= 12093	$\sum$= 566219			$\sum$= 232755.50	$\sum$= 17.5	$\sum$= 3425583898.83
$\overline{X}$ = 2015.5	$\overline{Y}$ = 94369.83					

Obtenemos el coeficiente de correlación:

$$r = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\sqrt{\Sigma(X-\overline{X})^2\Sigma(Y-\overline{Y})^2}}$$

$$r = \frac{232755.50}{\sqrt{(17.5)(3425583898.83)}}$$

El coefiente de correlación es:

$$r = 0.95$$

El centro de gravedad tiene coordenadas

$(\overline{X}, \overline{Y}) = (2015.5, 94369.83)$

La pendiente de la recta de regresión es:

$$m = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\Sigma(X-\overline{X})^2}$$

$$m = \frac{2327555.50}{17.7}$$

$m = 13300.31$

La ordenada al origen es:

$b = \overline{Y} - m\overline{X}$

$b = 94369.83 - (13300.31)(2015.5)$

$b = -26712404.98$

Entonces la ecuación de la recta de regresión es:

$y = 13300.31x - 26712413.98$

Para calcular **las ventas esperadas para el año 2024;**

$x = 2024$

Se sustituye en la ecuación

$y = 13300.31(2024) - 26712413.98$

$y = 207422.46$

Para calcular l**as ventas esperadas para el año 2025;**

$x = 2025$

Se sustituye en la ecuación

$y = 13300.31(2025) - 26712413.98$

$y = 220722.77$

Figura 22. Diagrama de dispersión con recta de regresión

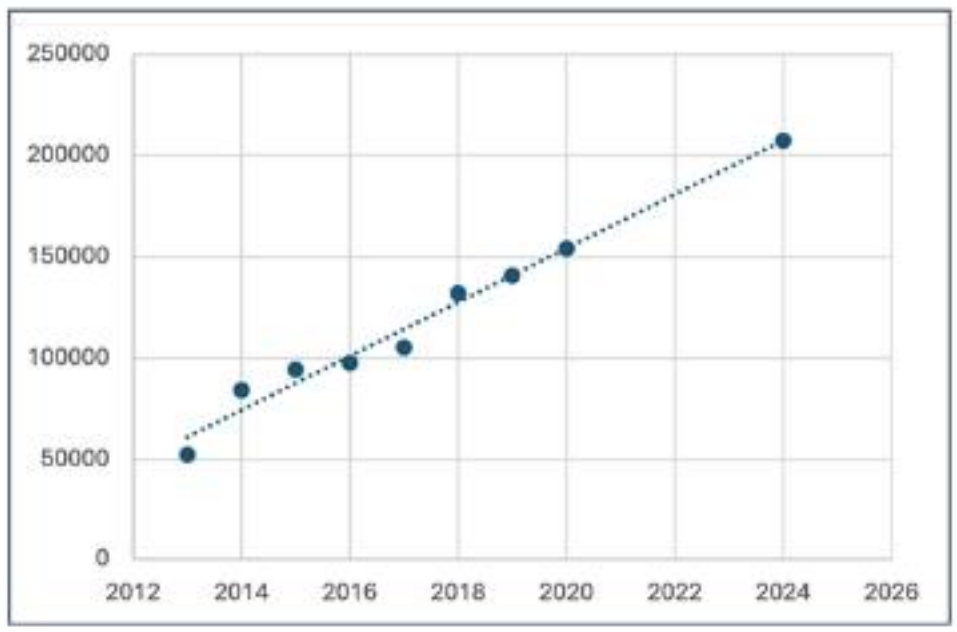
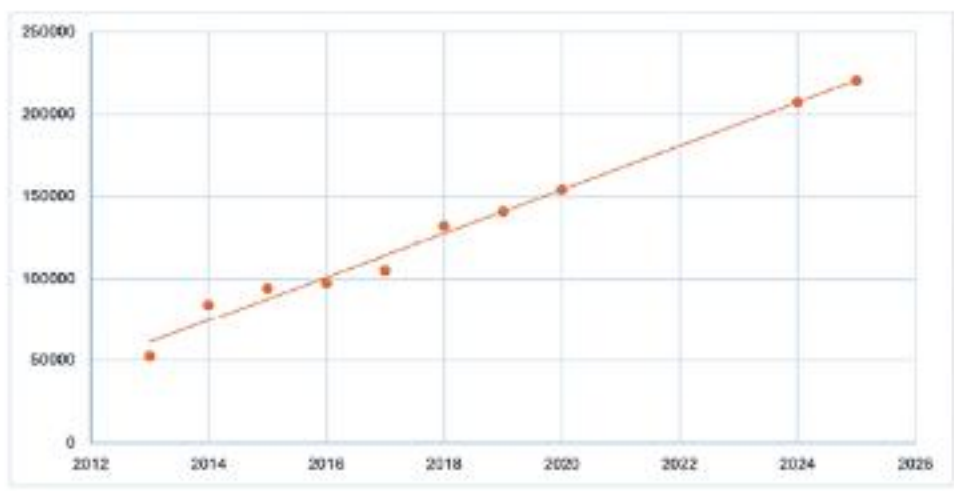
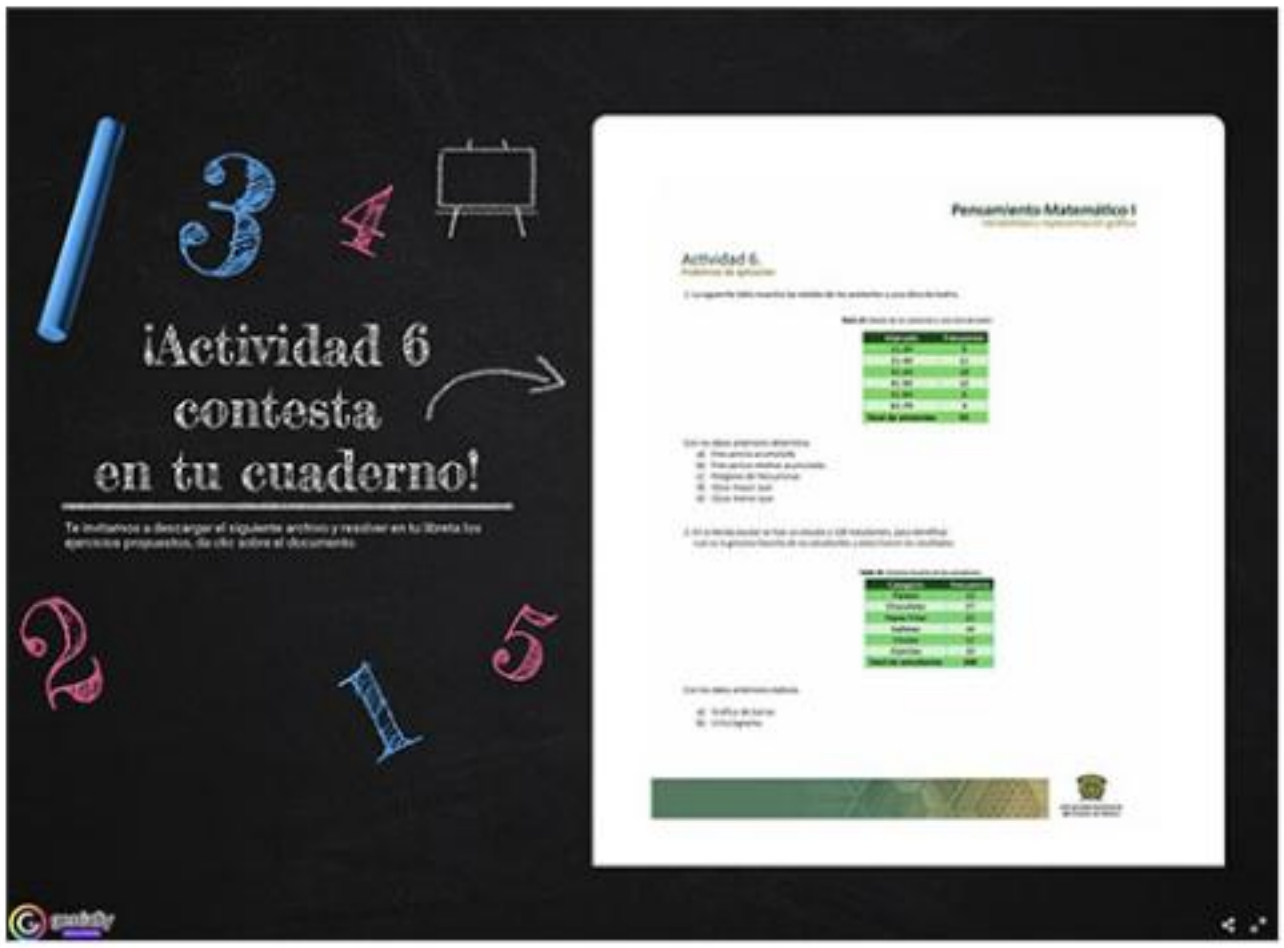


Figura 23. Diagrama de dispersión con recta de regresión (escala ampliada)



Así, la venta esperada para el año 2024 y 2025 son 207422.46 y 220722.77 unidades vendidas, respectivamente, según la recta de regresión.

2.7 Problemas de aplicación



Actividad 6.

Problemas de aplicación

Instrucciones:
Descarga el archivo adjunto, dando clic en la imagen, imprímelo y realiza el ejercicio.

1. La siguiente tabla muestra las edades de los asistentes a una obra de teatro.

Tabla 28. Edades de los asistentes a una obra de teatro

Intervalo	Frecuencia
11-20	5
21-30	11
31-40	18
41-50	12
51-60	6
61-70	3
Total de asistentes	55

Con los datos anteriores determina:

a. Frecuencia acumulada
b. Frecuencia relativa acumulada
c. Polígono de frecuencias
d. Ojiva mayor que
e. Ojiva menor que

2. En la tienda escolar se hizo un estudio a 100 estudiantes, para identificar cuál es la golosina favorita de los estudiantes y estos fueron los resultados:

Tabla 29. Golosina favorita de los estudiantes

Categoría	Frecuencia
Paletas	13
Chocolates	27
Papas fritas	22
Galletas	16
Chicles	12
Gomitas	10
Total de estudiantes	100

Con los datos anteriores elabora:

a. Gráfica de barras
b. Circulograma

3. Una empresa ofrece visitas semanales a clientes con el fin de dar a conocer sus productos alimenticios. Por cuestiones de sanidad, la empresa debe ofrecer un uniforme especial a los invitados, por lo que ha decidido hacer un estudio estadístico sobre la estatura de los 100 asistentes que tuvo durante una semana, obteniendo los siguientes resultados:

Tabla 30. Estatura de 100 asistentes a una visita en una empresa alimenticia

159	160	153	171	152	161	165	155	174	175
156	166	157	157	175	158	158	157	160	172
165	166	166	154	167	162	164	160	177	180
177	156	155	167	162	166	149	160	168	150
171	155	154	159	164	157	157	167	158	167
167	156	153	147	155	157	164	160	168	175
156	159	154	169	155	155	157	160	166	160
167	157	157	163	159	166	160	167	165	164
147	161	164	159	163	166	152	164	167	164
166	164	159	160	161	168	172	161	163	180

La empresa te ha asignado que organices los datos de la tabla con las siguientes solicitudes:

a. Elaborar una tabla de distribución de frecuencias con 8 intervalos
b. Realizar una gráfica de polígono de frecuencias y un circulograma
c. Elabora un reporte en el que expliques cuáles tallas conviene producir en mayor y en menor cantidad, en tu informe coloca el gráfico y explica por qué es el que mejor representa a los datos

4. En un supermercado se desea saber qué tipo de alimentos son los más consumidos por los clientes, si los productos frescos o los enlatados, para ello, se registró el número de clientes y el número de productos que eligieron a lo largo de una semana, obteniendo los siguientes datos:

Tabla 31. Productos elegidos por los clientes en un supermercado durante una semana

Día de la semana	Número de clientes del día	Número de clientes que llevaron productos frescos	Número de clientes que llevaron productos enlatados
Lunes	312	78	218
Martes	386	103	221
Miércoles	278	203	147
Jueves	216	158	74
Viernes	478	206	297
Sábado	516	278	314
Domingo	684	306	507

Con esta información realiza lo siguiente:

a. Una gráfica de dispersión con "número de clientes del día" y "número de clientes que llevaron productos frescos"

b. Una gráfica de dispersión con "número de clientes del día" y "número de clientes que llevaron productos enlatados"
c. Observa las gráficas y en un reporte describe cuáles productos tienen una mejor relación entre ellos.
d. En tu reporte describe con cuáles tipos de productos es más conveniente realizar una predicción de las ventas y por qué

5. El Plantel: "Dr. Ángel Ma. Garibay Kintana" de la Escuela Preparatoria de la Universidad Autónoma del Estado de México ha registrado el número de aspirantes y aceptados a lo largo de los últimos 5 periodos obteniendo la siguiente información **(https://spydi.uaemex.mx/docs/InfBasCon/PlantelDrAngelMGaribay/Informes/Periodo2019-2023/4to-Informe-Plantel-Dr.-Angel-Ma-Garibay.pdf).**

Figura 24. Alumnos aspirantes y aceptados en el Plantel: "Dr. Ángel Ma. Garibay Kintana" de la Escuela Preparatoria de la Universidad Autónoma del Estado de México



Con esta información realiza lo siguiente:

a. Describe en una tabla de frecuencias de datos bivariados el número de solicitudes recibidas (x) versus en número de inscritos (y)
b. Con los datos de la tabla elabora un diagrama de dispersión
c. Obtén el centro de gravedad de la tabla de frecuencias de datos bivariados
d. Calcula el coeficiente de correlación de los datos
e. Obtén la ecuación de la recta de regresión de los datos
f. Calcula cuántos posibles ingresos se tendrían cuando se reciben 1580 y 1900 solicitudes respectivamente
g. Interpreta los datos obtenidos en los incisos c), d) y f)

6. Obtener la ecuación de regresión lineal que se ajusta a los datos que recabaron de probar un prototipo de motocicleta en la Faculta de Ingeniería.

Tabla 32. Datos para aprobación de prototipo de motocicleta

Tiempo (segundos)	Distancia (metros)
1	20
2	25
3	30
4	35
5	40
6	45

7. Para la siguiente colección de datos encuentra:
a. El diagrama de dispersión
b. El coeficiente de correlación
c. La recta de regresión
d. La gráfica de la recta de regresión previamente calculada

Tabla 33. Registro de datos bivariados

X	Y
0.9	16.4
1.4	15.1
2.7	12.2
3.5	9.9
4.2	7.8
5.3	4.9
6	3.7

8. Una empresa chocolatera recabo los siguientes datos de las ventas realizadas en relación con el gasto en su publicidad en los últimos 7 años.

a. Traza el diagrama de dispersión
b. Encontrar el coeficiente de correlación
c. ¿Existe relación lineal entre las ventas de la empresa y los gastos en publicidad? En caso de serlo encuentra la ecuación de regresión
d. Traza la gráfica de la recta de regresión previamente calculada
e. ¿Qué volumen de gastos en publicidad se espera que invierta la empresa con una venta de 50 000?

Tabla 34. Datos de las ventas realizadas en relación con el gasto en su publicidad en los últimos 7 años.

Volumen de Ventas $	Gastos de Publicidad $
10 000	16 000
15 000	32 000
20 000	48 000
22 000	56 000
30 000	64 000
32 000	80 000
40 000	90 000

2.8 Interpretación gráfica

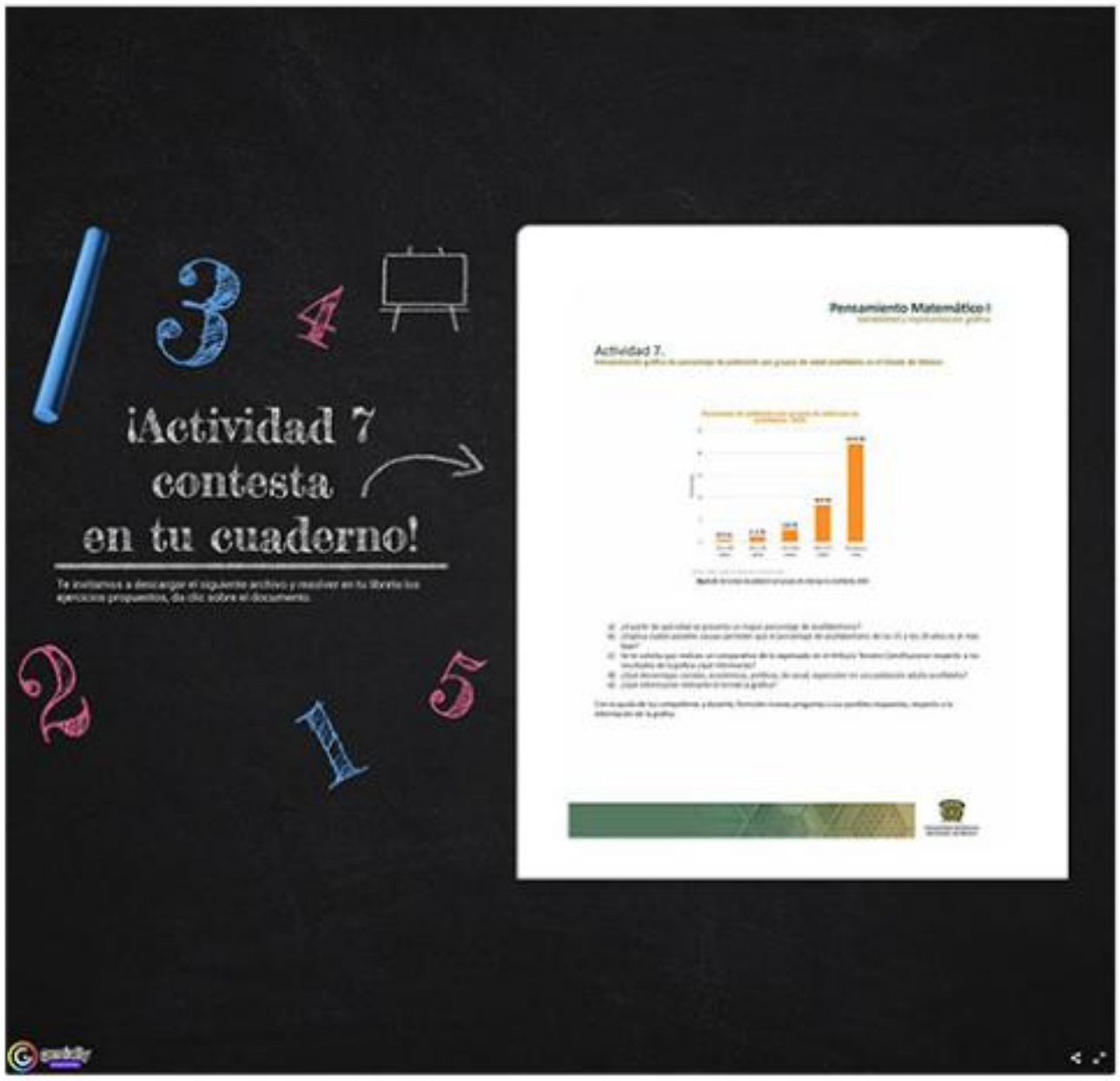
Uno de los fines del manejo de datos estadísticos, desde su obtención y organización, es la representación gráfica, cuyo objetivo es brindar información de fácil interpretación, tal como se cita a continuación:

> "Los gráficos estadísticos son potentes herramientas para la visualización de datos que permiten representar de manera accesible información compleja. Consiguen presentar la información al usuario o lector de manera clara y precisa, facilitando la comparación y la comprensión de la evolución de distintas variables." (UNIR; 2021)

La interpretación de tablas y gráficas estadísticas tiene diversas utilidades, al respecto Amanda Quiñones (2012) menciona:

Definición

- Comprender la información estadística que sobre diversos temas proporcionan los medios de comunicación y poder evaluarla de forma crítica.
- Entender la realidad social, económica y política.
- Llevar a cabo investigaciones que requieran interpretar datos.
- Discutir o comunicar las propias opiniones sobre la información que las estadísticas presentan.



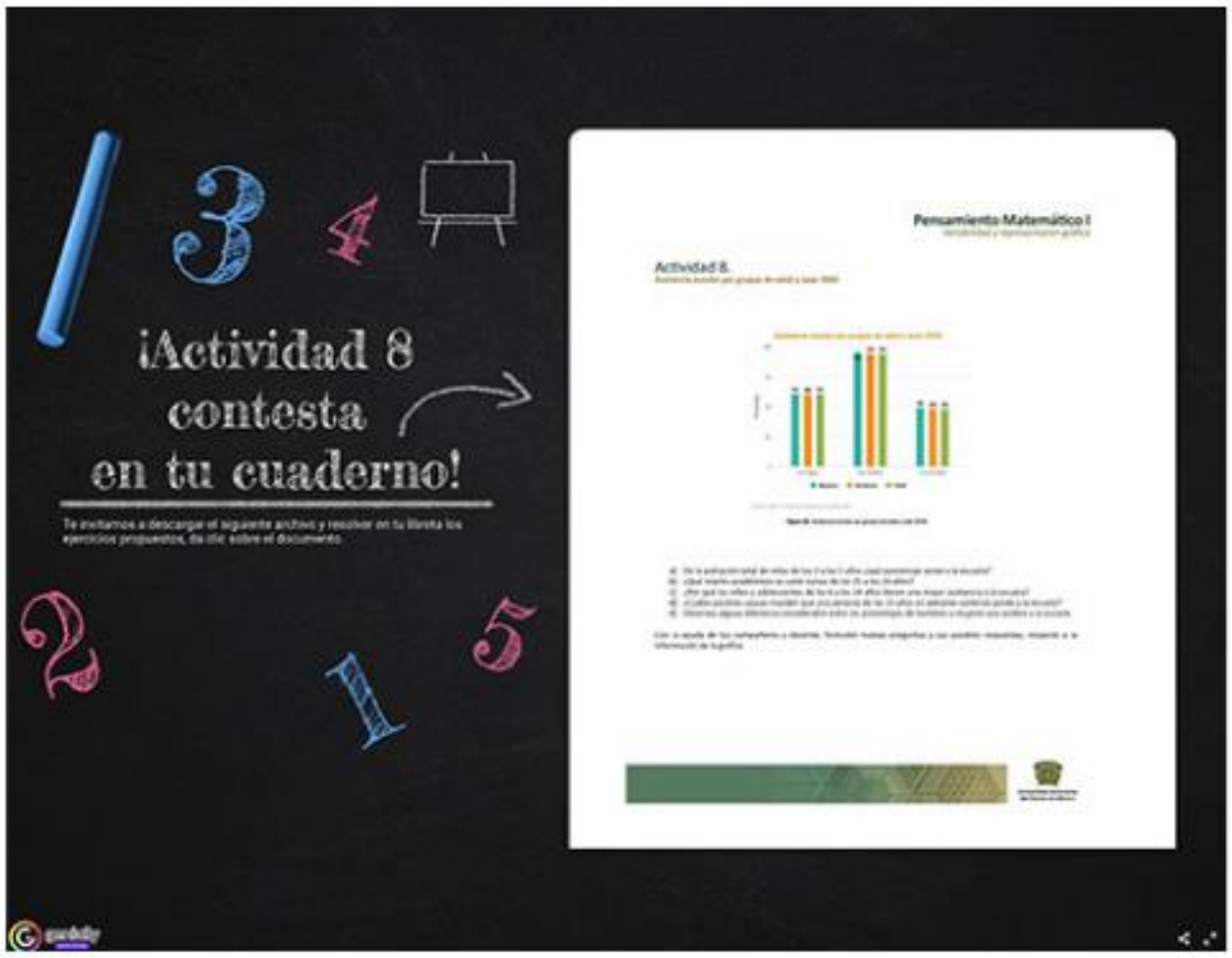
Actividad 7.



Interpretación gráfica de porcentaje de población por grupos de edad analfabeta en el Estado de México

Instrucciones:
Se muestra una gráfica con preguntas detonadoras para que junto con tus compañeros obtengan sus posibles conclusiones. En ella se muestra el porcentaje de población por grupos de edad que es analfabeta en el Estado de México, con datos obtenidos en el censo de 2020:

Descarga el archivo adjunto, dando clic en la imagen, imprímelo y realiza el ejercicio.



Actividad 8.

Asistencia escolar por grupos de edad y sexo 2020

Instrucciones:
Se muestra una gráfica con preguntas detonadoras para que junto con tus compañeros obtengan sus posibles conclusiones.

Descarga el archivo adjunto, dando clic en la imagen, imprímelo y realiza el ejercicio.



Actividad 9.

Variación porcentual anual del Producto Interno Bruto Turístico Trimestral

Instrucciones:
Se muestra una gráfica con preguntas detonadoras para que junto con tus compañeros obtengan sus posibles conclusiones.

Descarga el archivo adjunto, dando clic en la imagen, imprímelo y realiza el ejercicio.

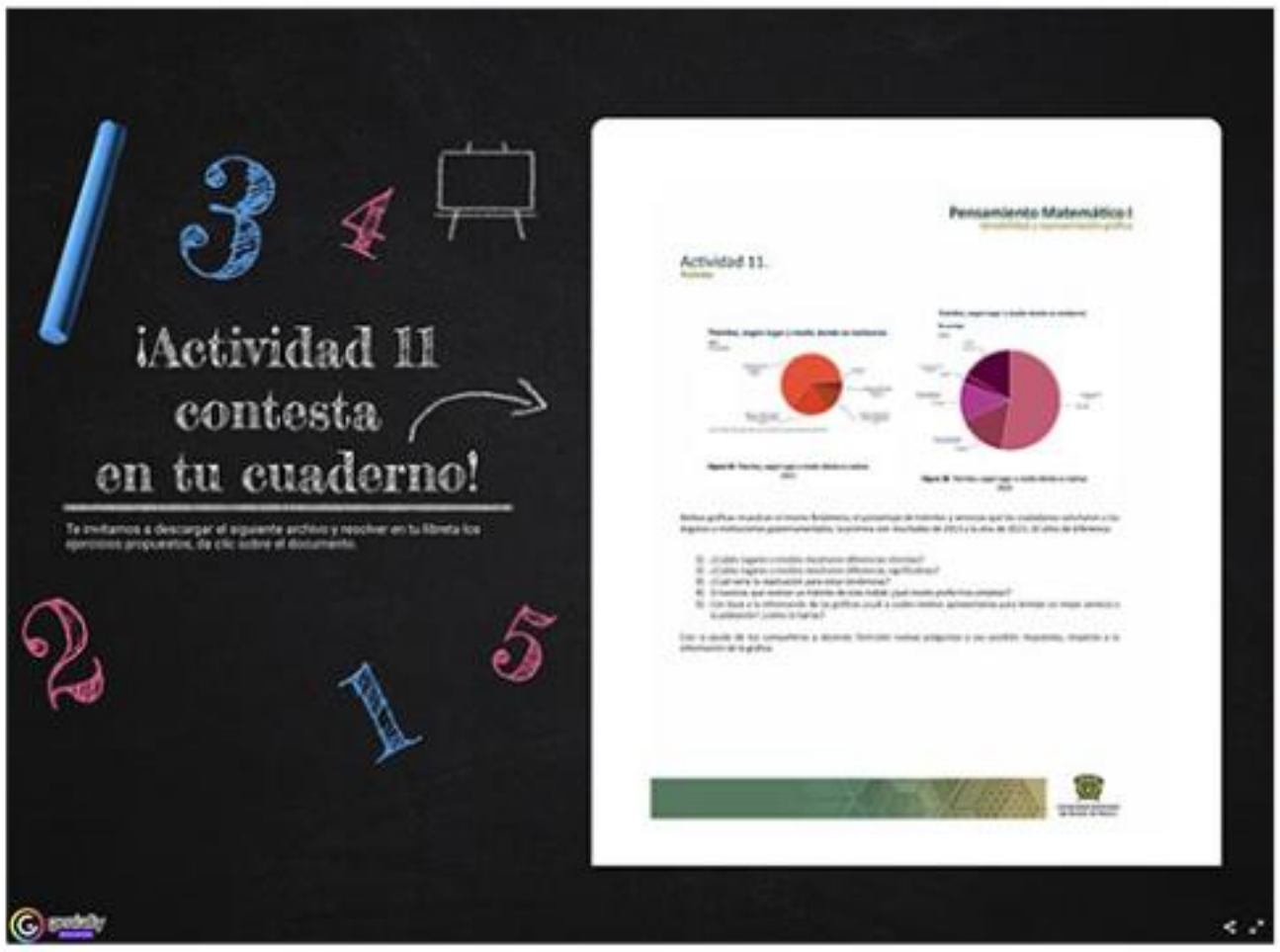


Actividad 10.

Población Total y tasa de crecimiento promedio anual

Instrucciones:
Se muestra una gráfica con preguntas detonadoras para que junto con tus compañeros obtengan sus posibles conclusiones (INEGI, 2017).

Descarga el archivo adjunto, dando clic en la imagen, imprímelo y realiza el ejercicio.

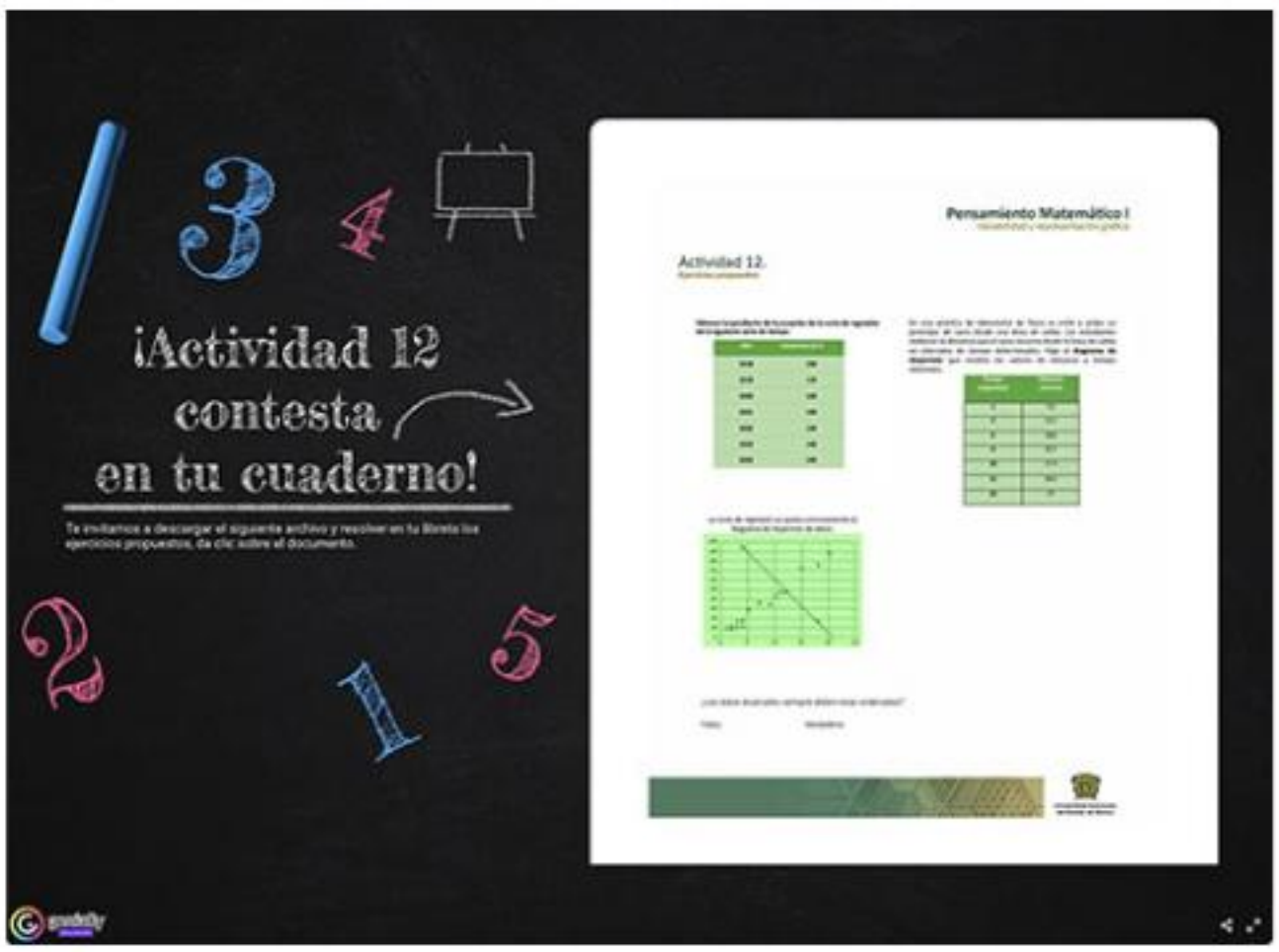


Actividad 11.

Trámites

Instrucciones:
Se muestra dos gráficas con preguntas detonadoras para que junto con tus compañeros obtengan sus posibles conclusiones (INEGI, 2017).

Descarga el archivo adjunto, dando clic en la imagen, imprímelo y realiza el ejercicio.



Actividad 12.

Ejercicios propuestos

Instrucciones:
Te invitamos a descargar el siguiente archivo y resolver en tu libreta los ejercicios propuestos.

Descarga el archivo adjunto, dando clic en la imagen, imprímelo y realiza el ejercicio.

Resumen del Módulo II

Hemos terminado lo referente a variabilidad y representación gráfica, comenzamos identificando que los datos estadísticos pueden ser variables cuantitativas o cualitativas y que son la base para crear modelos matemáticos que permiten realizar pronósticos o bien evaluar riesgos y lograr una acertada toma de decisiones en el análisis de una situación en particular.

Se buscó valorar la importancia de los datos estadísticos y se analizaron diferentes tipos de gráficas en situaciones del entorno detonando un análisis más exhaustivo que permitió realizar conclusiones propias, identificando cuando es conveniente realizar por ejemplo una gráfica de barras o un circulograma.
Los temas aquí expuestos y los ejemplos analizados mostraron las relaciones existentes entre dos variables, así como su representación gráfica e interpretación, se aprendió sobre el coeficiente de correlación y la recta de regresión como herramientas importantes que ayudan a definir la relación existente entre las variables, de manera que sustenten las decisiones futuras.

Sigamos adelante, en el siguiente módulo aprenderás sobre los métodos de muestreo y su aplicación en la solución de problemas contextualizados.

Actividad 13.



Cuestionario Autoevaluable

Instrucciones:
Como actividad final, se te pide realizar un cuestionario; para comenzar, revisa en la plataforma **SEDUCA**, la sección Actividades actividad rio Final.

Your desktop is currently shared with archundial@gmail.com Stop Sharing

Glosario

Circulograma: representación gráfica en forma circular, también conocida como gráfica de pastel, que muestra el valor de la frecuencia como un sector circular dentro de un círculo completo. Son útiles para hacer comparaciones entre datos, sobre todo si se expresa en forma de porcentaje.

Coeficiente de correlación: cantidad que permite conocer que tan bien se ajusta a un conjunto de pares ordenados a un comportamiento determinado.

Datos bivariados: se refieren a dos conjuntos de datos.

Diagrama de dispersión: herramienta empleada cuando queremos gráficamente realizar un análisis de datos que involucran a dos conjuntos de datos.

Gráfica de barras: diagrama que se conforman por barras rectangulares, cuya longitud es proporcional a los valores o cantidades representados, son muy útiles si se requiere presentar a una variable en dos situaciones distintas.

Histograma: representación gráfica en forma de barras, que a diferencia con la gráfica de barras en el histograma las barras están unidas, ya que indican continuidad.

Ojiva: representación gráfica de líneas que muestra relaciones entre factores.

Polígono de frecuencias: representación gráfica de líneas que se construye trazando una línea que una los puntos medios de los rectángulos o barras de un histograma.

Recta de regresión: representación gráfica de la regresión lineal que se define como un modelo estadístico que permite observar la tendencia lineal de un conjunto de puntos obtenida de los posibles resultados.

Serie de tiempo: conjunto o una colección de valores o fenómenos observados durante una serie de periodos temporales secuencialmente ordenada.

Variables cualitativas: describe cualidades, circunstancias o características de un objeto o persona que permiten expresar una característica, atributo, cualidad o categoría no numérica.

Variables cualitativas binaria: trabaja con valores específicos de tipo binario.

Variables cualitativas nominales: variable que se representa sin orden ni números, y matemáticamente es menos precisa.

Variables cualitativas ordinaria: conocida también como variable cuasi cuantitativa, es representada por una modalidad que no requiere números, sin embargo, sí de una jerarquía, orden o un puesto.

Variables cuantitativas: son variables que solo pueden expresarse mediante números. Es decir, solo puede tomar un valor numérico.

Variables cuantitativas continuas: esta variable puede asignar un valor cualquiera dentro de un intervalo o medición, es decir, siempre habrá opción de que haya más valores entre dos valores conocidos. Normalmente se representan por valores decimales para ser más específico.

Variables cuantitativas discretas: son aquellas que pueden adquirir un valor de un conjunto numérico dado, es decir no existen valores intermedios entre ellos, dicho de otra manera, es un valor entero.

Referencia Bibliográfica

- CEPAL (2009). Balance preliminar de las economías de América Latina y el Caribe. PP 121. ISBN 9789213233887
- Editorial Grudemi (2021). Media aritmética. Recuperado de Enciclopedia Iberoamericana (https://enciclopediaiberoamericana.com/media-aritmetica/). Última edición: febrero 2021. Consultado el 21 de mayo de 2024.
- INEGI (2017). Presentación de datos estadísticos en cuadros y gráficas : actualización 2017 / Instituto Nacional de Estadística y Geografía.-- México : INEGI, c2017. Disponible en línea:
- https://www.inegi.org.mx/contenidos/productos/prod_serv/contenidos/espanol/bvinegi/productos/nueva_estruc/702825094416.pdf
- Porras A., (s/f). Diplomado en Análisis de Información Geoespacial. CentroGEO. CONACYT. México, CDMX.
- Quiñónez, A. (2012). MATEMÁTICAS. Interpretación de tablas y gráficas para hacer inferencias de la vida cotidiana. Sexto grado del Nivel Primario. Guatemala: Dirección General de Evaluación e Investigación Educativa, Ministerio de Educación. Disponible en red: http://www.mineduc.gob.gt/DIGEDUCA
- Sagaró del Campo, N., & Zamora Matamoros, L. (2019). Técnicas estadísticas para identificar posibles relaciones bivariadas. Rev Cubana Anestesiología y Reanimación, 19(2).
- Recuperado de https://revanestesia.sld.cu/index.php/anestRean/article/view/603/909
- UNIR (2021). Gráficos estadísticos: tipos e importancia en la comunicación de datos. Disponible en red: https://www.unir.net/marketing-comunicacion/revista/graficos-estadisticos/#:~:text=Los%20gr%C3%A1ficos%20estad%C3%ADsticos%20son%20potentes,la%20evoluci%C3%B3n%20de%20distintas%20variables.
- Vargas, p. (2021, febrero). Estadística distribución de frecuencias y representación gráfica. Repositorio.usam.ac.cr



Módulo III

Métodos de Muestreo

Propósito de Módulo

Analiza y aplica las diferentes técnicas de muestreo en la solución de problemas contextualizados.

Introducción

Para nuestro tercer módulo nos centraremos en técnicas de muestreo probabilístico y no probabilístico, el estudio observacional y experimental.

El muestreo es un conjunto de técnicas y procedimientos que permitirán el análisis y la obtención de resultados acerca de un tema propuesto, gracias a la elección de una muestra que extrapolándola permitirá determinar la(s) características de toda la población de interés.

Por otra parte, el estudio observacional facilitará tener un enfoque real y más apropiado del fenómeno por investigar, lo que permitirá reducir el sesgo en los resultados. Los estudios observacionales permiten realizar análisis que otros métodos de investigación no permiten llevar a cabo.

Los estudios experimentales hacen posible que el o los investigadores controlen el factor de estudio, es decir, tendrán injerencia sobre dosis administradas, frecuencia de aplicación, duración del estudio, etc. Basándose en procedimientos preestablecidos para la investigación.

Actividad 1.



Evaluación Diagnóstica
(conocimientos previos)

Antes de comenzar con la revisión de los contenidos del Módulo 3 de Pensamiento Matemático I, te invitamos a responder la evaluación diagnóstica.

Eso te ayudará a descubrir los conocimientos previos que tienes sobre los temas a abordar y, al mismo tiempo, le permitirá a tu docente conocer tus fortalezas y debilidades.

Para comenzar con el ejercicio, revisa en la plataforma **SEDUCA** la sección Actividades (Estudiante) y responde a la actividad denominada: M3 Examen Diagnóstico.

3.1 Métodos de muestreo

A lo largo del libro se ha mencionado la importancia de la recolección de datos en las investigaciones científicas y en los sectores sociales para conocer de manera confiable y eficiente la información recabada. En algunos casos es casi imposible estudiar el total de la población, por lo que la investigación se realiza a solo una muestra de la población, la cual ya se ha estudiado con anterioridad; sin embargo, la elección de la muestra también es relevante.

Definición

En la estadística, el muestreo es un conjunto de técnicas que permiten el análisis y la obtención de conclusiones acerca de un determinado tema, gracias a la elección de la muestra. Además, las conclusiones pueden extrapolar o inferir de la población sobre ese tema.

En una investigación, las técnicas de muestreo permiten determinar los criterios de selección de los elementos a estudiar; se requiere establecer el tamaño de la muestra.

También es necesario considerar, los recursos humanos, de tiempo y económicos para llevar a buen término el estudio.



Figura 1. El muestreo

Existen dos métodos de muestreo: uno es aleatorio o probabilístico y el otro es no aleatorio o no probabilístico.

3.1.1 Probabilístico

Definición

Este tipo de muestreo permite diseñar un grupo muestra de manera aleatoria, reconoce en cada individuo iguales probabilidades de ser seleccionados, de manera que se tenga un enfoque confiable y riguroso, desde el punto de vista metodológico, se considera exento de cualquier sesgo que pueda comprometer los resultados del estudio. **El objetivo común es garantizar que la muestra sea representativa de la población total.**

Las técnicas de muestreo probabilístico se presentan en el siguiente esquema:

Figura 2. Técnicas de muestreo probabilístico.

SIMPLE



- Todos los individuos de la población tienen la misma probabilidad de ser seleccionados.
- Este método es muy sencillo y es el más utilizado.
- Es el menos eficaz en poblaciones muy extensas, porque el tamaño de la muestra es muy grande.

SISTEMÁTICO



- Se enumeran todos los individuos de la población, luego se elige el primero de la muestra de manera aleatoria y los siguientes de manera periódica.
- Pueden darse casos en los que en la población haya ciclos o periodicidades y que al utilizar este método haya una elevada homogeneidad en la muestra.
- Esta muestra puede que no sea representativa de la población.

ESTRATIFICADO



- Los individuos de la población se agrupan en estratos, en función de características esenciales de interés para el estudio a realizar.
- Se seleccionan los individuos de la muestra de cada uno de los estratos de manera aleatoria.
- El número de individuos de la muestra pertenecientes a cada estrato es proporcional al número de individuos de cada uno de ellos con respecto a la población total.
- Aumenta la representatividad de la muestra los individuos por lo cual los resultdos son más confiables.

CONGLOMERADOS



- Las poblaciones están formadas generalmente por determinados grupos o conglomerados.
- Esta técnica no selecciona a los individuos de manera directa; sino que se escogen de manera aleatoria algunos de los grupos o conglomerados.

Ejemplo 2. Técnicas de muestreo no probabilístico

Figura 3. Ejemplos de técnicas de muestreo probabilístico.

SIMPLE



Si tienes una lista de 1000 estudiantes y quieres seleccionar una muestra de 100, puedes usar un generador de números aleatorios para seleccionarlos y así cada uno tendrá una probabilidad de 1/1000 de ser seleccionado.

Ejemplo 2. Técnicas de muestreo no probabilístico

SISTEMÁTICO



Si tienes una lista de 1000 empleados y deseas seleccionar 100, puedes elegir un número aleatorio como punto de partida. Por ejemplo, el número 5, y luego seleccionar cada décimo empleado en la lista (5, 15, 25, etc.).

ESTRATIFICADO



Si una empresa tiene 1000 empleados y se dividen en tres departamentos (administración, ventas, producción), puedes realizar un muestreo aleatorio simple en cada uno de ellos.. Si el departamento de administración tiene 200 empleados, el de ventas 300 y el de producción 500, puedes seleccionar 20 del departamento de administración, 30 del de ventas y 50 del de producción.

CONGLOMERADOS



Si quieres estudiar el rendimiento académico de los estudiantes de una ciudad y en ella hay 50 escuelas, podrías seleccionar aleatoriamente 10 y estudiar a todos los estudiantes en esas escuelas.

3.1.2 No probabilístico

Definición

Son técnicas de muestreo en las que se establecen determinados criterios que no están basados en la equiprobabilidad, es decir, no todos los individuos tienen la misma oportunidad de ser seleccionados para la muestra. Estos métodos de muestreo son menos confiables que los aleatorios o probabilísticos, porque su nivel de representatividad de la población es menor y se utilizan cuando no es posible la realización de los otros debido a restricciones o cuando se está realizando una exploración inicial y se quiere obtener cierto grado de representatividad.

El **muestreo no probabilístico** se utiliza en estudios en los que no es posible extraer un muestreo de probabilidad aleatorio por limitaciones de tiempo y costo. Además, estos métodos pueden no proporcionar una muestra representativa de la población en general, pueden ser útiles en situaciones donde el muestreo probabilístico no es factible.

Figura 3. El muestreo



Las técnicas de muestreo no probabilístico se presentan en el siguiente esquema:

Figura 4. Técnicas de muestreo no probabilístico.

CUOTAS

- Para seleccionar los elementos de la muestra se fijan una serie de características o condiciones representativas del estudio o análisis a realizar.
- Se selecciona un determinado número de elementos que las cumplen, llamado tamaño de la muestra.
- Es necesario conocer a fondo la población objeto del estudio para seleccionar las variables o características adecuadas.
- Es utilizado en las encuestas de opinión.

INTENCIONAL

- Se determinan o eligen intencionadamente los elementos de la muestra o de grupos típicos de la población, ya que se entiende que la muestra será representativa de la población.
- Implica un gran conocimiento de las características esenciales de la población para la elección, se considera la técnica más empleada en las encuestas de intención de voto.
- Es utilizada en la investigación cualitativa.

CADENA

- Se seleccionan unos determinados elementos de la población y estos están en contacto o relacionados con otros elementos que se van agregando a la muestra hasta completar su tamaño.
- Se usan, en estudios de grupos sesgados o marginales, como relacionados con delincuencia, enfermedades, entre otros.

DISCRECIONAL

- Los elementos son escogidos deliberadamente por el responsable del estudio, ya que son los más adecuados para el objeto de investigación.
- Se utiliza cuando no existe mucho tiempo para llevar a cabo la investigación.

Ejemplo 2. Técnicas de muestreo no probabilístico

Figura 5. Ejemplos de técnicas de muestreo no probabilístico.

CUOTAS



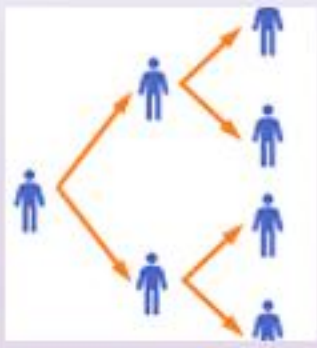
Un investigador desea estudiar las opiniones sobre un producto, para lograr su objetivo decide encuestar a 50 jóvenes, 50 adultos y 50 ancianos. Los individuos de cada grupo o rango de edad los seleccionó por conveniencia hasta alcanzar la cuota.

INTENCIONAL



Se quiere estudiar el impacto de una nueva técnica de enseñanza en estudiantes talentosos, el investigador puede seleccionar a los estudiantes que, según los profesores, tienen más talento en la materia.

CADENA



Para estudiar a una población difícil de alcanzar, como usuarios de drogas ilegales, el investigador podría comenzar por entrevistar a unos pocos usuarios conocidos, y luego pedirles que recomienden a otros usuarios que estén dispuestos a participar.

DISCRECIONAL



Si un investigador quiere estudiar el comportamiento de compra de los consumidores promedio en una tienda de ropa, puede seleccionar a individuos que aparenten ser compradores típicos; en lugar de seleccionar a individuos al azar.

Para concluir es necesario mencionar que las técnicas de muestreo y manejo de la muestra permiten obtener resultados satisfactorios de estudios, investigaciones científicas y encuestas de todo tipo según su fin u objetivo. Así la metodología aplicada en cada caso tiene que reflejar el compromiso y la responsabilidad social que se esperan de un trabajo serio.



Actividad 2.



Métodos de muestreo

Instrucciones:
te invitamos a resolver el siguiente juego referente a los métodos de muestreo, para iniciar da clic en la **imagen.**

3.2 Estudio observacional

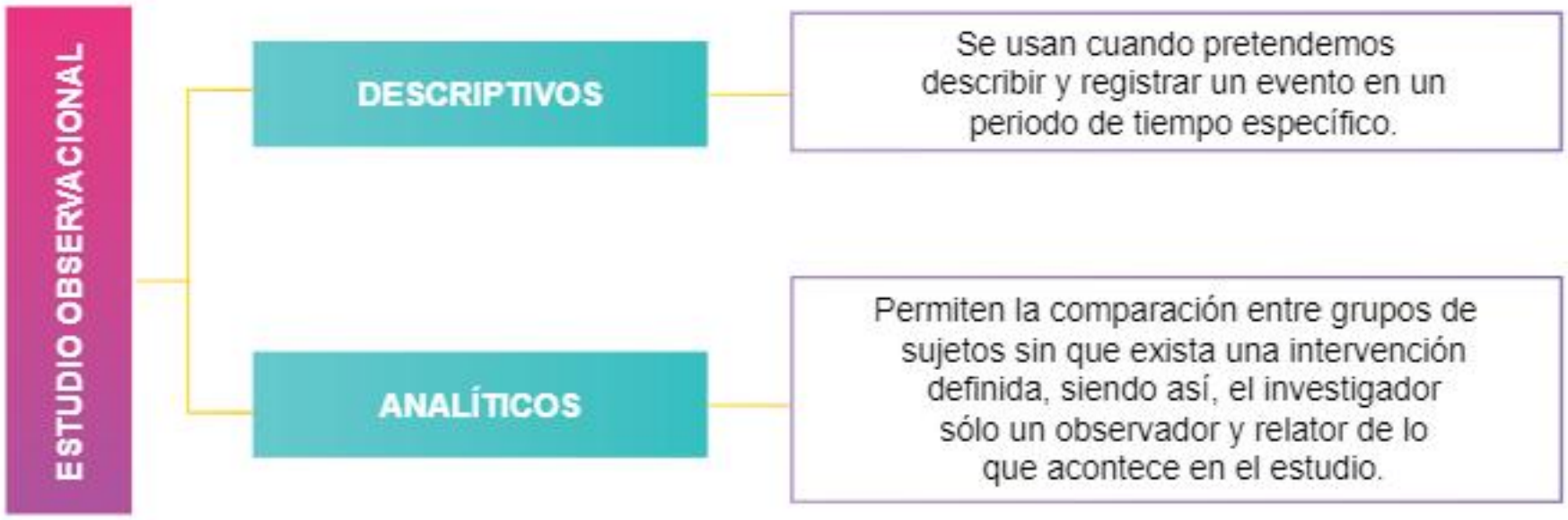
Definición

Un estudio observacional (EO) se refiere a diseños de búsqueda, se realiza con el objeto de "observar y registrar" eventos sin intervenir en el curso natural de estos. Se le considera de carácter estadístico o demográfico, las mediciones se pueden realizar a lo largo del tiempo de forma prospectiva, retrospectiva o única.

Un EO por su naturaleza misma será difícil o imposible de reproducir por un observador diferente, lo que lo hace único.

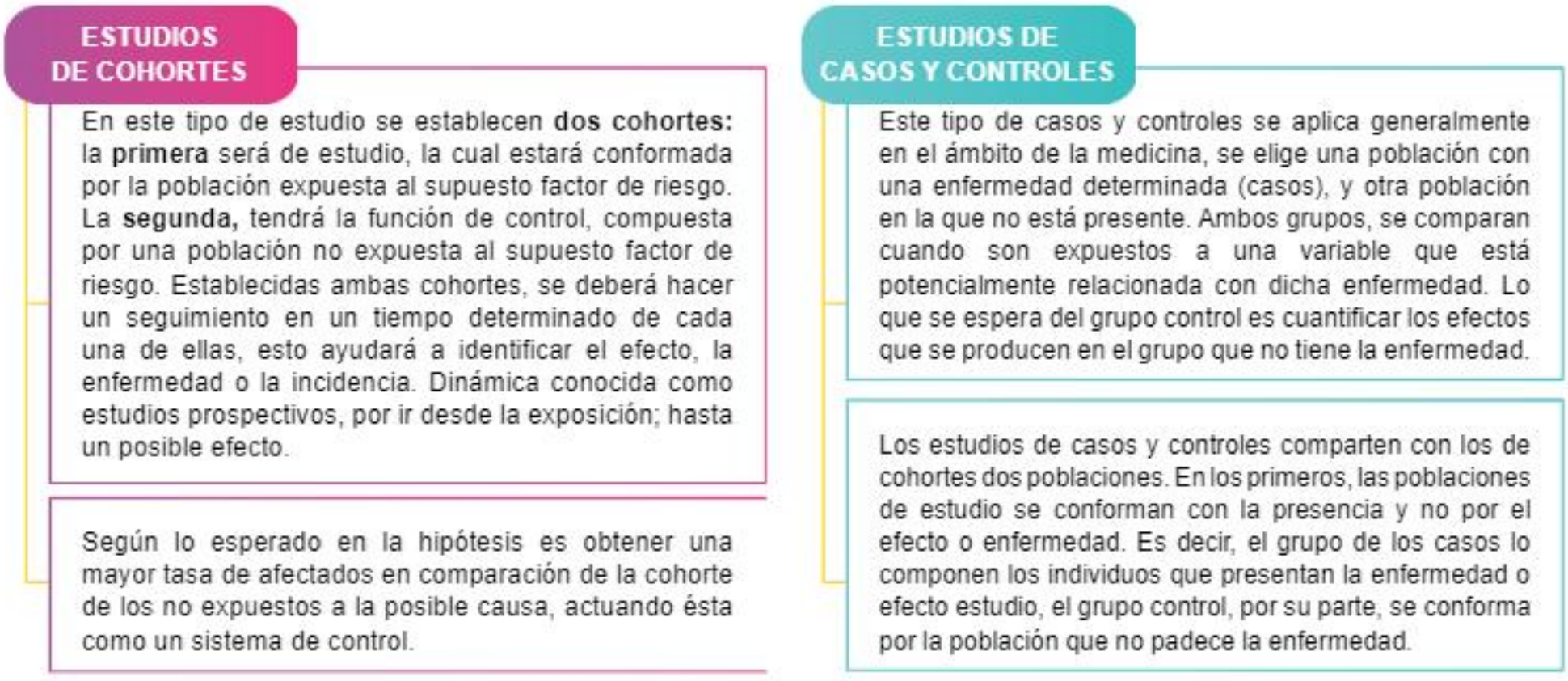
Normalmente se clasifican en dos tipos:

Figura 6. Clasificación del estudio observacional.



El estudio observacional analítico a su vez, se divide en dos tipos:

Figura 7. Clasificación del estudio observacional analítico.



En resumen, en un estudio de Casos y Controles lo que se considera es si con el factor de riesgo en la población enferma, la frecuencia del factor de riesgo es superior a la observada en nuestra población de control no enferma.

En los estudios de Cohortes se inicia desde la exposición al factor de riesgo para llegar a la enfermedad y en los estudios de Casos y Controles se parte de la enfermedad para intentar establecer una relación con el factor de riesgo.

Los estudios de Cohortes son estudios más fiables con un costo mayor que los estudios de Casos y Controles. En contraparte los estudios de Casos y Controles son más rápidos y económicos con la desventaja que deben establecerse los mecanismos adecuados para controlar los sesgos.

Figura 8. Clasificación del estudio observacional analítico.



Como todo tipo de diseño para obtener información, el EO tienen fortalezas y debilidades.

Fortalezas:
- Proporciona un enfoque real y pertinente del fenómeno.
- Reduce el sesgo del investigador.
- Facilitan realizar estudios que otros métodos de investigación no permiten llevar a cabo.
- Comprensión contextual: Permiten comprender el contexto de los comportamientos, esclareciendo el "¿por qué?" y el "¿cómo?" de ciertas acciones.

Debilidades:
- Proporcionan "un nivel de evidencia inferior" al de los estudios experimentales porque son susceptibles al sesgo y a la interpretación personal de los datos proporcionados.
- No pueden utilizarse para demostrar causa-efecto.
- Difícil o imposible de reproducir por un observador diferente.

Diseño de estudios observacionales

Para poder establecer el diseño de un estudio observacional se sugiere:

- Definir de forma clara el problema a investigar, estableciendo los objetivos a lograr,
- tener claros los criterios de selección para el cálculo de la muestra,

- establecer el tiempo de seguimiento,
- definir el efecto por estudiar y la recolección de datos,
- al igual, el análisis e interpretación de resultados,
- las conclusiones y
- las consideraciones éticas del estudio.

También se requiere considerar la **finalidad descriptiva o analítica** de los estudios observacionales, al respecto se dice que un estudio se considera analítico cuando su finalidad es evaluar una posible relación causal entre un factor y un efecto o resultado. (p. ej. El tabaco se cree que puede ocasionar diversas enfermedades).

Un **estudio** se considera **descriptivo** cuando no busca evaluar una presunta relación causa-efecto, sólo la información se utiliza para fines netamente descriptivos. Se emplea para generar hipótesis que serán contrastadas y suele ser útil para generar hipótesis que deberán contrastarse posteriormente con estudios analíticos.



La secuencia temporal de los estudios observacionales, puede ser:

Transversal
Los estudios se consideran transversales cuando los datos de cada sujeto representan específicamente un momento del tiempo, o bien donde se muestra la relación entre diferentes variables en un tiempo determinado. Al ser registradas las variables de forma simultánea, no puede dictaminarse que haya una relación temporal entre ellas, por lo cual, no permiten abordar el estudio en una relación causa-efecto; por naturaleza son descriptivos.

Longitudinal
En este tipo de estudio existe un lapso entre las distintas **variables** a evaluar para establecer una secuencia temporal entre ellas. Estas pueden ser de tipo descriptivo o analítico. Las variables analíticas toman en cuenta la dirección temporal, que puede ir de la causa al efecto (estudios experimentales y estudios de cohortes) o bien desde el efecto a la causa (estudios de casos y controles).

En resumen, un estudio es longitudinal si las observaciones corresponden a dos momentos en el tiempo, aún cuando la recopilación de información se realice de forma simultánea y considerando una secuencia temporal entre ellas.

Ejemplo 2. Estudios observacionales

- El recuento de plaquetas en la sangre: es una observación realizada en la medicina y que determina un recuento de plaquetas por microlitro de sangre. Un nivel normal es entre 150,000 y 400,000 plaquetas por microlitro de sangre en un paciente.

- El registro del comportamiento de la mariposa monarca: se aplica para la Zoología y Biología, se pretende observar a la mariposa monarca en su hábitat natural para determinar el recorrido promedio diario, el tipo de sociedad que conforman y su comportamiento social, alimentación, etc.

- La medición del crecimiento económico: es una observación que se utiliza en las ciencias económicas y permite monitorear el incremento del producto interno bruto (PIB) de un país.

- La medición de la temperatura ambiente: utilizada en meteorología y que, a través de distintos instrumentos de medición, permite registrar el estado del tiempo y la sensación térmica.

- El registro de las preferencias en el consumo de productos: observación que se realiza en publicidad y marketing para obtener información sobre los productos de mayor demanda por los consumidores.

- Un estudio investiga la posible relación causal entre el consumo de tabaco y la hipertensión arterial: para ello, se identifican enfermos con hipertensión arterial y un grupo de pacientes sin el padecimiento y en dicho momento se les cuestiona sobre sus antecedentes de consumo de tabaco. Este estudio es analítico porque evalúa una presunta relación causal y es observacional porque no se controla el factor de estudio. A pesar de que la información sobre el efecto y la supuesta causa se recogen en un mismo momento, se considera que dicha información es con base a un tiempo anterior a la enfermedad, por lo que este estudio puede clasificarse, como longitudinal, ya que la dirección del estudio es de efecto a causa y se considera como un estudio de casos y controles.



Actividad 3.

Estudio observacional

Instrucciones:
te invitamos a resolver el siguiente juego referente a los estudios observacionales, para iniciar da clic en la **imagen.**

3.3 Estudio experimental

Los fenómenos naturales y sociales suelen involucrar un gran número de variables en su entorno, lo que los convierte en situaciones complejas para su estudio; tal como menciona Miguel Ángel Sabadell (2023) Astrofísico y divulgador científico respecto a la predicción del clima:

> Por desgracia, la predicción del tiempo no es tan sencilla como mirar la Luna; de hecho, resulta verdaderamente complicado. Ahora bien, no pensemos que no conocemos las leyes físicas que lo gobiernan; nada de eso, son bien conocidas, pero las ecuaciones que las traducen al lenguaje matemático son muy complejas y solo se pueden obtener soluciones exactas en un puñado de casos tan simples que no sirven para mucho. Los matemáticos conocen bien estas ecuaciones, que han bautizado con el enigmático nombre de **ecuaciones diferenciales no lineales en derivadas parciales.**

Razones como estas impiden que hoy exista una solución formal o científica a muchos de los fenómenos que acontecen en el día a día, por lo que, la comunidad científica suele darse a la tarea de investigar y estudiar fenómenos donde se involucren apenas unas cuantas variables. Si bien es cierto que también puede ser imposible eliminar otras variables, se toman las precauciones necesarias para identificar los cambios o relaciones en las variables de interés del fenómeno de estudio.

Definición

Un estudio experimental es aquel en el que el investigador manipula a la variable de interés (variable independiente) y comprueba el efecto que tiene sobre otra (variable dependiente) (Palacio, 2019).

Respecto a los estudios experimentales Saavedra (2016) indica:

> En estos tipos de estudio el investigador manipula las condiciones de la investigación, es por esto que son reconocidos como los más fuertes (mayor nivel de evidencia), por definición, son prospectivos. De todos los diseños en investigación, el estudio experimental es el que permite acercarse a la noción de causalidad de manera más directa.

Además, entre otras características, a diferencia de los estudios observacionales, en los que los investigadores se limitan a identificar el comportamiento natural de un evento o fenómeno sin alterarlo, el estudio experimental requiere la intervención del experimentador; esto es porque en este tipo de estudio, para reafirmar el efecto que tiene una variable sobre otra, la población o muestra se divide aleatoriamente en dos grupos:

Grupo experimental:

"Es aquel que es expuesto a la condición, variable, o estímulo experimental" (Adis, s/f). Es el grupo en el que se aplica la acción de la variable independiente para medir el efecto o consecuencia de la variable dependiente.

Grupo control:

"Es aquel que es utilizado para propósitos comparativos, no siendo expuesto a la condición, variable, o estímulo experimental" (Adis, s/f).

Su función consiste en verificar que la variable independiente realmente produce el cambio esperado de la variable dependiente; es decir, existe una causalidad de la acción y consecuencia, con ello, se evita la ambigüedad, el azar, las coincidencias u otros factores que pudieran intervenir en una consecuencia similar.

Para que el estudio experimental sea confiable y consistente con los resultados se requiere que (Zurita-Cruz, 2018):

- La población o muestra debe dividirse de forma estrictamente aleatoria en los grupos de control y experimental.
- Ningún integrante de cada grupo debe saber a qué grupo pertenece (experimento ciego).
- Para verificar la eficacia de un estudio experimental debe existir una diferencia significativa del grupo experimental respecto al grupo control conforme a las variables del experimento.

Los estudios experimentales suelen ser empleados en el área de la salud para someter a pruebas científicas el uso de nuevos medicamentos, tratamientos o intervenciones médicas; sin embargo, dado que los individuos deben saber que son parte de un experimento, no pueden ser elegidos de forma aleatoria, tampoco pueden ser coaccionados, además de que deben firmar un consentimiento expreso en el que se especifican los detalles y posibles consecuencias del estudio, por estas razones, quiénes participan en un estudio experimental son generalmente voluntarios.

Cabe resaltar que los individuos involucrados en estudios experimentales en el ámbito de la Medicina son conscientes de que participan en un experimento; es decir, la muestra no es una representación aleatoria de la población, por lo que sus efectos no pueden describir el comportamiento general de la población, sino que solamente describe los efectos en la muestra en estudio.

Analiza los siguientes ejemplos, identifica los casos en los que se trata de un estudio experimental y contrasta las características de los que no lo son:

Ejemplo 4. Identificar la relación que existe entre la cantidad de huevos que puede producir una gallina al escuchar música clásica.

Un grupo de investigadores decidió identificar la relación que existe entre la cantidad de huevos que puede producir una gallina al escuchar música clásica. De un total de 80 gallinas eligieron de forma aleatoria a 40 de ellas como grupo de control y lo colocaron en un gallinero. Al resto lo definieron como grupo experimental y lo colocaron en un gallinero independiente y separado del primero. El grupo control continuó su desarrollo como lo hacía habitualmente, el grupo experimental tuvo la única diferencia de que se les colocó música clásica en ciertos momentos del día, sobre todo a la hora de las ovoposiciones. Se contó la cantidad de huevos que colocaba cada grupo antes, durante y después del experimento. Se observó que el grupo experimental tenía un conteo 20% superior de huevos que los contados en el grupo control.

Sí, se trata de un estudio experimental correctamente aplicado, esto debido a que:

- la muestra se dividió en un grupo experimental y uno de control,
- las gallinas no sabían a qué grupo pertenecían,
- cada grupo fue sometido de forma independiente a las variables de estudio y
- el conteo de los huevos permitió identificar si existía mayor, menor o nula diferencia en la producción entre un grupo y otro.

Se concluyó que, al escuchar música clásica, las gallinas podían poner más huevos que sin ella.

Descargo de responsabilidad
Los siguientes ejemplos son ficticios, con fines educativos del área de matemáticas, en ningún caso representan ni se fundamentan en casos reales, ni de conocimientos técnicos del área de la salud o de ninguna otra área del conocimiento. No deben ser considerados como referencias a problemas de salud ni de ninguna otra índole, ni a sus tratamientos o curas o similares.

Ejemplo 5. Verificación de la eficacia de un nuevo fármaco para reducir los niveles de azúcar en la sangre.

Un grupo de investigadores desean verificar la eficacia de un nuevo fármaco para reducir los niveles de azúcar en la sangre. De una muestra de 100 personas con diabetes Tipo I, seleccionaron de forma aleatoria a la mitad para pertenecer al grupo experimental y los otros 50 al grupo control. El grupo experimental recibió una dosis del fármaco experimental, mientras que el grupo control recibió una dosis de placebo. Ningún individuo tenía conocimiento de si tomaba el fármaco o el placebo. Ambos grupos fueron sometidos a estudios clínicos antes, durante y después del estudio experimental. Los resultados mostraron que tanto el grupo control como el experimental mostraban niveles similares de azúcar en la sangre.

Sí se trata de un estudio experimental correctamente aplicado porque:

- la muestra se dividió en un grupo experimental y uno de control,
- ningún individuo sabía a qué grupo pertenecía,
- cada grupo fue sometido de forma independiente a la administración del fármaco o al placebo y
- los estudios clínicos permitieron medir los niveles de sangre a lo largo del estudio experimental.

Si bien es cierto el resultado mostró que no existía diferencia significativa de la eficacia del fármaco en ambos grupos, el estudio experimental permite concluir que el fármaco no cumple con el beneficio esperado.

Ejemplo 6. Saber la eficiencia que tiene el uso de aplicaciones móviles para el aprendizaje de ecuaciones lineales en la clase de matemáticas.

En un estudio experimental se desea saber la eficiencia que tiene el uso de aplicaciones móviles para el aprendizaje de ecuaciones lineales en la clase de matemáticas, por lo que los investigadores eligieron como grupo experimental a los estudiantes que tenían algún dispositivo móvil con datos, mientras que el grupo control se conformó con aquellos que no tenían datos o no tenían dispositivo móvil. El grupo control tomó clase en el aula con un docente como mediador del aprendizaje, mientras que el grupo experimental tomó las clases mediante las aplicaciones móviles. Al final del experimento se preguntó a los estudiantes cuánto tiempo dedicaron al uso y aprendizaje en las aplicaciones móviles, encontrando que en promedio la empleaban 5 horas a la semana.

No se trata de un estudio experimental correctamente aplicado porque:

- si bien es cierto que la población fue distribuida en un grupo experimental y uno de control, esta elección NO fue aleatoria, por lo que NO se puede establecer una conclusión de causalidad.
- por otra parte, los miembros de cada grupo sabían a cuál de los grupos pertenecían, lo cual no debería ocurrir, ya que no es un experimento ciego y
- preguntar cuántas horas dedicaban al uso de las aplicaciones móviles, NO permite identificar su eficiencia en el aprendizaje de ecuaciones lineales y los resultados no corresponden a la relación planteada en el estudio experimental.

En conclusión, el estudio experimental fue aplicado de forma incorrecta, lo cual no permite establecer ninguna relación de causalidad entre el uso de aplicaciones móviles y el aprendizaje de ecuaciones lineales en la clase de matemáticas.

Ejemplo 7. Determinar el efecto que tiene tomar café por las noches y la dificultad que tienen las personas para dormir.

Mediante una encuesta, un grupo de investigadores desea determinar el efecto que tiene tomar café por las noches y la dificultad que tienen las personas para dormir, por lo que realizaron llamadas telefónicas a 200 personas. Se denominó grupo control a quienes contestaron que tomaban café por las noches y podían dormir sin problema, mientras que se denominó **grupo experimental** a quienes tomaban café por las noches y mencionaron que tenían dificultades para dormir. Los investigadores observaron en sus resultados que el grupo experimental resultó ser 8% mayor que el **grupo control.**

No se trata de un estudio experimental correctamente aplicado porque:

- la elección del grupo control y del grupo experimental fue incorrecta, porque no depende de que cumplan o no con alguna variable del experimento, la selección debe ser aleatoria, por lo que NO se puede establecer una conclusión de causalidad; además

- no existe prueba ni evidencia alguna de que el café haya sido la causa de dormir o de la falta de sueño.

Es incorrecto concluir que tomar café por las noches es causa de la dificultad que tienen las personas para dormir, no se realizaron pruebas de rigor médico o científico en ninguno de los grupos que demostraran que tomar café fuera la causa de dormir bien o de la falta de sueño.

Ejemplo 8. Determinar si un trabajador es más eficiente al desayunar antes de ir al trabajo en comparación de uno que decide hacerlo o no.

Un estudio reciente mostró que las personas que no consumían desayuno antes de ir a trabajar eran menos eficientes en sus actividades diarias. El estudio implicó analizar los datos de 2500 empleados de cuatro empresas diferentes, dos de esas empresas fueron asignadas como grupo experimental y las otras dos como grupo control. Al grupo control se le exigió desayunar antes de llegar al trabajo, mientras que el grupo experimental era libre de decidir si desayunaba o no. Los resultados mostraron que las personas que desayunaban antes de ir al trabajo eran 15% más eficientes que las que no lo hacían, independientemente de la empresa en la que trabajaran.

No se trata de un estudio experimental correctamente aplicado porque:

- la elección del grupo control y del grupo experimental fue incorrecta porque las características de una empresa no son necesariamente iguales a las de otra, por tal situación los grupos pueden no tener características homogéneas.
- Los resultados muestran que las personas que desayunaban antes de ir al trabajo eran 15% más eficientes que las que no lo hacían, sin embargo, al mencionar que esto ocurre independientemente de la empresa en la que trabajan, no se puede establecer una relación de causalidad, a lo mucho se podría establecer que existe una relación de esas variables; pero no necesariamente de causa y efecto.

En conclusión, no se puede establecer una relación de causalidad entre consumir el desayuno y ser más eficiente en el trabajo.

Actividad 4.



Estudio experimental

Instrucciones:
te invitamos a **descargar** el archivo adjunto, dando clic en la imagen, imprímelo y realiza el ejercicio.





Actividad 5.



Repaso de conceptos

Instrucciones:
ya casi hemos llegado al final del módulo III, donde has aprendido sobre los diferentes métodos de muestreo, antes de pasar al siguiente tema mide tu aprendizaje y para ello te invitamos a resolver el siguiente ejercicio, da clic en la **imagen.**

3.4 Problemas de aplicación

Ejemplo 1. Programa de capacitación

Un investigador está estudiando los efectos de un programa de capacitación específica en el rendimiento de los empleados, selecciona a 20 empleados que han participado en el programa porque considera que tienen el conocimiento y la experiencia necesarios para proporcionar información relevante.

Preguntas:

a. ¿Qué tipo de muestreo se lleva a cabo? Muestreo de cuotas

b. ¿En qué se basa la selección de los empleados en este método? En que todos participaron del programa de capacitación

c. ¿Qué ventajas tiene este tipo de muestreo para estudios cualitativos? Es sencillo y rápido, ya que utiliza la participación de una cuota para la creación de las muestras

d. ¿Qué limitaciones puede tener este método? El no incluir una cuota que sea relevante para el estudio, podría ocasionar que los resultados se distorsionen

Ejemplo 2. Comportamiento de una subcultura urbana

Un investigador está estudiando el comportamiento de una subcultura urbana que es difícil de acceder, comienza entrevistando a unos pocos miembros conocidos de esta subcultura y luego les pide que recomienden a otros miembros para participar en el estudio.

Preguntas:

a. ¿Qué tipo de muestreo se lleva a cabo? Muestreo de cadena

b. ¿Cómo se expande la muestra en este método? Por recomendación

c. ¿Cuáles son las ventajas de usar el muestreo bola de nieve? Permite al investigador acceder a poblaciones difíciles de forma simple y rentable

d. ¿Qué problemas puede presentar este método en términos de representatividad y sesgo? La representatividad de la muestra no está garantizada. El investigador no tiene ni idea de la verdadera distribución de la población ni de la muestra

Ejemplo 3. Preferencias de los consumidores sobre un nuevo sabor de helado.

Un investigador de mercado quiere conocer las preferencias de los consumidores sobre un nuevo sabor de helado en diferentes partes de la ciudad, selecciona 10 tiendas de helados en distintas áreas de la ciudad basándose en su experiencia y conocimiento sobre cuáles son las más representativas.

Preguntas:

a. ¿En qué se basa la selección de las tiendas en este método? Basada en el juicio y experiencia del investigador

Descargo de responsabilidad

Los siguientes ejemplos son ficticios y con fines educativos del área de matemáticas, en ningún caso representan ni se fundamentan en casos reales ni de conocimientos técnicos del área de la salud o de ninguna otra área del conocimiento. No deben ser considerados como referencias a problemas de salud ni de ninguna otra índole, ni a sus tratamientos o curas o similares.

b. ¿Qué beneficios aporta el conocimiento del investigador en este tipo de muestreo? Puede ser útil cuando el investigador tiene un buen conocimiento del campo y puede identificar tiendas clave que son representativas de la población objetivo

c. ¿Qué desventajas puede tener este método en términos de representatividad? Puede introducir sesgo personal; la muestra puede no ser representativa si el juicio del investigador es incorrecto o sesgado

Ejemplo 4. Sabor favorito de jugo

Un investigador le pregunto a 20 alumnos de la escuela primaria ¿Cuál era su sabor favorito de jugo? de esta encuesta se seleccionó la respuesta de 5 alumnos, quedando seleccionados los alumnos 3,7,11,15 y 19

a. Si el muestreo reconoce en cada individuo iguales probabilidades de ser seleccionado y garantiza que la muestra sea representativa de la población total es: Probabilístico

b. Si el muestreo elige la muestra de manera periódica es: Sistemático

c. ¿Qué ventajas tiene este tipo de muestreo? Garantiza una selección perfectamente equitativa de la población, lo que resulta muy útil en la distinción de grupos

d. ¿Qué limitaciones puede tener este método? El orden en el que se coloca a la población seleccionada puede llegar a tener un tipo de periodicidad oculta, la cual puede coincidir con el intervalo seleccionado, ocasionando una muestra sesgada

Ejemplo 5. Actividades que realizan los estudiantes en el tiempo dedicado al ocio.

En una escuela se hace un estudio para conocer qué tipo de actividades realizan los estudiantes en el tiempo dedicado al ocio. De un total de 1500 estudiantes, 750 son de primer año, 300 de segundo año y 450 de tercer año, se eligen 300 estudiantes, de los cuales 150 son de primer año, 60 de segundo año y 90 de tercer año.

a. Si el muestreo reconoce en cada individuo iguales probabilidades de ser seleccionado y garantiza que la muestra sea representativa de la población total es: Probabilístico

b. Si el muestreo elige la muestra de cada uno de los estratos de manera proporcional: Estratificado

c. ¿Qué ventajas tiene este tipo de muestreo? Se tienen estimaciones más precisas

d. ¿Qué limitaciones puede tener este método? Se ha de conocer la distribución en la población de las variables utilizadas para la estratificación

Ejemplo 6. Evaluación de un nuevo fármaco.

Un estudio con margen de error máximo esperado del 3% se realizó de la siguiente manera:

Una farmacéutica solicitó a 300 voluntarios de la Ciudad de Toluca de Lerdo que asisten de forma regular a una clínica, para que participaran en un estudio sobre la eficacia de un nuevo fármaco experimental cuya función es evitar contagiarse del virus del resfriado común. Se solicitó a los voluntarios que se inscribieran en el grupo experimental o grupo control, dependiendo a su criterio en cuál querían participar. El grupo experimental recibió una dosis del fármaco, mientras que el grupo de control recibió una dosis de placebo. El experimento se llevó a cabo durante 30 días y se obtuvieron los siguientes resultados:

	Total de individuos	Número de personas que se reportaron enfermas de resfriado común
Grupo experimental	150	18
Grupo control	150	19

Tabla 1. Número de personas que se reportaron enfermas de resfriado común.

Describe en detalle si se trata o no de un estudio experimental correctamente aplicado, en cualquier caso, justifica el porqué.

Solución:

No se trata de un estudio experimental por las siguientes razones:

a. No se cumple el margen de muestra mínimo para la población en estudio respecto al margen de error esperado, ya que como lo informa Data México, la población de la Ciudad de Toluca de Lerdo es de 910,608 habitantes, por lo que la muestra debería ser de al menos 1110 habitantes, tal como se describe a continuación:

$$n = \frac{910608}{1 + (910608)\,(0.03)^2}$$

$$n = 1109.75$$

(Se utiliza la fórmula para calcular el tamaño de muestra definida en el tema 1.6 del módulo I)

$$n = \frac{N}{1 + N(e)^2}$$

b. Debido a que la muestra fue elegida solamente de una clínica, podría ser que los individuos no representen a las características de la población en general.

c. Los integrantes del grupo experimental y grupo control NO fueron elegidos de forma aleatoria, ellos eligieron a cuál grupo pertenecían.

d. No se trató de un experimento ciego, ya que los integrantes sabían a qué grupo pertenecían, ya que ellos lo eligieron.

e. La diferencia entre la cantidad de personas enfermas de resfriado común del grupo experimental y del grupo control no es significativa ya que solo existe una persona de diferencia.

f. No existe evidencia de que quienes manifestaron estar enfermos de resfriado común estuvieran diagnosticados de ese padecimiento, ya que existen afecciones con síntomas similares como la influenza y otros problemas de índole respiratorios.

g. Por las razones anteriores, no es posible concluir una relación causal entre el consumo del nuevo fármaco experimental y evitar contagiarse del virus del resfriado común.

Ejemplo 7. ¿Dormir con una mascota influye o no en la calidad del sueño?

Un grupo de investigadores desea identificar si el hecho de dormir con una mascota entre perro o gato influía o no en la calidad del sueño. Para ello eligieron una pequeña población de 5200 personas, en la cual primero averiguaron mediante una encuesta quiénes dormían con perro, gato, ambos o ninguno de ellos. Obteniendo los siguientes resultados:

Tabla 2. Número de personas que duermen con perro, gato, ambos o ninguno de ellos.

Mascota	Frecuencia
Gato, perro o ambos	3264
Ninguno	1936
Total	5200

Tiempo después seleccionaron una muestra aleatoria representativa de cada grupo, a los seleccionados que dormían con gato, perro o ambos se les denominó grupos experimentales; a los que no dormían con ninguna de esas mascotas se les denominó grupo control. A los seleccionados se les comentó que se deseaba conocer la calidad de su sueño por lo que se les ofreció una banda de medición que puede medir la hora de dormir, las horas de sueño profundo, sueño liviano y sueño REM, las veces y horarios en las que se despertaron, la calidad de su respiración, entre otros factores; evaluando de 0 a 100 la calidad de sueño, donde 0 es el resultado menos favorable y 100 el resultado más favorable. A los seleccionados no se les informó que habían sido asignados a un grupo debido a si dormían con mascota o no. Al cabo de 30 días, los investigadores recopilaron todos los datos obtenidos con las bandas, obteniendo los siguientes resultados:

Tabla 3. Resultados obtenidos con la banda de medición para medir la calidad de sueño.

Mascota	Puntaje
Gato, perro o ambos	87
Ninguno	73

Con la información anterior, contesta:

a. ¿Se trata de un estudio experimental?
b. ¿Cuál es la variable independiente del estudio?
c. ¿Cuál es la variable dependiente del estudio?
d. ¿Qué tipo de muestreo se empleó en el experimento?
e. ¿Qué técnica de muestreo fue aplicada?
f. ¿Cuál es la muestra mínima de cada grupo si se desea un error máximo del 5%?
g. ¿Qué evidencia permite determinar la calidad del sueño?
h. ¿El estudio es válido para toda la población?
i. ¿Se puede concluir una relación de causalidad?

Solución:

a. ¿Se trata de un estudio experimental?

Respuesta: Sí, porque las muestras representativas del grupo experimental y del grupo control fueron seleccionadas al azar, además, debido a que el estudio fue hecho en dos momentos, solo los investigadores sabían quiénes pertenecían a cada grupo.

b. ¿Cuál es la variable independiente del estudio?

Respuesta: Dormir con gato, perro o ambos.

c. ¿Cuál es la variable dependiente del estudio?

Respuesta: La calidad del sueño.

d. ¿Qué tipo de muestreo se empleó en el experimento?

Respuesta: Muestreo probabilístico, porque la muestra fue seleccionada al azar.

e. ¿Qué técnica de muestreo fue aplicada?

Respuesta: Muestreo estratificado, porque la población fue dividida en dos estratos o grupos, después, de cada uno de ellos se eligió a una muestra aleatoria representativa.

f. ¿Cuál es la muestra mínima de cada grupo si se desea un error máximo del 5%?

Respuesta:

Grupo experimental

$$n = \frac{3264}{1 + (3264)\,(0.05)^2}$$

$$n = 356.33$$

Grupo control

$$n = \frac{1936}{1 + (1936)\,(0.05)^2}$$

$$n = 331.51$$

(Se utiliza la fórmula para calcular el tamaño de muestra definida en el tema 1.6 del módulo I) $n = \frac{N}{1 + N(e)^2}$

Del grupo experimental se requiere una muestra representativa de 357 individuos.

Del grupo control se requiere una muestra representativa de 332 individuos.

g. ¿Qué evidencia permite determinar la calidad del sueño?

Respuesta: La banda de medición. De lo contrario, la medición mediante una encuesta, test o cuestionario podría no reflejar un resultado verídico de la calidad del sueño, sino más bien una percepción subjetiva de la calidad del sueño.

h. ¿El estudio es válido para toda la población?

Respuesta: Sí, porque los individuos del grupo experimental y grupo control fueron elegidos de forma aleatoria de una muestra representativa de la población.

i. ¿Se puede concluir una relación de causalidad?

Respuesta: Sí, además de que se eligieron de forma aleatoria a los integrantes del grupo experimental y grupo control, existe evidencia objetiva respecto a las variables de estudio debido a las bandas de medición, la cual muestra que dormir con gato, perro o ambos logró un puntaje de 87 puntos respecto a los 73 puntos de aquellos que no duermen con estas mascotas, obteniendo así, una diferencia significativa de 14 puntos en la escala empleada en el estudio. Por lo tanto, el hecho de dormir con una mascota (perro o gato) si influye en la calidad del sueño.

Resumen del Módulo III

En este módulo has aprendido sobre técnicas de muestreo y el correcto manejo de la muestra para obtener resultados satisfactorios en estudios, investigaciones científicas y encuestas de cualquier tipo.

Aprendiste que hay diferentes tipos de muestreo, así como las fortalezas y debilidades de cada de uno de ellos, de manera que ahora cuentas con más herramientas para determinar correctamente los criterios de selección para la realización de un estudio.

Los diferentes ejemplos analizados en este módulo te muestran cómo la metodología aplicada te permite obtener resultados de un tema específico y al extrapolarlos puedes determinar las características de toda la población sujeta de interés, cuidando que el diseño reflejé el compromiso de realizar un trabajo serio como parte de nuestra responsabilidad social.

Valoraste que en el ámbito de la Medicina se emplean tanto estudios observacionales como los estudios experimentales, los primeros con el objeto de comparar grupos de sujetos y los segundos para someter a pruebas científicas el uso de nuevos medicamentos, tratamientos o intervenciones médicas.

Todas las técnicas de muestreo deben ser confiables y consistentes con los resultados obtenidos, de ahí la importancia de aplicar correctamente la metodología de cada una de ellas, estamos seguros de que los temas vistos en este módulo te han proporcionado los elementos necesarios para elegir correctamente la técnica apropiada al estudio que se desea realizar.

Actividad 6.



Ejercicio de autoevaluación

Instrucciones:

Como actividad final, se te pide realizar un cuestionario; para comenzar, revisa en la plataforma **SEDUCA** la sección Actividades (Estudiante) y responde a la actividad denominada: M3 Cuestionario Final.

Glosario

Aleatoriamente: que depende del azar o no sigue una pauta definida.

Coaccionados: fuerza o violencia que se hace a alguien para obligarlo a que diga o ejecute algo.

Consentimiento expreso: consentimiento que una persona otorga mediante una acción que lo expresa clara y de forma determinante. Requiere de registro en el que manifiesta su voluntad de consentir. Lo fundamental en este registro es que el consentimiento sea claro, directo y no genere dudas.

Dosis: toma de medicina que se da al enfermo cada vez.

Ecuaciones diferenciales no lineales en derivadas parciales: tipo de ecuación matemática que resulta del estudio del cálculo diferencial. La no linealidad representa sistemas cuyo comportamiento no es expresable como la suma de los comportamientos de sus causas.

Estudio experimental: es aquel en el que el investigador manipula a la variable de interés (variable independiente) y comprueba el efecto que tiene sobre otra (variable dependiente) (Palacio, 2019).

Estudio observacional: concierne a diseños de búsqueda con el objetivo de "observar y registrar" eventos sin intervenir en el curso natural de estos, es de carácter estadístico o demográfico.

Estudio observacional analítico: permiten la comparación entre grupos de sujetos sin que exista una intervención definida, siendo así, el investigador sólo un observador y relator de lo que acontece en el estudio.

Estudio observacional descriptivo: utilizado cuando pretendemos describir y registrar un evento en un periodo de tiempo específico.

Grupo control: "Es aquel que es utilizado para propósitos comparativos, no siendo expuesto a la condición, variable, o estímulo experimental" (Adis, s/f). Su función consiste en verificar que la variable independiente realmente produce el cambio esperado de la variable dependiente.

Grupo experimental: "Es aquel que es expuesto a la condición, variable, o estímulo experimental" (Adis, s/f). Es el grupo en el que se aplica la acción de la variable independiente para medir el efecto o consecuencia de la variable dependiente.

Placebo: sustancia o un procedimiento que CARECE de una actividad específica objetiva contra la enfermedad o contra la condición en la que se emplea. Incluyen varios tipos de sustancias inertes (que pueden ser ingeridas, inyectadas, inhaladas o aplicadas en la superficie externa del cuerpo) y procedimientos inactivos (cirugías simuladas, intervenciones psicológicas, entre otros).

Prospectivo: conjunto de análisis y estudios realizados con el fin de explorar o de predecir el futuro en una determinada materia.

Muestreo: conjunto de técnicas que permiten el análisis y la obtención de conclusiones acerca de un determinado tema, gracias a la elección de la muestra.

muestreo no probabilístico: son técnicas de muestreo en las que se establecen determinados criterios que no están basados en la equiprobabilidad para la selección de los individuos de la muestra.

Muestreo probabilístico: permite diseñar un grupo muestra de manera aleatoria, reconoce en cada individuo iguales probabilidades de ser seleccionados.

Muestreo probabilístico conglomerado: las poblaciones están formadas generalmente por determinados grupos o conglomerados.

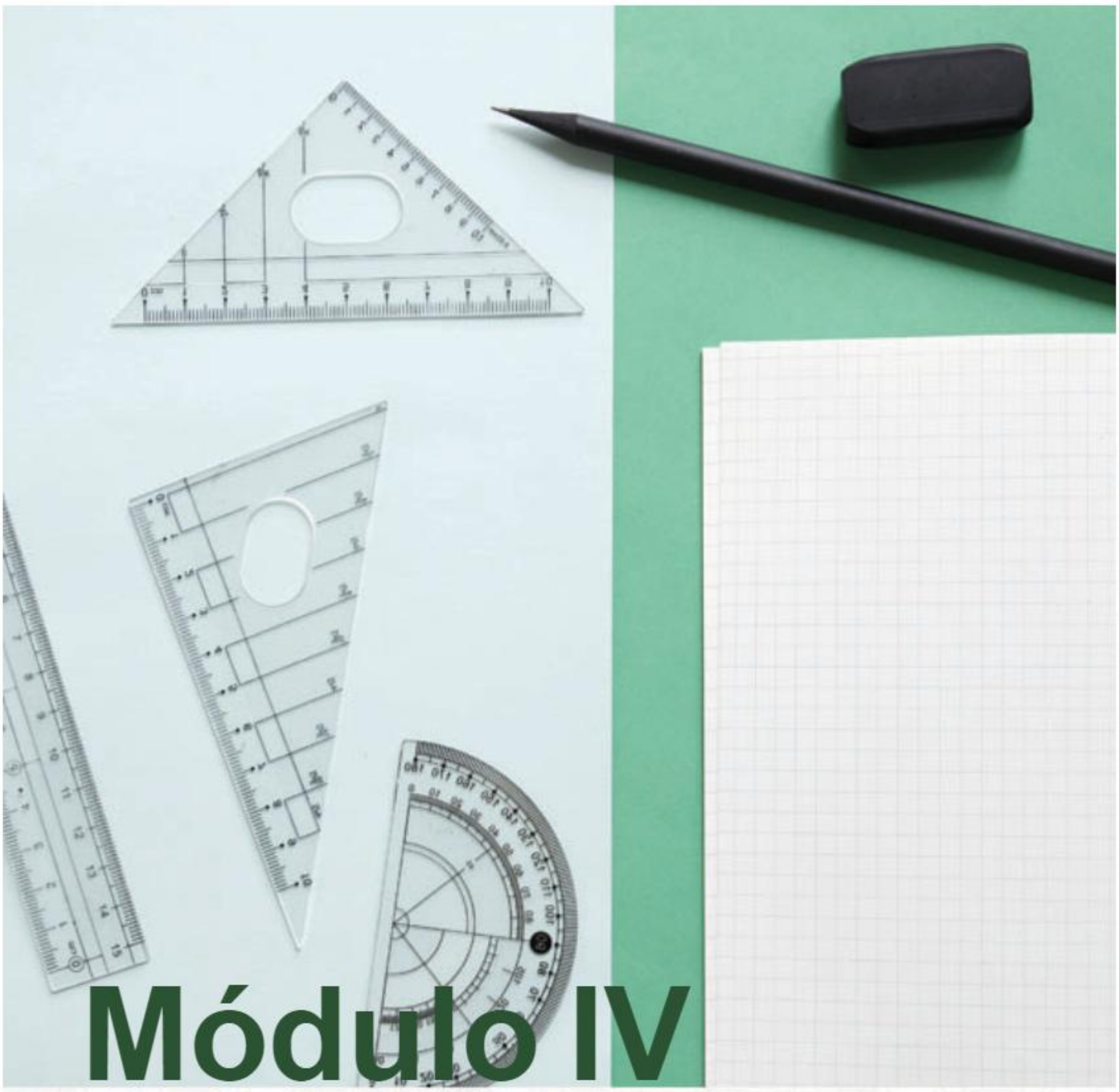
Muestreo probabilístico estratificado: los individuos de la población se agrupan en estratos, en función de características esenciales de interés para el estudio a realizar. Se seleccionan los individuos de la muestra de cada uno de los estratos de manera aleatoria.

Muestreo probabilístico simple: todos los individuos de la población tienen la misma probabilidad de ser seleccionados.

Muestreo probabilístico sistemático: se enumeran todos los individuos de la población, luego se elige el primero de la muestra de manera aleatoria y los siguientes de manera periódica.

Referencia Bibliográfica

- Adis, G., (s/f). CONCEPTOS BASICOS DE INVESTIGACIÓN. Recuperado de https://www.binasss.sa.cr/bibliotecas/bhp/cupula/v6n12/art1.pdf
- Data México, (2020). Toluca municipio del Estado de México. Gobierno de México. Disponible en: https://www.economia.gob.mx/datamexico/es/profile/geo/toluca#:~:text=Acerca%20de%20Toluca&text=En%202020%2C%20la%20poblaci%C3%B3n%20en,2.04%25%20respecto%20al%20a%C3%B1o%20anterior.
- Palacios, I., et.al. (2019). Diccionario electrónico de enseñanza y aprendizaje de lenguas. ISBN 978-84-09-10971-5. Disponible en línea en https://www.dicenlen.eu/es/diccionario/entradas/estudio-experimental
- Saavedra, P. (2016). Estudios Experimentales. Síntesis de Conocimientos. Biblioteca para estudiantes y profesionales de la salud. Recuperado de https://sintesis.med.uchile.cl/condiciones-clinicas/condiciones-clinicas-salud-publica/12548-estudios-experimentales
- Sadabell, M., (2023). ¿Por qué es tan complicado predecir el tiempo atmosférico? Revista MUY Interesante, en línea. Disponible en https://www.muyinteresante.com/ciencia/60028.html
- Zurita-Cruz, J., et.al., (2018). Estudios experimentales: diseños de investigación para la evaluación de intervenciones en la clínica. *Revista alergia México, 65*(2), 178-186. https://doi.org/10.29262/ram.v65i2.376



Módulo IV

Medidas numéricas

Propósito de Módulo

Determina las medidas de tendencia central, posición y dispersión de un conjunto de datos.

Introducción

La estadística es una rama fundamental de las matemáticas que se ocupa de la recolección, análisis, interpretación y presentación de datos. En este marco, existen varias medidas y herramientas que permiten comprender y describir diferentes aspectos de los datos. En este módulo se estudiarán las medidas de tendencia central con el objetivo de identificar datos típicos y atípicos en una distribución de datos, aprenderás también sobre las diferentes medidas de posición como son los cuartiles, deciles y percentiles que aplicados a la vida cotidiana te permiten evaluar rápidamente la dispersión y tomar decisiones oportunas y certeras en los diferentes estudios de investigación que se realicen.

Este módulo pretende proporcionarte los aspectos teóricos y prácticos para hacer fácil el estudio y la comprensión del pensamiento matemático; hoy en día es necesario el estudio sistemático de los conceptos básicos de la probabilidad y estadística a fin de disponer de las técnicas y herramientas necesarias para resolver con eficacia los problemas de la vida actual.

También se estudia de forma introductoria a la Estadística inferencial, que utiliza técnicas para hacer generalizaciones o inferencias sobre una población a partir de una muestra de datos. Esto incluye estimar parámetros, probar hipótesis y hacer predicciones.

Estos conceptos y herramientas son fundamentales para el análisis y comprensión de datos en diversos campos, desde la investigación científica hasta la toma de decisiones en negocios.



Actividad 1.



Evaluación Diagnóstica
(conocimientos previos)

Antes de comenzar con la revisión de los contenidos del Módulo 4 de Pensamiento Matemático I, te invitamos a responder la evaluación diagnóstica.

Eso te ayudará a descubrir los conocimientos previos que tienes sobre los temas a abordar y, al mismo tiempo, le permitirá a tu docente conocer tus fortalezas y debilidades.

Para comenzar con el ejercicio, da clic en la **imagen** y responde a la actividad.

4.1 Medidas de tendencia central

Imagina que un vendedor de paletas hace un registro en una tabla de frecuencias, en ella coloca: ¿cuántas paletas vende cada día en una semana? y ¿qué sabores vende más? obteniendo los siguientes resultados:

Tabla 1. Cantidad de paletas vendidas a la semana

Día de la semana	Cantidad de paletas vendidas
Lunes	124
Martes	87
Miércoles	145
Jueves	218
Viernes	315
Sábado	420
Domingo	532
Total	1841

Tabla 2. Cantidad de paletas vendidas por sabor

Sabor	Cantidad de paletas vendidas
Pistache	145
Limón	346
Fresa	321
Vainilla	216
Mamey	184
Chocolate	537
Zarzamora	92
Total	1841

Al final de la semana el vendedor expresa: **¡En esta semana, en promedio, vendí 263 paletas diarias y el sabor que más se vendió fue chocolate!**

Piensa **¿qué quiere decir con esa expresión?** comenta con tus compañeros la respuesta a esta pregunta y realicen una conclusión con ayuda del profesor.

Definición

Las **medidas de tendencia central** son medidas estadísticas que pretenden resumir en un solo valor a un conjunto de valores, representan un centro en torno al cual se encuentra ubicado el conjunto de los datos; las más utilizadas son: **media aritmética, mediana y moda** (Quevedo, 2011).

"Según la UNESCO la enseñanza de la estadística es fundamental para el ciudadano al favorecer la adquisición de conocimientos y habilidades que le permiten acceder a la sociedad de la información, en este sentido, la estadística juega un rol fundamental para el ejercicio de la ciudadanía en un mundo cada vez más globalizado, debido a que el sujeto moderno es demandado a poseer competencias para leer e interpretar tablas y gráficos que se presentan en los medios informativos para la toma de decisiones; como así también aquella información que se resume mediante indicadores numéricos o representaciones gráficas" (Alveal, 2016).

Regresando al ejemplo de las paletas: **¿Por qué crees que es importante que el vendedor sepa cuál es el número promedio de ventasdel día? y ¿por qué es importante conocer el sabor que más se vende?**

Comenta con tus compañeros las respuestas a estas preguntas y realicen una conclusión con ayuda del profesor.

Es importante considerar que las medidas de tendencia central tienen el mismo significado si se trata de datos no agrupados o datos agrupados; sin embargo, la forma operacional sí presenta diferencias, por lo que en cada una de ellas se explicarán ambos procedimientos para su obtención.

4.1.1 Media aritmética

Definición

También conocida como promedio aritmético, la media aritmética consiste en sumar todos los valores individuales entre el número total de valores; representa el punto de equilibrio de la distribución de los datos (Rendon-Macías, 2016).

Símbolo de la media aritmética: $\overline{X}$

MEDIA ARITMÉTICA PARA DATOS NO AGRUPADOS

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Donde:

$\overline{X}$ = Media aritmética
x_i = Valor del *i-ésimo* dato
n = Total de datos

Ejemplo 1. Calcula la media aritmética del siguiente conjunto de datos.

18, 24, 35, 17, 21, 46, 27, 32

Solución:

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\overline{X} = \frac{\sum_{i=1}^{8} x_i}{8} = \frac{18+24+35+17+21+46+27+32}{8} = \frac{220}{8} = 27.5$$

La media aritmética del conjunto de datos es $\overline{X}$ = 27.5

Ejemplo 2. Un vendedor de paletas registró las ventas de una semana obteniendo los datos de la siguiente tabla, determina:

a. En promedio ¿cuántas paletas vendió al día esa semana?
b. Escribe al menos una ventaja de saber el promedio de paletas vendidas al día.

Tabla 3. Cantidad de paletas vendidas a la semana

Día de la semana	Cantidad de paletas vendidas
Lunes	124
Martes	87
Miércoles	145
Jueves	218
Viernes	315
Sábado	420
Domingo	532

Solución

a. En promedio ¿cuántas paletas vendió al día esa semana?

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\overline{X} = \frac{\sum_{i=1}^{7} x_i}{7} = \frac{124+87+145+218+315+420+532}{7} = \frac{1841}{7} = 263$$

En promedio vendió $\overline{X}$ = 263 paletas diarias.

b. Escribe al menos una ventaja de saber el promedio de paletas vendidas al día.

Aunque no todos los días tiene las mismas ventas, sabe que en un día puede existir una cantidad aproximada de las ventas. Esto le permite identificar la cantidad de utensilios, materiales, personal, entre otros elementos, que debería tener disponibles.

Ejemplo 3. Las calificaciones del primer parcial de un estudiante son las mostradas en la siguiente tabla, ¿cuál es la **media aritmética** en ellas?

Tabla 4. Calificaciones del primer parcial de un estudiante

Asignatura	Calificación
Pensamiento matemático I	9.0
Cultura digital I	9.2
Inglés I	8.6
Lengua y comunicación I	9.4
Química I	8.7
Humanidades I	9.8
Ciencias sociales I	9.6
Recursos y ámbitos socioemocionales I	9.3

Solución

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\overline{X} = \frac{\sum_{i=1}^{8} x_i}{8} = \frac{9.0+9.2+8.6+9.4+8.7+9.8+9.6+9.3}{8} = \frac{73.6}{8} = 9.2$$

El promedio de las calificaciones mostradas del primer parcial es de $\overline{X}$ = 9.2

MEDIA ARITMÉTICA PARA DATOS AGRUPADOS

$$\overline{X} = \frac{\sum_{i=1}^{k} f_i M_i}{n}$$

Donde:

$\overline{X}$ = Media aritmética
f_i = Frecuencia del i-ésimo intervalo
M_i = Marca de clase del i-ésimo intervalo
k = Total de intervalos
n = Total de datos (suma de las frecuencias)

Ejemplo 4. Calcula la media aritmética del siguiente conjunto de datos.

Tabla 5. Conjunto de datos

Intervalo	Lim inferior	Lim superior	Frecuencia
1	6	10	11
2	11	15	14
3	16	20	8
4	21	25	16
5	26	30	9
6	31	35	15
7	36	40	7

Solución

Primero complementamos la tabla 5 con una columna que represente a la marca de clase (revisa el tema 2.2.3, polígono de frecuencias) y, una columna más donde se multiplique la frecuencia por la marca de clase de cada intervalo:

$$M_i = \frac{límite\ superior + límite\ inferior}{2}$$

Tabla 6. Cálculo de la media (1)

Intervalo	Lim inferior	Lim superior	f_i	M_i	f_iM_i
1	6	10	11	$= \frac{10+6}{2} = 8$	= (11)(8) = 88
2	11	15	14	$= \frac{15+11}{2} = 13$	= (14)(8) = 182
3	16	20	8	$= \frac{20+16}{2} = 18$	= (8)(18) = 144
4	21	25	16	23	368
5	26	30	9	28	252
6	31	35	15	33	495
7	36	40	7	38	266
		n =	80	SUMA =	1795

$$\overline{X} = \frac{\sum_{i=1}^{n} f_iM_i}{n}$$

$$\overline{X} = \frac{\sum_{i=1}^{7} f_iM_i}{n} = \frac{1795}{80} = 22.44$$

La media aritmética o promedio del conjunto de datos es $\overline{X}$ = 22.44

Ejemplo 5. Un negocio de envíos registró en la siguiente tabla los movimientos realizados a lo largo de 100 días.

39	58	58	67	58	32	35	30	43	57
61	42	57	54	39	54	30	74	65	55
74	72	56	80	56	50	48	31	55	36
56	30	66	57	27	34	79	53	71	54
80	33	50	75	75	65	68	57	28	27
31	79	61	73	63	66	69	41	73	66
41	46	57	80	76	43	60	78	28	34
58	68	37	53	52	76	43	25	54	33
76	29	54	55	52	50	25	26	62	52
45	61	45	40	53	72	44	34	54	67

Con esta información realiza y contesta lo siguiente:

a. Elabora una tabla de distribución de frecuencias con 7 intervalos
b. ¿Cuál es la cantidad promedio de movimientos al día?

Solución

a. Elabora una tabla de distribución de frecuencias con 7 intervalos

Dato mayor = 80
Dato menor = 25
Rango = 80 − 25 = 55
Variación = 1
Tamaño del intervalo = $\frac{55+1}{7} = \frac{56}{7} = 8$

La tabla de distribución de frecuencias con 7 intervalos, incluyendo columna de marca de clase y producto de frecuencia por marca de clase, es:

Intervalo	Lim inferior	Lim superior	f_i	M_i	f_iM_i
1	25	32	14	28.5	399
2	33	40	11	36.5	401.5
3	41	48	11	44.5	489.5
4	49	56	21	52.5	1102.5
5	57	64	15	60.5	907.5
6	65	72	13	68.5	890.5
7	73	80	15	76.5	1147.5
		n =	100	SUMA =	5338

Tabla 7. Cálculo de la media (2)

b. ¿Cuál es la cantidad promedio de movimientos al día?

$$\overline{X} = \frac{\sum_{i=1}^{n} f_i M_i}{n}$$

$$\overline{X} = \frac{\sum_{i=1}^{n} f_i M_i}{n} = \frac{5338}{100} = 53.38$$

En promedio, el negocio realizó $\overline{X}$ = 53.38 movimientos al día.

4.1.2 Mediana

Definición

Representa la cifra o valor que divide la muestra en dos mitades, es decir, el valor donde 50% de la población está por debajo o arriba del mismo (Rendon-Macías, 2016), en otras palabras, es el valor central de los datos.

Símbolo de la mediana: $\tilde{X}$

MEDIA ARITMÉTICA PARA DATOS NO AGRUPADOS

Caso 1. Número de datos impar
- Ordenar los datos de forma ascendente o descendente
- La mediana es el dato central

Caso 2. Número de datos par
- Ordenar los datos de forma ascendente o descendente
- La mediana es el promedio de los datos centrales

Ejemplo 6. Calcular la mediana del siguiente conjunto de datos:

18, 24, 35, 17, 21, 46, 27, 32, 58

Solución

Para obtener el resultado ordenamos los datos de manera ascendente:

17	18	21	24	27	32	35	46	58

↓

Dato central

La mediana del conjunto de datos es $\tilde{X}$ = 27

Ejemplo 7. Calcular la mediana del siguiente conjunto de datos:

8, 9, 2, 3, 5, 6, 7, 7, 8, 3, 9, 2, 8, 2

Solución

Ordenamos los datos de forma ascendente, dando como resultado lo siguiente:

2	2	2	3	3	5	6	7	7	8	8	8	9	9

↓

Datos centrales

En este caso los datos centrales se promedian: $\frac{6+7}{2} = \frac{13}{2} = 6.5$

La mediana del conjunto de datos es $\tilde{X}$ = 6.5

MEDIANA PARA DATOS AGRUPADOS

$$\tilde{X} = L_i + \left(\frac{\frac{n}{2} - Fa}{f_{\tilde{x}}} \right)(C)$$

Donde:

$\tilde{X}$ = Mediana
L_i = Límite real inferior del intervalo que contiene a la mediana
n = Total de datos (suma de las frecuencias)
F_a = Frecuencia acumulada del intervalo anterior que contiene a la mediana
$f_{\tilde{x}}$ = Frecuencia del intervalo que contiene a la mediana
C = Tamaño del intervalo

Ejemplo 8. Calcular la mediana del siguiente conjunto de datos:

Tabla 8. Conjunto de datos para calcular la mediana

Intervalo	Lim inferior	Lim superior	Frecuencia
1	6	10	11
2	11	15	14
3	16	20	8
4	21	25	16
5	26	30	9
6	31	35	15
7	36	40	7
		n =	80

Solución

Se agrega la columna de la frecuencia acumulada (ver tema 2.2.4, ojiva).

Tabla 9. Conjunto de datos para calcular la mediana

Intervalo	Lim inferior	Lim superior	f_i	Fa
	-	-	-	0
1	6	10	11	= 0 + 11 = 11
2	11	15	14	= 14 + 11 = 25
3	16	20	8	= 25 + 8 = 33
4	21	25	16	= 33 + 16 = 49
5	26	30	9	= 49 + 9 = 58
6	31	35	15	= 58 + 15 = 73
7	36	40	7	= 73 + 7 = 80
		n =	80	

De la definición de mediana se sabe que es el dato central, por lo que el total de datos n se divide entre 2:

$$\frac{n}{2} = 40$$

Se busca la fila de la frecuencia acumulada que incluya al dato número 40:

Tabla 10. Ubicación de la mediana

Intervalo	F_a		
	0	Esta fila no incluye datos	
1	11	Esta fila incluye los datos 1 al 11	
2	25	Esta fila incluye los datos 12 al 25	
3	33	Esta fila incluye los datos 26 al 33	
4	49	Esta fila incluye los datos 34 al 49	Aquí está el dato 40
5	58	Esta fila incluye los datos 50 al 58	
6	73	Esta fila incluye los datos 59 al 73	
7	80	Esta fila incluye los datos 74 al 80	

Por lo que el intervalo que contiene a la mediana es el 4.

Tabla 11. Información del intervalo para el cálculo de la mediana

	Intervalo	Lim inferior	Lim superior	f_i	F_a
		-	-	-	0
	1	6	10	11	11
	2	11	15	14	25
	3	16	20	8	33
$\tilde{X}$	4	21	25	16	49
	5	26	30	9	58
	6	31	35	15	73
	7	36	40	7	80
			n =	80	

F_a

$f_{\tilde{X}}$

De ese mismo intervalo se obtiene la siguiente información.

Límite real inferior

$$L_i = Límite\ inferior - \frac{variación}{2} = 21 - 0.5 = 20.5$$

Tamaño del intervalo

$$C = Límite\ superior - Límite\ inferior + variación = 25 - 21 + 1 = 5$$

Con estos datos ya se puede **calcular la mediana:**

$$\tilde{X} = L_i + \left(\frac{\frac{n}{2} - F_a}{f_{\tilde{x}}}\right)(C)$$

$$\tilde{X} = 20.5 + \left(\frac{40 - 33}{16}\right)(5)$$

$$\tilde{X} = 20.5 + \left(\frac{13}{16}\right)(5)$$

$$\tilde{X} = 20.5 + 4.06$$

$$\tilde{X} = \mathbf{24.56}$$

La mediana del conjunto de datos es $\tilde{X}$ = 24.56

Ejemplo 9. Una tienda registró las edades de 300 clientes, los cuales asistieron a hacer compras los fines de semana a lo largo de un mes, obteniendo los siguientes datos.

14	23	53	21	26	13	42	21	28	31
43	19	27	35	14	43	31	53	18	29
33	14	51	15	17	57	16	20	27	17
53	31	53	33	34	58	52	29	15	54
43	41	52	14	38	39	38	24	55	34
43	22	40	46	30	55	13	16	24	38
17	34	39	60	57	18	49	50	27	26
13	60	39	37	39	22	41	50	24	30
60	18	42	43	46	24	23	49	57	27
15	15	32	32	43	25	19	31	27	58
26	19	35	47	58	13	47	58	14	15
17	26	22	31	42	57	51	27	33	48
41	35	40	32	43	42	30	27	20	51
50	32	16	45	34	43	24	22	34	28
27	28	33	44	36	30	45	36	13	13
34	31	33	42	36	28	38	26	51	21
14	28	27	33	40	34	25	43	55	21
18	42	19	39	41	38	37	26	18	40
33	20	23	24	37	44	38	44	42	33
32	32	27	35	28	31	39	31	32	20
47	24	30	26	32	23	38	23	30	49
38	20	42	38	38	29	39	45	30	21
55	16	29	32	27	36	38	33	36	52
23	29	30	28	24	41	30	20	14	18
13	20	36	45	27	38	27	38	44	31
31	32	20	36	30	29	32	35	27	20
51	60	40	29	26	44	23	20	25	24
37	18	23	32	38	43	44	36	58	20
50	32	31	50	58	44	43	18	25	20
42	30	37	47	31	44	50	15	14	22

Con esta información realiza y contesta lo siguiente:

a) Elabora una tabla de distribución de frecuencias con 8 intervalos
b) ¿Cuál es la mediana de la edad de los asistentes a la tienda?

Solución

a. Elabora una tabla de distribución de frecuencias con 8 intervalos

Dato mayor = 60
Dato menor = 13
Rango = 60 − 13 = 47
Variación = 1
Tamaño del intervalo = $\frac{47+1}{8} = \frac{48}{8} = 6$

La tabla de distribución de frecuencias con 8 intervalos, incluyendo frecuencia acumulada es:

Tabla 12. Distribución de frecuencias con 8 intervalos

Intervalo	Lim inferior	Lim superior	fi	F_a
	-	-	-	0
1	13	18	37	37
2	19	24	43	80
3	25	30	51	131
4	31	36	54	185
5	37	42	45	230
6	43	48	30	260
7	49	54	22	282
9	55	60	18	300
		n =	300	

b. ¿Cuál es la mediana de la edad de los asistentes a la tienda?

Como $\frac{n}{2} = \frac{300}{2}$ el intervalo que contiene a la mediana es el número 4

Tabla 13. Información del intervalo para el cálculo de la mediana

	Intervalo	Lim inferior	Lim superior	f_i	F_a
		-	-	-	0
	1	13	18	37	37
	2	19	24	43	80
	3	25	30	51	131
$\tilde{X}$	4	31	36	54	185
	5	37	42	45	230
	6	43	48	30	260
	7	49	54	22	282
	9	55	60	18	300
			n =	300	

F_a

$f_{\tilde{X}}$

De ese mismo intervalo se obtiene la siguiente información:

Límite real inferior

$$L_i = Límite\ inferior - \frac{variación}{2} = 31 - \frac{1}{2} = 31 - 0.5 = 30.5$$

Tamaño del intervalo

$$C = Límite\ superior - Límite\ inferior + variación = 36 - 31 + 1 = 6$$

Con estos datos ya se puede **calcular la mediana**

$$\tilde{X} = L_i + \left(\frac{\frac{n}{2} - Fa}{f_{\tilde{x}}}\right)(C)$$

$$\tilde{X} = 30.5 + \left(\frac{150 - 131}{54}\right)(6)$$

$$\tilde{X} = 30.5 + \left(\frac{19}{54}\right)(6)$$

$$\tilde{X} = 20.5 + 2.11$$

$$\tilde{X} = \mathbf{22.61}$$

La mediana de las edades para el conjunto de datos es $\tilde{X}$ = 22.61 años.

Ejemplo 10. Calcular la media aritmética y la mediana del siguiente conjunto de datos:

Tabla 14. Conjunto de datos agrupados

Intervalo	Lim inferior	Lim superior	Frecuencia
1	12	20	5
2	21	29	8
3	30	38	16
4	39	47	24
5	48	56	38
6	57	65	45
7	66	74	42
8	75	83	36
9	84	92	28
10	93	101	8
		n =	250

Solución:
Se agregan las columnas de marca de clase, producto de la frecuencia del intervalo por la marca de clase y frecuencia acumulada.

Tabla 15. Cálculo de la media aritmética y la mediana para el conjunto de datos agrupados.

Intervalo	Lim inferior	Lim superior	f_i	M_i	$f_i M_i$	F_a
1	12	20	5	16	80	5
2	21	29	8	25	200	13
3	30	38	16	34	544	29
4	39	47	24	43	1032	53
5	48	56	38	52	1976	91
6	57	65	45	61	2745	136
7	66	74	42	70	2940	178
8	75	83	36	79	2844	214
9	84	92	28	88	2464	242
10	93	101	8	97	776	250
		n =	250	Suma =	15601	

Como $\frac{n}{2} = \frac{250}{2} = 125$ el intervalo que contiene a la mediana es el número 6:

Tabla 16. Información del intervalo para el cálculo de la mediana

	Intervalo	Lim inferior	Lim superior	f_i	M_i	$f_i M_i$	F_a
	1	12	20	5	16	80	5
	2	21	29	8	25	200	13
	3	30	38	16	34	544	29
	4	39	47	24	43	1032	53
	5	48	56	38	52	1976	91
$\bar{X}$	6	57	65	45	61	2745	136
	7	66	74	42	70	2940	178
	8	75	83	36	79	2844	214
	9	84	92	28	88	2464	242
	10	93	101	8	97	776	250
			n =	250	Suma =	15601	

F_a

$f_{\bar{X}}$

De ese mismo intervalo se obtiene la siguiente información:

Límite real inferior

$$L_i = Límite\ inferior - \frac{variación}{2} = 57 - \frac{1}{2} = 57 - 0.5 = 56.5$$

Tamaño del intervalo

$$C = Límite\ superior - Límite\ inferior + variación = 65 - 57 + 1 = 9$$

Con estos datos ya se pueden **calcular la media y la mediana.**

$$\overline{X} = \frac{\sum_{i=1}^{n} f_i M_i}{n}$$

$$\overline{X} = \frac{\sum_{i=1}^{10} f_i M_i}{250}$$

$$\overline{X} = \frac{15601}{250}$$

$$\overline{X} = \mathbf{62.4}$$

$$\tilde{X} = L_i + \left(\frac{\frac{n}{2} - F_a}{f_{\tilde{x}}}\right)(C)$$

$$\tilde{X} = 56.5 + \left(\frac{125 - 91}{45}\right)(9)$$

$$\tilde{X} = 56.5 + \left(\frac{34}{45}\right)(9)$$

$$\tilde{X} = 56.5 + 6.8$$

$$\tilde{X} = \mathbf{63.3}$$

La media aritmética del conjunto de datos es $\overline{X}$ = 62.4, mientras que la mediana es $\tilde{X}$ = 63.3

4.1.3 Moda

Definición

Valor más frecuentemente encontrado en las mediciones (Rendon-Macías, 2016); es decir, es el valor que más se repite.

Símbolo de la moda: $\widehat{X}$

MODA ARITMÉTICA PARA DATOS NO AGRUPADOS

La moda corresponde al dato que más se repite del conjunto de datos; sin embargo, pueden ocurrir los siguientes casos:

- Amodal: no se repite ningún dato.
- Unimodal: un solo dato se repite el mayor número de veces.
- Bimodal: dos datos se repiten de forma equivalente el mayor número de veces.
- Trimodal: tres datos se repiten de forma equivalente el mayor número de veces.
- Esto puede continuar de forma sucesiva.

Ejemplo 11. La siguiente tabla muestra las ventas de un negocio de paletas según el sabor elegido por los clientes.

Sabor	Cantidad de paletas vendidas
Pistache	145
Limón	346
Fresa	321
Vainilla	216
Mamey	184
Chocolate	537
Zarzamora	92
Total	1841

Tabla 17. Registro de ventas de un negocio de paletas

Con esta información determina:

a) ¿Cuál sabor representa la moda de los datos?
b) Escribe al menos una ventaja de conocer cuál sabor representa la moda de los datos.

Solución

a) ¿Cuál sabor representa la moda de los datos?

Si bien los datos están en una tabla, se trata de un conjunto de datos NO agrupados, ya que no están expresados por intervalos, por lo tanto, la moda corresponde al sabor que posee la mayor frecuencia o número de ventas, que en este caso es el sabor Chocolate.

b) Escribe al menos una ventaja de conocer cuál sabor representa la moda de los datos.

Conocer cuál es el sabor que representa la moda de los datos equivale a decir cuál es el sabor que más se vende, esto es útil porque permite tomar la decisión de comprar más ingredientes para ese sabor.

Ejemplo 12. ¿Cuál es la moda de la siguiente lista de datos?

12,26,18,36,45,18,13,20,32,27,18,25,18

Solución

Se puede observar que, de la lista de datos, el número 18 se repite cuatro veces, por lo que el conjunto de datos es unimodal y el valor de la moda es 18:

25,18,12,26,18,36,45,18,13,20,32,27,18

La moda del conjunto de datos es $\widehat{X}$ = 18.

Ejemplo 13. ¿Cuál es la moda de la siguiente lista de datos?

139,124,136,145,136,141,124,138,127,143,136,152,124,127,141,138

Solución

Se puede observar que, de la lista de datos, los números 124 y 136 se repiten cada uno tres veces, por lo que el conjunto de datos es bimodal y el valor de la moda es 124 y 136:

139,124,136,145,136,141,124,138,127,143,136,152,124,127,141,138

La moda del conjunto de datos es $\widehat{X}$ = 124 y 136

MODA PARA DATOS AGRUPADOS

$$\widehat{X} = L_i + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right)(C)$$

Donde:

$\widehat{X}$ = Moda
L_i = Límite real inferior del intervalo que contiene a la moda
Δ_1 = Diferencia entre la frecuencia del intervalo que contiene a la moda y el intervalo anterior
Δ_2 = Diferencia entre la frecuencia del intervalo que contiene a la moda y el intervalo siguiente
C = Tamaño del intervalo

Ejemplo 14. Calcular la moda del siguiente conjunto de datos:

Tabla 18. Conjunto de datos agrupados

Intervalo	Límite inferior	Límite superior	Frecuencia
1	6	10	11
2	11	15	14
3	16	20	8
4	21	25	16
5	26	30	9
6	31	35	15
7	36	40	7
		n=	80

Solución

Como la definición de moda es el dato que más se repite, el intervalo que se selecciona para los datos de la fórmula será aquel que tenga la mayor frecuencia, que en este caso es el número 4.

Tabla 19. Información del intervalo para el cálculo de la moda

	Intervalo	Lim inferior	Lim superior	Frecuencia	
	1	6	10	11	
	2	11	15	14	
	3	16	20	8	$\Delta_1 = 16 - 8 = 8$
$\hat{X}$	4	21	25	16	Mayor frecuencia
	5	26	30	9	$\Delta_2 = 16 - 9 = 7$
	6	31	35	15	
	7	36	40	7	
			n =	80	

De ese mismo intervalo se obtiene la siguiente información:

Límite real inferior

$$L_i = Limite\ inferior - \frac{variación}{2} = 21 - \frac{1}{2} = 21 - 0.5 = 20.5$$

Tamaño del intervalo

$$C = Límite\ superior - Límite\ inferior + variación = 25 - 21 + 1 = 5$$

$$\Delta_1 = 16 - 8 = 8$$

$$\Delta_2 = 16 - 9 = 7$$

Con estos datos ya se puede **calcular la moda**

$$\widehat{X} = 20.5 + \left(\frac{8}{8+7}\right)(5)$$

$$\widehat{X} = 20.5 + \left(\frac{8}{15}\right)(5)$$

$$\widehat{X} = 20.5 + 2.66$$

$$\widehat{X} = 23.16$$

La moda del conjunto de datos es $\widehat{X}$ = 23.16

Ejemplo 15. El encargado de un juego mecánico fue comisionado para registrar las edades de las personas que hacían uso del juego, los 120 registros se muestran en la tabla siguiente.

18	29	14	39	29	37	8	24
39	15	22	31	14	27	38	9
11	19	10	11	29	18	13	24
21	13	15	38	22	18	23	15
28	29	34	13	19	30	27	8
36	24	20	33	25	24	19	19
34	20	34	29	26	32	31	9
40	9	15	26	33	37	9	31
25	13	18	23	27	34	15	9
18	31	19	31	27	8	28	40
28	17	38	23	26	9	30	9
34	40	22	14	35	13	26	14
28	35	33	32	22	18	30	28
34	19	29	9	17	15	16	27
15	12	35	43	39	31	9	40

Con esta información realice y conteste lo siguiente:

a) Elabora una tabla de distribución de frecuencias con 6 intervalos
b) ¿Cuál es la moda de la edad de los usuarios del juego mecánico?

Solución

a) Elabora una tabla de distribución de frecuencias con 6 intervalos

Dato mayor = 43
Dato menor = 8
Rango = 43 − 8 = 35
Variación = 1
Tamaño del intervalo = $\frac{35+1}{6} = \frac{36}{6} = 6$

La tabla de distribución de frecuencias con 6 intervalos es:

Tabla 20. Distribución de frecuencias con 6 intervalos.

Intervalo	Lim inferior	Lim superior	f_i
1	8	13	21
2	14	19	26
3	20	25	16
4	26	31	29
5	32	37	17
6	38	43	11
		$n =$	120

b) ¿Cuál es la moda de la edad de los usuarios del juego mecánico?

Como la definición de moda es el dato que más se repite, el intervalo que se selecciona para los datos de la fórmula será aquel que tenga la mayor frecuencia, que en este caso es el número 4:

Tabla 21. Información del intervalo para el cálculo de la moda.

	Intervalo	Lim inferior	Lim superior	f_i	
	1	8	13	21	
	2	14	19	26	
	3	20	25	16	$\Delta_1 = 29 - 16 = 13$
$\widehat{X}$	4	26	31	29	Mayor frecuencia
	5	32	37	17	$\Delta_2 = 29 - 17 = 12$
	6	38	43	11	
			n =	120	

De ese mismo intervalo se obtiene la siguiente información:

Límite real inferior

$$L_i = Límite\ inferior - \frac{variación}{2} = 26 - \frac{1}{2} = 26 - 0.5 = 25.5$$

Tamaño del intervalo

$$C = Límite\ superior - Límite\ inferior + variación = 31 - 26 + 1 = 6$$

$$\Delta_1 = 29 - 16 = 13$$

$$\Delta_2 = 29 - 17 = 12$$

Con estos datos ya se puede **calcular la moda**

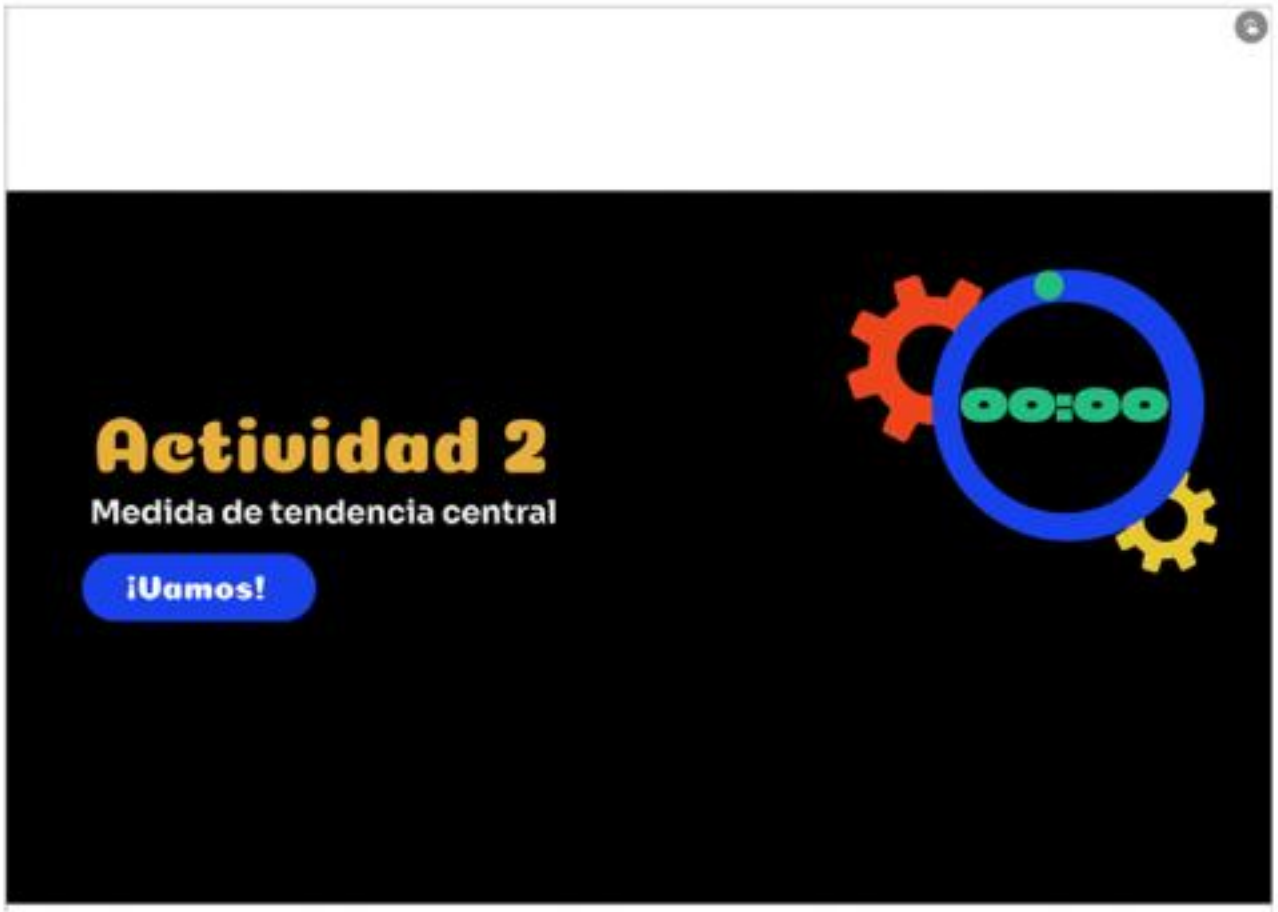
$$\widehat{X} = 25.5 + \left(\frac{13}{13+12}\right)(6)$$

$$\widehat{X} = 25.5 + \left(\frac{13}{25}\right)(6)$$

$$\widehat{X} = 25.5 + 3.12$$

$$\widehat{X} = \mathbf{28.62}$$

La moda del conjunto de datos es $\widehat{X}$ = 28.62



Actividad 2.

Ejercicios de medida de tendencia central

Instrucciones:
Es momento de poner en práctica lo aprendido, te invitamos a resolver los siguientes ejercicios, para comenzar da clic en la **imagen.**

4.2 Medidas de posición

Definición

Las medidas de posición son aquellas en donde podemos dividir los datos en cuatro partes iguales y se considerarán cuartiles, al dividir en diez partes iguales se les considerará deciles y finalmente se puede dividir en 100 partes iguales, y se les llamará percentiles.

Utilizamos las medidas de posición en una serie de datos o en una distribución de frecuencias para encontrar valores específicos, aunado a que proporcionan información sintetizada de la variable en estudio.

Figura 1. Medidas de posición



4.2.1 Cuartiles

PARA DATOS NO AGRUPADOS

Fórmula

Para encontrar la posición

$$Q_1 = \frac{(n+1)}{4},\quad Q_2 = \frac{2(n+1)}{4},\quad Q_3 = \frac{3(n+1)}{4}$$

Donde:

$Q_{1,2\ y\ 3} = Cuartil \quad y \quad n = total\ de\ datos$

Hay que considerar que Q_2 equivale a la mediana.

Ejemplo 1. Del grupo 405 del plantel "Dr. Pablo González Casanova" de la Escuela Preparatoria, se obtienen las primeras 20 calificaciones de la lista. Calcular el tercer cuartil Q_3

Las calificaciones obtenidas son:

9	9	5	5	10	6	8	9	4	8	5	9	10	6	6	8	6	9	5	7

Ordenamos los datos de forma ascendente

4	5	5	5	5	6	6	6	6	7	8	8	8	9	9	9	9	9	10	10

$Cuartil\ Q_3 = \frac{3(n+1)}{4} = \frac{3(20+1)}{4} = 15.75$ **(Posición)**

Cuartil Q_3 corresponde al valor 9

4	5	5	5	5	6	6	6	6	7	8	8	8	9	9	9	9	9	10	10
														9					

Como las posiciones 15 y 16 corresponden al valor 9, este será la respuesta al cuartil 3.

PARA DATOS AGRUPADOS

Antes de hacer uso de la fórmula general, encontraremos la posición en una distribución de frecuencias de la siguiente manera:

$$Q_1 = \frac{n}{4},\ Q_2 = \frac{2(n)}{4},\ Q_3 = \frac{3(n)}{4}$$

El valor de esta posición obtenida corresponde al de f_{Q_n} en la siguiente fórmula

Fórmula

$$Q_n = L_i + \left[\frac{f_{Qn} - f_{i-1}}{f_i - f_{i-1}}\right] * C$$

Donde:

- Qn Cuartil que desea calcular
- L_i = Limite real inferior donde se encuentra la frecuencia del cuartil.
- f_i = Frecuencia acumulada anterior a la f_{Q_n}
- f_{i-1} = Frecuencia de la clase cuartil donde se localiza (posición)
- C = Amplitud

Ejemplo 2. En un centro comercial se tomó la muestra del tiempo en que 40 personas tardan en ser atendidos en las cajas de cobro, para el centro comercial es de suma importancia atender a sus clientes lo más rápido posible. Desean saber de las cuarenta personas que tiempo se tardan en atender al 25%. Para esto hay que calcular las medidas de posición.

Tabla 22. Distribución de frecuencias para los datos del centro comercial.

Intervalo de clase	Límites reales	Frecuencia	fa
1.0-2.0	0.92-2.05	9	9
2.1-3.1	2.05-3.15	11	20
3.2-4.2	3.15-4.25	8	28
4.3-5.3	4.25-5.35	7	35
5.4-6.4	5.35-6.45	1	36
6.5-7.5	6.45-7.55	2	38
7.6-8.6	7.55-8.65	1	39
8.7-9.7	8.65-9.75	1	40
	Total	40	

Para obtener el tiempo de atención al 25% de los clientes, aplicaremos el primer cuartil.

$Q_1 = \frac{n}{4} = \frac{40}{4} = 10\ posición$

$L_i = 2.1$

$f_i = 20$

$f_{i-1} = 9$

$C = 1.1$

$Q_1 = 2.1 + \left[\frac{10-9}{20-9}\right] * 1.1 = 2.1 + 0.09(1.1) = 2.1 + 0.1 = \mathbf{2.2}$

El tiempo estimado para realizar sus pagos en las cajas de la tienda es en 2.2 minutos.

4.2.2 Deciles

PARA DATOS NO AGRUPADOS

Fórmula

Para encontrar la posición

$$D_1 = \frac{(n+1)}{10}, \ldots, \quad D_5 = \frac{5(n+1)}{10}, \ldots, \quad D_9 = \frac{9(n+1)}{10}$$

Donde:

$D_{1,2,\ldots,9} = Decil \quad y \quad n = total\ de\ datos$

Hay que considerar que D_5 equivale a la mediana.

Ejemplo 3. Del grupo 405 del plantel "Dr. Pablo González Casanova" de la Escuela Preparatoria, se obtienen las primeras 20 calificaciones de la lista. Calcular el cuarto decil D_4

Las calificaciones obtenidas son:

9	9	5	5	10	6	8	9	4	8	5	9	10	6	6	8	6	9	5	7

Ordenamos los datos de forma ascendente

4	5	5	5	5	6	6	6	6	7	8	8	8	9	9	9	9	9	10	10

$$Decil\ D_4 = \frac{4(n+1)}{10} = \frac{4(20+1)}{10} = 8.4$$

Decil D_4 corresponde al valor **6**

4	5	5	5	5	6	6	6	6	7	8	8	8	9	9	9	9	9	10	10

6

Como la posición 8 y 9 corresponde al valor 6, este será la respuesta al decil 4.

PARA DATOS AGRUPADOS

Antes de hacer uso de la fórmula general, encontraremos la posición en una distribución de frecuencias de la siguiente manera:

$$D_1 = \frac{n}{10}, \ldots, D_5 = \frac{5(n)}{10}, \ldots, D_9 = \frac{9(n)}{10}$$

El valor de esta posición obtenida corresponde al de f_{Dn} en la siguiente fórmula

Fórmula

$$D_n = L_i + \left[\frac{f_{Dn} - f_{i-1}}{f_i - f_{i-1}}\right] * C$$

Donde:

- Dn Decil que desea calcular
- L_i = Limite real inferior donde se encuentra la frecuencia del decil.
- f_i = Frecuencia acumulada anterior a la f_{Dn}
- f_{i-1} = Frecuencia de la clase cuartil donde se localiza (posición)
- C = Amplitud

Ejemplo 4. En un centro comercial se tomó la muestra del tiempo en que 40 personas tardan en ser atendidos en las cajas de cobro, para el centro comercial es de suma importancia atender a sus clientes lo más rápido posible. Desean saber de las cuarenta personas que tiempo se tardan en atender al 50%. Para esto hay que calcular las medidas de posición.

Tabla 23. Distribución de frecuencias para los datos del centro comercial.

Intervalo de clase	Límites reales	Frecuencia	fa
1.0-2.0	0.95-2.05	9	9
2.1-3.1	2.05-3.15	11	20
3.2-4.2	3.15-4.25	8	28
4.3-5.3	4.25-5.35	7	35
5.4-6.4	5.35-6.45	1	36
6.5-7.5	6.45-7.55	2	38
7.6-8.6	7.55-8.65	1	39
8.7-9.7	8.65-9.75	1	40
Total	Total	40	

• Para obtener el tiempo de atención al 50% de los clientes, aplicaremos el quinto decil.

$$D_5 = \frac{5(40)}{10} = \frac{80}{10} = 20\ posición$$

Como la posición se encuentra en la frecuencia acumulada, el valor del decil 5 será lo que equivale el límite superior del intervalo, es decir 3.1

El tiempo estimado para realizar sus pagos en las cajas de la tienda es en 3.1 minutos.

4.2.3 Percentiles

PARA DATOS NO AGRUPADOS

Fórmula

Para encontrar la posición

$$P_1 = \frac{(n+1)}{100}, \ldots, P_{50} = \frac{50(n+1)}{100}, \ldots, P_{99} = \frac{99(n+1)}{100}$$

Donde:

$P_{1,2,\ldots,99} = Percentil \quad y \quad n = total\ de\ datos$

Hay que considerar que P_{50} equivale a la mediana.

Ejemplo 5. Del grupo 405 del plantel "Dr. Pablo González Casanova" de la Escuela Preparatoria, se obtienen las primeras 20 calificaciones de la lista. Calcular el percentil cincuenta P_{50}

Las calificaciones obtenidas son:

9	9	5	5	10	6	8	9	4	8	5	9	10	6	6	8	6	9	5	7

Ordenamos los datos de forma ascendente

4	5	5	5	5	6	6	6	6	7	8	8	8	9	9	9	9	9	10	10

$$Percentil\ P_{50} = \frac{50(n+1)}{100} = \frac{50(20+1)}{100} = 10.5$$

Para obtener el valor del percentil (puede aplicarse al cuartil o decil) cuando la posición se encuentra con decimal o entre dos posiciones exactas.

a) Identificamos el valor de la posición inferior y superior, en este caso posición 10 y 11

Valor posición 10 en los datos = 7	Valor posición 11 en los datos = 8

b) Restamos valor de la posición 11 – valor posición 10

8 – 7 = 1

c) Multiplicamos el resultado de la resta por el valor decimal de la posición

1 * 0.5 = 0.5

d) Finalmente, para obtener el valor del percentil sumamos la cantidad anterior con el valor que se ubica en la posición inferior

7 + 0.5 = 7.5

De acuerdo con el resultado anterior, el valor de la posición 10.5 es 7.5

4	5	5	5	5	6	6	6	6	7	8	8	8	9	9	9	9	9	10	10

7.5

PARA DATOS AGRUPADOS

Antes de hacer uso de la fórmula general, encontraremos la posición en una distribución de frecuencias de la siguiente manera:

$$P_1 = \frac{n}{100}, \ldots, P_{50} = \frac{50(n)}{100}, \ldots, P_{99} = \frac{99(n)}{100}$$

El valor de esta posición obtenida corresponde al de f_{P_n} en la siguiente fórmula

Fórmula

$$P_n = L_i + \left[\frac{f_{Pn} - f_{i-1}}{f_i - f_{i-1}}\right] * C$$

Donde:

- Pn Percentil que desea calcular
- L_i = Limite real inferior donde se encuentra la frecuencia del decil.
- f_i = Frecuencia acumulada anterior a la f_{Pn}
- f_{i-1} = Frecuencia de la clase cuartil donde se localiza (posición)
- C = Amplitud

Ejemplo 6. En un centro comercial se tomó la muestra del tiempo en que 40 personas tardan en ser atendidos en las cajas de cobro, para el centro comercial es de suma importancia atender a sus clientes lo más rápido posible. Desean saber de las cuarenta personas que tiempo se tardan en atender al 75%. Para esto hay que calcular las medidas de posición.

Intervalo de clase	Límites reales	Frecuencia	fa
1.0-2.0	0.95-2.05	9	9
2.1-3.1	2.05-3.15	11	20
3.2-4.2	3.15-4.25	8	28
4.3-5.3	4.25-5.35	7	35
5.4-6.4	5.35-6.45	1	36
6.5-7.5	6.45-7.55	2	38
7.6-8.6	7.55-8.65	1	39
8.7-9.7	8.65-9.75	1	40
Total	Total	40	

Tabla 24. Distribución de frecuencias para los datos del centro comercial.

Para obtener el tiempo de atención al 75% de los clientes, aplicaremos el septuagésimo quinto percentil.

$$P_{75} = \frac{75(40)}{100} = \frac{3000}{10} = 30 \; posición$$

$L_i = 4.3$

$f_i = 35$

$f_{i-1} = 28$

$C = 1.1$

$$Q_1 = 4.3 + \left[\frac{30-28}{35-28}\right] * 1.1 = 4.3 + 0.285(1.1) = 2.1 + .31 = \mathbf{4.6}$$

El tiempo estimado para realizar sus pagos en las cajas de la tienda es en 4.6 minutos.

En resumen, las medidas de posición nos permiten encontrar puntos específicos en datos agrupados y no agrupados, la mediana es una medida que representa el 50% de los datos, misma que los divide en dos partes iguales, por lo cual se le considera también una medida de posición.



Actividad 3.

Ejercicios de medidas de posición

Instrucciones:
para practicar lo revisado en el tema, te invitamos a resolver los siguientes ejercicios, para comenzar da clic en la **imagen.**

4.3 Desviación media

Definición

La **desviación media** se define como la media aritmética de los valores absolutos de las desviaciones de las variables respecto a la media aritmética.

Para datos no agrupados matemáticamente se expresa por la ecuación:

$$DM = \frac{\sum_{i=1}^{n} |x_i - \overline{X}|}{n}$$

Donde:

x_i: es el i-ésimo dato

n: es número total de datos

$\overline{X}$: Media aritmética de los datos

Ejemplo 1. El dueño de un restaurante desea evaluar el número de clientes que visitan su establecimiento durante cinco días. Los datos registrados son: 45, 50, 47, 55 y 53. Calcula la desviación media del número de clientes.

a) Calcular la media aritmética $\overline{X}$

$$\overline{X} = \frac{45 + 50 + 47 + 55 + 53}{5} = 50$$

b) Calcula la suma de las desviaciones absolutas

$$\sum_{j=1}^{5} |x - \overline{X}| = |45 - 50| + |50 - 50| + |47 - 50| + |55 - 50| + |53 - 50| = 5 + 0 + 3 + 5 + 3 = 16$$

c) Calcular la desviación media

$$DM = \frac{16}{5} = 3.2$$

La desviación media con respecto a la media es 3.2, esto significa que los datos están a una distancia de 3.2 unidades por debajo o por arriba de la media aritmética en promedio.

Ejemplo 2. Un entrenador de atletismo desea analizar los tiempos (en segundos) de carrera de sus atletas en una competencia. Los tiempos son: 12.5, 13.0, 12.7, 12.9, 13.2, 12.0. Calcula la desviación media de los tiempos de carrera.

a) Calcular la media aritmética

$$\overline{X} = \frac{12.5 + 13.0 + 12.7 + 12.9 + 13.2 + 12.0}{6} = \frac{76.3}{6} = 12.717$$

b) Calcula la suma de las desviaciones absolutas

$$\sum_{j=1}^{5} |x - \overline{X}| = |12.5 - 12.717| + |13.0 - 12.717| + |12.7 - 12.717| + |12.9 - 12.717| + |13.2 - 12.717| + |12.0 - 12.717|$$

$$= 0.217 + 0.283 + 0.017 + 0.183 + 0.483 + 0 \ -0.717$$

$$= 1.9$$

c) Calcular la desviación media

$$DM = \frac{1.9}{6} = 0.317$$

La desviación media con respecto a la media es 0.317, esto significa que los datos están a una distancia de 0.317 unidades por debajo o por arriba de la media aritmética en promedio.

Otra forma práctica de escribir es en tabla. Veamos el siguiente ejercicio.

Ejemplo 3. Un economista está evaluando el precio de las acciones de una empresa durante ocho días. Los precios (en dólares) de cierre de cada día son: 150, 155, 160, 162, 158, 145, 190,181. Calcula la desviación media de los precios de las acciones.

Tabla 25. Cálculo de desviación media.

Precios en dolares	$\lvert x_i - \overline{X} \rvert$
150	|150 - 162.625| = 12.625
155	|155 - 162.625| = 7.625
160	|160 - 162.625| = 2.625
162	|162 - 162.625| = 0.625
158	|158 - 162.625| = 4.625
145	|145 - 162.625| = 17.625
190	|190 - 162.625| = 27.375
181	|181 - 162.625| = 18.375
Total 1301	91.5
$\overline{X} = \frac{1301}{8}$ $\overline{X} = 162.625$	$\mathbf{DM} = \frac{91.5}{8} = 11.4375$

La desviación media con respecto a la media es 13.8125, esto significa que los datos están a una distancia de 13.8125 unidades por debajo o por arriba de la media aritmética en promedio.

Para datos agrupados la desviación media o desviación promedio se determina mediante la fórmula:

$$\mathrm{DM} = \frac{\sum_{i=1}^{n} \mathrm{fi}\lvert \mathrm{Mi} - \overline{\mathrm{X}} \rvert}{n}$$

Donde:

- f_i : frecuencia del i-ésimo intervalo
- M_i : marca de clase del i-ésimo intervalo
- n : es número total de datos
- $\overline{X}$: Media aritmética de los datos

En los datos agrupados la forma más práctica es mediante una tabla de frecuencias.

Ejemplo 4. La siguiente tabla muestra las horas dedicadas al trabajo en una semana de un grupo de mujeres junto con la frecuencia de cada intervalo.

Tabla 26. Horas dedicadas al trabajo en una semana y su frecuencia.

No. intervalo	Intervalo (hrs)	Frecuencia
1	10-20	3
2	20-30	7
3	30-40	10
4	40-50	5
5	50-60	2

Primero obtenemos la media aritmética $\overline{X} = \frac{\sum \text{fiMi}}{n}$

Tabla 27. Horas dedicadas al trabajo en una semana con su frecuencia y marca de clase.

No. intervalo	Intervalo	Frecuencia	M_i	f_iM_i
1	10-20	3	15	3(15) = 45
2	20-30	7	25	7(25) =73
3	30-40	10	35	10(35) =350
4	40-50	5	45	5(45) =225
5	50-60	2	55	2(27) =110
Total		27		905

$$\overline{X} = \frac{905}{27} = 33.52$$

Para obtener la desviación media completamos la siguiente tabla

Tabla 28. Cálculo de la desviación media para las horas dedicadas al trabajo en una semana.

No. intervalo	Intervalo	Frecuencia	M_i	$\lvert M_i - \overline{X}\rvert$	$f_i\lvert M_i - \overline{X}\rvert$
1	10-20	3	15	$\lvert 15-33.52\rvert=18.52$	55.56
2	20-30	7	25	$\lvert 25-33.52\rvert=8.52$	59.64
3	30-40	10	35	$\lvert 35-33.52\rvert=1.48$	14.8
4	40-50	5	45	$\lvert 45-33.52\rvert=11.48$	57.4
5	50-60	2	55	$\lvert 55-33.52\rvert=21.48$	42.96
Total		27			230.36

Entonces la desviación media está dada por:

$$DM = \frac{\sum_{i=1}^{n} fi|Mi - \overline{X}|}{n}$$

$$DM = \frac{230.36}{27} = 8.53$$

La desviación media con respecto a la media es 8.53

Los procedimientos anteriores se pueden realizar en una sola tabla, completando todas las columnas en ella, como se procederá en el siguiente ejemplo.

Ejemplo 5. La siguiente tabla muestra los ingresos semanales (en dólares) de un grupo de trabajadores junto con la frecuencia de cada intervalo.

Tabla 29. Ingresos semanales en dólares de un grupo de trabajadores y su frecuencia.

No. intervalo	Intervalo de ingresos	Frecuencia
1	100-200	5
2	200-300	8
3	300-400	12
4	400-500	15
5	500-600	10
6	600-700	6
7	700-800	4

Procedemos a completar la tabla de distribución, observando que los intervalos están dados por límites reales, es decir, no hay variación.

Tabla 30. Tabla de distribución de los Ingresos semanales en dólares de un grupo de trabajadores.

No. intervalo	Intervalo de ingresos	Frecuencia	M_i	f_iM_i	$\lvert M_i - \overline{X}\rvert$	$f_i\lvert M_i - \overline{X}\rvert$
1	100-200	5	150	750	285	1425
2	200-300	8	250	2000	185	1480
3	300-400	12	350	4200	85	1020
4	400-500	15	450	6750	15	225
5	500-600	10	550	5500	115	1150
6	600-700	6	650	3900	215	1290
7	700-800	4	750	3000	315	1260
Total		$\sum = 60$		$\sum = 26100$		$\sum = 7850$

$$\overline{X} = \frac{26100}{60} = 435$$

Entonces la desviación media se calcula por:

$$DM = \frac{\sum_{i=1}^{n} fi|Mi - \overline{X}|}{n}$$

$$DM = \frac{7850}{60} = 130.83$$

La desviación media con respecto a la media es 130. 83

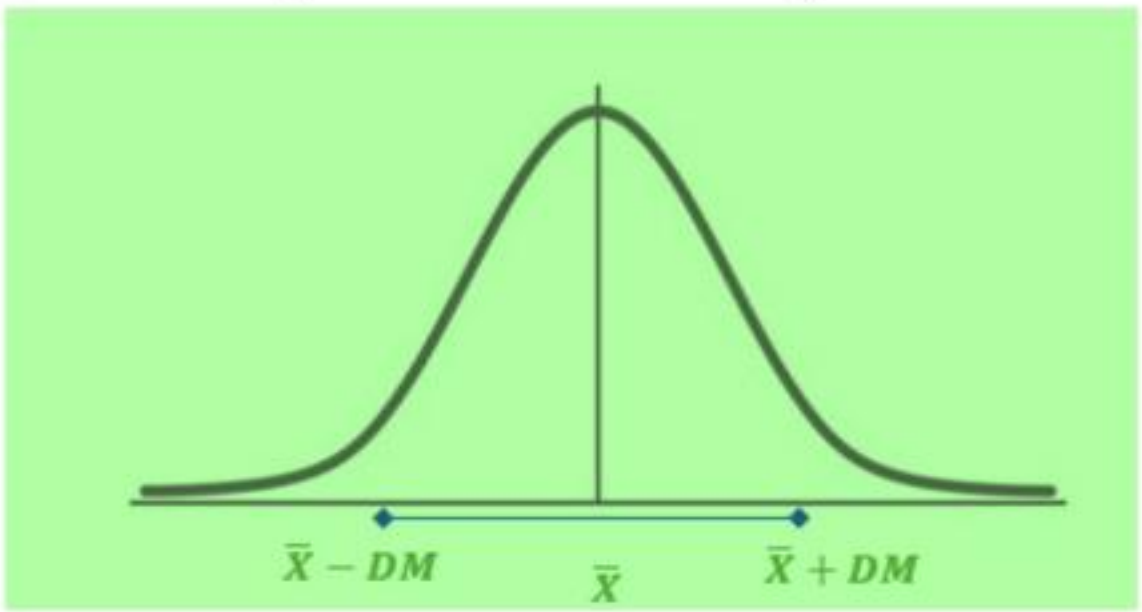
4.3.1 Interpretación de la desviación media

Definición

La desviación media o promedio presenta desventajas en cuanto a las medidas de tendencia central, una de ellas es cuando los signos de las desviaciones no se toman en cuenta (por ser valores absolutos) ya que no se muestra si la desviación está por arriba o por debajo de la media.

Si la distribución de frecuencias presenta forma simétrica y gráficamente tiene forma de campana o normal (Véase Figura 2), donde se hace notar por ejemplo que el 65% de las observaciones caen en el intervalo $\overline{X} \pm DM$, esto es que más de la mitad de las observaciones se concentran dentro del intervalo y el resto a uno y otro lado de la media aritmética.

Figura 2. Distribución de frecuencias de forma gráfica



Retomaremos el Ejemplo 1 sobre el dueño de un restaurante quien desea evaluar el número de clientes que visitan su restaurante durante cinco días. Los números son: 45, 50, 47, 55, 53. Se obtuvo una desviación media de 3.2, es decir que los datos están en promedio a una distancia de 3.2 respecto a la media $\overline{X}$ = 50

También podemos determinar que en el intervalo (calculado como la media menos la desviación media para el extremo izquierdo y la media más la desviación media para el extremo derecho), es decir, 46.8 y 53.2 contiene el 60% de los datos porque los datos de 45 y 55 clientes registrados quedan fuera del intervalo.

En el ejemplo 3 de datos agrupados sobre los ingresos semanales (en dólares) de un grupo de trabajadores (ver tabla 31).

Tabla 31. Ingresos semanales en dólares de un grupo de trabajadores y su frecuencia.

No. intervalo	Intervalo de ingresos	Frecuencia
1	100-200	5
2	200-300	8
3	300-400	12
4	400-500	15
5	500-600	10
6	600-700	6
7	700-800	4

Se obtuvo una desviación media 130.83 de, por lo que los datos están en promedio a 130.83 unidades de distancia con respecto a la media.

4.3.2 Varianza y desviación estándar

Definición

Para datos no agrupados, la varianza es la suma de los cuadrados de las diferencias del i-ésimo dato con la media aritmética y se divide entre el total de datos. Se denota por σ^2 y su fórmula es:

$$\sigma^2 = \frac{\sum (x_i - \overline{X})^2}{n}$$

Donde:

x_i: es el i-ésimo dato

n: es número total de datos

$\overline{X}$: Media aritmética de los datos

Cabe mencionar que si en la fórmula en lugar de dividir por n se divide por $n - 1$, entonces se está calculando o estimando de una población muestra. Se sabe que si $n > 30$ entonces no hay diferencia entre los denominadores.

Ejemplo 6. Varianza en el estudio del dueño de un restaurante que desea evaluar el número de clientes que visitan su restaurante durante cinco días.

La cantidad de clientes registrados es: 45, 50, 47, 55 y 53. Se cálculo que $\overline{X} = 50$. Ahora calcularemos la varianza

$$\sigma^2 = \frac{\sum (x_i - \overline{X})^2}{n}$$

$$= \frac{(45-50)^2 + (50-50)^2 + (47-50)^2 + (55-50)^2 + (53-50)^2}{5}$$

$$= \frac{25+0+9+25+9}{5} = \frac{68}{8} = 13.6$$

La varianza del número de clientes es 13.6, como esta cantidad es pequeña se puede decir que los datos no están muy dispersos. En particular todos los datos están en el intervalo [36.4, 63.6]

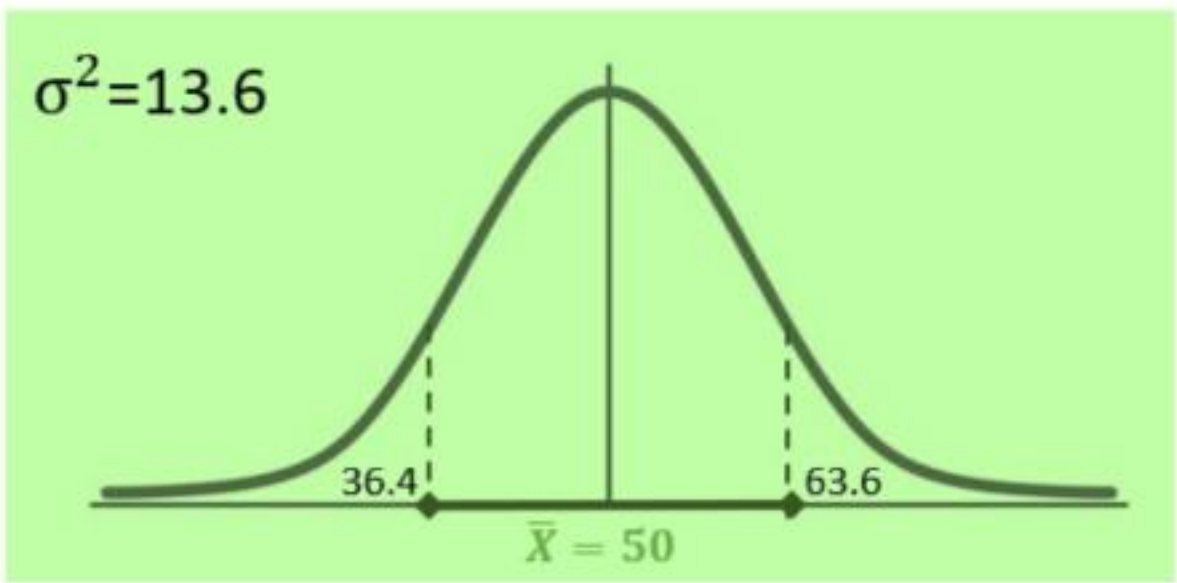


Figura 3. Datos están en el intervalo [36.4, 63.6]

Ejemplo 7. Varianza en el estudio del economista que evalúa el precio de las acciones de una empresa durante ocho días.

Los precios (en dólares) de cierre de cada día son: 150, 155, 160, 162, 158, 145, 190, 181.

Para calcular la varianza tenemos lo siguiente:

	Precios en dólares	$(X_i - \overline{X})^2$
	150	$(150 - 162.625)^2 = 159.390625$
	155	$(155 - 162.625)^2 = 58.140625$
	160	$(160 - 162.625)^2 = 6.890625$
	162	$(162 - 162.625)^2 = 0.390625$
	158	$(158 - 162.625)^2 = 21.390625$
	145	$(145 - 162.625)^2 = 310.640625$
	190	$(190 - 162.625)^2 = 749.390625$
	181	$(181 - 162.625)^2 = 337.640625$
Total	**1301**	**1643.875**

Tabla 32. Cálculo de la varianza para el estudio del economista.

De la tabla se tiene que $\Sigma x_i = 1301$, entonces

$$\overline{X} = \frac{1301}{8} = 162.625$$

Entonces la varianza se obtiene así:

$$\sigma^2 = \frac{\Sigma(x_i - \overline{X})2}{8} = \frac{1643.875}{8} = 205.484375$$

La varianza de los precios de las acciones en dólares es 361.67, indica que los precios de las acciones tienen una dispersión moderada alrededor de la media de 150.125 USD.

Definición

Para datos agrupados la varianza se calcula mediante la fórmula:

$$\sigma^2 = \frac{\sum f_i (M_i - \overline{X})^2}{n}$$

Donde:
f_i: frecuencia del i-ésimo intervalo
M_i: marca de clase del i-ésimo intervalo
n : número total de datos
$\overline{X}$: Media aritmética de los datos

tienen una dispersión moderada alrededor de la media de 150.125 USD.

No. intervalo	Intervalo	Frecuencia
1	60-70	5
2	70-80	8
3	80-90	12
4	90-100	5

Tabla 33. Calificaciones agrupadas en intervalos reales con sus respectivas frecuencias.

Ejemplo 8. Un profesor quiere calcular la varianza de las calificaciones de un grupo de estudiantes. Las calificaciones se agrupan en los siguientes intervalos con sus respectivas frecuencias.

Para calcular la varianza completamos la siguiente tabla (los intervalos ya son reales)

No. intervalo	Intervalo (horas)	Frecuencia (número de alumnos)	M_i	$f_i M_i$	$(M_i - \bar{X})^2$	$f_i(M_i - \bar{X})^2$
1	60-70	5	65	325	245.60	1228
2	70-80	8	75	600	32.11	256.88
3	80-90	12	85	1020	18.68	224.16
4	90-100	5	95	475	204.02	1020.10
Total		$\sum = 30$		$\sum = 2420$		$\sum = 2729.14$

Tabla 34. Cálculo de la varianza para las calificaciones.

Calculamos la varianza M_i

$$\sigma^2 = \frac{\sum f_i (M_i - \bar{X})^2}{n}$$

$$= \frac{2729.14}{30}$$

$$\sigma^2 = 90.97$$

La varianza de las calificaciones es 90.97, indica que las calificaciones de los estudiantes varían considerablemente alrededor de la media de 80.67

Ejemplo 9. Un jardinero quiere analizar la dispersión de las alturas de las plantas en su jardín. Las alturas (en cm) se agrupan en los siguientes intervalos con sus respectivas frecuencias.

No. intervalo	Intervalo	Frecuencia
1	10-11	3
2	11-12	7
3	12-13	10
4	13-14	5
5	14-15	4
6	15-16	1

Tabla 35. Alturas de las plantas con sus respectivas frecuencias

Para calcula la varianza completaremos la siguiente tabla

No. intervalo	Intervalo (horas)	Frecuencia (número de alumnos)	Freciencia Acumulada	M_i	$f_i M_i$	$(M_i - \overline{X})^2$	$f_i (M_i - \overline{X})^2$
1	10-11	3	3	10.5	31.8	4.41	13.23
2	11-12	7	10	11.5	80.5	1.21	8.47
3	12-13	10	20	12.5	125	0.01	0.1
4	13-14	5	25	13.5	67.5	0.81	4.05
5	14-15	4	29	14.5	58	3.61	14.44
6	15-16	1	30	15.5	15.5	8.41	8.41
Total		$\sum = 30$			$\sum = 378$		$\sum = 48.7$

Tabla 36. Cálculo de la varianza para las alturas de las plantas

Calculamos la varianza

$$\sigma^2 = \frac{\sum f_i (M_i - \overline{X})^2}{n}$$

$$= \frac{48.7}{30}$$

$$\sigma^2 = 1.62$$

La varianza de las alturas de las plantas es de 1.62, indica que las alturas de las plantas tienen una dispersión moderada alrededor de la media de 12.6 cm. En otras palabras, hay cierta variabilidad en las alturas de las plantas, sin embargo, la dispersión no es muy amplia.

Desviación típica o estándar

Definición

Para un conjunto de datos no agrupados la desviación típica o estándar se define como la raíz cuadrada del cuadrado de la media de los cuadrados de las desviaciones respecto a la media, esto es la raíz de la varianza. La fórmula es:

$$\sigma = \sqrt{\frac{\sum (x_i - \overline{X})^2}{n}}$$

Donde:
x_i: el i-ésimo dato
n: número total de datos
$\overline{X}$: Media aritmética de los datos

Definición

Para datos agrupados se calcula como:

$$\sigma = \sqrt{\frac{\sum fi (M_i - \overline{X})^2}{n}}$$

Donde:
f_i: frecuencia del i-ésimo intervalo
M_i: marca de clase del i-ésimo intervalo
n: número total de datos
$\overline{X}$: Media aritmética de los datos

El número de clientes registrados durante cinco días es: 45, 50, 47, 55 y 53. Se obtuvo **la varianza del número de clientes con un valor de 13.6.**

$$\sigma^2 = 13.6$$

Entonces **la desviación típica es:**

$$\sigma^2 = \sqrt{13.6} = 3.69$$

Ejemplo 10. Calcular la desviación típica o estándar del ejemplo referente al dueño del restaurante que desea evaluar el número de clientes que visitan su establecimiento.

Ejemplo 11. Calcular la desviación típica o estándar del ejemplo referente al economista que evalúa el precio de las acciones de una empresa durante ocho días.

Los precios (en dólares) de cierre de cada día son: 150, 155, 160, 162, 158, 145, 190 y 181. Se obtuvo **la varianza de los precios en dólares con un valor de 361.67**

$$\sigma^2 = 205.48$$

Entonces **la desviación típica es:**

$$\sigma = \sqrt{205.48} = 14.33$$

PARA DATOS NO AGRUPADOS

Ejemplo 12. Calcular la desviación típica o estándar del ejemplo referente al profesor que calcula la varianza de las calificaciones de un grupo de estudiantes.

Tabla 37. Calificaciones de los estudiantes agrupados en intervalos con sus respectivas frecuencias

No. intervalo	Intervalo	Frecuencia
1	60-70	5
2	70-80	8
3	80-90	12
4	90-100	5

Se obtuvo que **la varianza de las calificaciones es 90.97**

$$\sigma^2 = 90.97$$

Entonces **la desviación típica es:**

$$\sigma = \sqrt{90.97} = 9.06$$

Ejemplo 13. Calcular la desviación típica o estándar del ejemplo referente al jardinero que desea analizar la dispersión de las alturas de las plantas en su jardín.

Tabla 38. Alturas (en cm) agrupadas en intervalos con sus respectivas frecuencias

No. intervalo	Intervalo	Frecuencia

µTorrent

A new stable version of µTorrent 3.6.0 (build 47134) is available. Would you like to download and install it?

New in this version:
- Stability improvements
- Security fixes

By clicking "Yes", you acknowledge that you have read and agree to the End User License Agreement and Privacy Policy.

This will also download the latest translations.

Always check for updates to this version

Yes

No

Se obtuvo

Entonces

central

4.3.2.1

La varianz

que propo

ce a la

datos en re

del contex

indica may

endidas

dispersión

n típica

mayor dive

tmética.

una varianza baja sugiere una mayor homogeneidad o menor dispersión.

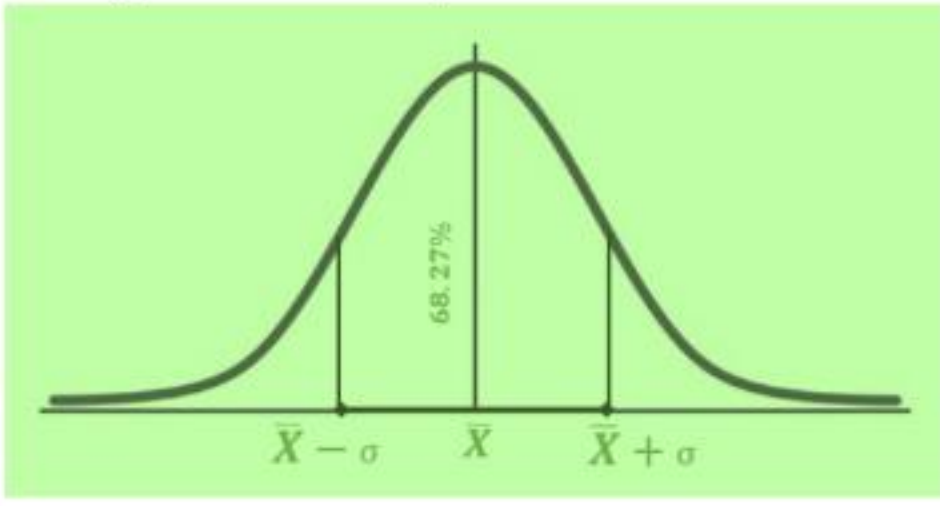
Además, representa la base para calcular la desviación estándar al eliminar las cantidades elevadas al cuadrado por medio de la raíz.

4.3.2.2 Interpretación de la desviación estándar

Interpretación de la Desviación Estándar en los Intervalos:

Para la **interpretación de la desviación estándar** es importante hacer referencia a la distribución normal que se caracteriza por ser un patrón estadístico cuando un conjunto de datos se distribuye de manera uniforme

Figura 4. Distribución normal para el 68.27% de las observaciones



- El 95.45% de las observaciones quedan comprendidas en $\overline{X} \pm 2\sigma$; es decir, el valor de la desviación típica o estándar a ambos lados de la media aritmética. Gráficamente se representa como:

Figura 5. Distribución normal para el 95.45% de las observaciones

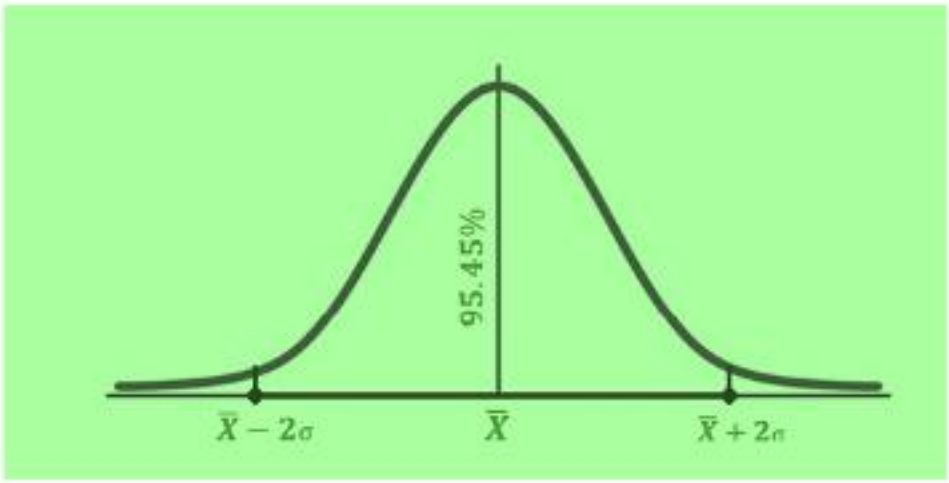
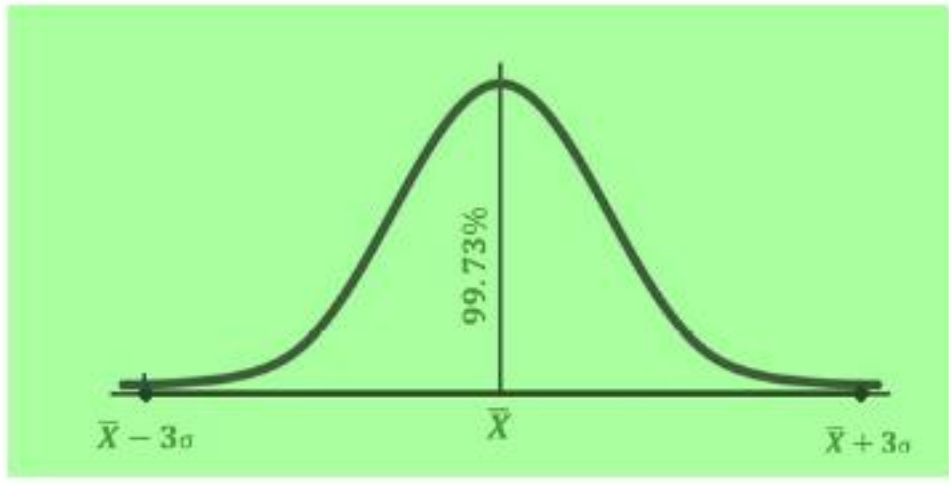


Figura 6. Distribución normal para el 99.73% de las observaciones



- El 99.73% de las observaciones quedan comprendidas en $\overline{X} \pm 3\sigma$; es decir, el valor de la desviación típica o estándar a ambos lados de la media aritmética. Gráficamente se representa como:

En los siguientes ejemplos encontraremos los intervalos que contienen el 68.27%, 95.45% y 99.73% de según la media y desviación que se obtenga.

Ejemplo 14. Supongamos que tenemos el siguiente conjunto de datos no agrupados 5,7,8,9,10

Calcular la media aritmética $\overline{X}$

$$\overline{\mathbf{X}} = \frac{5+7+8+9+10}{5} = 7.8$$

- **Calcular la suma de las desviaciones absolutas**

$$\sum_{j=1}^{5} |x - \overline{X}| = |5-7.8| + |7-7.8| + |8-7.8| + |9-7.8| + |10-7.8| = 28 + 0.8 + 0.2 + 1.2 + 22$$

- **Calcular la desviación media**

$$\mathbf{DM} = \frac{\mathbf{7.2}}{\mathbf{5}} = \mathbf{1.44}$$

La desviación media con respecto a la media es 1.44

- **Calcular la varianza**

$$\sigma^2 = \frac{\Sigma(x_i - \overline{X})^2}{n} = \frac{(5-7.8)^2 + (7-7.8)^2 + (8-7.8)^2 + (9-7.8)^2 + (10-7.8)^2}{5}$$

$$= \frac{7.84 + 0.64 + 0.04 + 1.44 + 4.84}{5} = \frac{14.8}{8} = \mathbf{2.96}$$

- **Calcular la desviación estándar**

$$\sigma = \sqrt{\mathbf{2.96}} = \mathbf{1.72}$$

Interpretación de la Desviación Estándar

Como $\overline{X} = 7.8$ y $\sigma = 1.72$ entonces el:

a. 68.27% de los datos (aproximadamente) están en el intervalo 7.8±1.72, es decir, [6.08,9.52].
b. 95.45% de los datos (aproximadamente) están en el intervalo 7.8±2(1.72), es decir, [4.36,11.24].
c. 99.73% de los datos (aproximadamente) están en el intervalo 7.8±3(1.72), es decir, [2.64,12.96].

Ejemplo 10. Retomemos el ejemplo del jardinero que desea analizar la dispersión de las alturas de las plantas en su jardín. Las alturas (en cm) se agrupan en los siguientes intervalos con sus respectivas frecuencias:

Tabla 39. Alturas (en cm) agrupadas en intervalos con sus respectivas frecuencias

No. intervalo	Intervalo	Frecuencia
1	10-11	3
2	11-12	7
3	12-13	10
4	13-14	5
5	14-15	4
6	15-16	1

Se obtuvo que **la varianza** de las alturas de las plantas **es de 1.62**

$$\sigma^2 = 1.62$$

Entonces **la desviación típica es:**

$$\sigma = \sqrt{1.62} = 1.27$$

Interpretación de la Desviación Estándar en los Intervalos:

Como se obtuvo que la $\overline{X} = 12.6$ y la $\sigma = 1.27$ entonces el:

a. 68.27% de los datos (aproximadamente) están en el intervalo 12.6 ±1.27, es decir, [11.33,13.87].
b. 95.45% de los datos (aproximadamente) están en el intervalo 12.6 ± 2(1.27), es decir, [10.06,15.14].
c. 99.73% de los datos (aproximadamente) están en el intervalo 12.6±3(1.27), es decir, [8.79,16.41].

Esto ayuda a entender que la mayoría de los datos en una distribución normal estarán cerca de la media, y podemos usar la desviación estándar para cuantificar y predecir esta dispersión.

Actividad 4.



Ejercicios de desviación media

Instrucciones:
te invitamos a **descargar** el archivo adjunto, dando clic en la imagen, imprímelo y encuentra los intervalos que contiene los porcentajes aproximados 68.27%, 95.45% y 99.73% del total de los datos.



4.4 Diagrama de dispersión y centro de gravedad

Definición

Como ya se ha mencionado en el tema 2.4 diagramas de dispersión, este tipo de gráficos permiten identificar visualmente el tipo de correlación que existe entre las variables, como se muestra a continuación:

Figura 9. Diagramas de dispersión por la fuerza de la correlación

a. Por la correlación entre las variables

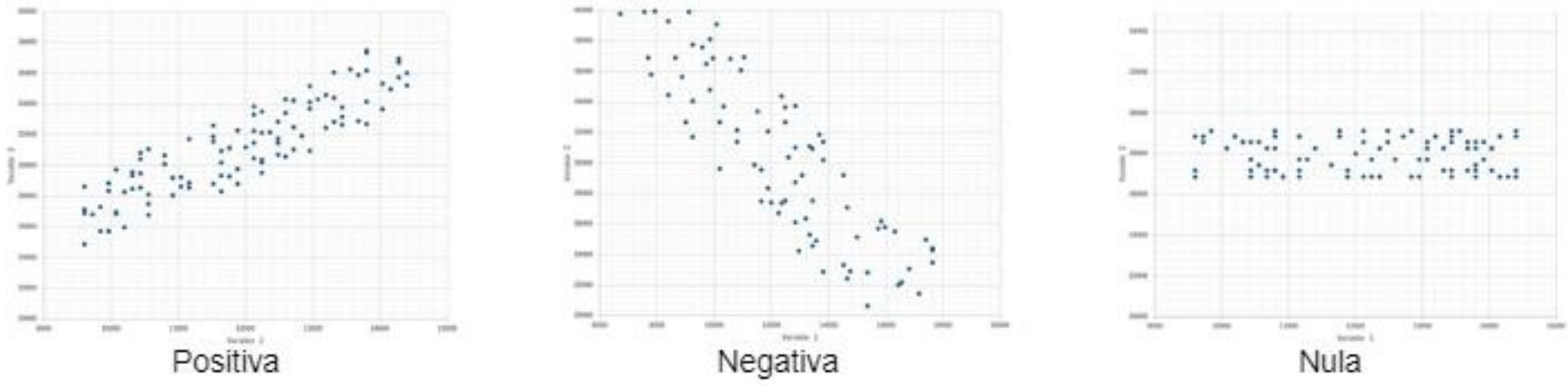


Figura 7. Diagramas de dispersión por la correlación entre las variables

b. Por el tipo de correlación

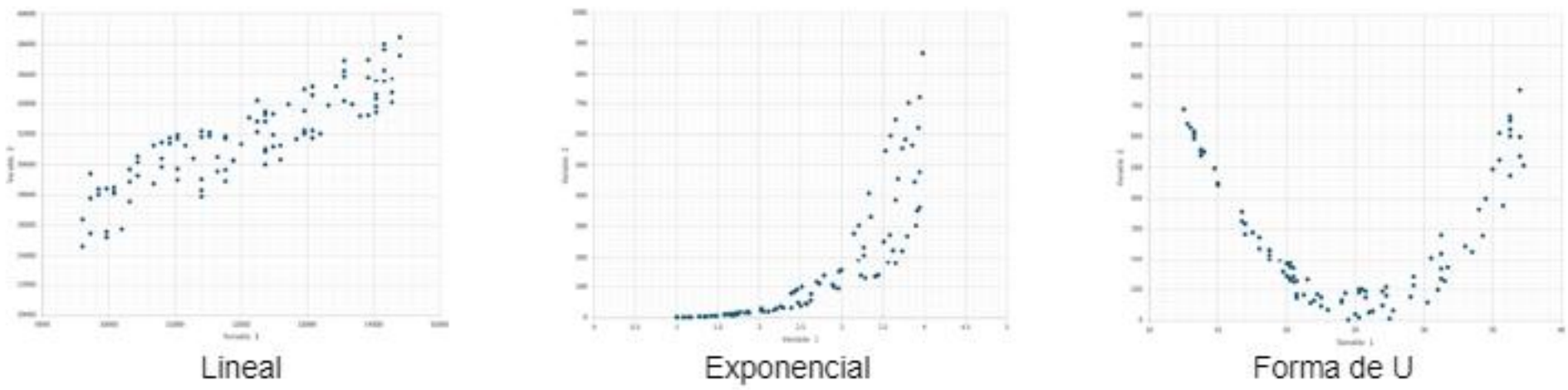
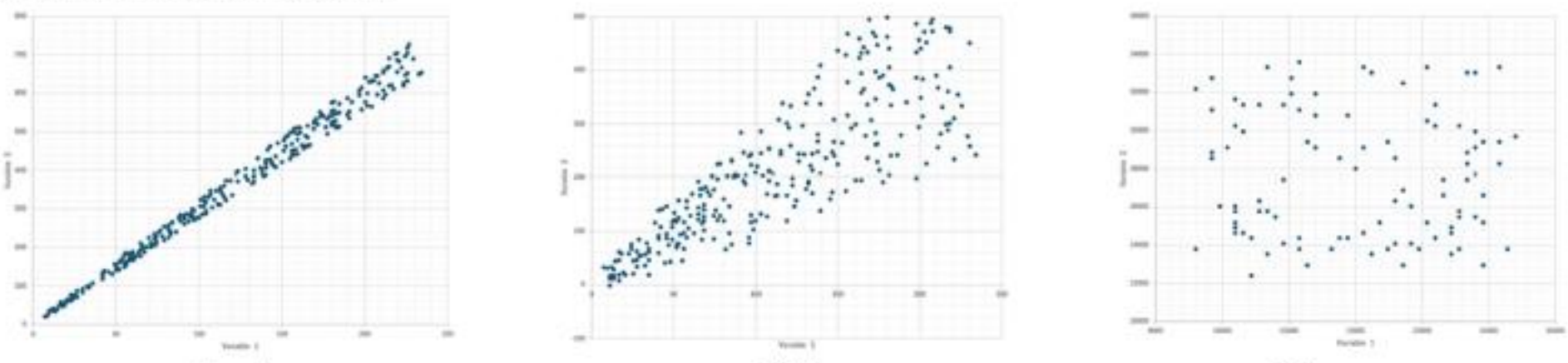


Figura 8. Diagramas de dispersión por el tipo de correlación

b. Por el tipo de correlación

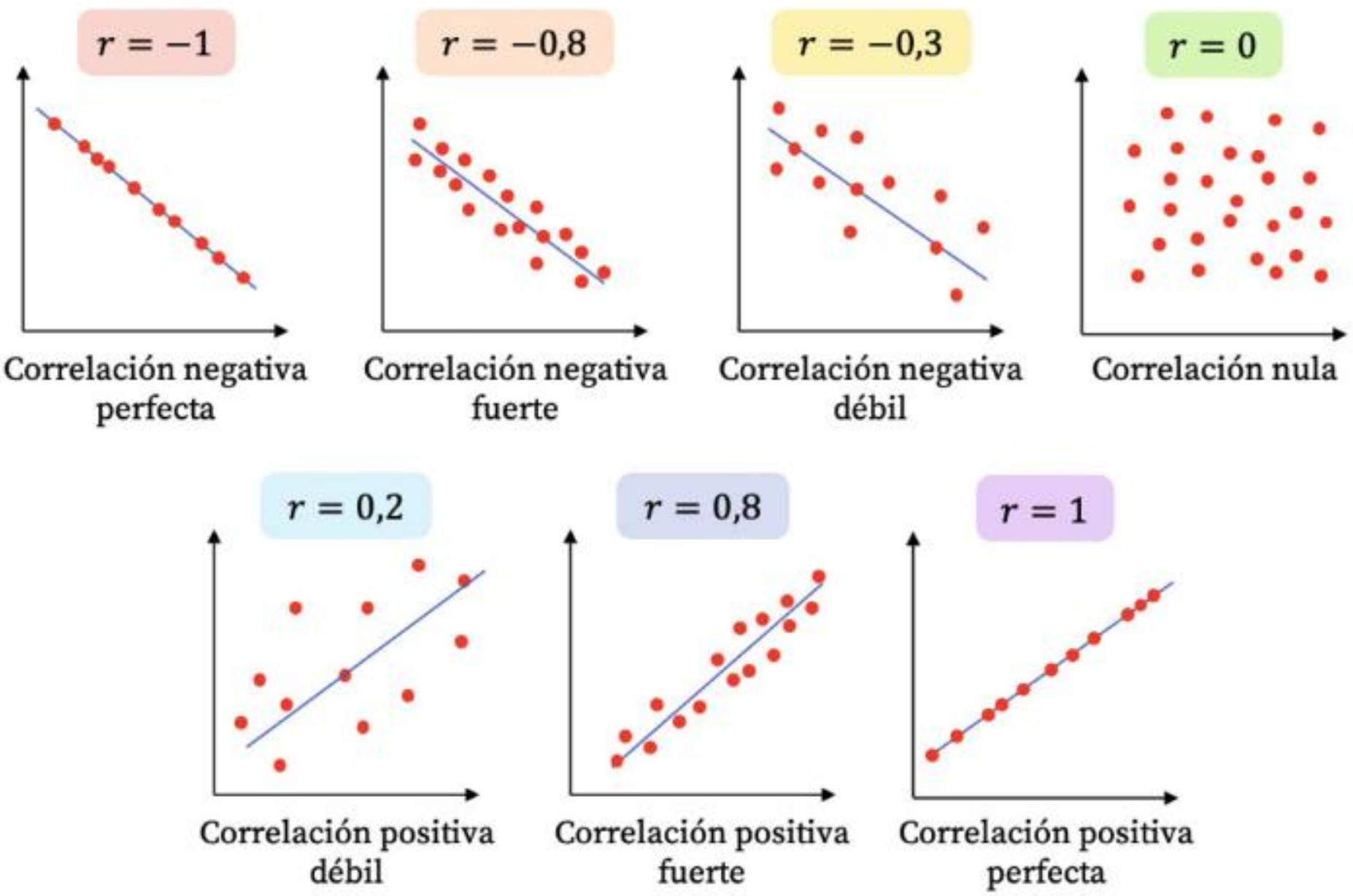


El coeficiente de correlación se puede calcular con la fórmula vista en el tema 2.5. Coeficiente de correlación:

$$r = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\sqrt{\Sigma(X-\overline{X})^2\Sigma(Y-\overline{Y})^2}}$$

Para un conjunto de datos con relación lineal, el coeficiente de correlación se puede interpretar de la siguiente forma:

Figura 10. Coeficiente de correlación



Recordemos también que, a partir de la tabla de datos con comportamiento lineal, se puede calcular la recta de regresión que los representa con la siguiente expresión:

Ecuación de la recta de regresión

$$y = mx + b$$

Donde:

$$m = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\Sigma(X-\overline{X})^2} \qquad b = \overline{Y} - m\overline{X}$$

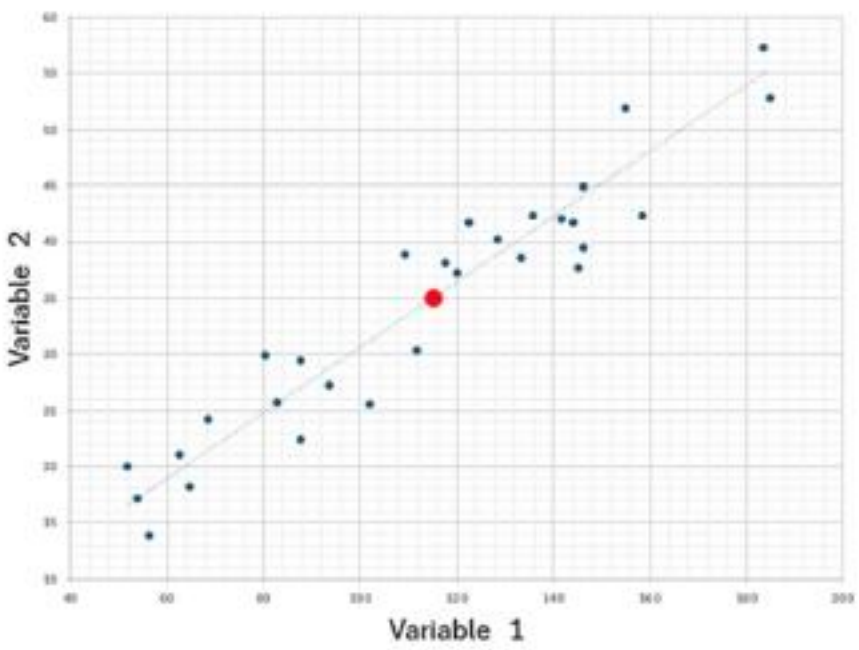
Por otra parte, el centro de gravedad corresponde al punto cuyas coordenadas son la media aritmética de los datos de cada variable y por este punto pasa la recta de regresión.

Coordenadas del centro de gravedad

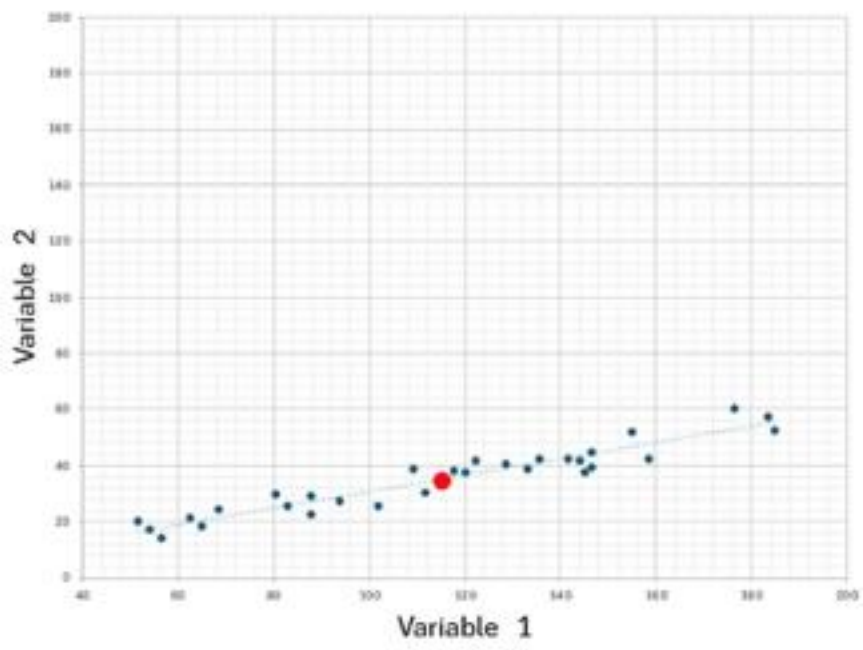
$$(\overline{X}, \overline{Y})$$

Para trazar un diagrama de dispersión se emplea un plano cartesiano en el cual se debe elegir la escala correcta para las variables, ya que una escala mal elegida puede causar una incorrecta interpretación de los datos como se muestra a continuación:

Figura 11. Diagramas de dispersión a escala



Con la escala empleada de forma correcta se observa una correlación positiva fuerte.



Con la escala empleada de forma incorrecta se observa una correlación cercana a nula.

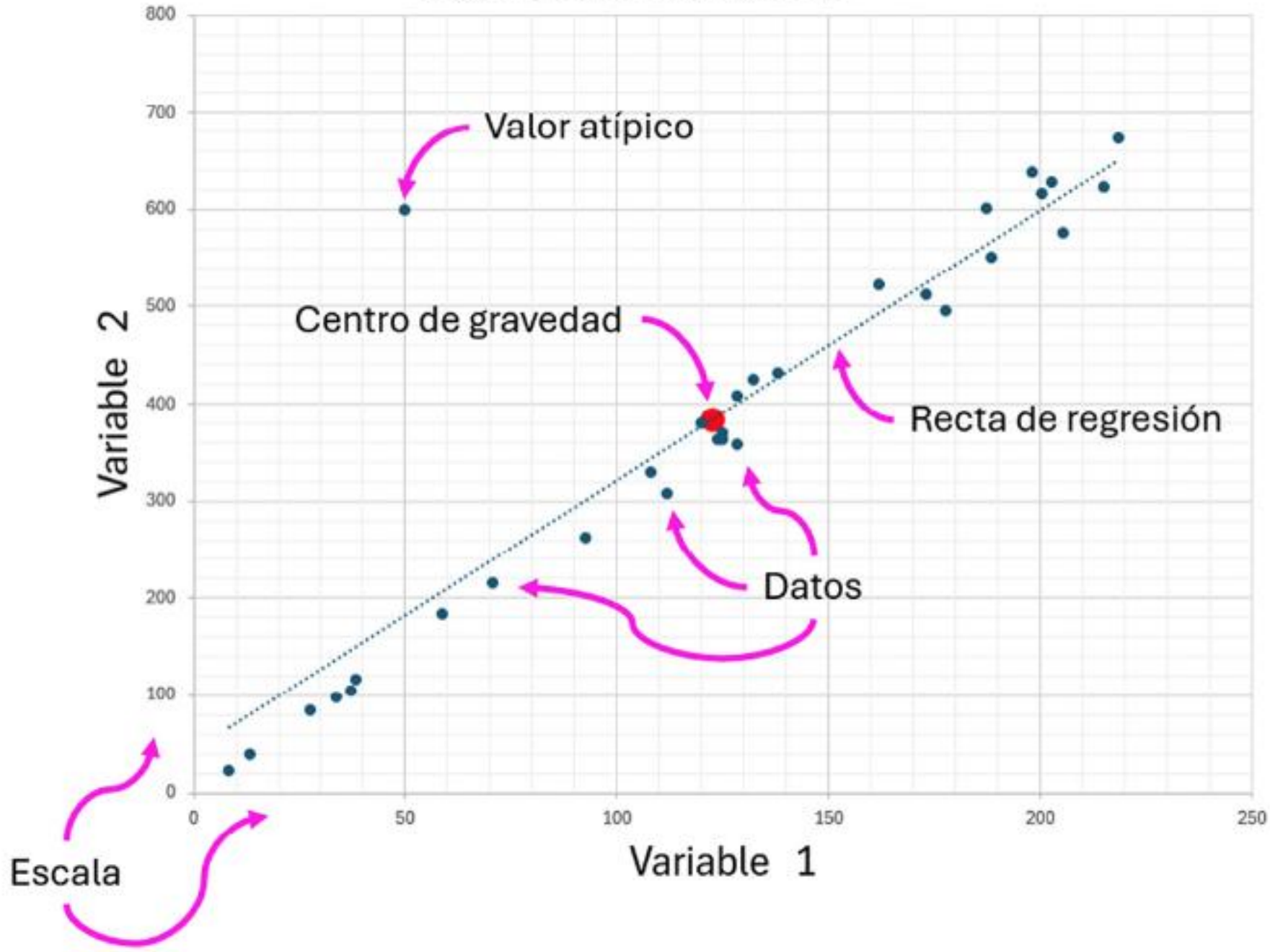
Ocasionalmente se observan valores atípicos, que son aquellos que distan del resto y se alejan de forma excesiva de la recta de regresión; las causas de los valores atípicos pueden ser:

- errores en la medición
- comportamiento anómalo de un individuo en el estudio
- acontecimientos extraordinarios
- causas desconocidas

Los valores atípicos suelen ser únicos, por lo que algunos estudios los descartan; sin embargo, en caso de ser repetitivos, es importante identificar la causa y darle un tratamiento más adecuado al comportamiento de las variables.

Con todos estos antecedentes, un diagrama de dispersión puede incluir generalmente los siguientes elementos:

Figura 12. Elementos de un diagrama de dispersión



Ejemplo 1. Una constructora desea participar en una licitación pública de un conjunto residencial. Un requisito consiste en definir en cuánto tiempo su equipo de trabajo podría concluir la construcción de una casa. Para ello, tomó un registro de las últimas 16 casas construidas por su personal, obteniendo los siguientes datos:

Tabla 40. Números de trabajadores vs días en construir una casa.

Número de trabajadores	Días en construir una casa
10	70
11	68
10	77
9	78
11	66
11	64
9	76
15	52
14	50
8	80

Número de trabajadores	Días en construir una casa
9	75
12	60
15	47
10	69
14	50
12	54

Con esta información determina:

a. ¿Influye que en algunas casas han participado más o menos trabajadores?

b. ¿Cuál sería el número de trabajadores y el tiempo que tardarían en construir una casa que reportarías para concursar en la licitación pública?

c. Representa el diagrama de dispersión y el centro de gravedad del conjunto de datos.

Solución

a. ¿Influye que en algunas casas han participado más o menos trabajadores?

Para determinar esta relación se determinará el coeficiente de correlación de los datos, para ello se realiza la tabla que contiene las siguientes columnas, tal y como se analizó en el tema 2.5, coeficiente de correlación.

Tabla 41. Cálculo del coeficiente de correlación

X	Y	$X-\overline{X}$	$Y-\overline{Y}$	$(X-\overline{X})(Y-\overline{Y})$	$(X-\overline{X})^2$	$(Y-\overline{Y})^2$
10	70	-1.3	5.3	-6.5625	1.5625	27.5625
11	68	-0.3	3.3	-0.8125	0.0625	10.5625
10	77	-1.3	12.3	-15.3125	1.5625	150.0625
9	78	-2.3	13.3	-29.8125	5.0625	175.5625
11	66	-0.3	1.3	-0.3125	0.0625	1.5625
11	64	-0.3	-0.8	0.1875	0.0625	0.5625
9	76	-2.3	11.3	-25.3125	5.0625	126.5625
15	52	3.8	-12.8	-47.8125	14.0625	162.5625
14	50	2.8	-14.8	-40.5625	7.5625	217.5625
8	80	-3.3	15.3	-49.5625	10.5625	232.5625
9	75	-2.3	10.3	-23.0625	5.0625	105.0625
12	60	0.8	-4.8	-3.5625	0.5625	22.5625
15	47	3.8	-17.8	-66.5625	14.0625	315.0625
10	69	-1.3	4.3	-5.3125	1.5625	18.0625
14	50	2.8	-14.8	-40.5625	7.5625	217.5625
12	54	0.8	-10.8	-8.0625	0.5625	115.5625
$\sum = 180$	$\sum = 1036$			$\sum = -363$	$\sum = 75$	$\sum = 1899$

El coeficiente de correlación se calcula con la siguiente fórmula:

$$r = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\sqrt{\Sigma(X-\overline{X})^2\Sigma(Y-\overline{Y})^2}}$$

$$r = \frac{-363}{\sqrt{(75)(1899)}}$$

$$r = \frac{-363}{\sqrt{142425}}$$

$$r = \frac{-363}{377.3924}$$

$$\mathbf{r = -0.9619}$$

Este valor muestra una fuerte correlación lineal negativa, por lo que, independientemente de que no en todas las casas han participado el mismo número de trabajadores, ellos trabajan a un ritmo similar. El valor negativo expresa que, si trabajan más personas en la obra, entonces tardan menos días en terminar una casa, de forma análoga, si trabajan menos personas tardan más días.

b. ¿Cuál sería el número de trabajadores y el tiempo que tardarían en construir una casa que reportarías para concursar en la licitación pública?

Dado que la correlación de los datos es fuerte, el valor más conveniente a reportar es el promedio de cada variable, es decir, el centro de gravedad.

Centro de gravedad = $(\overline{X}, \overline{Y})$

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{180}{16} = 11.25$$

$$\overline{Y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{1036}{16} = 64.75$$

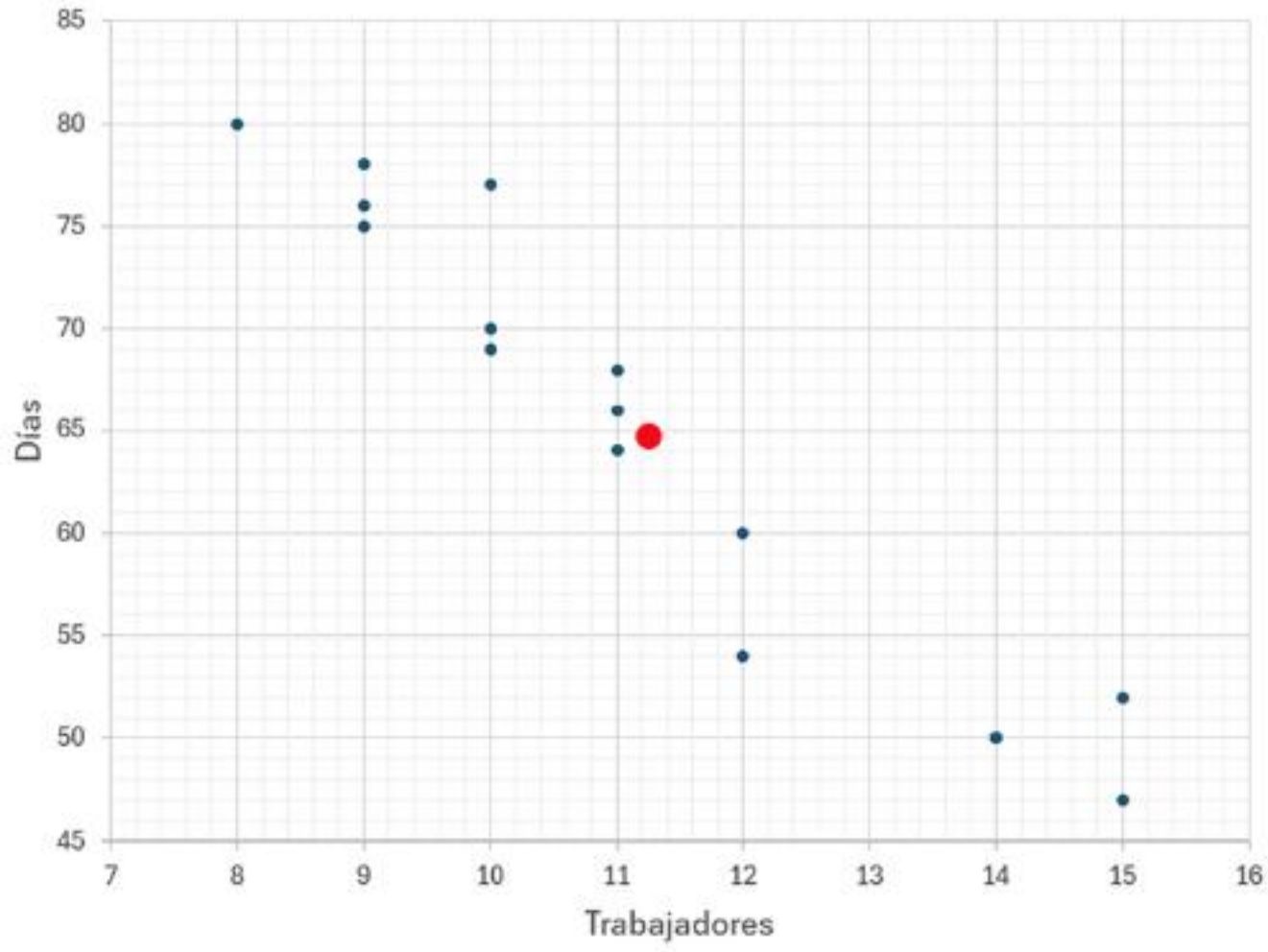
$$(\overline{X}, \overline{Y}) = (\mathbf{11.25, 64.75})$$

Considerando que una cantidad de personas es un valor discreto, es decir, no puede haber media persona o dos tercios de persona, el número de trabajadores se redondea a 12, mientras que el número de días, de forma similar, se redondea a 65 días.

Conviene reportar que se requieren 12 personas para terminar una casa en 65 días.

c. Representa el diagrama de dispersión y el centro de gravedad del conjunto de datos.

Figura 13. Diagrama de dispersión y centro de gravedad



Ejemplo 2. Una empresa de lácteos registró la cantidad de kilogramos que consumía cada una de sus 20 vacas y la cantidad de litros de leche que producían. Los datos obtenidos se muestran en la siguiente tabla:

Tabla 42. Cantidad de forraje vs leche producida por vacas.

Identificador	Forraje (kg)	Leche producida (L)
1	30.9	32.30
2	28.0	28.38
3	30.4	30.00
4	28.5	28.41
5	27.6	27.79
6	29.7	30.83
7	26.1	26.82
8	29.1	29.80
9	31.9	31.58
10	28.3	28.59
11	29.9	30.82
12	27.5	27.78
13	27.3	28.05
14	29.2	21.52
15	27.8	28.56
16	31.4	31.72
17	26.3	25.95
18	27.3	26.76
19	32.0	31.37
20	26.7	28.40

Con esta información realiza y contesta lo que se te solicita:

a. Calcula las coordenadas del centro de gravedad.
b. Obtén la recta de regresión que representa al conjunto de datos.
c. Con la escala adecuada representa los datos en un diagrama de dispersión que incluya al centro de gravedad y la recta de regresión.
d. Determina si existe un valor atípico y, de ser así, a cuál identificador pertenece.
e. Describe cómo se correlacionan los datos.
f. A partir de la ecuación de la recta de regresión, determina ¿Cuántos litros de leche podría producir una vaca que consume 27 kg de forraje?

Solución

Para poder cubrir algunos requisitos del problema se realiza la tabla que contiene las columnas siguientes, tal y como se analizó en el tema 2.5, coeficiente de correlación.

Tabla 43. Cálculo del coeficiente de correlación.

X	Y	$X-\overline{X}$	$Y-\overline{Y}$	$(X-\overline{X})(Y-\overline{Y})$	$(X-\overline{X})^2$	$(Y-\overline{Y})^2$
30.9	32.30	2.1	3.5	7.4274	4.4310	12.4499
28.0	28.38	-0.8	-0.4	0.3123	0.6320	0.1543
30.4	30.00	1.6	1.2	1.9722	2.5760	1.5099
28.5	28.41	-0.3	-0.4	0.1079	0.0870	0.1339
27.6	27.79	-1.2	-1.0	1.1782	1.4280	0.9721
29.7	30.83	0.9	2.1	1.8636	0.8190	4.2405
26.1	26.82	-2.7	-2.0	5.2718	7.2630	3.8265
29.1	29.80	0.3	1.0	0.3123	0.0930	1.0486
31.9	31.58	3.1	2.8	8.7342	9.6410	7.9127
28.3	28.59	-0.5	-0.2	0.0917	0.2450	0.0344
29.9	30.82	1.1	2.1	2.2692	1.2210	4.2170
27.5	27.78	-1.3	-1.0	1.2865	1.6770	0.9869
27.3	28.05	-1.5	-0.7	1.0813	2.2350	0.5231
29.2	21.52	0.4	-7.3	-2.9367	0.1640	52.5800
27.8	28.56	-1.0	-0.2	0.2085	0.9900	0.0439
31.4	31.72	2.6	2.9	7.6742	6.7860	8.6787
26.3	25.95	-2.5	-2.8	7.0291	6.2250	7.9369
27.3	26.76	-1.5	-2.0	2.9997	2.2350	4.0260
32.0	31.37	3.2	2.6	8.3373	10.2720	6.7670
26.7	28.40	-2.1	-0.4	0.7688	4.3890	0.1347
$\sum = 575.9$	$\sum = 575.4$			$\sum = 56.0$	$\sum = 63.4$	$\sum = 118.2$

a. Calcula las coordenadas del centro de gravedad

$$(\overline{X},\overline{Y}) = \left(\frac{575.9}{20},\frac{575.4}{20}\right) = (28.8, 28.8)$$

b. Obtén la recta de regresión que representa al conjunto de datos

$$m = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\Sigma(X-\overline{X})^2} = \frac{56}{63.4} = 0.8833$$

$$b = \overline{Y} - m\overline{X} = 28.8 - (0.8833)(28.8) = 3.361$$

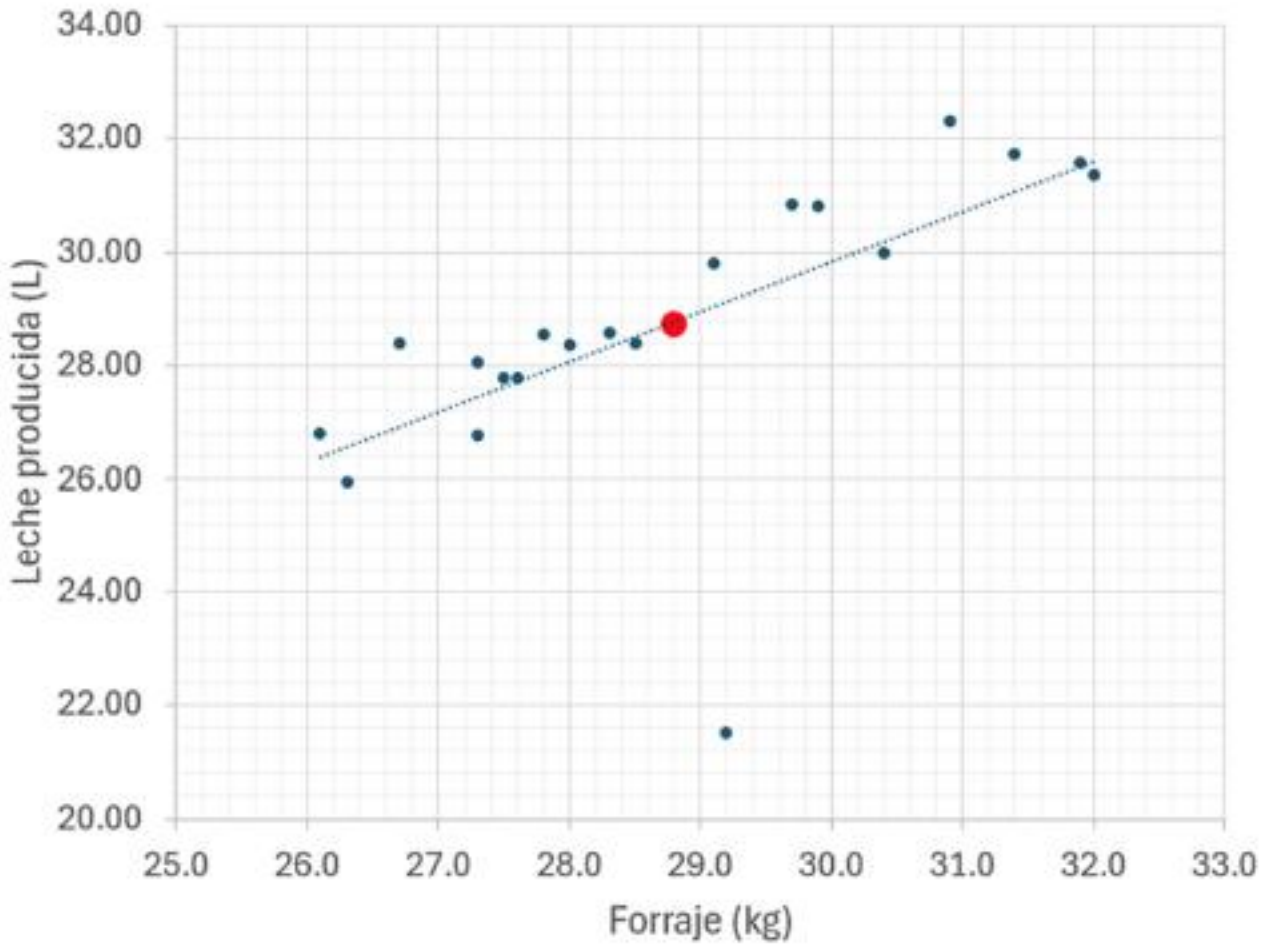
De donde:

$$y = mx + b$$

$$\mathbf{y = 0.8833x + 3.361}$$

c. Con la escala adecuada representa los datos en un diagrama de dispersión que incluya al centro de gravedad y la recta de regresión

Figura 14. Diagrama de dispersión y centro de gravedad



d. Determina si existe un valor atípico y, de ser así, a cuál identificador pertenece.

Sí existe un valor atípico, pertenece a la vaca marcada con el identificador número 14. Sin embargo, aun con la presencia de este dato, la recta de regresión y el conjunto de datos muestran una buena aproximación.

e. Describe cómo se correlacionan los datos.

El coeficiente de correlación se calcula con la siguiente fórmula:

$$r = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\sqrt{\Sigma(X-\overline{X})^2\Sigma(Y-\overline{Y})^2}}$$

$$r = \frac{56}{\sqrt{(63.4)(118.2)}}$$

$$r = \frac{56}{\sqrt{7493.88}}$$

$$r = \frac{56}{86.5672}$$

$$\mathbf{r = -0.6469}$$

Es importante resaltar que en este caso particular el valor atípico influye en el valor de coeficiente de correlación, no obstante, con base a este resultado y a la gráfica se observa que es una correlación positiva, lineal fuerte. Esto permite que los datos puedan ser empleados para predecir el comportamiento de las variables, como se requiere en el inciso f).

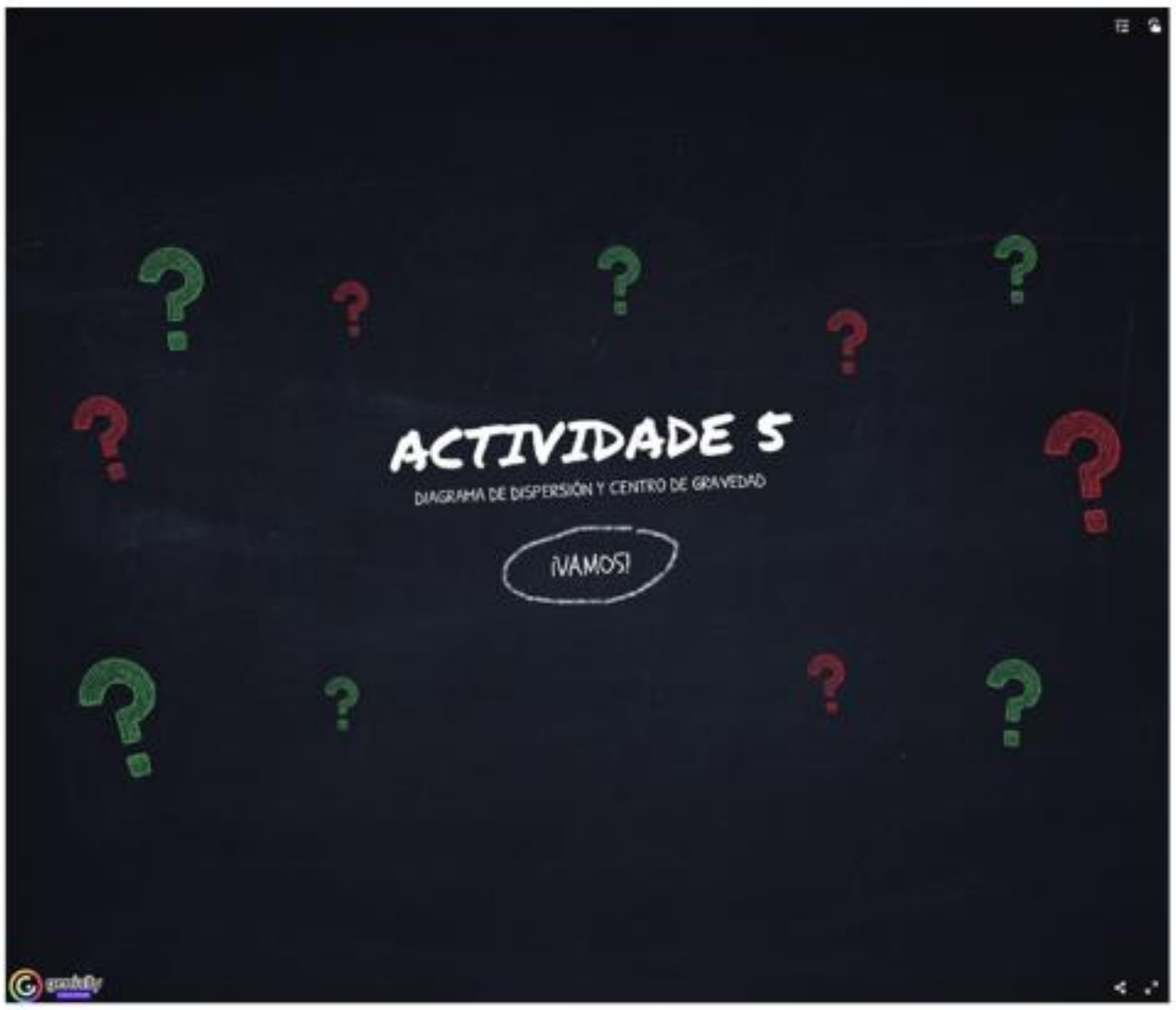
NOTA: Se deja como actividad al estudiante calcular cuál sería el valor del coeficiente de correlación si se elimina el valor atípico, considerando únicamente la producción de 19 vacas.

f. A partir de la ecuación de la recta de regresión, determina ¿Cuántos litros de leche podría producir una vaca que consume 27 kg de forraje?

Como la ecuación de la recta de regresión es basta sustituir el valor de kg de forraje:

$$y = 0.8833x + 3.361$$
$$y = 0.8833(27) + 3.361$$
$$y = 23.8491 + 3.361$$
$$y = 27.21$$

En conclusión, si una vaca consumiera 27 kg de forraje, probablemente produciría 27.21 L de leche.



Actividad 5.



Ejercicios de diagrama de dispersión y centro de gravedad

Instrucciones:
para practicar lo aprendido, te invitamos a resolver los siguientes ejercicios, para comenzar da clic en la **imagen.**

4.5 Covarianza

Definición

Covarianza
Es el valor que refleja en qué cantidad dos variables aleatorias varían de forma conjunta respecto a sus medias aritméticas. Es decir, nos permite conocer el comportamiento de las variables en cuestión respecto de otras variables.

La fórmula para determinar la covarianza es:

$$Cov(X,Y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n}$$

Donde:
$\bar{y}$: media de la variable Y
$\bar{x}$: media de la variable X
"i": posición de la observación
"n": número total de observaciones

- Ayuda a entender la relación entre dos variables, en el mercado bursátil es primordial monitorear el precio de las acciones para la toma de decisiones en finanzas e inversiones.
- Puede ser positiva (las variables son directamente proporcionales), negativa (las variables son inversamente proporcionales), o nula (no existe relación lineal entre las variables).

Covarianza

$$Cov(X,Y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n}$$

Covarianza (S_{xy})

Media aritmética o promedio $\bar{x}, \bar{y}$

$$S_{xy} = \frac{\sum xy - n\bar{x}\bar{y}}{n-1}$$

La covarianza mide la cantidad de relación lineal entre las variables y el sentido de esta, de la forma:

- $S_{xy} > 0$, relación lineal positiva (si crece una variable, la otra también)
- $S_{xy} < 0$, relación lineal negativa (si crece una variable, la otra decrece).
- $S_{xy} = 0$, no hay relación lineal entre las variables.

Propiedades de la covarianza

Sean "X" la primera variable en el estudio, "Y" nuestra segunda variable a analizar y "b y c" dos constantes.

- $Cov(X,Y) > 0$ cuando los valores de X incrementan y los de Y también, tendremos una relación positiva.
- $Cov(X,Y) < 0$ cuando los valores de X incrementan y los de Y decrementan, tendremos una relación negativa.

- $Cov\,(X,b) = 0$.
- $Cov\,(X,X) = Var\,(X)$ es decir, la covarianza de una variable y de sí misma es igual a la varianza de la variable.
- $Cov\,(X,Y) = Cov\,(Y,X)$ la covarianza no cambia si invertimos el orden de las variables en la fórmula.
- $Cov\,(bX,cY) = (cb)\,Cov\,(X,Y)$. La covarianza de dos variables multiplicadas por dos constantes cualesquiera es igual a la covarianza de las dos variables multiplicada por la multiplicación de dichas constantes.
- $Cov\,(b+X,c+Y) = Cov\,(X,Y)$ sumar dos constantes cualesquiera a cada variable, no modifica el resultado de la covarianza.

Ejemplo 1. En los cables eléctricos AWG se tiene por NOM (Norma Oficial Mexicana) que cada calibre puede hacer circular una corriente eléctrica máxima.

Los datos son los siguientes:

Tabla 44. Calibres de los conductores vs corriente máxima.

Calibre del conductor AWG	Corriente máxima (Amperes)
1	250
3	150
6	55
10	30
12	20
14	15

Determinar la covarianza entre ambas variables

Tabla 45. Cálculo de la covarianza.

Calibre del conductor AWG (X)	Corriente máxima (Amperes) (Y)	$X-\overline{X}$	$Y-\overline{Y}$	$(X-\overline{X})^2$	$(Y-\overline{Y})^2$	$(X-\overline{X})(Y-\overline{Y})$
1	250	-6.66666667	163.34	44.44	26678	-1088.88
3	150	-4.66666667	63.34	21.77	4.011	-295.55
6	55	-1.66666667	-31.66	2.77	1.003	52.77
10	30	2.333333333	-56.66	5.44	3.211	-132.22
12	20	4.333333333	-66.66	18.77	4.444	-288.88
14	15	6.333333333	-71.66	225	5.136	-453.88
				$\sum = 133.33$	$\sum = 44483$	$\sum = -2206.66$
$\overline{X} = 7.66$	$\overline{Y} = 86.66$			$\overline{x^2} = 22.22$	$\overline{y^2} = 7.41$	$Cov(X,Y) = -367.77$
$r = -0.9$						

$$Cov(X,Y) = \frac{\sum(X-\overline{X})(Y-\overline{Y})}{n} = \frac{-2206.66}{6} = -367.77$$

El coefiente de correlación es:

$$r = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\sqrt{\Sigma(X-\overline{X})^2\Sigma(Y-\overline{Y})^2}} \qquad r = \frac{-2206.66}{\sqrt{(133.3)(44483)}} = -0.9$$

El valor resultante $Cov\ (X,Y) = -367.77$ nos indica, al ser menor que cero, que las dos variables tienen una relación negativa, es importante resaltar que nos interesa el signo y no el valor obtenido. Para saber qué tan estrecha es dicha relación, hacemos referencia al coeficiente de correlación $r = -0.9$ obtenido en la tabla y observamos que, si existe una correlación lineal, todos los puntos en el gráfico se alinearán en una línea con pendiente negativa.

Figura 15. Gráfica de corriente máxima permitida en conductores AWG



Ejemplo 2. De acuerdo con la página **statista,** los ingresos anuales estimados de Shein entre 2016 y 2023 estimados en miles de millones de dólares fueron los siguientes:

Tabla 46. Ingresos anuales estimados de Shein entre 2016 y 2023

Año	Ingresos anuales estimados en MMD (Miles de millones de dólares)
2016	0.61
2017	1.51
2018	1.99
2019	3.15
2020	9.81
2021	15.7
2022	22.7
2023	36.5

Determinar la covarianza entre ambas variables

Tabla 47. Cálculo de la covarianza.

Año (X)	Ingreso estimado en MMD (Y)	$X-\overline{X}$	$Y-\overline{Y}$	$(X-\overline{X})^2$	$(Y-\overline{Y})^2$	$(X-\overline{X})(Y-\overline{Y})$
2016	0.61	-3.5	-10.89	12.25	119	38.101875
2017	1.51	-2.5	-9.99	6.25	100	24.965625
2018	1.99	-1.5	-9.51	2.25	90	14.259375
2019	3.15	-0.5	-8.35	0.25	70	4.173125
2020	9.81	0.5	-1.69	0.25	3	-0.843125
2021	15.7	1.5	4.20	2.25	18	6.305625
2022	22.7	2.5	11.20	6.25	126	28.009375
2023	36.5	3.5	25.00	12.25	625	87.513125
				$\sum = 42$	$\sum = 1151$	$\sum = 202.4$
$\overline{X} = 2019.5$	$\overline{Y} = 11.49$			$\overline{x^2} = 5.18$	$\overline{y^2} = 135$	$Cov(X,Y) = 25.3$
$r = -0.92$						

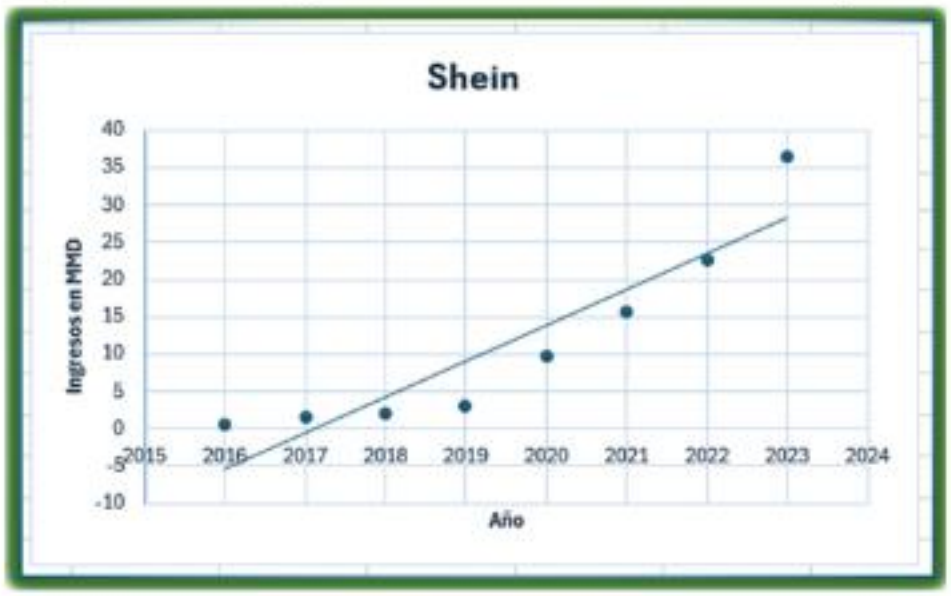
$$Cov(X,Y) = \frac{\sum\left(X-\overline{X}\right)\left(Y-\overline{Y}\right)}{n} = \frac{202.4}{8} = \mathbf{25.3}$$

El coefiente de correlación es:

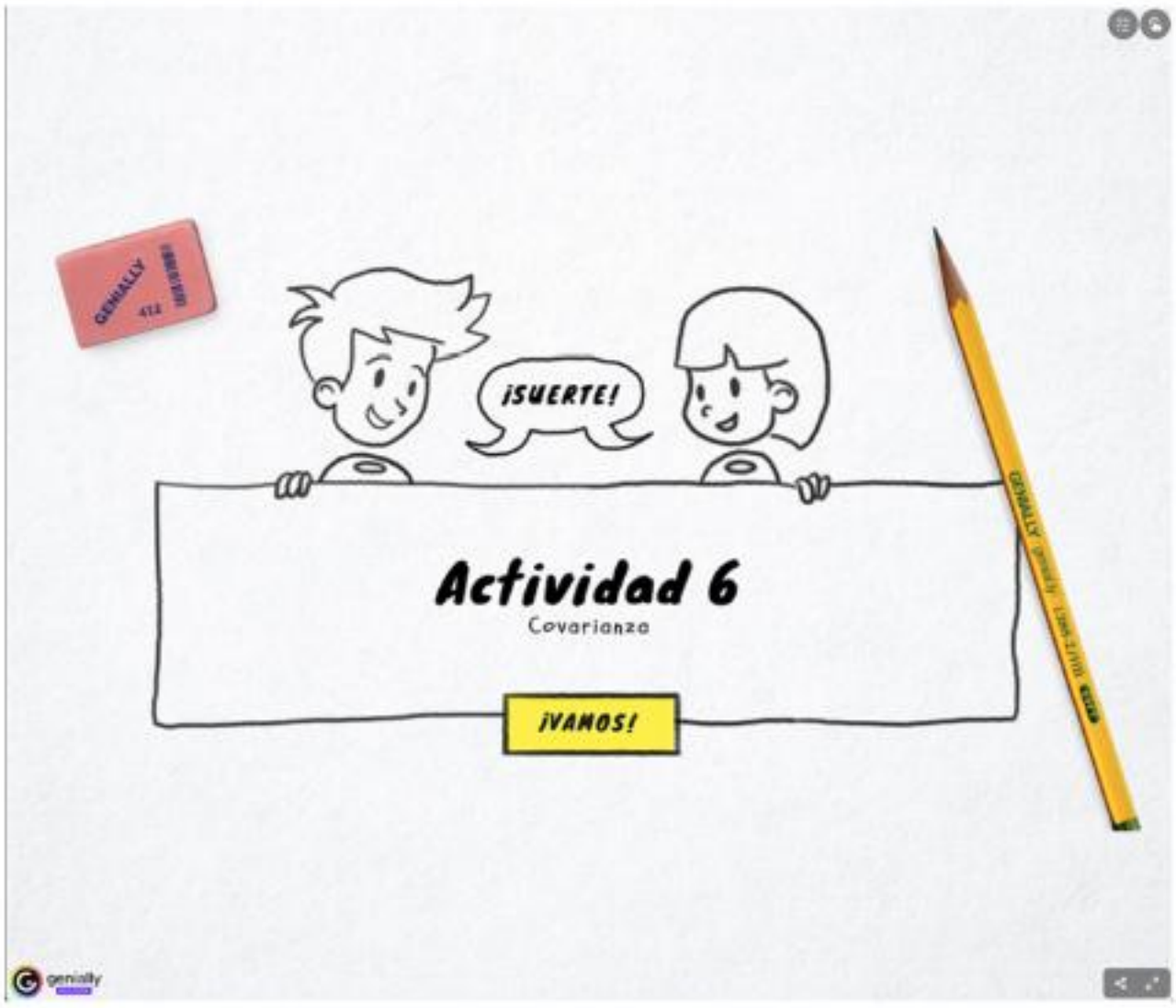
$$r = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\sqrt{\Sigma(X-\overline{X})^2\Sigma(Y-\overline{Y})^2}} \qquad r = \frac{202.4}{\sqrt{(42)(1151)}} = \mathbf{0.92}$$

El valor resultante $Cov\,(X,Y) = 25.3$ nos india, al ser mayor que cero, que las dos variables tienen una relación positiva, recordemos que nos interesa el signo y no el valor obtenido. Para saber qué tan estrecha es dicha relación hacemos referencia al coeficiente de correlación $r = -0.92$ obtenido en la tabla y observamos que, si existe una correlación lineal, todos los puntos en el gráfico se alinearán en una línea con pendiente positiva.

Figura 16. Gráfica de ingresos anuales estimados de Shein entre 2016 y 2023.



La covarianza se utiliza en relaciones lineales, por lo que es importante hacer un gráfico para ver si existe o no dicha relación lineal.



Actividad 6.

Ejercicios de diagrama de dispersión y centro de gravedad

Instrucciones:
para practicar lo aprendido, te invitamos a resolver los siguientes ejercicios, para comenzar da clic en la **imagen.**

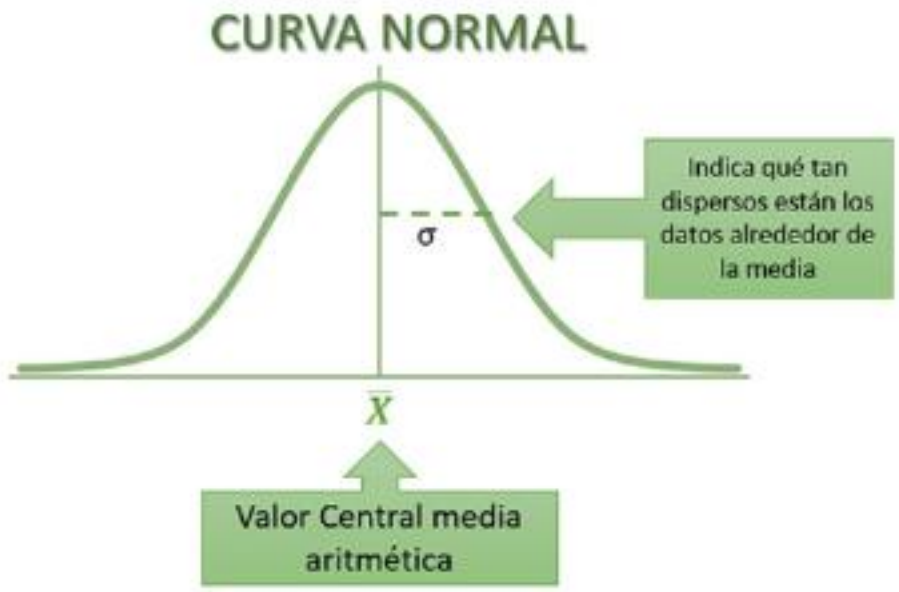
4.6 Distribución normal

Definición

La distribución normal es una distribución de probabilidad continua que se caracteriza por ser un patrón estadístico cuando un conjunto de datos se distribuye de manera uniforme alrededor de un valor central. Esto significa que los datos u observaciones se agrupan a un valor central (generalmente es la media aritmética) y lo valores menos comunes se alejan de este valor central.

Su gráfica se denomina curva normal, es la curva con forma de campana y se usa para describir de manera aproximada muchos fenómenos que ocurren en la naturaleza, la industria e investigación

Una variable aleatoria continua que tiene la distribución en forma de campana se denomina variable aleatoria normal. La ecuación de la probabilidad de la variable normal depende de los parámetros $\overline{X}$ y σ, la media aritmética y su desviación estándar respectivamente.



ALGUNAS PROPIEDADES DE LA DISTRIBUCIÓN NORMAL

- Es simétrica. El valor de la media, la mediana y la moda coinciden. Matemáticamente,

  Media = Mediana = Moda

- 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.

Analizaremos un ejemplo de datos no agrupados que sigue una distribución normal.

Ejemplo 1. El conjunto de datos son los pesos (en kg) de 15 paquetes enviados por una empresa.

Los datos son: 54,57,52,59,58,56,60,53,55,57,58,59,60,61,54

Observemos que si acomodamos los datos observados tenemos que su gráfica se parece a la curva normal.

Esto es: 52,54,56,57,58,59,60,61,60,59,58,57,55,54,53



Figura 17. Gráfica de pesos en (kg) de paquetes enviados.

Análisis:

a. Obtenemos la media $\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$

$$\overline{X} = 54 + 57 + 52 + 59 + 58 + 56 + 60 + 53 + 55 + 57 + 58 + 59 + 60 + 61 + 54$$

$$= \frac{853}{15}$$

$$\overline{X} = 56.86$$

b. Obtenemos la suma de las desviaciones al cuadrado

	Precios en dólares	$(X_i - \overline{X})^2$
	54	$(54 - 56.86)^2 = 8.1796$
	57	$(57 - 56.86)^2 = 0.0196$
	52	$(52 - 56.86)^2 = 23.6196$
	59	$(59 - 56.86)^2 = 4.5796$
	58	$(58 - 56.86)^2 = 1.2996$
	56	$(56 - 56.86)^2 = 0.7396$
	60	$(60 - 56.86)^2 = 9.8596$
	53	$(53 - 56.86)^2 = 14.8996$
	55	$(55 - 56.86)^2 = 3.4596$
	57	$(57 - 56.86)^2 = 0.0196$
	58	$(58 - 56.86)^2 = 1.2996$
	59	$(59 - 56.86)^2 = 4.5796$
	60	$(60 - 56.86)^2 = 9.8596$
	61	$(61 - 56.86)^2 = 17.1396$
	54	$(54 - 56.86)^2 = 8.1796$
Total	**853**	**107.734**
	$\overline{X} = \frac{853}{15} = 56.86$	

Tabla 48. Cálculo de las desviaciones al cuadrado

c. Calculamos la Varianza

$$\sigma^2 = \frac{107.73}{15}$$

$$\sigma^2 = 11.58$$

d. Calculamos la desviación estándar

$$\sigma = \sqrt{\frac{\sum (x_i - \overline{X})^2}{n}}$$

$$\sigma = \sqrt{11.58}$$

$$\sigma = 3.4$$

Dado que presenta una distribución normal, entonces se tiene que la siguiente i**nterpretación de la Desviación Estándar en los Intervalos:**

- 68.27% de los datos (aproximadamente) están en el intervalo 56.86±2.67, es decir, [54.19,59.53].
- 95.45% de los datos (aproximadamente) están en el intervalo 56.86± 2(2.67), es decir, [51.52,62.2].
- 99.73% de los datos (aproximadamente) están en el intervalo 56.86± 3(2.67), es decir, [48.85,64.87].

Podemos ver que la mayoría de los pesos de los paquetes están concentrados alrededor de la media (56.86 kg), con una dispersión según la desviación estándar (2.67 kg).

4.7 Clasificación de la probabilidad

Hay tres formas básicas para clasificar la teoría de probabilidad:

Definición

1. Probabilidad clásica

Se basa en el conocimiento previo del proceso; define la probabilidad de que un evento suceda, con la fórmula:

$$P(A) = \frac{Resultados\ favorables}{Resultados\ posibles}$$

Para que esta expresión sea válida, todas las formas en que puede presentarse el fenómeno deben ser igualmente posibles.

Este tipo de probabilidad se utiliza en juegos de azar, sin embargo, es muy complicado utilizar este tipo de probabilidad en problemas de decisiones menos previsibles y más riesgosas, como los problemas de administración.





2. Probabilidad objetiva (a partir del concepto de frecuencia relativa).

Cuando un fenómeno toma valores distintos, la frecuencia relativa de ocurrencia de cada uno de ellos tiende a coincidir con su probabilidad de ocurrencia cuando el número de veces en que se presenta el evento es lo apreciablemente grande. Esta forma de cálculo utiliza este concepto de frecuencia relativa de situaciones previas como probabilidad. De esta manera, se analiza qué tan frecuente ha ocurrido una situación en el pasado y esos datos los utilizaremos para prever con qué probabilidad dicha situación podrá ocurrir en el futuro.

Probabilidad subjetiva

Se sostiene bajo las opiniones o presunciones de los encargados a realizar el estudio de probabilidad. Se puede considerar como la probabilidad sobre la ocurrencia de un hecho que un individuo le otorga, sostenida en la evidencia que, sobre el mismo hecho, disponga de oportunidades anteriores.

La probabilidad basada en el aspecto subjetivo nos otorga mayor flexibilidad, que las otras probabilidades estudiadas no logran. Este tipo de probabilidad es más común en los eventos que ocurren una vez o con poca frecuencia.



Cálculo de probabilidades

La probabilidad es una medida en fracción o porcentaje que tiene un evento de ocurrir, el evento puede ser lanzar una moneda, tirar un dado, girar una ruleta, sacar una pelota de una bolsa, etc.

Si la probabilidad se acerca al 100% quiere decir que es muy probable que ocurra y si se acerca al 0% es muy probable que nunca ocurra.

Un evento es el resultado posible de un experimento y es la unidad mínima de análisis para efectos de cálculo de probabilidades.

Figura 18. Clasificación de eventos



4.7.1 Eventos mutuamente excluyentes o disjuntos

Eventos mutuamente excluyentes: No pueden ocurrir al mismo tiempo: es uno u otro.

Eventos mutuamente excluyentes

Concepto

Dos o mas eventos son mutuamente excluyentes, si la ocurrencia de uno de ellos excluye la posibilidad de ocurrencia de otro.



Fórmula

Los eventos mutuamente excluyentes no pueden ocurrir de forma simultanea, es decir, al mismo tiempo.

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



Ejemplo

Al lanzar una moneda. ¿Cuál es la probabilidad de que caiga sol o águila?

El evento A es que caiga Sol

El evento B es que caiga Águila

$P(A) = 1/2 = .50 = 50\%$

$P(B) = 1/2 = .50 = 50\%$





Procedimiento

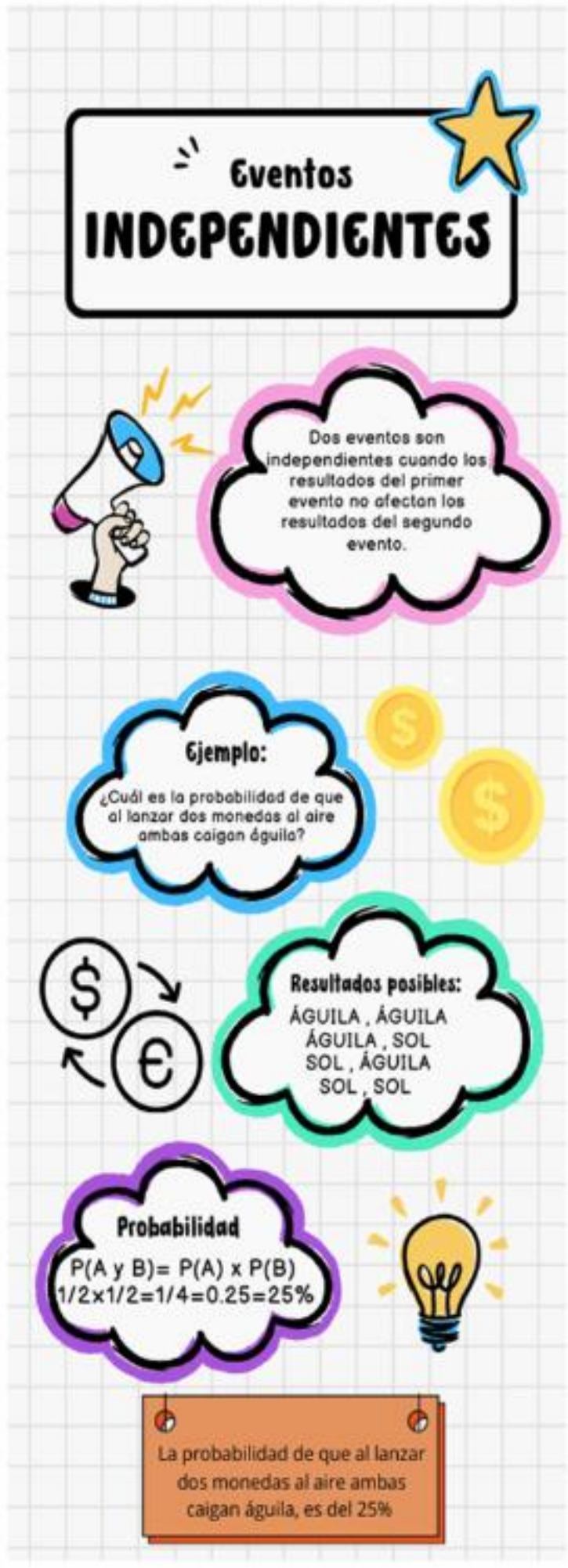
El evento A y el evento B son mutuamente excluyentes porque cuando ocurre uno elimina la posibilidad de que ocurra otro, pues en una moneda solo puede caer águila o sol.

$P(A \text{ o } B) = 1/2 + 1/2 = 1$

Ambos eventos no pueden ocurrir al mismo tiempo, por lo que la probabilidad de que salga sol o águila es del 100%

4.7.2 Eventos Independientes

Eventos independientes: No se ven afectados por otros.



4.7.3 Eventos dependientes

Eventos dependientes: Cuando un evento afecta la probabilidad de que suceda otro



4.7.4 Eventos no excluyentes entre si

Eventos no excluyentes entre sí: Cuando la ocurrencia de un evento no impide que suceda otro.





Actividad 7.



Ejercicios de clasificación de la probabilidad

Instrucciones:
te invitamos a resolver los siguientes ejercicios, para comenzar da clic en la imagen.

4.8 Estadística inferencial

Definición

Es una rama de la estadística que se enfoca en hacer predicciones o generalizaciones sobre una población a partir de una muestra de datos. A diferencia de la estadística descriptiva, que se ocupa de resumir y describir los datos obtenidos, la estadística inferencial utiliza estos datos para inferir propiedades y hacer estimaciones sobre la población completa.

Las principales actividades de la estadística inferencial incluyen:

Figura 19. Actividades de la Estadística Inferencial

ESTIMACIÓN	PRUEBAS DE HIPÓTESIS	ANÁLISIS DE REGRESIÓN	ANÁLISIS DE VARIANZA (ANOVA)
Calcular estimaciones de parámetros de la población, como la media o la proporción, basándose en los datos de la muestra. Esto incluye el uso de intervalos de confianza para proporcionar un rango de valores dentro del cual es probable que se encuentre el parámetro poblacional.	Evaluar afirmaciones o hipótesis sobre una población utilizando datos muestrales. Esto implica formular una hipótesis nula (que no hay efecto o diferencia) y una hipótesis alternativa (que sí hay efecto o diferencia), y luego usar técnicas estadísticas para determinar cuál es más probable en función de los datos disponibles.	Examinar las relaciones entre variables y construir modelos predictivos. Esto ayuda a entender cómo una variable puede predecir otra y a cuantificar la fuerza y la dirección de estas relaciones.	Comparar las medias de tres o más grupos para determinar si al menos uno de ellos es significativamente diferente de los otros.

Las técnicas utilizadas en la estadística inferencial se basan en la teoría de probabilidad y en modelos matemáticos que permiten evaluar la variabilidad y la incertidumbre en los datos muestrales, proporcionando una base sólida para la toma de decisiones informadas en diversas áreas como la investigación científica, la economía, la medicina y las ciencias sociales.

Resumen del Módulo IV

El pensamiento matemático es necesario para tomar decisiones acertadas en varias áreas de la vida, por ejemplo, administrativas, educativas, empresariales, entre otras, los temas explicados en este módulo referentes a las diferentes medidas numéricas de la estadística tienen como finalidad proporcionarte fundamentos sólidos para la vida profesional futura, esperamos se haya logrado el propósito.

Al conocer y saber determinar las medidas de tendencia central, posición y dispersión de un conjunto de datos, tienes más herramientas para tomar decisiones correctas de aras de mejorar la calidad y la productividad por ejemplo de algún negocio, debido a que las medidas de dispersión son herramientas estadísticas que describen la variabilidad o extensión de un conjunto de datos, en este caso la desviación media, la varianza como base de la desviación estándar proporcionan información sobre la distribución de los datos alrededor de un punto central.

Estas medidas son esenciales en estadística para resumir y describir las características de un conjunto de datos permitiendo comparaciones y análisis más profundos y son ampliamente utilizadas en diversas áreas como la economía las ciencias sociales y las ciencias naturales para comprender la variabilidad y la incertidumbre en los datos.

Estás por terminar el primer semestre de tu vida académica en el nivel medio superior y estamos seguros que has adquirido buenas bases para continuar y avanzar con tus estudios en el área de las matemáticas

Actividad 9.



Ejercicio Autoevaluación

Instrucciones:
Como actividad final, se te pide realizar un cuestionario; para comenzar, revisa en la plataforma **SEDUCA** la sección Actividades (Estudiante) y responde a la actividad denominada: M4 Cuestionario Final.

Glosario

Centro de gravedad: es el promedio de cada variable.

Coeficiente de correlación: cantidad que permite conocer que tan bien se ajusta a un conjunto de pares ordenados a un comportamiento determinado.

Covarianza: es el valor que refleja en qué cantidad dos variables aleatorias varían de forma conjunta respecto a sus medias aritméticas.

Cuartil: es la división de los datos en cuatro partes iguales.

Decil: es la división de los datos en diez partes iguales.

Diagrama de dispersión: es una representación gráfica de dos variables que muestra la relación entre ellas.

Desviación media: se define como la media aritmética de los valores absolutos de las desviaciones de las variables respecto a la media aritmética.

Desviación típica o estándar: se define como la raíz cuadrada del cuadrado de la media de los cuadrados de las desviaciones respecto a la media, esto es la raíz de la varianza.

Distribución normal: se caracteriza por ser un patrón estadístico cuando un conjunto de datos se distribuye de manera uniforme alrededor de un valor central.

Estadística inferencial: rama de la estadística que se enfoca en hacer predicciones o generalizaciones sobre una población a partir de una muestra de datos.

Eventos dependientes: cuando un evento afecta la probabilidad de que suceda otro.

Eventos independientes: no se ven afectados por otros.

Eventos mutuamente excluyentes: no pueden ocurrir al mismo tiempo: es uno u otro.

Eventos no excluyentes entre sí: cuando la ocurrencia de un evento no impide que suceda otro.

Globalizado: proceso histórico de integración mundial en los ámbitos económico, político, tecnológico, social y cultural, que ha convertido al mundo en un lugar cada vez más interconectado. En ese sentido, se dice que este proceso ha hecho del mundo una aldea global.

Licitación pública: todo procedimiento administrativo para la adquisición de suministros, contratación de servicios o la ejecución de obras (construcciones) que organicen entes, organismos y entidades que formen parte del Sector Público.

Medidas de tendencia central: representan un centro en torno al cual se encuentra ubicado el conjunto de los datos.

Media aritmética: es el promedio obtenido al sumar todos los valores individuales entre el número total de valores; representa el punto de equilibrio de la distribución de los datos.

Mediana: es el valor central de los datos.

Medidas de posición: son aquellas en donde podemos dividir los datos en partes iguales.

Moda: Es el valor que más se repite en las mediciones.

Percentil: es la división de los datos en 100 partes iguales.

Probabilidad clásica: se basa en el conocimiento previo del proceso; define la probabilidad de que un evento suceda.

Probabilidad objetiva: es cuando un fenómeno toma valores distintos, la frecuencia relativa de ocurrencia de cada uno de ellos tiende a coincidir con su probabilidad de ocurrencia cuando el número de veces en que se presenta el evento es lo apreciablemente grande.

Probabilidad subjetiva: se sostiene bajo las opiniones o presunciones de los encargados a realizar el estudio de probabilidad.

UNESCO: (por sus siglas en inglés) Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura.

Varianza: es la suma de los cuadrados de las diferencias del i-ésimo dato con la media aritmética y se divide entre el total de datos.

Referencias Electrónicas

- https://es.statista.com/temas/10559/shein/#topicOverview
- Rodriguez Alveal, F. E., Maldonado Fuentes, A. C., & Sandoval Rubilar, P. R. (2016). Comprensión de las medidas de tendencia central: un estudio comparativo en estudiantes de pedagogía en matemática en dos instituciones formadoras chilenas. Avaliação: Revista da Avaliação da Educação Superior, 21(3), 929-952. Disponible en: https://www.redalyc.org/pdf/2191/219148307013.pdf
- Quevedo, F., (2011). Medidas de tendencia central y dispersión. Medwave. Año XI, No. 3, Marzo 2011. Open Access, Creative Commons. Estadística Aplicada a la Investigación en Salud. Departamento de Educación en Ciencias de la Salud, Facultad de Medicina, Universidad de Chile. Disponible en: https://dsp.facmed.unam.mx/wp-content/uploads/2013/12/Quevedo-F.-Medidas-de-tendencia-central-y-dispersion.-Medwave-2011-Ma-113..pdf
- Rendón-Macías, M. E., Villasís-Keeve, M. Á., & Miranda-Novales, M. G. (2016). Estadística descriptiva. Revista Alergia México, 63(4), 397-407. Disponible en: https://www.redalyc.org/pdf/4867/486755026009.pdf

Resumen general

En un mundo cada vez más complejo y cambiante, la preparación académica se convierte en un pilar del progreso y la superación de la sociedad. La educación no solo se trata de adquirir conocimientos, sino también de desarrollar habilidades, valores y actitudes. Cada tema, cada tarea, cada actividad, es una oportunidad para crecer, aprender a pensar críticamente, establecer mecanismos de comunicación efectiva y, entre otras cualidades más, desarrollar habilidades para trabajar en equipo.

Con la finalidad de lograr estos objetivos, en el **módulo I** se incluyó una descripción de los símbolos matemáticos, los números reales y sus operaciones; lo cual es importante porque te permitirá identificar en el bello lenguaje matemático, la forma en que las situaciones de la vida cotidiana se describen con símbolos para su estudio y análisis. En nuestro día a día se desarrollan acontecimientos que nos pueden brindar información útil para la toma de decisiones, por lo que se abarcó además el diseño de encuestas, la forma de determinar una muestra y algunas opciones para la adquisición de datos estadísticos. De igual forma se definieron las bases de la probabilidad, que es fundamental en la toma de decisiones y la predicción de eventos. Los jóvenes deben entender conceptos como el espacio muestral, eventos, probabilidades condicionales y teorema de Bayes. La estadística te permite evaluar riesgos, calcular probabilidades y tomar decisiones basadas en evidencia.

En el **módulo II** se contempló la importancia del análisis de datos y las formas gráficas de representarlos, por lo que podrás encontrar conceptos como gráfica de barras, histograma, circulograma, entre otros, así como la forma en que los datos se relacionan entre ellos apoyándose con diagramas de dispersión y la recta de regresión representativa de un conjunto de datos. En otras palabras, no basta saber que los fenómenos existen y que de ellos podemos extraer información sino, además, es igualmente importante organizarlos y representarlos en formas más fáciles de asimilar para la toma de decisiones.

Hasta este punto se te mostró la importancia de la recolección de datos y su análisis, no obstante, en el **módulo III** se describió la importancia de los tipos de muestreo, si bien, los datos son representativos del fenómeno, la forma en los adquiridos permiten identificar si representan, o no, a una población en general; en otras palabras, aprenderás a identificar si por las características de un estudio, ya sea observacional o experimental, podrás determinar una relación entre las variables de estudio, o en su caso, una causalidad entre una y otra variable que sean aplicables a un grupo, una comunidad, una población, o inclusive, a un país entero.

En el **módulo IV** se abordaron las medidas de tendencia central, como la media, la mediana y la moda, que te permitirán resumir grandes conjuntos de datos en un solo valor representativo; por ejemplo, la media nos dice el valor promedio, la mediana nos muestra el valor central y la moda representa el dato más frecuente, estas medidas son esenciales para interpretar datos y tomar decisiones informadas. De igual forma se incluyeron las medidas de dispersión, como la desviación media y la desviación estándar, que indican cuánto se alejan los datos individuales del valor central. Comprender la variabilidad en los datos es crucial para evaluar la consistencia y la confiabilidad de los resultados.

El E-Book de Pensamiento Matemático busca que los estudiantes identifiquen a la probabilidad y a la estadística como apoyo para la interpretación de datos y resultados de manera significativa, en lugar de lidiar con números aislados, ya que se pueden identificar patrones, tendencias y relaciones. Al seguir métodos estadísticos, se evitan sesgos personales y las decisiones se basan en evidencia sólida. Comparadas con otras formas de análisis (como el juicio personal o la intuición), la probabilidad y la estadística son más prácticas y confiables, puesto que permiten cuantificar incertidumbre y evaluar resultados de manera más precisa. Por tales razones, este E-Book se diseñó con propuestas innovadoras que captarán tu atención desde el primer instante; con las que tendrás la oportunidad de apropiarte de todos estos nuevos saberes, además, de ponerlos a prueba para medir tu progreso. Se buscó implementar actividades dinámicas que salgan de lo rutinario del texto plano y tedioso, por lo que se te invita a resolverlas todas, cada actividad terminada satisfactoriamente te llevará un paso más cerca a ese estudiante crítico e informado que nuestro país necesita.

Doctor en Tecnología e Innovación de la Educación
Arturo Mejía Zamora
Plantel "Lic. Adolfo López Mateos"

Maestro en Ciencias de la Educación
Erick Vieyra Méndez
Plantel "Nezahualcóyotl"

Maestro en Administración Pública y Gobierno
Raúl Juárez Toledo
Plantel "Cuauhtémoc"

Doctora en Ciencias de la Educación
María del Rocío García de León Pastrana
Plantel "Ignacio Ramírez Calzada"

Doctor en Ciencias de la Educación
Alejandro Morales Velázquez
Plantel "Dr. Ángel Ma. Garibay Kintana"

Maestro en Educación con Orientación en Innovación y Tecnología Educativa
Sergio Luis Roldán Vásquez
Plantel "Texcoco"

Maestro en Docencia
Erick Rodolfo Sánchez Cortés
Plantel "Sor Juana Inés de la Cruz"

Química Farmacéutica Bióloga
Janette Jaimes García
Encargada del Despacho de la Dirección del Plantel "Dr. Pablo González Casanova"

Doctora en Educación
María del Rosario Sotelo Hurtado
Encargada del Despacho de la Dirección del Plantel "Isidro Fabela Alfaro"

Licenciada en Comunicación
Ana Vianey Suárez Castro
Encargada del Despacho de la Dirección del Plantel " Mtro. José Ignacio Pichardo Pagaza"

Universidad Autónoma
del Estado de México

ADMINISTRACIÓN
UNIVERSITARIA
2021 - 2025

Universidad Autónoma
del Estado de México

ADMINISTRACIÓN
UNIVERSITARIA
2021 - 2025

Universidad Autónoma
del Estado de México

ADMINISTRACIÓN
UNIVERSITARIA
2021 - 2025

Universidad Autónoma
del Estado de México

ADMINISTRACIÓN
UNIVERSITARIA
2021 - 2025

Universidad Autónoma
del Estado de México

ADMINISTRACIÓN
UNIVERSITARIA
2021 - 2025

Universidad Autónoma
del Estado de México

ADMINISTRACIÓN
UNIVERSITARIA
2021 - 2025

Universidad Autónoma
del Estado de México

ADMINISTRACIÓN
UNIVERSITARIA
2021 - 2025

Universidad Autónoma
del Estado de México

ADMINISTRACIÓN
UNIVERSITARIA
2021 - 2025

Universidad Autónoma
del Estado de México

ADMINISTRACIÓN
UNIVERSITARIA
2021 - 2025

Universidad Autónoma
del Estado de México

ADMINISTRACIÓN
UNIVERSITARIA
2021 - 2025

Universidad Autónoma
del Estado de México

ADMINISTRACIÓN
UNIVERSITARIA
2021 - 2025

Universidad Autónoma
del Estado de México

ADMINISTRACIÓN
UNIVERSITARIA
2021 - 2025