



**Universidad Autónoma
del Estado de México**



Estadística

Libro de texto basado en competencias

Xiomara Rodríguez Mondragón · María Guadalupe Jardón Segura

6°
Semestre

Estadística

Libro de texto basado en competencias

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



Libro de texto basado en competencias

Xiomara Rodríguez Mondragón
María Guadalupe Jardón Segura

Primera edición, febrero 2022

ESTADÍSTICA

Libro de texto basado en competencias

Autores

Xiomara Rodríguez Mondragón

Plantel “Dr. Pablo González Casanova” de la Escuela Preparatoria de la UAEM

María Guadalupe Jardón Segura

Plantel “Dr. Pablo González Casanova” de la Escuela Preparatoria de la UAEM

Universidad Autónoma del Estado de México

Av. Instituto Literario 100 Ote.

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, así como el tratamiento basado en competencias realizado a los libros de texto. 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, 2018

ISBN: 978-607-633-415-7

Hecho en México

Responsable del Programa de Diseño y Producción Editorial: L. D. G. Miguel Angel Conzuelo Endeje

Corrección de estilo: Érika Carmona

Portada e interiores: Karla López



Mensaje

A través de su historia, la Universidad Autónoma del Estado de México ha tenido como uno de sus objetivos fundamentales lograr que el estudiantado reciba una educación basada en el libre pensamiento, apelando a un sentido científico. La Escuela Preparatoria de la UAEM continuará mostrándose como un punto de inflexión para la comunidad estudiantil por la calidad de la educación que se ofrece al tiempo que se enseña a adquirir conocimientos, valores, habilidades y destrezas que favorezcan el entendimiento de nuestro entorno social y ambiental.

La responsabilidad de la Universidad Autónoma del Estado de México es asegurar que la educación que se imparte en la Escuela Preparatoria contribuya a ampliar el horizonte personal y profesional de los estudiantes. El enfoque para la apropiación del conocimiento parte de aprovechar la experiencia de vida y desde la autonomía de pensamiento y la reflexión lograr una comprensión más amplia de la realidad.

El material educativo que tienes en tus manos ha sido elaborado colegiadamente entre especialistas de cada una de las áreas del conocimiento. La creatividad, el pensamiento crítico y reflexivo, así como el espíritu científico del claustro de docencia e investigación se reúne en esta antología con el fin de que te acerques al conocimiento con una mirada crítica. Este ejemplar ha sido elaborado con el fin de establecer un aprendizaje dialógico que estimule tu curiosidad, que te sirva como guía, y que invariablemente te provoque pasión por el conocimiento.

El objetivo es lograr que recibas una educación integral e incluyente que incida de manera positiva en tu trayectoria escolar. Por ello es importante que asumas un compromiso personal, empeño y dedicación que te permita apropiarte del conocimiento.

Patria, Ciencia y Trabajo

Doctor en Ciencias e Ingeniería Ambientales
Carlos Eduardo Barrera Díaz

Rector



ÍNDICE

Presentación.....	10
Módulo Uno. La estadística y su aplicación.....	12
Introducción	14
Evaluación diagnóstica.....	15
1. Conceptualización y clasificación de la estadística.....	16
1.1 Dato estadístico.....	19
1.1.1 Población y muestra	25
1.2 Registro y ordenamiento de datos	27
1.3 Tabla de distribución de frecuencias	28
1.4 Representación gráfica de datos	40
Ejercicios.....	54
Autoevaluación	59
Formulario del módulo	60
Módulo Dos. Medidas de tendencia central y medidas de posición.....	61
Introducción	63
Evaluación diagnóstica.....	63
2.1 Medidas de tendencia central	64
2.1.1 Media aritmética.....	64
2.1.2 Mediana.....	66
2.1.3 Moda	69
2.2. Medidas de posición.....	71
2.2.1 Cuartiles.....	71
2.2.2 Deciles	73
2.2.3 Percentiles	76
Ejercicios.....	78
Autoevaluación	80
Formulario del módulo	81

Módulo Tres. Las medidas de dispersión en la estadística descriptiva y recta de regresión	83
Introducción	86
Evaluación diagnóstica.....	86
3.1 Desviación media	86
3.1.1 Interpretación de la desviación media.....	88
3.2 Varianza y desviación estándar	89
3.2.1 Interpretación de la varianza	93
3.2.2 Interpretación de la desviación estándar y sesgo.....	96
3.3 Diagrama de dispersión y centro de gravedad.....	100
3.4 Covarianza	102
3.5 Coeficiente de correlación	105
3.6 Ecuación de la recta de regresión y aplicación a series de tiempo.....	110
Ejercicios.....	111
Autoevaluación	118
Formulario del módulo	119
Módulo Cuatro. Introducción a la probabilidad.....	121
Introducción	123
4.1 El papel de la probabilidad.....	124
4.1.1 Evento y espacio muestral	124
4.2 Problemas de conteo	130
4.2.1 Factorial, combinaciones y permutaciones.....	130
4.2.1.1 Diagrama de árbol.....	142
4.3 Clasificación de la probabilidad.....	143
4.3.1 Eventos mutuamente excluyentes o disjuntos.....	144
4.3.2 Eventos independientes.....	145
4.3.3 Eventos dependientes.....	145
4.3.4 Eventos no excluyentes entre sí.....	146
Ejercicios.....	149
Autoevaluación	151
Formulario del módulo	152
Glosario	154
Referencias	156



PRESENTACIÓN

Feb	Mar	Apr
3.0	1.0	2.0
4.0	3.0	1.0
7.0	4.0	1.0
10.0	13.0	1.0
15.0	11.0	1.0

100%	100.0
Dolo	quis nos exercitat ullam co

11.0	7.0
	10.0
	15.0

voluptatem, ugiat pariur.

ex ea commodo consequat contro. r. Cumtrollim persum illo derneque lpa qui officia deserunt mollitan.



Los usos de la estadística se registran desde la época de los babilonios (5000 a. C.), quienes llevaban un control de sus soldados y esclavos en tablillas. También se sabe que los chinos y los egipcios realizaban censos en los años 2000-2500 a. C.; ello dio pauta a diversas transformaciones y aplicaciones de la estadística en la actualidad, como en el caso de las elecciones ciudadanas, las calificaciones de escuelas, los resultados de los equipos de nuestro deporte favorito, el banco y en áreas de índole social, político o económico.

El término estadística, proviene del alemán *statistik* y éste del italiano *statista*, que significa “hombre de estado” (RAE, 2014). A Godofredo Achenwall, profesor y economista alemán, es a quien se le atribuye el desarrollo de la estadística.

A través del tiempo, el hombre ha tenido la necesidad de controlar las actividades económicas, políticas y sociales de su entorno para lograr los objetivos establecidos, ya sean de carácter personal o grupal, para lo cual se ha servido de la estadística. Su aplicación es muy amplia, la podemos observar en el área de la salud, actuaría, biología, psicología, ciencias ambientales, física, informática, derecho, geografía, sociología, entre algunas otras. Actividades en las que se usan las medidas de tendencia central, las medidas de posición, la regresión lineal, así como las combinaciones, permutaciones y medidas de posición.

La aplicación del método científico en diversos ámbitos ya sea matemático, social, económico y más, da pauta a la observación, la formulación de hipótesis, la experimentación, comprobación y conclusión para desarrollar investigaciones que apoyan sus análisis y conclusiones en los fundamentos de la estadística. (Chong, 2016).

El programa de estadística que se oferta en el sexto semestre del bachillerato universitario tiene como propósito formar a los estudiantes de manera integral, por medio de la construcción de conocimientos, habilidades, actitudes y valores para su ingreso al nivel superior. Por ello, este libro presenta las bases para lograr las competencias establecidas en el programa de estudio del CBU 2015, al integrar teoría y práctica de forma clara y precisa, con material de apoyo que fomente el desarrollo de aprendizajes previos a un andamiaje más amplio dentro de sus competencias.

El desarrollo de cada módulo permite conocer la conceptualización y aplicación de la estadística descriptiva e inferencial a través de un acercamiento paulatino a temas básicos, como la identificación del dato estadístico, la clasificación de las variables, la elaboración de tablas de frecuencia y el cálculo de las medidas de tendencia central, posición y dispersión. Asimismo, se incluye un primer acercamiento a la probabilidad, que permitirá a los estudiantes identificar las aplicaciones de la estadística en su vida cotidiana.

BIENVENIDOS

A su último semestre de nivel medio superior y éxito en cada una de sus actividades.

Propósito general

Relaciona la estadística descriptiva e inferencial, además de la probabilidad para dar solución a problemas de su vida cotidiana.

Temas de los cuatro módulos



MÓDULO UNO

La estadística y su aplicación



Módulo Uno. La estadística y su aplicación

Competencias disciplinares

Matemáticas básicas y extendidas

3. Explica e interpreta los resultados obtenidos mediante procedimientos matemáticos y los contrasta con modelos establecidos en situaciones reales.
4. Argumenta la solución obtenida de un problema, con métodos numéricos, gráficos, analíticos o variacionales, mediante el lenguaje verbal, matemático y el uso de las tecnologías de la información y la comunicación.
5. Analiza las relaciones entre dos o más variables de un proceso social o natural para determinar o estimar su comportamiento.
8. Interpreta tablas, gráficas, mapas, diagramas y textos con símbolos matemáticos y científicos.

Competencias genéricas

1. Se conoce y valora a sí mismo y aborda problemas y retos teniendo en cuenta los objetivos que persigue.
 - 1.1 Enfrenta las dificultades que se le presentan y es consciente de sus valores, fortalezas y debilidades.
4. Escucha, interpreta y emite mensajes pertinentes en distintos contextos mediante la utilización de medios, códigos y herramientas apropiados.
 - 4.1 Expresa ideas y conceptos mediante representaciones lingüísticas, matemáticas o gráficas.
5. Desarrolla innovaciones y propone soluciones a problemas a partir de métodos establecidos.
 - 5.1 Sigue instrucciones y procedimientos de manera reflexiva, comprendiendo cómo cada uno de sus pasos contribuye al alcance de un objetivo.
 - 5.2 Ordena información de acuerdo a categorías, jerarquías y relaciones.
 - 5.6 Utiliza las tecnologías de la información y comunicación para procesar e interpretar información.
8. Participa y colabora de manera efectiva en equipos diversos.
 - 8.1 Propone maneras de solucionar un problema o desarrollar un proyecto en equipo, definiendo un curso de acción con pasos específicos.
 - 8.2 Aporta puntos de vista con apertura y considera los de otras personas de manera reflexiva.
 - 8.3 Asume una actitud constructiva, congruente con los conocimientos y habilidades con los que cuenta dentro de distintos equipos de trabajo.



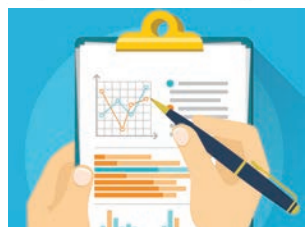
Propósito

Identifica la estadística descriptiva e inferencial con el propósito de seleccionar y organizar las variables de la población o muestra para el análisis, interpretación y elaboración de gráficas.

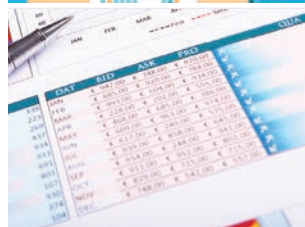
Temas del módulo uno



1. Conceptualización y clasificación de la estadística
 - 1.1. Dato estadístico
 - 1.1.1. Población y muestra



- 1.2. Registro y ordenamiento de datos



- 1.3. Tabla de distribución de frecuencias



- 1.4. Representación gráfica de datos

Introducción

En este módulo se presenta la base teórica y conceptual de la estadística, lo cual te permitirá identificar, definir, analizar, y elaborar tablas de frecuencia y graficar con variables cualitativas o cuantitativas. Se revisan los conceptos de estadística, dato, variable, clasificación de las variables y la construcción de tablas de frecuencia, que es la base para determinar las medidas de tendencia central, y posición; temas que serán abordados en los módulos dos y tres del libro.

También se revisa la ecuación para determinar la población y la muestra. En la primera parte de este módulo se desarrollan los métodos de recolección de los datos estadísticos para su organización y posterior elaboración de las gráficas, lo cual se verá en la segunda parte. Se abordarán también la estadística descriptiva y la inferencial.

En cada uno de los módulos se encuentra una sección denominada actividad autónoma, marcada con las variables (AA) que representa las actividades a desarrollar en cada uno de los apartados del libro.



Recuerda que el éxito depende de ti.

Evaluación diagnóstica

El cuento de las cuentas o el valor real de la estadística

Antonio Toca

Cuando en el mundo del deporte se habla de números, por lo general la gente mira para otro lado. Sin contar el fútbol y todo lo que se mueve alrededor de él, donde los datos sólo tienen sentido para el universo de las apuestas deportivas (y curiosamente en Europa es el deporte que mueve más dinero). Por esa razón me gusta la afirmación de Paul Shirley en su columna *Estadísticas y falsedades*: “en el mundo de las estadísticas nadie tiene toda la razón sobre nada”. Y subrayo yo, se necesita a alguien que sepa de verdad analizarlas (como ese término tan en boga hoy en día en el universo de los medios *online* llamado *analítica web*, que clarifica el enorme caudal de datos informativos que genera una web, cuyo valor interpretativo es imprescindible para plantear la estrategia futura).

Así que, el problema reside en la interpretación de esos datos numéricos. No la hay o nadie la realiza de la manera que se debe llevar a cabo. ¿Por qué? Entiendo que por aburrimiento. En realidad, es una tarea que no te apasiona, o a quien hayan señalado para realizarla, la terminará dejando en el montón de trabajos pendientes. Sin embargo, la importancia verdadera reside en el análisis o interpretación de esos datos (como el analista web del que hablaba antes), en las estadísticas y en la información importantísima que genera la combinación de los mismos. Números, estadística, datos... en el fondo es información para preparar una estrategia. Y esto pocos lo ven.

Llevándolo al terreno del fútbol, y saltándome una regla sagrada de esta columna, supondría dar mayor importancia a lo realizado por aquel jugador que se lleva todos los titulares por destacar en uno o dos partidos al año (los mediáticos, en terminología de los “buenos” narradores de fútbol de este país), mientras existe otro jugador que, al calor de los datos, muestra que su rendimiento es muy válido y notorio partido tras partido, independientemente del valor mediático del mismo (usemos el ejemplo de este gran FC Barcelona de Guardiola, y salen a la palestra varios nombres).

Regreso otra vez a la citada columna de Paul Shirley (de lectura obligatoria todos los lunes), en este caso planteando un ejemplo en el baloncesto, en concreto, sobre la importancia de un jugador cuando está en la cancha: “la estadística, en esencia, mide el impacto de un jugador sobre el resultado final haciendo un seguimiento de los puntos marcados cuando el jugador está en la cancha y cuando está en el banquillo (...) quieren evaluar por qué sus jugadores son mejores o peores que otros y, en consecuencia, si seguir pagándoles o no”.

Recuperado de <https://www.sportyou.es/noticias/author/antonio-toca/page/2>



Responde las siguientes preguntas.

1. ¿Qué es la estadística?
2. ¿Cuál consideras que es la aplicación de la estadística?
3. ¿En dónde has visto el uso de la estadística?
4. ¿Sabes qué es una tabla de frecuencias?
5. ¿Qué es una variable?

1. Conceptualización y clasificación de la estadística

Estadística

La estadística se encarga de la recolección, organización, tratamiento y análisis de datos para extraer conclusiones y tomar decisiones de la situación que se plantea.

Su aplicación se encuentra inmersa en diversos fenómenos de estudio, para lo cual es necesario seguir diversos pasos para el logro de los resultados. La aplicación de la estadística es metódica y concreta.

El Diccionario de la Lengua Española (DLE, 2017) la define como el estudio que reúne, clasifica y recuenta todos los hechos que tienen una característica en común, para poder llegar a conclusiones a partir de los datos numéricos extraídos.

Spiegel, M. (2009), la define como el estudio de los métodos científicos para recoger, organizar, resumir y analizar datos, así como para sacar conclusiones válidas y tomar decisiones razonables basadas en tal análisis.

Mora, R. (2000) la describe como una ciencia que proporciona un conjunto de métodos para ser aplicados. Por medio de la recolección, resumen, clasificación, análisis e interpretación del comportamiento de los datos con respecto a una característica, materia de estudio o investigación.



De acuerdo con los conceptos de estadística antes mencionados, escribe tres actividades que observes en tu entorno que hacen uso de la ella.



1. _____

2. _____

3. _____

Clasificación

La estadística se divide en:



a) Descriptiva o deductiva



b) Inferencial o inductiva



Estadística **descriptiva o deductiva**: corresponde a las primeras acciones que se realizan para el análisis de datos. En ella se aplican las técnicas de recolección, organización y tratamiento de los datos: recabar, organizar y graficar la información seleccionada.

Ejemplo

En una pregunta de investigación: ¿cuál es el uso que dan a la telefonía celular los estudiantes del Plantel "Dr. Pablo González Casanova" del municipio de Tenancingo en el Estado de México?

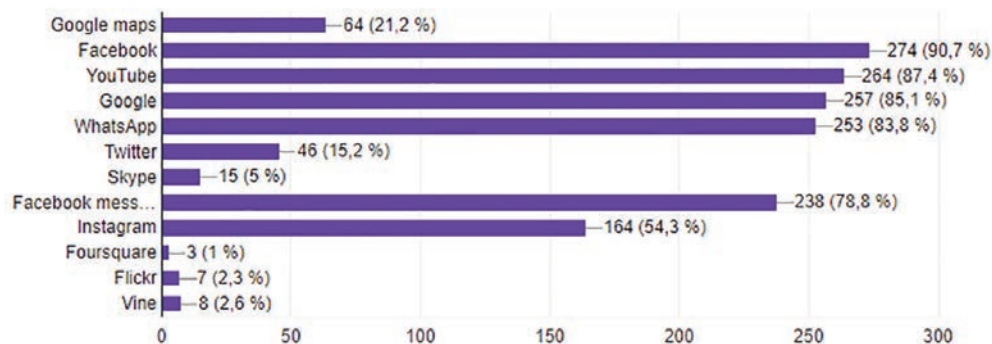
Primero, se aplica un cuestionario con preguntas abiertas y cerradas para identificar qué usos se les da a este tipo de tecnología.

Las respuestas se unifican y los datos se organizan por medio de tablas y gráficas para la comprensión de la información, la cual es analizada a través de medidas de tendencia central, posición, varianza, entre otras.



5. De las siguientes aplicaciones, seleccionas las que usualmente utilizas

302 respuestas



Respuestas

Fuente: <https://docs.google.com/forms/d/1b1bwqsS1HEVVR E3tOh4W59qsN MhODG9IS1dCl Rlv8zk/edit>



Actividad Autónoma

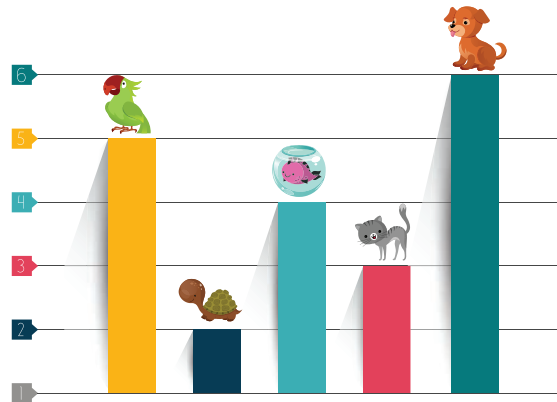
Estadística inferencial o inductiva: interpreta los datos recabados; es decir, se encarga de analizar la información para elaborar conclusiones y tomar decisiones.

Observa las respuestas y escribe una conclusión.

¿Qué sugerencia consideras pertinente con relación a la información del uso del celular?

1.1 Dato estadístico

Es un valor, medida o característica que permite ser observada o contada para su posterior análisis.



Ejemplo

Un dato estadístico son los automóviles que circulan en la ciudad, y los valores, características o medidas que pueden ser analizadas son color, tamaño y marca.

Se desea analizar la preferencia de marca y colores para adquirir un auto en mujeres de 18 a 20 años en el municipio de Metepec. Para este caso, los datos estadísticos con los que se cuenta son marcas y colores de automóviles.

Escribe un dato estadístico con el valor, medida o característica que deseas identificar.



La representación de datos estadísticos se realiza mediante las variables estadísticas.

Variable: presenta un conjunto de datos que se encuentran integrados por la misma característica, pueden ser indicados por diversas letras, como x , y , z , m , b , entre otras.

Ejemplo

Se desea identificar ¿cuáles son las mascotas preferidas en niños de 10 a 12 años? El dato estadístico es "perro, raza labrador, color miel".

El dato estadístico se encuentra representado por las mascotas; en el caso particular de los perros, éstos se pueden distinguir según la raza, el tamaño y alimentación.



X_{11} representa la raza de la mascota 11

Y_{10} representa el tamaño de la mascota 10

H_{20} representa el tipo de alimentación de la mascota 20

La variable X_{11} corresponde a la raza del perro número 11 que estamos analizando.



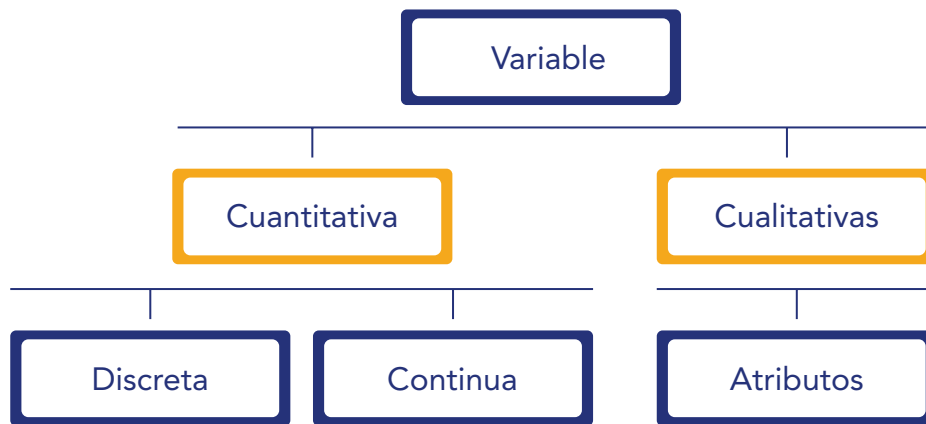
Como se observa, contamos con variables relacionadas con raza, tamaño y alimentación del dato estadístico de las mascotas "perros".

Si colocamos X_{11}, Y_{10}, H_{20}

¿Qué es lo que representa?

Por lo tanto, la variable ayuda a identificar las características del dato mediante letras o símbolos.

Clasificación de las variables



Variables cualitativas: se integra por cualidades o atributos (características como color, tamaño, entre otros).

Variables cuantitativas: representan valores numéricos; se clasifican en discretas y continuas.

Variable discreta: es aquella que toma un número finito o valor de conteo en el cual intervienen números enteros exclusivamente; número de hijos, integrantes del grupo de estudio, cantidad de mascotas de una familia, número probable al lanzar un dado. En todos los casos podrán ser 2 o 3, nunca números decimales o intermedios; no es posible tener 3.2 peces.

Variable continua: puede tomar infinitamente varios valores y se asocia a un proceso de medición: la estatura, el peso, los kilómetros recorridos en bicicleta; comprende valores decimales o intermedios. El peso en los menores de 5 años oscila entre 19.400 gramos y 20.100 gramos en los estándares de la salud.

En los siguientes casos identifica con una "X", si corresponde a una variable cualitativa o una cuantitativa (continua o discreta).



Variable	Cualitativa	Cuantitativa Discreta	Cuantitativa Continua
Medida del calzado			
Número de piezas dentales cariadas en menores de 18 años			
Porcentaje de votación en las pasadas elecciones			
Tipo de música más escuchada en Tenancingo			
Ingreso mensual familiar			
Estatura de los estudiantes de segundo año de nivel primaria			
Color de suéter utilizado en las escuelas secundarias			
Número de latidos de corazón por minuto en personas de 70 a 80 años			
Número de integrantes de las familias de la ciudad de Metepec			

Clasificación de los números

Para analizar los niveles o escalas de medición de la estadística es importante identificar la aplicación y características de los números cardinales, ordinales y nominales.

Número cardinal: indica la cantidad de elementos de un conjunto; pueden ser números enteros o decimales.

Ejemplo:

1 kilo de tortillas, 10 centímetros de listón, 1 pulgada de grosor, 10 niños (Portus, 1998).

Número ordinal: denota la posición o el orden de un elemento dentro de una sucesión con base en tres premisas: A es mayor que B , A es menor que B y A es igual que B .



Ejemplo

En un conjunto de datos de 10 números enteros que van del 110 al 120, tenemos que 115 es mayor que 112, 110 es menor que 120 y que 115 es igual a 115; otra manera de ordenar los datos es de forma creciente o decreciente.

Número nominales: se usan para identificar o etiquetar algo; no representan ningún resultado al sumar, dividir o multiplicar.

Ejemplo

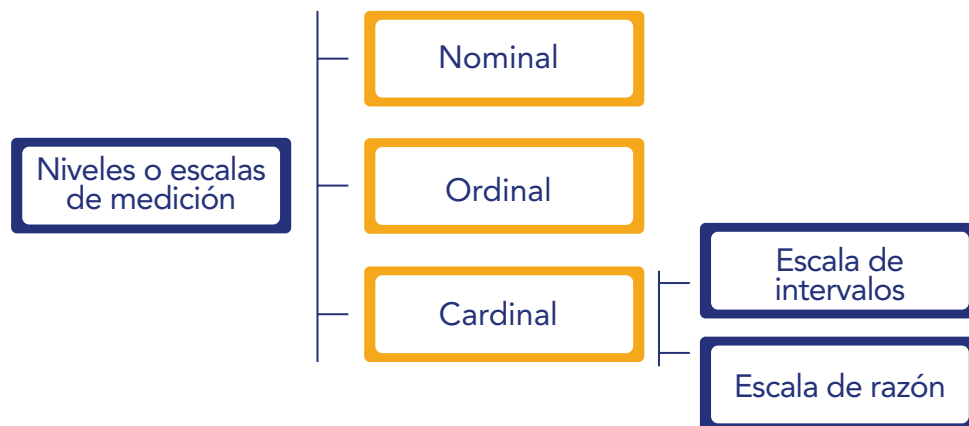
Los números de cuenta de los estudiantes o la matrícula de los automóviles.



Anota dentro del paréntesis el número que corresponda.

1. Ordinal	()	Número de celular Número de lotería CURP
2. Nominal	()	Mi edad son 18 años Pares (0, 2, 4, 6 y 8) impares (1, 3, 5, 7 y 9) Abril tiene 30 días
3. Cardinal	()	Jorge ganó el primer lugar A,B,C,D

Niveles o escalas de medición



Escala nominal: se utiliza para categorizar haciendo uso de nombres o números nominales. Ejemplos: número telefónico, división de pantalones con defecto y sin defectos de fábrica, asistencias e inasistencias.

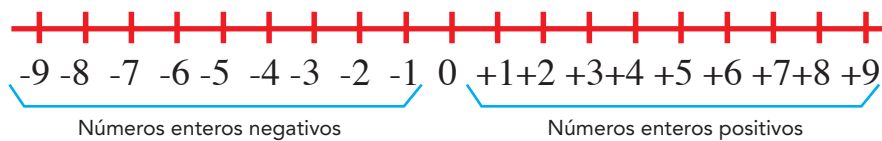
Escala ordinal: se ordena con base en algún criterio y se tiene la existencia de un orden mayor que la escala nominal. Su principal característica reside en la existencia de una

relación entre las variables y el orden. Ejemplos: la relación entre orden (1^{er}, 2^{do} y 3^{ro}), clasificación (peso de 30-40, de 40-50), tamaño (chico, mediano, grande), talla (mediana, grande, extragrande). Los integrantes de un grupo de acuerdo al semestre (primero, tercero y quinto) representan un ejemplo de este tipo de variable.

Escala cardinal: está asignada a la cuantificación y se asocia a las operaciones aritméticas; se dividen en **escalas de intervalos** y **escalas de razón**.

Un ejemplo de escala cardinal es la distancia entre dos números; es decir, de 3 a 5 siempre van a estar contenidos 3 números, lo mismo que de 14 a 16.

Recta Numérica



Escalas de intervalos: en este tipo de medida es importante identificar que representa un atributo a los valores de la escala. Y el cero es un valor arbitrario que ayuda a la asociación de características de los números con los que se pueden realizar operaciones aritméticas de suma, resta, multiplicación y división.

Ejemplo

La estatura de los hermanos Fabián y Yael. Tomando como base la estatura de Yael se marca una línea para identificar su altura, posteriormente se marca otra línea con la de Fabián y con una regla se mide la diferencia. El cero arbitrario está representado por la estatura de Yael, y a partir de ella se determina la diferencia de tamaño entre los hermanos. Se establece que la diferencia entre $80-40=40$ cm. Por lo tanto, Fabián es 40 cm más alto que Yael. Existiendo una relación de $80:40=2$, es decir, Fabián es el doble de centímetros más alto que Yael.



Para el caso del tiempo, se cuenta con el valor arbitrario de antes y después de Cristo. Con relación a la temperatura, al ser medida en grados centígrados, el cero representa un valor relativo y por lo tanto es un intervalo entre la presencia y ausencia de temperatura.



Escala de razón: en esta escala el cero es un valor real y no indica ningún atributo de los números; por lo tanto, la medición del cero es absoluta, como es el caso del peso en gramos o el sueldo pagado a los trabajadores.

Siguiendo con el ejemplo de los hermanos, se toma una cinta métrica y se mide a partir del cero que marca la cinta en el suelo identificando que Yael mide 1.40 cm; si Fabián mide 40 cm más que Yael, entonces $1.40 + 40 \text{ cm} = 1.80 \text{ cm}$. En este caso se observa que el cero es el número real a partir del cual se registra la estatura. Se observa una relación entre las estaturas de $180:140 = 1.2857$

De donde la estatura de Fabián es 1. 2857 de la estatura de Yael.



Sujeto 1
Estatura: 1.70 m
Peso: 86.7
IMC: 30
Grasa corporal: 9%



Sujeto 2
Estatura: 1.70 m
Peso: 86.7
IMC: 30
Grasa corporal: 35%

Los datos estadísticos integrados en las variables cualitativas se ubican en la escala o nivel nominal u ordinal, mientras que las cuantitativas en intervalos o de razón.



Identifica el tipo de nivel o la escala de medición.

Ejemplo	Nominal	Ordinal	Intervalar	De razón
Azul				
18 cm				
Teléfono: 1406830				
Juan tiene \$4 pesos más que Pepe				
80.25 kg				
Soltero				
Luis llegó en tercer lugar				

1. Si las variables son cualitativas, la escala se clasifica como _____
2. En la escala de intervalos el cero es un valor _____; esto quiere decir _____
3. En la escala de razón el cero es un valor _____; esto quiere decir _____
4. En una medición cualitativa, las variables se agrupan en _____
5. En una variable cuantitativa, las variables se agrupan en _____

1.1.1 Población y muestra

Según la RAE (2017), el término universo o población hace referencia al conjunto de personas que habitan un determinado lugar y a aquellos individuos de la misma especie que ocupan un espacio geográfico.

En estadística, la población o universo es el conjunto de datos recolectados que reúnen determinadas características; pueden estar integrados por individuos u objetos. A veces las poblaciones pueden ser muy grandes, lo que hace necesario considerar una muestra representativa que permita analizar los datos.

La **muestra** es la parte proporcional de la población que será analizada, contando con la característica de confiabilidad para generar información necesaria.

Ejemplo

El Plantel “Dr. Pablo González Casanova” de la UAEM, cuenta con una población estudiantil total de 1 050 personas, se desea medir y pesar a los estudiantes para corroborar que no existan riesgos de salud en los adolescentes. Debido a la magnitud de la población se hace necesario determinar una muestra representativa con el propósito de usar información confiable para ahorrar tiempo en la obtención y el análisis de resultados.



Ecuación para determinar la muestra:

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



Donde:

n = tamaño de la muestra

N = número de elementos de la población

e = error máximo

Ejemplo

Se determina la selección de la muestra de 1 050 personas que integran la población del Plantel "Dr. Pablo González Casanova" para identificar la cantidad de estudiantes que es necesario pesar y medir considerando un margen de error de 5%.

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

$$n = \frac{1050}{1 + 1050 (.05)^2}$$

$$n = \frac{1050}{1 + 1050 (.0025)}$$

$$n = \frac{1050}{1 + 2.625}$$

$$n = \frac{1050}{3.625}$$

$$n = 289.65$$

Por tanto, para una población de 1 050 estudiantes, y tomando como error máximo 5%, se tiene que la muestra representativa es de 290 estudiantes.



Determina la muestra representativa de tu salón de clases, tomando como margen de error 5%.

1.2 Registro y ordenamiento de datos

Un dato estadístico es un valor, medida o característica que permite ser observada o contabilizada para su posterior análisis, ya que si no están organizados no hay forma de analizarlos. Para la obtención de datos estadísticos se usan diversos instrumentos, los cuales deben de ser veraces y confiables; los principales instrumentos para la recolección de datos son los cuestionarios (abiertos o cerrados), las entrevistas y las escalas de actitudes, entre otros.

Después de obtener la información es necesario clasificarla u ordenarla para que sea uniforme.

La acción de **registrar** corresponde a las unidades de información que pueden medirse o cuantificarse.

La acción de **ordenar** es organizar los datos de una forma conveniente para la investigación que se está realizando.

Los pasos para elaborar un orden y un registro en una tabla son los siguientes:

1. Se elabora un cuestionario (cualitativo o cuantitativo) y se aplica a la muestra representativa o al total de la población.
2. Se obtiene la información de los cuestionarios aplicados.
3. Se registran y ordenan según las respuestas.

Con base en esta información se elabora la tabla de frecuencias.

1. Cuestionario



2. Aplicación de cuestionario



Imagen derecha
Fuente: <https://www.questionpro.com/blog/es/cualitativa-cuantitativa-o-ambas-haz-investigacion-de-mercados/>

3. Registro y orden de respuestas cualitativas

PREGUNTAS	RESPUESTAS	302
11. ¿Cuáles son las páginas de internet que más utilizas para tareas escolares?	302 respuestas	
	Redalyc (24)	
	Wikipedia (18)	
	Google (17)	
	Wikipedia (14)	
	google (12)	
	wikipedia (10)	
	redalyc (7)	
	PDF (6)	
	Redalyc (6)	
	Google (6)	
	Yahoo (3)	
	wikipedia (3)	

4. Registro y orden de respuestas cuantitativas

PREGUNTAS	RESPUESTAS	302
36. ¿Cuántas veces aproximadamente al semestre hiciste uso de los servicios de la biblioteca virtual?	302 respuestas	
	2 (27)	
	5 (18)	
	1 (18)	
	3 (16)	
	4 (15)	
	6 (11)	
	ninguna (8)	
	2 veces (6)	
	10 (6)	
	0 (6)	
	Ninguna (8)	
	2 o 3 (5)	


 M
1

 Actividad
Autónoma

Pregunta el número de integrantes de su familia a cinco de tus compañeros; registra y ordena las respuestas en orden alfabético, iniciando por apellido paterno, materno y nombre.

Compañero	Respuesta

1.3 Tabla de distribución de frecuencias

La tabla de frecuencias o tabla de distribución de frecuencias es el cuadro en el que se registran de forma organizada, en clases, categorías o intervalos, los datos estadísticos obtenidos, ya sean cualitativos o cuantitativos.

Los **datos cualitativos** se clasifican en clases o categorías de acuerdo con la información recabada, posteriormente se realiza el conteo de cada categoría que se registra en la tabulación para determinar su frecuencia.

Frecuencia: es el número de veces que se repite un elemento en cada categoría del conjunto de datos estadísticos.

Ejemplo

Dato estadístico: zapato

Variable cualitativa: color de zapatos

La tabla muestra los colores de zapatos y la cantidad o frecuencia, que llevan puestos los estudiantes del grupo 307 de un plantel de la UAEM.

Color (categoría)	Tabulación	Frecuencia
Azul	III	3
Rojo	HHH	5
Blanco	HHH HHH	10
Negro	HHH HHH HHH HHH	20
Morado	II	2
Verde	I	1
Café	HHH HHH	10
Naranja	IIII	4
Total		55

Se observa que el color de zapatos predominante es el negro, cuya frecuencia es de 20 estudiantes.

Los estudiantes de un plantel tienen la intención de vender paletas de sabores, por lo cual preguntaron a 30 de sus compañeros ¿cuál es su sabor preferido?

Las siguientes fueron las respuestas obtenidas.

Piña	Fresa	Chocolate	Vainilla	Limón	Vainilla	Chocolate	Fresa	Piña	Vainilla
Limón	Piña	Fresa	Chocolate	Vainilla	Fresa	Chocolate	Vainilla	Limón	Limón
Limón	Piña	Fresa	Chocolate	Vainilla	Limón	Piña	Fresa	Chocolate	Vainilla

Realiza la tabla de frecuencia y escribe cuál es el sabor preferido; posteriormente tabula para obtener la frecuencia.



Categoría	Tabulación	Frecuencia
-----------	------------	------------

Para el caso de **datos cuantitativos**, los valores se unifican de acuerdo con clases o categorías, denominadas intervalos de clase.

Ejemplo

La siguiente corresponde a una tabla de frecuencias con datos cuantitativos.

La cantidad de viajes que han realizado a playas mexicanas durante su vida 111 personas del municipio de Metepec entre 10 y 44 años de edad.

Intervalo de clase	Tabulación	Frecuencia
10 – 16	HHH HHH	10
17 – 23	HHH HHH HHH III	18
24 – 30	HHH HHH HHH HHH III	23
31 – 37	HHH HHH HHH HHH HHH IIII	29
38 – 44	HHH HHH HHH HHH HHH HHH I	31
Total		111



Se observa que las personas entre 38 y 44 años han viajado a estas playas 31 veces, mientras que los niños y los adolescentes son los que menos lo han hecho, con una frecuencia de 10.

La construcción de la tabla de frecuencias para datos cuantitativos se realiza de la siguiente manera:

1. Se determina el rango
2. Se determina la variación entre los datos estadísticos
3. Se determina el tamaño de intervalo o número de intervalos
4. Se elabora la tabla de frecuencia con sus límites de clases

Rango: es la diferencia entre el valor mayor y el menor de un conjunto de datos estadísticos.

Ecuación:

$$\text{Rango} = \text{dato mayor} - \text{dato menor}$$

Dato mayor: denominación o valor máximo que se ubica en la serie de datos.

Dato menor: denominación o valor mínimo en el conjunto de datos.

Variación: representa la diferencia mínima entre los datos estadísticos.

En el siguiente conjunto de números se identifica la diferencia mínima entre ellos para determinar la variación que existe.

1, 9, 7, 12, 17, 10, 6, 1, 2 la diferencia mínima es $10 - 9 = 1 \therefore = 1$ variación

3, 15, 21, 15, 27, 9, 30 la diferencia mínima es $30 - 27 = 3 \therefore = 3$ variación

2.1, 3.5, 4.5, 5.4, 3.4, 9.7, 10.2, 9.7, la diferencia mínima es =



Tamaño de intervalo: es la longitud de los intervalos y representa la diferencia entre el límite superior y el límite inferior; se identifica con la letra "C".

Ecuación:

$$\text{Tamaño de intervalo} = \frac{\text{rango} + \text{variación}}{\text{número de intervalos}}$$

Número de intervalo: cantidad de límites que va a contener una tabla de frecuencias, lo que va a delimitar la acumulación o conteo de datos estadísticos.

Ecuación:

$$\text{Número de intervalo} = \frac{\text{rango} + \text{variación}}{\text{tamaño de intervalos}}$$

Núm.	Intervalo	Frecuencia f_i
1	5.1 – 5.5	7
2	5.6 – 6.0	8
3	6.1 – 6.5	13

Observamos una tabla de frecuencia que cuenta con tres renglones, a cada uno se les nombra intervalo

El intervalo va del 5.1 al 5.5 (esto es el tamaño del intervalo)

Cada intervalo se encuentra integrado por un límite inferior y uno superior. En el caso de la tabla que se integra por tres intervalos, los números 5.1, 5.6 y 6.1 son los límites inferiores y 5.5, 6.0 y 6.5 los superiores.

Es decir, el intervalo número dos se encuentra integrado por el límite inferior 5.6 y el límite superior 6.0.

Te sugerimos ordenar de menor a mayor el conjunto de datos para que sea más simple elaborar la tabla de frecuencias.

Ejemplo

Se entrevistó a un conjunto de 40 personas para identificar los minutos que caminan de forma diaria de su casa al trabajo. Se obtuvieron los siguientes resultados:

Elaborar la tabla de frecuencias con 5 intervalos

59	57	58	57
51	55	54	56
10	12	24	35
20	30	25	36
23	13	28	37
19	32	16	17
33	40	26	27
33	50	26	38
48	38	39	41
30	39	45	42

$Rango = \text{dato mayor} - \text{dato menor}$

Rango = $59 - 10 = 49$

$Variación = 25 - 24 = 1$

La variación: 1, los datos no agrupados aumentan de uno en uno.

Número de intervalos = 5

$Tamaño\ del\ intervalo = \frac{rango + variación}{número\ de\ intervalos} = \frac{49 + 1}{5} = 10$

Ya conocemos que la tabla va a contener 5 intervalos de tamaño 10.

La construcción de la tabla de frecuencias se inicia colocando el dato menor (10) en el primer intervalo y en el intervalo número cinco se finaliza con el dato mayor (59).

El número 10 representa el límite inferior del primer intervalo de clase:



Número de intervalos	Intervalo de clase	
	Límite inferior	Límite superior
1	10	
2		
3		
4		
5		59

Para determinar el límite superior del primer intervalo de clase se realiza la siguiente ecuación:

Límite superior = (límite inferior + tamaño de intervalo) – variación

Límite superior = $(10 + 10) - 1 = 19$

Número de intervalos	Intervalo de clase
1	10 – 19
2	20
3	
4	
5	– 19

Para conocer el límite inferior del segundo intervalo es necesario sumar la variación al límite superior del primer intervalo:

Límite inferior del segundo intervalo es = Límite superior del intervalo anterior más la variación

Límite inferior del segundo intervalo = $19 + 1 = 20$

Límite superior del segundo intervalo = $(20 + 10) - 1 = 29$

Límite inferior del tercer intervalo = $29 + 1 = 30$

Con base en los datos de la página 31, completa la información en las columnas de tabulación y frecuencia.

Intervalo de clase	Tabulación	Frecuencia
10 – 19		
20 – 29		
30 – 39		

Intervalo de clase	Tabulación	Frecuencia
40 – 49		
50 – 59		
Total		40

Para comprobar que la tabla de frecuencias está bien elaborada, debes de iniciar con el dato menor y finalizar con el dato mayor.

En la tabla de frecuencia que se ha elaborado se observa que los datos son discretos, es decir, son números enteros. ¿Cómo se elabora una tabla de frecuencias cuando los datos son continuos (peso, estatura, dinero que traes en la bolsa)?

Se consideran los decimales para realizar la tabla de frecuencias como a continuación se describe en el ejercicio.

Se hizo un análisis de la cantidad de kilómetros recorridos por galón de gasolina según marca y año a 16 vehículos y se obtuvieron los siguientes resultados.

28.4	28.6	28.5	28.2	28.3	29.9	30.1	28.7
29.7	29.1	29.0	29.8	27.6	28.9	30.2	30.0

Con base en las fórmulas, determinamos rango, variación, número y tamaño de intervalos. Recuerda que existen diversas formas de elaborar una tabla de frecuencias considerando para cada intervalo los decimales o no; así como el tamaño del intervalo y el número de intervalos de acuerdo a la cantidad y característica de los datos estadísticos.

Caso 1. Se consideran los decimales:

$$\text{Rango} = \text{dato mayor} - \text{dato menor}$$

$$\text{Rango} = 30.2 - 28.2 = 2.0$$

$$\text{Variación} = 0.1$$

Se decide que el número de intervalos sea 3 por la cantidad de datos:

$$\text{Tamaño de intervalo} = \frac{\text{rango} + \text{variación}}{\text{números de intervalos}} = \frac{2.0 + 0.1}{3} = 0.7$$

$$\text{Número de intervalos} = \frac{\text{rango} + \text{variación}}{\text{tamaño de intervalo}} = 3 \text{ (dato asignado)}$$



La tabla de distribución de frecuencias queda de la siguiente forma:

Número de Intervalos	Límite inferior	Límite superior
1	28.2	28.8
2	28.9	29.5
3	29.6	30.2

El cálculo del límite superior para el primer intervalo se realiza de forma normal, como ya fue explicado.

Límite superior = (límite inferior + tamaño de intervalo) – variación

Límite superior = $(28.2 + 0.7) - 0.1 = 28.8$

Límite inferior del intervalo dos = $28.8 + 0.1 = 28.9$

Otra forma de calcular el tamaño del intervalo es $C = (LS - LI) + V$

$C = (\text{Límite superior} - \text{Límite inferior}) + \text{variación}$

$C = (28.8 - 28.2) + 0.01$

$C = 0.7$



El docente de la asignatura de Cálculo Diferencial necesita identificar las calificaciones de sus dos grupos, integrados por 80 estudiantes.

Elabora la tabla de frecuencias con 5 intervalos.

5.1	6.1	5.3	8.1	5.5	8.5	5.2	5.9	6.0
5.6	7.6	6.2	6.3	7.7	8.5	7.1	7.2	7.8
8.1	7.1	8.3	5.6	8.4	5.8	7.2	5.8	7.9
7.2	5.2	6.7	6.4	9.0	6.4	7.0	6.5	6.1
8.2	5.6	7.6	6.7	7.7	6.6	7.3	6.1	7.8
7.2	8.2	6.6	5.3	7.8	5.1	7.4	6.2	7.4
6.6	6.8	6.9	5.7	7.5	8.0	6.8	8.0	6.8
7.3	7.2	7.4	6.8	7.0	6.6	6.9	6.8	7.0
8.6	6.9	6.1	6.2	6.3	6.4	7.5	6.8	

Los pasos para elaborar la tabla de frecuencias son:

1. Rango
2. Variación
3. Tamaño de intervalo
4. Número de intervalo

$Rango = \text{dato mayor} - \text{dato menor}$

$Variación =$

$Tamaño\ de\ intervalo = \frac{rango + variación}{número\ de\ intervalos}$

$$\text{Número de intervalos} = \frac{\text{rango} + \text{variación}}{\text{tamaño de intervalo}}$$

Número de intervalos	Intervalo de clase		Tabulación	Frecuencia
	Límite inferior	Límite superior		

Intervalos reales de clase

Se integran con los intervalos de clase inferior y superior; permiten ubicar con exactitud los datos estadísticos en el intervalo correspondiente cuando éstos contienen números con decimales.

Lo primero que se debe de identificar es la variación entre los intervalos.

Los **límites reales de clase** se calculan restando y sumando la mitad de la variación al límite inferior y superior de los intervalos.

Límite inferior – ½ variación = límite real inferior

Límite superior + ½ variación = límite real superior

Se observa que la variación corresponde a uno, es la diferencia entre el límite inferior y el límite superior del intervalo anterior.

Ejemplo

El límite inferior del tercer intervalo corresponde a 30 y el límite superior del intervalo anterior, que se ubica en el intervalo 2 corresponde al número 29 $\therefore 30 - 29 = 1$

La variación es 1.

Intervalo de clase		Límites reales de clase	
Límite inferior	Límite superior	Límite real inferior LRI	Límite real superior LRS
10	19	9.5	19.5
20	29		
30	39		39.5



Intervalo de clase		Límites reales de clase	
Límite inferior	Límite superior	Límite real inferior LRI	Límite real superior LRS
40	49	39.5	49.5
50	59	49.5	

La mitad de la variación corresponde a $1 \div 2 = 0.5$

Límite real inferior = límite inferior - mitad de la variación

$$10 - 0.5 = 9.5 \text{ LRI}$$

Límite real superior = límite superior + mitad de la variación

$$19 + 0.5 = 19.5 \text{ LRS}$$



Otra forma de calcular el tamaño de intervalo, es haciendo uso de los LRC de la siguiente forma

$$C = \text{LRS} - \text{LRI}$$

$$C = 49.5 - 39.5$$

$$C = 10$$

De la tabla anterior, finaliza las operaciones para calcular los límites reales de clase (inferior y superior).

Marca de clase

Es el punto central o medio de un intervalo de clase y se puede obtener mediante las siguientes ecuaciones:

$$\text{Marca de clase (Mi)} = \frac{\text{límite inferior} + \text{límite superior}}{2}$$

$$\text{Marca de clase (Mi)} = \frac{\text{límite real inferior} + \text{límite real superior}}{2}$$

Aplicando las dos ecuaciones vamos a determinar la marca de clase para el intervalo número tres de la siguiente tabla.

$$\text{Marca de clase} = \frac{\text{límite inferior} + \text{límite superior}}{2}$$

$$Mi = \frac{LI + LS}{2}$$

$$\text{Marca de clase (Mi)} = \frac{6.1 + 6.5}{2}$$

$$Mi = 6.30$$

$$\text{Marca de clase (Mi)} = \frac{\text{límite real inferior} + \text{límite real superior}}{2}$$

$$\text{Marca de clase } (Mi) = \frac{6.05 + 6.55}{2}$$

$$Mi = 6.30$$

Núm.	Intervalo	Frecuencia f_i	Límite inferior	Límite superior	Límite real inferior	Límite real superior	Marca de clase Mi
1	5.1 - 5.5	7	5.1	5.5	5.05	5.55	5.30
2	5.6 - 6.0	8	5.6	6.0	5.55	6.05	5.80
3	6.1 - 6.5	13	6.1	6.5	6.05	6.55	6.30
4	6.6 - 7.0	17	6.6	7.0	6.55	7.05	
5	7.1 - 7.5	15	7.1	7.5	7.05	7.55	7.30
6	7.6 - 8.0	10	7.6	8.0	7.55	8.05	
7	8.1 - 8.5	8	8.1	8.5	8.05	8.55	
8	8.6 - 9.0	2	8.6	9.0	8.55	9.05	8.80
	Total	80					

Complementa la tabla de frecuencias y calcula la marca de clase para los intervalos 4, 6 y 7.



Frecuencia acumulada, frecuencia relativa y frecuencia relativa acumulada

La frecuencia (f_i) representa la cantidad de veces que se repitió un dato estadístico en un intervalo.

Frecuencia acumulada: es la sumatoria o acumulación de la frecuencia normal de cada intervalo con la frecuencia del intervalo anterior.

Se obtiene sumando la frecuencia del intervalo uno con la frecuencia del intervalo dos para obtener la frecuencia del intervalo dos.

Núm.	Intervalo	Frecuencia f_i	Frecuencia acumulada f_a
1	5.1 - 5.5	7	7
2	5.6 - 6.0	8	15
3	6.1 - 6.5	13	28
4	6.6 - 7.0	17	
5	7.1 - 7.5	15	
6	7.6 - 8.0	10	
7	8.1 - 8.5	8	
8	8.6 - 9.0	2	80
	Total	80	

Para determinar la f_a en el primer intervalo, se coloca el número de la f_i

Para la frecuencia acumulada del intervalo dos, se suma la f_i del intervalo uno + f_i del intervalo dos: $7 + 8 = 15$


 M
1

 Actividad
Autónoma

Complementa la tabla de frecuencias anterior y calcula la frecuencia acumulada para los intervalos 4 al 7.

Frecuencia relativa: es la proporción que representa para cada intervalo la cantidad de datos que se ubican en él.

Se determina al dividir la frecuencia normal (f_i) entre el total de datos.

Núm.	Intervalo	Frecuencia (f_i)	Frecuencia acumulada (f_a)	Frecuencia relativa
1	5.1 - 5.5	7	7	0.087
2	5.6 - 6.0	8	15	0.100
3	6.1 - 6.5	13	28	
4	6.6 - 7.0	17	45	
5	7.1 - 7.5	15		0.1875
6	7.6 - 8.0	10		
7	8.1 - 8.5	8		
8	8.6 - 9.0	2	80	0.025
	Total	80		1.000

Se divide la frecuencia normal (f_i) ÷ la frecuencia total de datos (80)
Operaciones:
 $7 \div 80 = 0.087$
 $8 \div 80 = 0.100$


 Actividad
Autónoma

Complementa la tabla de frecuencias anterior, calculando los datos de la frecuencia acumulada y de la frecuencia relativa.

Frecuencia relativa acumulada: para obtener este dato es necesario dividir la frecuencia acumulada de cada intervalo entre el número total de datos.

Núm.	Intervalo	Frecuencia (f_i)	Frecuencia acumulada (f_a)	Frecuencia relativa	Frecuencia relativa acumulada
1	5.1 - 5.5	7	7	0.087	0.087
2	5.6 - 6.0	8	15	0.100	0.187
3	6.1 - 6.5	13	28		
4	6.6 - 7.0	17	45		0.562
5	7.1 - 7.5	15		0.187	
6	7.6 - 8.0	10			
7	8.1 - 8.5	8			
8	8.6 - 9.0	2	80	0.025	1.000
	Total	80		1.000	

Se divide la f_a entre el total de datos
 $7 \div 80 = 0.087$
 $15 \div 80 = 0.187$

Una vez completada la tabla de la página 38, con las columnas de frecuencia acumulada y frecuencia relativa, calcula la columna de frecuencia relativa acumulada.



Frecuencia relativa acumulada
0.087
0.187
0.562
1.000

Una empresa dedicada al diseño de modas preguntó a sus 44 trabajadores la cantidad de modelos que realizaron durante el año de acuerdo con las tendencias de moda, y obtuvo los siguientes resultados.



Elabora una tabla de frecuencias considerando un tamaño de intervalo 6.

60	75	92	73	98	65	55	78
67	62	54	69	65	87	48	92
83	88	72	94	75	73	68	60
81	83	72	86	75	68	89	98
45	73	77	85	83	84	70	59
62	79	55	56				

Rango = dato mayor – dato menor

Variación =

Tamaño de intervalo = $\frac{\text{rango} + \text{variación}}{\text{número de intervalos}}$

Núm.	Intervalo	Frecuencia (fi)	Límite real inferior	Límite real superior	Marca de clase (Mi)	Frecuencia acumulada (fa)	Frecuencia relativa	Frecuencia relativa acumulada



1.4 Representación gráfica de datos

La presentación del conjunto de datos que se han agrupado, ya sean cualitativos o cuantitativos, se puede realizar mediante gráficas, las cuales muestran visualmente cómo se encuentra distribuida la información.

Las gráficas más usuales son:

- a) Barra
- b) Histograma de frecuencias (relativa)
- c) Polígono de frecuencias
- d) Ojiva (mayor que y menor que)
- e) Circulograma

Gráfica de barras

Es una forma de representar gráficamente un conjunto de datos o valores; se conforman por **barras** rectangulares, cuya longitud es proporcional a los valores o cantidades representados.

Representación para datos cuantitativos

La presentación en el plano cartesiano se ubica de la siguiente manera:

Eje x : intervalos de datos cuantitativos

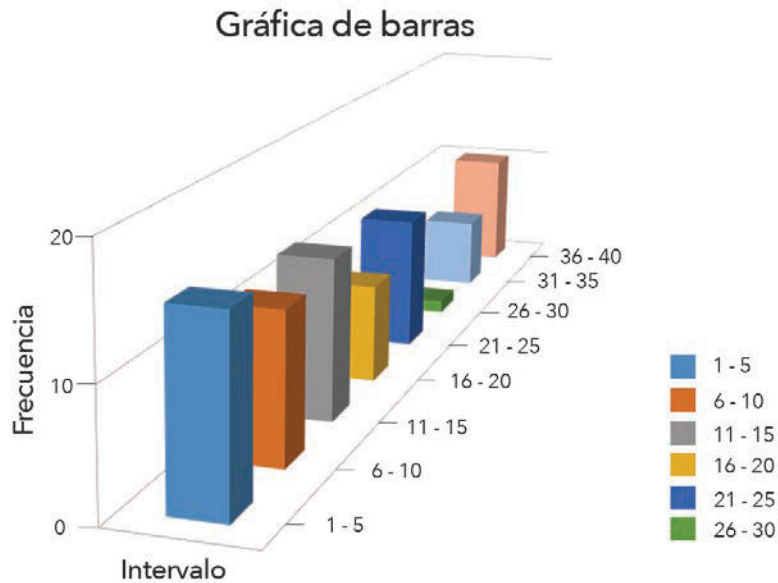
Eje y : la frecuencia normal f_i

Ejemplo

Durante las funciones de 10:00 a 16:00 horas del día domingo en un cine reconocido de la ciudad de Metepec, se entrevistó a 40 familias para identificar las edades de los asistentes. Obteniendo los siguientes resultados.

Intervalo	Frecuencia
1 – 5	15
6 – 10	12
11 – 15	13
16 – 20	8
21 – 25	11
26 – 30	7
31 – 35	6
36 – 40	10
Suma	82

Quedando la gráfica de barras de la siguiente manera:



En la gráfica de barras cuantitativa no se presentan pegadas las barras. Como se observa en la primera barra el tamaño del intervalo va de 1-5, para la segunda barra es de 6-10, mostrando un espacio de uno que representa la variación entre los intervalos; el cual también se debe mostrar en la representación gráfica.

Representación para datos cualitativos

Eje x : categoría o características

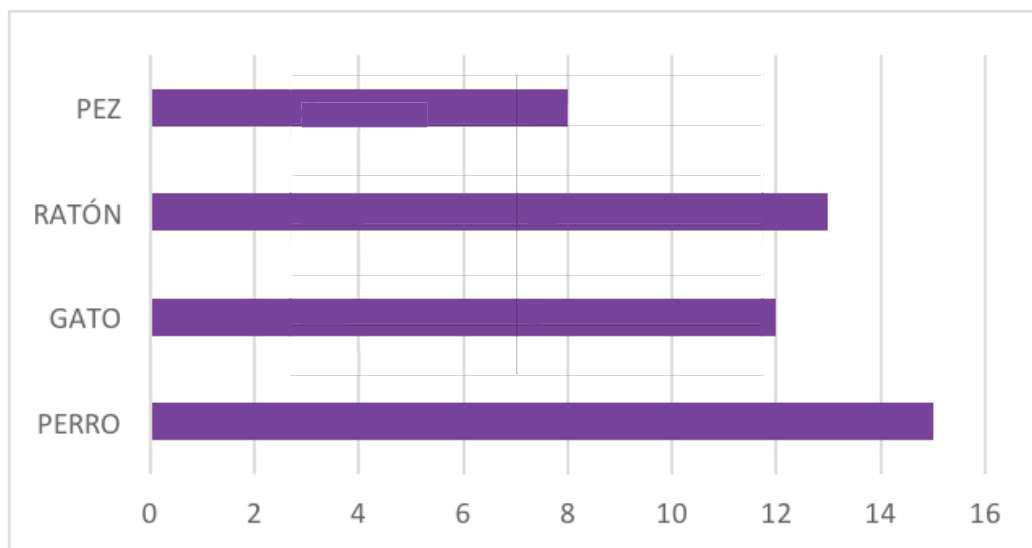
Eje y : frecuencia f_i

Ejemplo

Un veterinario de la colonia Xinantécatl preguntó a una muestra representativa de 48 niños el tipo de mascota que tienen en su casa para identificar el alimento que debe tener en su tienda. Obteniendo los siguientes resultados.

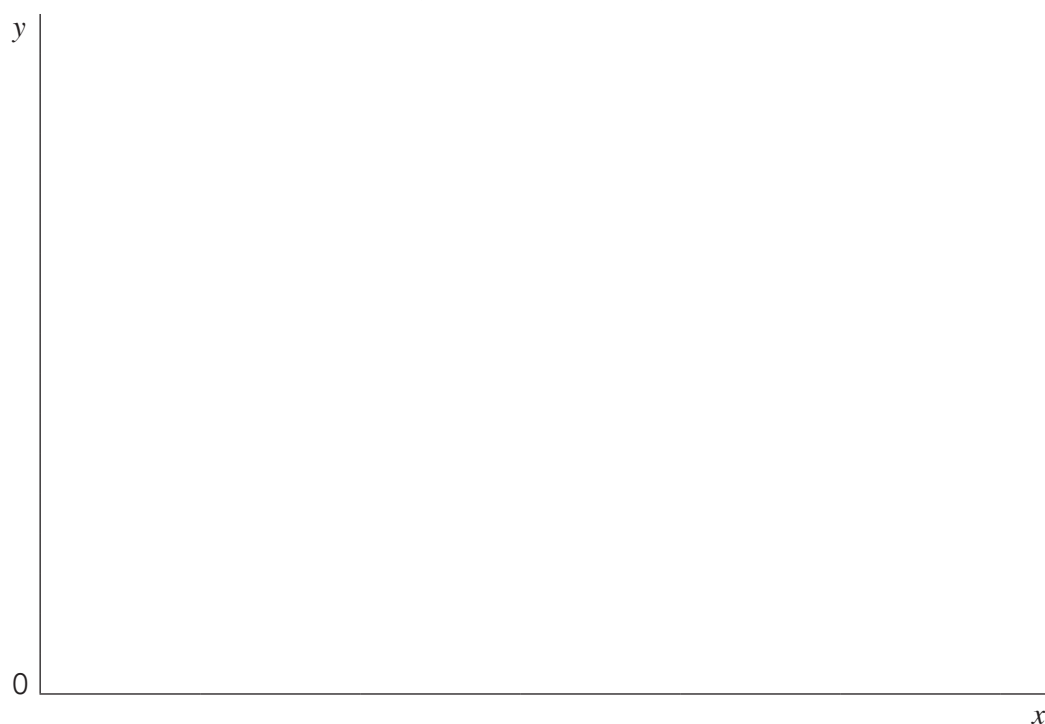
Categoría	Frecuencia
Perro	15
Gato	12
Ratón	13
Pez	8
Suma	48

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



La siguiente tabla de frecuencias muestra la cantidad de horas diarias que pasan los jóvenes entre 12 y 17 años en las redes sociales. Construye la gráfica de barras.

Núm.	Intervalo (horas)	Frecuencia f_i
1	5.1 – 5.5	7
2	5.6 – 6.0	8
3	6.1 – 6.5	13
4	6.6 – 7.0	17
5	7.1 – 7.5	15
Total		60



Histograma de frecuencias

Es una representación gráfica de una variable en forma de barras, en la que la superficie de cada barra es proporcional a la frecuencia de los valores representados.

En cada plano cartesiano se ubica la siguiente información:

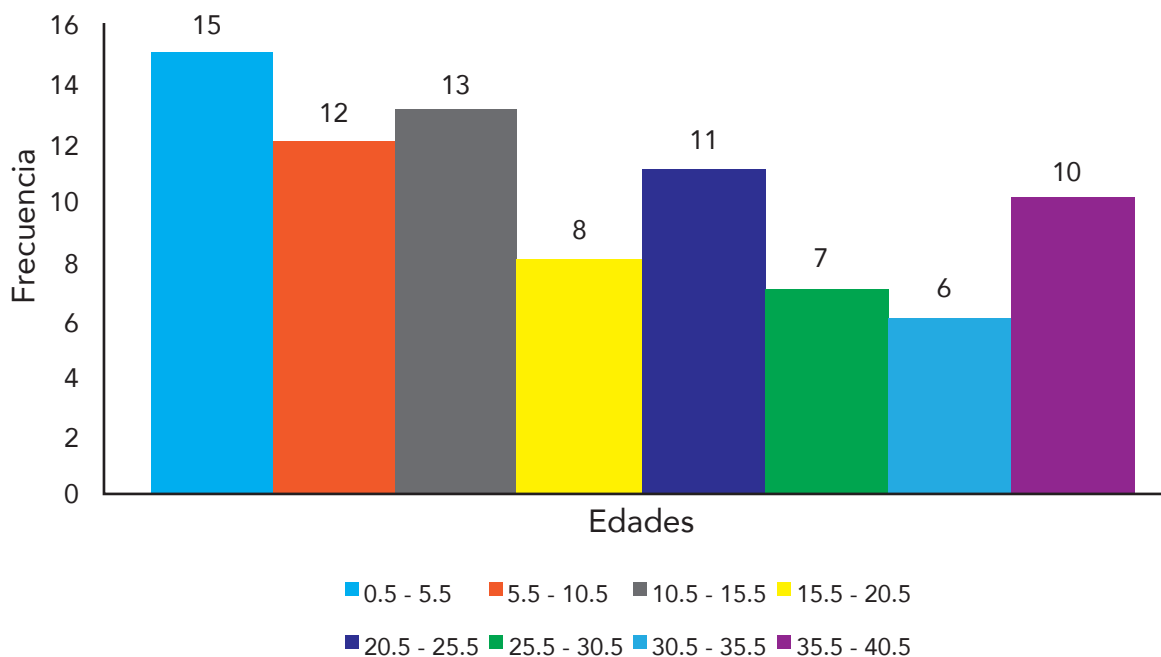
Eje x , límites reales de clase

Eje y , frecuencia normal f_i

Durante las funciones de 10:00 a 16:00 horas del día domingo en un cine reconocido de la ciudad de Metepec, se entrevistó a 40 familias para identificar las edades de los asistentes. Obteniendo los siguientes resultados.

Intervalo (edades)	Frecuencia	Lím. real de clase
1 – 5	15	0.5 – 5.5
6 – 10	12	5.5 – 10.5
11 – 15	13	10.5 – 15.5
16 – 20	8	15.5 – 20.5
21 – 25	11	20.5 – 25.5
26 – 30	7	25.5 – 30.5
31 – 35	6	30.5 – 35.5
36 – 40	10	35.5 – 40.5
Suma	82	

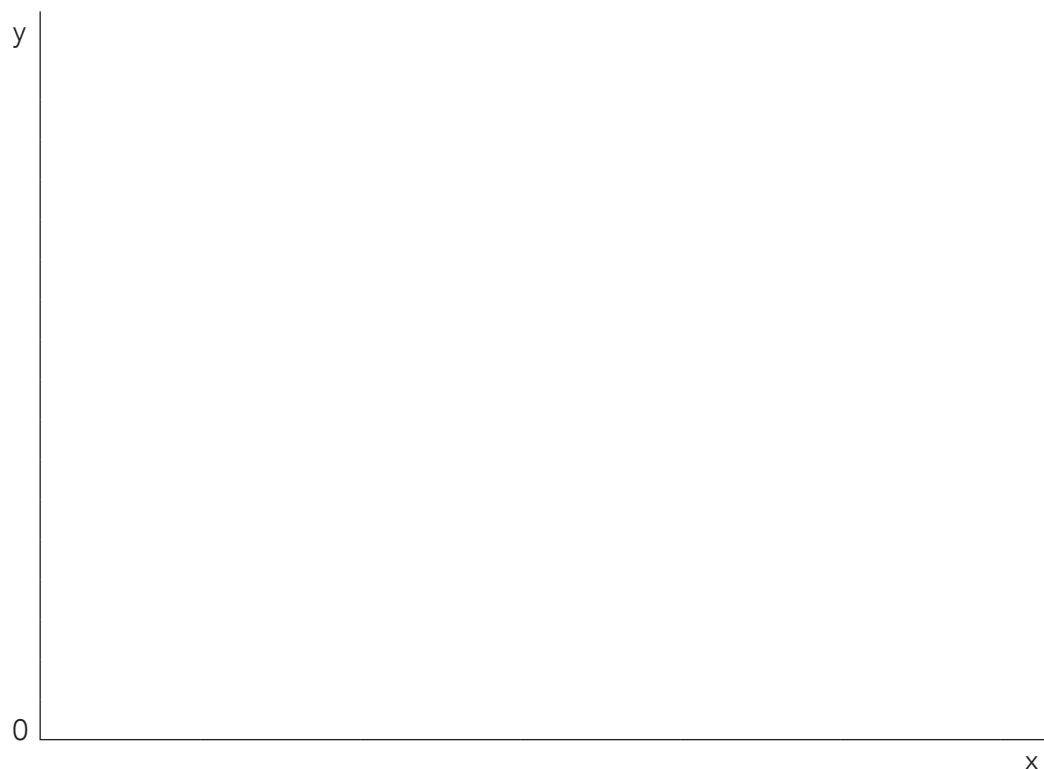
Histograma de frecuencias





La siguiente tabla de frecuencias muestra el consumo de productos lácteos en personas de 5 a 30 años del municipio de Zinacantepec. Construye el histograma de frecuencias.

Límites reales de clase	Intervalos de clase	Frecuencia
	1 – 5	3
	6 – 10	8
	11 – 15	7
	16 – 20	2
	21 – 25	9
	26 – 30	5
	Suma	34



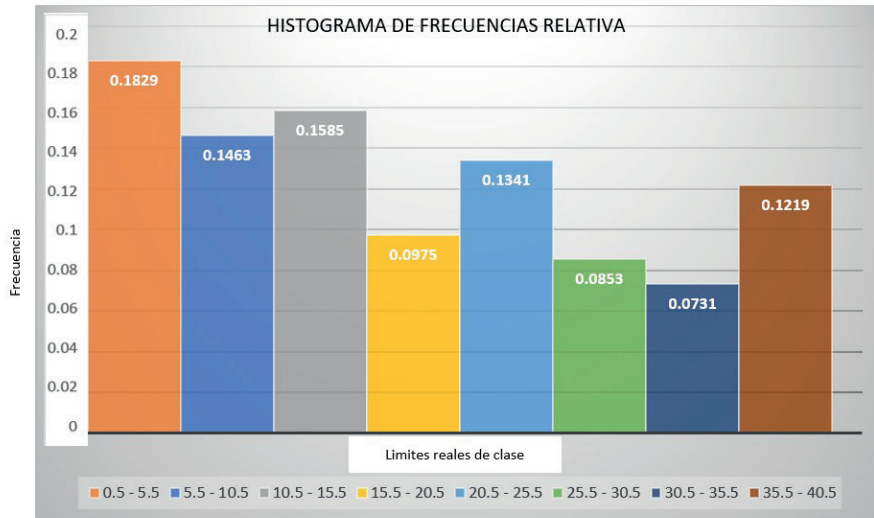
Histograma de frecuencia relativa

Es una representación gráfica de una variable en forma de barras, en la que la superficie de cada barra es proporcional a la frecuencia relativa de los valores representados.

Se elabora considerando la frecuencia relativa; por lo tanto, el eje y (donde se localiza la frecuencia) no podrá ser mayor a 1.

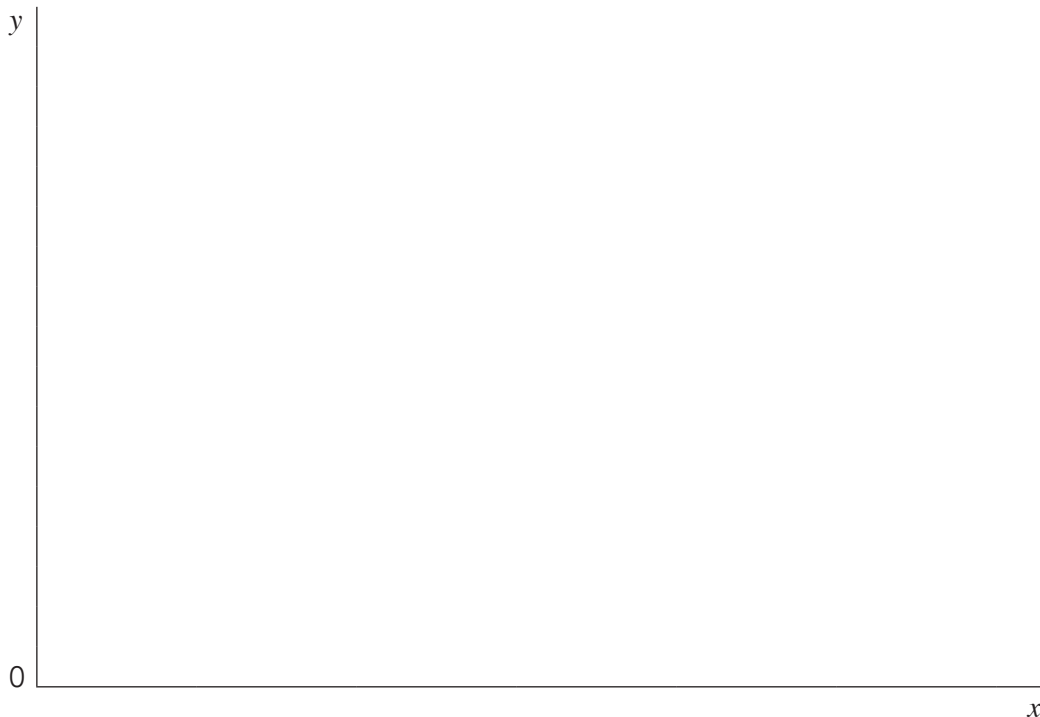
Eje x : límites reales de clase

Eje y : frecuencia relativa



En la siguiente tabla de frecuencias se muestra la información de 40 automovilistas que transitaban por arriba de 100 km/h en la autopista Cuernavaca-Chilpancingo en un horario de 8:00 am a 10:00 am. Construye el histograma de frecuencia relativa.

Núm.	Límites reales	Frecuencia	Frecuencia relativa
1	99.5 – 105.5	5	0.125
2	105.5 – 111.5	7	0.175
3	111.5 – 117.5	10	0.250
4	117.5 – 123.5	18	0.450
Total		40	1.000





Polígono de frecuencias

Es la representación gráfica por medio de líneas que se trazan uniendo las marcas de clase.

Este tipo de gráfica puede incluir barras o bien sólo líneas; es decir, puede presentar de base un histograma de frecuencias.

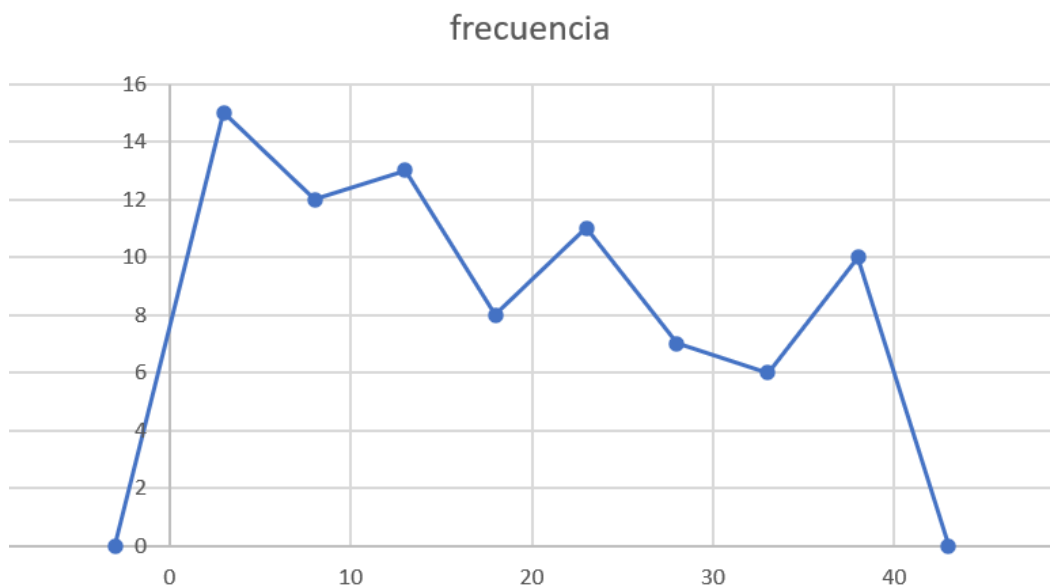
Para realizarla es necesario anexar un intervalo antes y otro al final con frecuencia cero para unir las líneas en el eje x .

Eje x : marca de clase

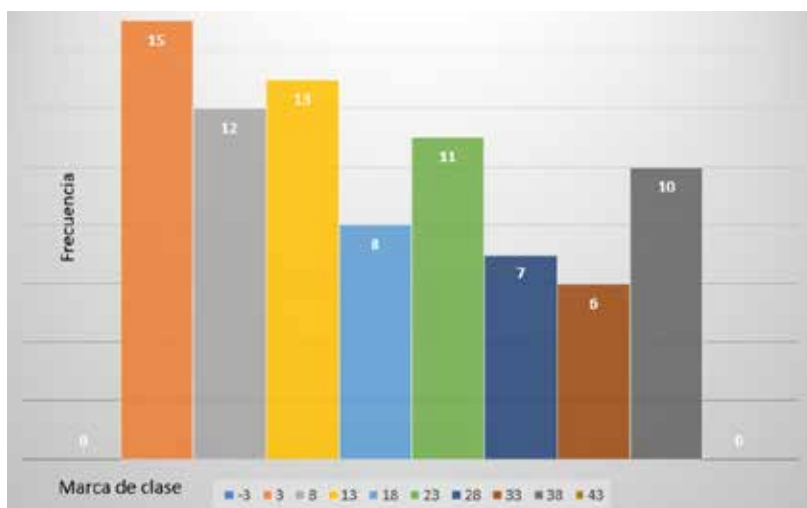
Eje y : frecuencia (fi)

Intervalo	Frecuencia	Lím. real de clase	Marca de clase
	0		-3
1 – 5	15	0.5 - 5.5	3
6 – 10	12	5.5 - 10.5	8
11 – 15	13	10.5 - 15.5	13
16 – 20	8	15.5 - 20.5	18
21 – 25	11	20.5 - 25.5	23
26 – 30	7	25.5 - 30.5	28
31 – 35	6	30.5 - 35.5	33
36 – 40	10	35.5 - 40.5	38
	0		43
Suma	82		

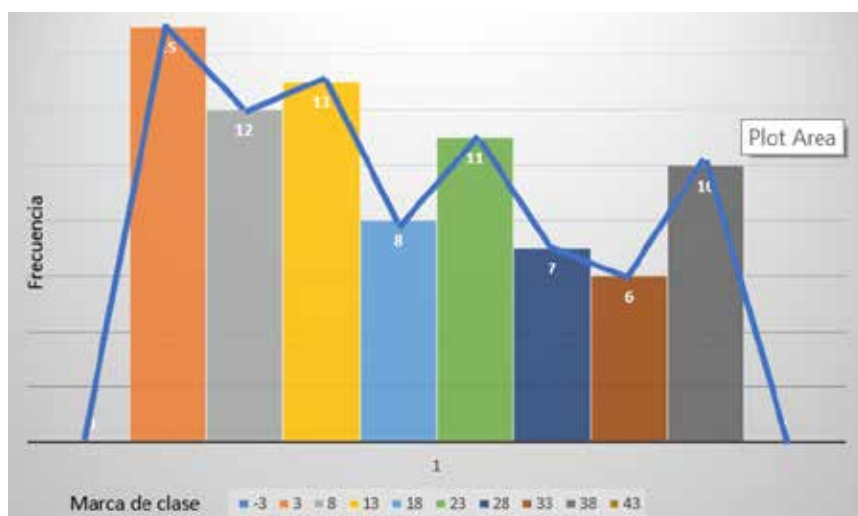
Representación en líneas



Representación con barras



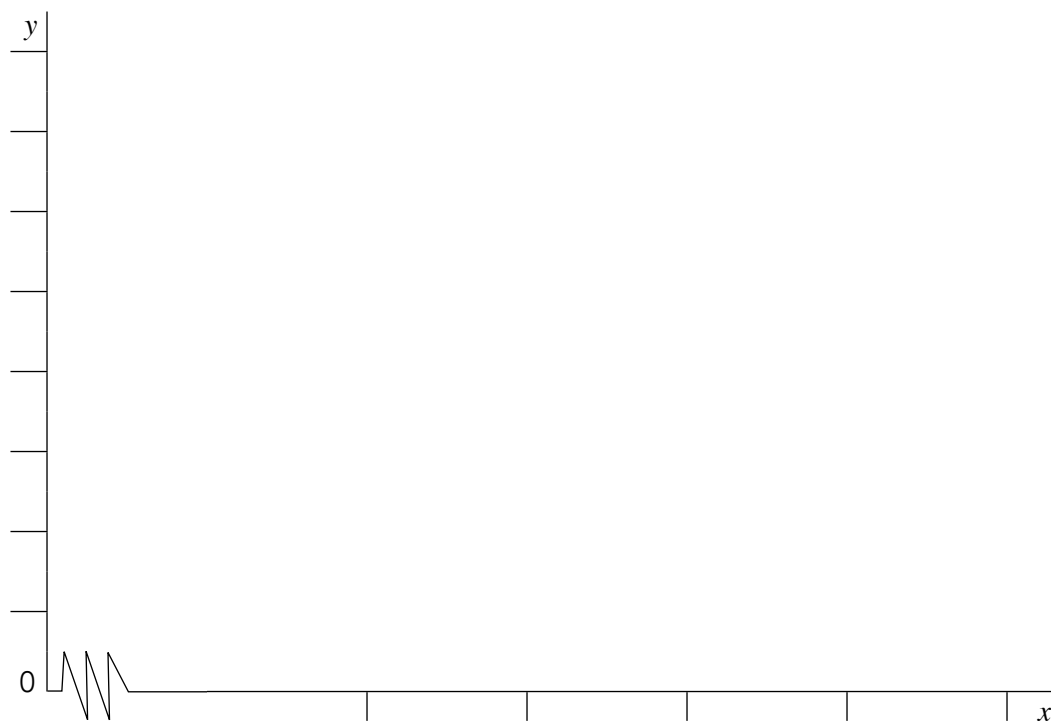
Representación con líneas y barras



En la siguiente tabla de frecuencias se muestra la información de 40 automovilistas que transitaban por arriba de los 100 km/h en la autopista Cuernavaca – Chilpancingo en un horario de 8:00 am a 10:00 am. Construye la gráfica de polígono de frecuencias.



Núm.	Límites reales km./h	Frecuencia	Frecuencia relativa	Marca de clase
1	99.5 – 105.5	5	0.125	
2	105.5 – 111.5	7	0.175	
3	111.5 – 117.5	10	0.250	
4	117.5 – 123.5	18	0.450	
	Total	40	1.000	



Nota: cuando el primer intervalo de frecuencia se localiza muy alejado del cero en el plano cartesiano, se hace un corte en el eje “x” para hacer referencia de que se inicia con el primer intervalo y no se visualice tan alejado del origen.

Ojiva

Es una representación gráfica de líneas; puede representarse por la ojiva mayor que o la ojiva menor que

Ojiva mayor que

Eje x: límite real inferior

Eje y: se localiza la frecuencia acumulada, posteriormente se resta a la suma de la frecuencia la frecuencia normal.

La nueva frecuencia acumulada se obtiene restando al número total de datos la frecuencia de cada intervalo como se muestra a continuación. Se anexa un intervalo al final con frecuencia cero.

Núm.	Intervalo	Frecuencia F_i	Frecuencia acumulada F_a	Eje x Límite real inferior	Frecuencia acumulada
1	1 – 5	7	7	0.5	80
2	6 – 10	8	15	5.5	73

Operaciones para calcular la nueva frecuencia acumulada:

$$80 - 7 = 73$$

$$73 - 8 = 65$$

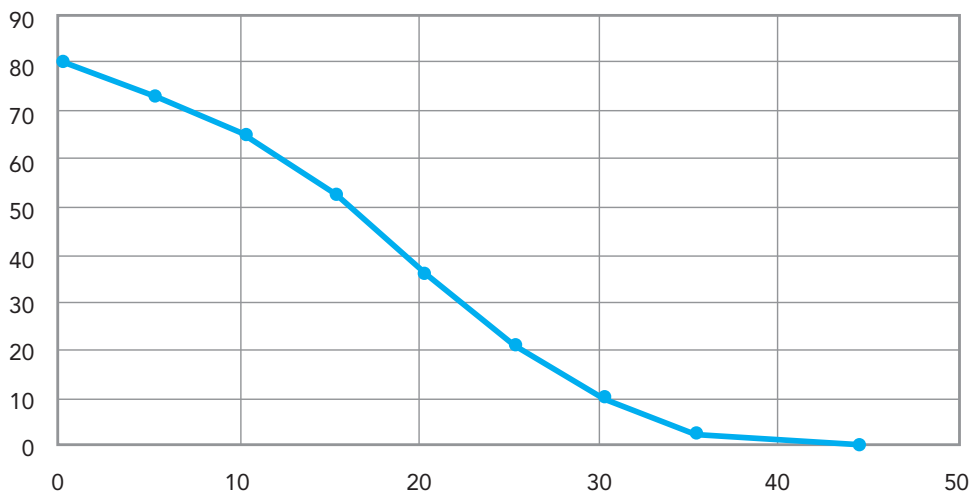
$$65 - 13 = 52$$

Núm.	Intervalo	Frecuencia F_i	Frecuencia acumulada F_a	Eje x Límite real inferior	Frecuencia acumulada y
3	11 – 15	13	28	10.5	65
4	16 – 20	17	45	15.5	52
5	21 – 25	15	60	20.5	35
6	26 – 30	10	70	25.5	20
7	31 – 35	8	78	30.5	10
8	36 – 40	2	80	35.5	2
Sumas		80			
9	41 – 45			45.5	0

Se anexa el renglón 9 con la frecuencia cero para elaborar la gráfica.

La forma de calcular el eje “y” se realiza iniciando con la suma de la frecuencia total, que en este caso es 80, y se le resta la frecuencia normal (7) para obtener el segundo dato de la frecuencia acumulada que será utilizada para elaborar la ojiva mayor que. Se realizan estas operaciones sucesivamente.

Ojiva “mayor que”



Ojiva menor que

Eje x : límite real superior

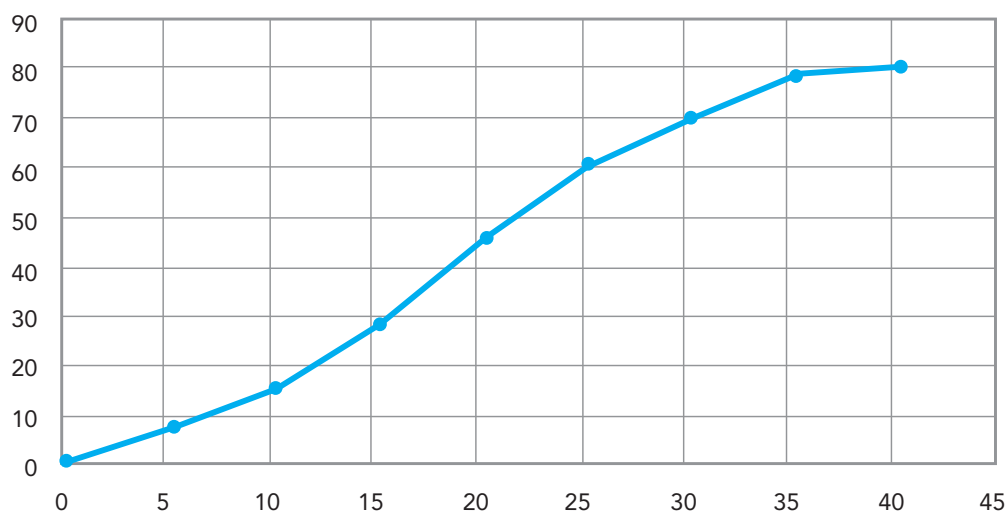
Eje y : se localiza la frecuencia acumulada, posteriormente se resta a la frecuencia total la frecuencia normal.

Se anexa al principio un intervalo con frecuencia cero.



Núm.	Intervalo	Frecuencia f_i	Eje x Límite real superior	Frecuencia acumulada
			0.5	0
1	1 - 5	7	5.5	7
2	6 - 10	8	10.5	15
3	11 - 15	13	15.5	28
4	16 - 20	17	20.5	45
5	21 - 25	15	25.5	60
6	26 - 30	10	30.5	70
7	31 - 35	8	35.5	78
8	36 - 40	2	40.5	80

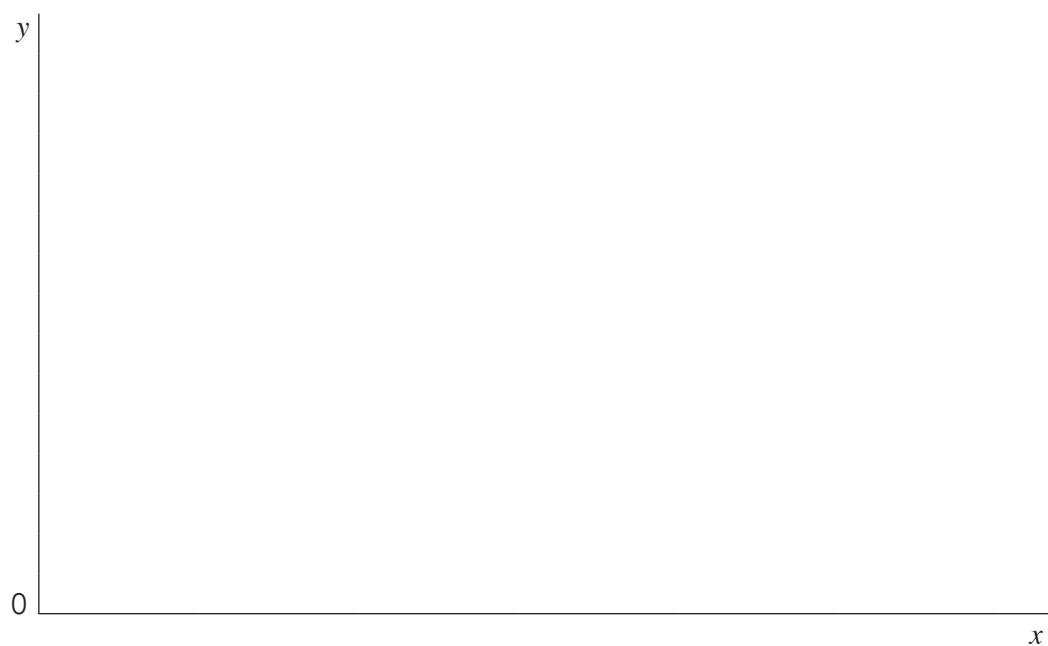
Ojiva "menor que"



Una empresa desea conocer qué tan eficientes son sus operadores telefónicos para responder a sus clientes las dudas relacionadas con sus productos; para ello se revisan las estadísticas de forma aleatoria de 132 clientes. Se obtiene la siguiente información.

Núm.	Intervalo (minutos)	Frecuencia f_i	Frecuencia acumulada f_a	Eje x Límite real superior
			0	
1	0 – 2	10		
2	2 – 4	22		
3	4 – 6	45		
4	6 – 8	50		
5	8 – 10	5		

Elabora la gráfica de ojiva menor que.



Circulograma

También conocida como gráfica de pastel, es aquella representación circular que muestra la frecuencia con base en el ángulo del círculo y el registro del porcentaje de la frecuencia.

Ecuaciones:

$$\text{ángulo} = \frac{360^\circ (f_i)}{n}$$

$$\text{porcentaje} = \frac{(f_i) 100}{n}$$



Tabla para determinar ángulo y porcentaje con datos cualitativos aplicando las ecuaciones.

	Color	Frecuencia	Ángulo	Porcentaje
1	Azul claro	21	94.5	26.25
2	Naranja	12	54	15
3	Lila	35	157.5	43.75
4	Amarillo	3	13.5	3.75
5	Azul turquesa	9	40.5	11.25
	Total	80	360	100

Circulograma (datos cualitativos)

Color de playeras preferidas

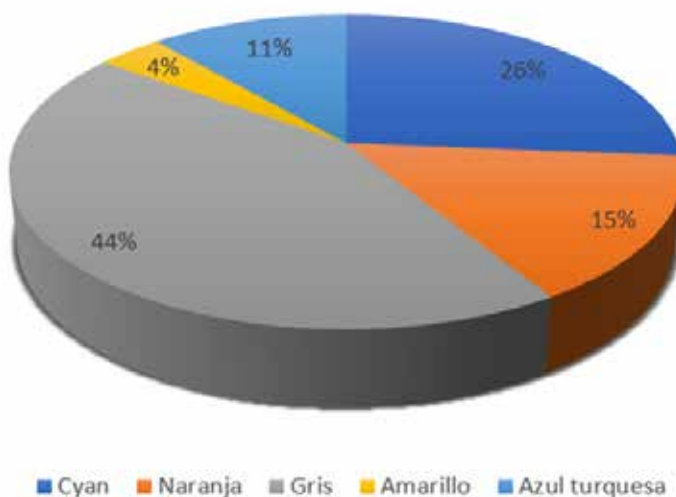
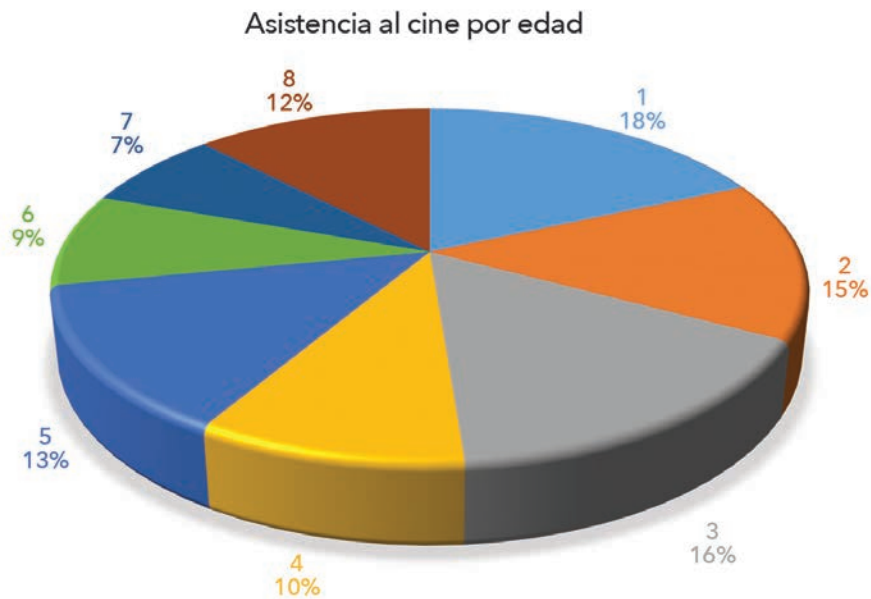


Tabla para determinar ángulo y porcentaje con datos cuantitativos

	Intervalo	Frecuencia	Ángulo	Porcentaje
1	1 – 5	15	65.85	18.29
2	6 – 10	12	52.68	14.63
3	11 – 15	13	57.07	15.85
4	16 – 20	8	35.12	9.76
5	21 – 25	11	48.29	13.41
6	26 – 30	7	30.73	8.54
7	31 – 35	6	26.34	7.32
8	36 – 40	10	43.90	12.20
	Suma	82	360	100

Circulograma (datos cuantitativos)



La siguiente tabla muestra la cantidad de páginas en internet que utilizan los estudiantes de nivel medio superior del grupo 502 para realizar sus trabajos de investigación durante un semestre. Elabora la gráfica de pastel o circulograma.

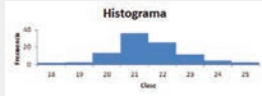
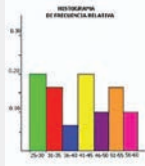
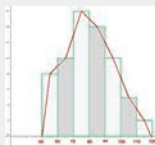
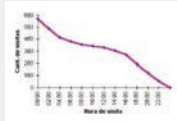
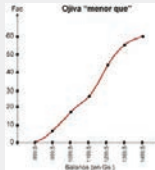


Intervalo (páginas de internet)	Frecuencia	Ángulo	Porcentaje
1 – 10	2		
11 – 20	15		
21 – 30	21		
31 – 40	1		
41 – 50	1		
Suma	40	360	100

Resumen de los datos necesarios para cada eje o ecuaciones para elaborar las gráficas

Gráfico	Eje x horizontal	Eje y vertical	Imagen
Barra	Límites de clase	Frecuencia	



Gráfico	Eje x horizontal	Eje y vertical	Imagen
Histograma de frecuencias	Límites reales de clase	Frecuencia f_i	
Histograma de frecuencia relativa	Límites reales de clase	Frecuencia relativa	
Polígono de frecuencias	Marca de clase	Frecuencia f_i	
Ojiva mayor que	Límite real inferior	Frecuencia acumulada f_a	
Ojiva menor que	Límite real superior	Frecuencia acumulada f_a	
Circulograma	$\text{ángulo} = \frac{360^\circ f_i}{n}$	$\text{porcentaje} = \frac{f_i 100}{n}$	

Ejercicios

1. La cafetería del plantel está realizando un estudio para identificar cuál es el platillo que los estudiantes consumen más, con el objetivo de cubrir la demanda de sus comensales. La muestra representativa es de 40 estudiantes.

Elabora la tabla de frecuencias.

Tortas	Hamburguesa	Enchiladas	Empanadas	Tacos de guisado
Tacos de guisado	Tacos de guisado	Tacos de guisado	Tortas	Empanadas
Empanadas	Empanadas	Enchiladas	Empanadas	Hamburguesa
Tacos de guisado	Tacos de guisado	Empanadas	Tacos de guisado	Tortas

Enchiladas	Tortas	Empanadas	Tortas	Empanadas
Hamburguesa	Empanadas	Hamburguesa	Tacos de guisado	Tacos de guisado
Enchiladas	Tortas	Tacos de guisado	Empanadas	Hamburguesa
Tacos de guisado	Tacos de guisado	Enchiladas	Enchiladas	Enchiladas

2. La siguiente información corresponde a la cantidad de accesorios que compran al mes las mujeres entre 15 a 18 años de edad. Construye la tabla de frecuencias con tamaño de intervalo 10. La tabla debe contener la siguiente información: intervalos, frecuencia, intervalos reales de clase, frecuencia relativa, frecuencia acumulada y frecuencia relativa acumulada.

18	20	45	21	10	26	30	23	12	25
60	10	26	25	27	21	31	28	27	17
18	23	36	10	22	48	11	25	35	26
24	25	20	24	28	22	28	11	29	18
19	29	19	27	33	26	34	21	28	22



3. El club deportivo “Los potros” ha recibido en el mes de octubre un promedio de 40 visitas diarias; se decide tomar el día 23 como muestra para identificar el tiempo (minutos) que pasan los socios dentro de las instalaciones. Con los datos obtenidos, elabora la tabla de frecuencias con 5 intervalos de clase. La tabla deberá contener los intervalos normales, frecuencia, límites reales de clase y frecuencia acumulada.

60	85	120	35	50	135	120	40	90	120
120	40	60	50	90	45	80	50	55	85
45	85	70	120	80	85	55	40	120	120
85	70	135	95	100	120	100	100	135	85

Responde las siguientes preguntas.

a) ¿Cuál es el intervalo que tiene mayor frecuencia?

b) ¿Cuál es el intervalo que tiene menor frecuencia?

4. Se cuenta con la tabla de frecuencias del índice de masa corporal en mujeres. Determina los intervalos reales de clase y la frecuencia relativa acumulada.

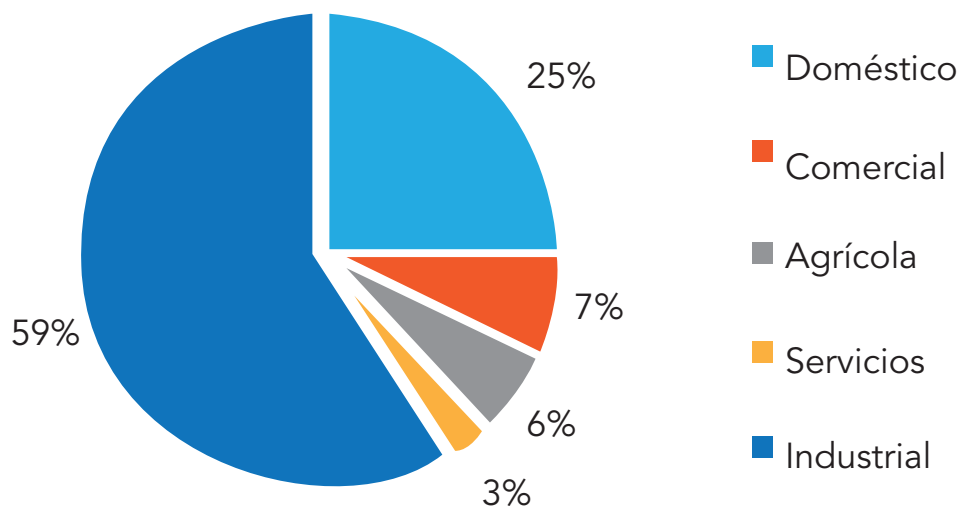
Intervalo Índice de masa corporal en mujeres	Frecuencia
15.0 – 20.9	7
21.0 – 26.9	12
27.0 – 32.9	14
33.0 – 38.9	4
39.0 – 44.9	1

5. Relación de columnas.

a. Histograma	()	Es un gráfico de líneas donde se presenta generalmente la frecuencia relativa acumulada.
b. Distribución de frecuencia	()	Son los valores obtenidos mediante un proceso de conteo.
c. Ojiva	()	Se utiliza para representar una distribución de frecuencias (absolutas) de datos discretos o categorías.
d. Polígono de frecuencias	()	Representa el número de veces que se observó cierto valor o cierta categoría.
e. Gráfica de barras	()	Son los valores obtenidos mediante un proceso de medición.
f. Circulograma o gráfica de pastel	()	Representa el % de cierto valor o categoría.
g. Frecuencia absoluta	()	Se utiliza para representar una distribución de frecuencias (relativas de datos discretos o categóricos).
h. Frecuencia relativa	()	Tablas que representan todos los posibles valores de una variable, junto con el número de veces que cada uno de ellos es observado.
i. Datos discretos	()	Es una gráfica de columnas que representa la distribución de frecuencias de datos continuos.
j. Datos continuos	()	Es una gráfica de líneas que se utiliza para ver la forma de la distribución, así como donde se localiza la mayor concentración de datos.

Fuente: Gutiérrez (2012)

6. La siguiente gráfica de pastel muestra el consumo de energía eléctrica en el Estado de México por sector del año 201__ al año 201__. Contesta las siguientes preguntas.





¿Cuál de los factores de energía es el que consume menos?

¿Cuál de los factores de energía es el que consume más?

¿Qué consecuencias consideras que tiene para el medio ambiente el consumo de energía eléctrica con base en los porcentajes que observas en la gráfica anterior?

7. Los siguientes datos corresponden a 90 calificaciones de la asignatura de matemáticas de estudiantes del tercer grado de secundaria.

Construye la tabla de frecuencias con 8 intervalos que contenga:

- a) Límites reales de clase
- b) Marca de clase
- c) Frecuencia acumulada
- d) Frecuencia relativa acumulada
- e) Ojiva mayor que
- f) Ojiva menor que
- g) Circulograma

La variación corresponde a 0.1

3.1	6.7	10.0	8.9	9.5	10.0	8.9	6.0	8.9
4.5	7.5	2.1	9.0	7.1	5.9	8.5	7.9	8.9
4.7	7.4	1.6	9.1	7.9	1.1	9.1	9.1	1.1
5.5	8.1	6.7	9.0	8.1	9.0	2.2	8.5	5.9
5.9	9.2	7.7	7.7	8.2	9.1	7.7	1.5	9.1
6.2	9.6	7.8	9.0	9.3	9.0	8.5	10.0	2.5
1.0	4.5	7.9	8.8	2.5	8.8	2.5	8.8	1.1
3.3	10.0	10.0	5.9	2.2	3.3	9.0	3.3	6.0
4.5	7.7	8.8	6.0	9.2	6.9	9.1	6.0	6.9
10.0	7.9	9.0	7.9	9.0	9.2	9.0	1.5	10.0

¿Cuántos estudiantes obtuvieron calificación aprobatoria?

¿Cuántos estudiantes obtuvieron calificación reprobatoria?

¿Qué estrategia consideras pertinente que el docente aplique con los estudiantes de calificación reprobatoria?

Autoevaluación

Marca en el recuadro correspondiente la respuesta sí o no a los siguientes indicadores.



Módulo 1

Núm.	Indicadores	Sí	No
1	Identificas la aplicación de la estadística descriptiva e inferencial		
2	Identificas la clasificación de las variables		
3	Reconoces los niveles o escalas de la estadística		
4	Sabes calcular la muestra		
5	Elaboras de forma autónoma la tabla de frecuencias		
6	Identificas y elaboras de forma correcta las diversas gráficas		

Para el caso de las respuestas negativas, escribe cuáles son las actividades o acciones que debes reforzar.

1	
2	
3	
4	
5	
6	

Escribe cuál es la razón por la que se te facilitó o dificultó el proceso de aprendizaje del módulo uno.



Fácil	Difícil

Formulario del módulo uno

Muestra:

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

Elaborar la tabla de frecuencias:

$$\text{Rango} = \text{dato mayor} - \text{dato menor}$$

$$\text{Tamaño de intervalo} = \frac{\text{rango} + \text{variación}}{\text{número de intervalos}} \quad \begin{aligned} C &= (LS - LI) + V \\ C &= LRS - LRI \end{aligned}$$

$$\text{Número de intervalos} = \frac{\text{rango} + \text{variación}}{\text{tamaño de intervalo}}$$

$$\text{Límite inferior} - \frac{1}{2} \text{variación} = \text{límite real inferior}$$

$$\text{Límite superior} + \frac{1}{2} \text{variación} = \text{límite real superior}$$

$$\text{Marca de clase (Mi)} = \frac{\text{límite inferior} + \text{límite superior}}{2} \quad Mi = \frac{LI + LS}{2}$$

$$\text{Marca de clase (Mi)} = \frac{\text{límite real inferior} + \text{límite real superior}}{2}$$

Circulograma:

$$\text{ángulo} = \frac{360^\circ (fi)}{n}$$

$$\text{porcentaje} = \frac{(fi) 100}{n}$$



MÓ DU LO DOS

Medidas de tendencia central
y medidas de posición



Competencias disciplinares

Matemáticas básicas y extendidas

3. Explica e interpreta los resultados obtenidos mediante procedimientos matemáticos y los contrasta con modelos establecidos en situaciones reales.
4. Argumenta la solución obtenida de un problema, con métodos numéricos, gráficos, analíticos o variacionales, mediante el lenguaje verbal, matemático y el uso de las tecnologías de la información y la comunicación.
5. Analiza las relaciones entre dos o más variables de un proceso social o natural para determinar o estimar su comportamiento.
8. Interpreta tablas, gráficas, mapas, diagramas y textos con símbolos matemáticos y científicos.

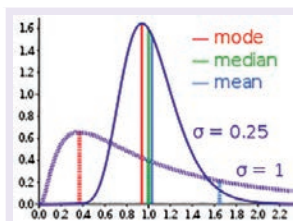
Competencias genéricas

1. Se conoce y valora a sí mismo y aborda problemas y retos teniendo en cuenta los objetivos que persigue.
 - 1.1 Enfrenta las dificultades que se le presentan y es consciente de sus valores, fortalezas y debilidades.
4. Escucha, interpreta y emite mensajes pertinentes en distintos contextos mediante la utilización de medios, códigos y herramientas apropiados.
 - 4.1 Expresa ideas y conceptos mediante representaciones lingüísticas, matemáticas o gráficas.
5. Desarrolla innovaciones y propone soluciones a problemas a partir de métodos establecidos.
 - 5.1 Sigue instrucciones y procedimientos de manera reflexiva, comprendiendo cómo cada uno de sus pasos contribuye al alcance de un objetivo.
 - 5.2 Ordena información de acuerdo a categorías, jerarquías y relaciones.
 - 5.6 Utiliza las tecnologías de la información y comunicación para procesar e interpretar información.
8. Participa y colabora de manera efectiva en equipos diversos.
 - 8.1 Propone maneras de solucionar un problema o desarrollar un proyecto en equipo, definiendo un curso de acción con pasos específicos.
 - 8.2 Aporta puntos de vista con apertura y considera los de otras personas de manera reflexiva.
 - 8.3 Asume una actitud constructiva, congruente con los conocimientos y habilidades con los que cuenta dentro de distintos equipos de trabajo.

Propósito

Calcula las medidas de tendencia central y posición para resolver problemas de su entorno.

Temas del módulo dos

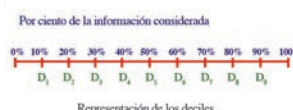


2.1. Medidas de tendencia central

2.1.1 Media aritmética

2.1.2 Mediana

2.1.3 Moda



2.2. Medidas de posición

2.2.1 Cuartiles

2.2.2 Deciles

2.2.3 Percentiles

Introducción

Para el análisis general de las tablas de frecuencias es necesaria la aplicación de diversas medidas, como es el caso de las de tendencia central y posición, las cuales permiten identificar cómo se distribuyen los datos para la toma de decisiones.

En este segundo módulo, los temas a desarrollar son las medidas de tendencia central, que se encuentran integradas por la media, mediana y moda, para conjuntos de datos agrupados y no agrupados. También se expondrán las medidas de posición, entre las que se localizan los cuartiles, deciles y percentiles, que ayudan a determinar la conglomeración de datos de acuerdo con su posición al dividir los datos en cuatro, diez o cien partes.

Evaluación diagnóstica

1. Sandra, una estudiante de excelencia del sexto semestre del Plantel "Dr. Pablo González Casanova", sufre actualmente "mal de amores", razón por la cual sus calificaciones fueron las siguientes:

Asignatura	Calificación
Sociología	8
Psicología	6
Ciudadanía y Civismo	6
Expresión del Arte	7



Asignatura	Calificación
Desarrollo Empresarial	7
Administración (optativa)	9
Economía (optativa)	6

1. ¿Cuál es el promedio de sus calificaciones?
2. ¿Cuál es la moda en sus calificaciones?
3. ¿Cómo se determina un cuartil y qué es lo que representa?

2.1. Medidas de tendencia central

Las medidas de tendencia central o centralización permiten obtener valores que representan el punto central de los datos; es decir, determinan el valor más representativo de la variable que estamos utilizando. Las medidas de tendencia central más utilizadas son media, mediana y moda.

2.1.1. Media aritmética

Es una medida de tendencia central también conocida como promedio; generalmente se representa mediante la variable $\bar{x}$, y se calcula mediante la siguiente ecuación:

Datos no agrupados

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

Donde:

$\bar{x}$ = media del conjunto de datos no agrupados

x_n = valor del i - ésimo dato

n = número total de datos

Ejemplo

Calcula la media de la muestra siguiente:

Muestra núm. 1	3	4	6	3	5	7
----------------	---	---	---	---	---	---

Calcula el promedio de tus calificaciones del quinto semestre. Registra la asignatura y la calificación.



Datos agrupados

Conociendo la tabla de distribución de frecuencia:

$$\bar{x} = \frac{\sum fiMi}{n}$$

Donde

$\bar{x}$ = Media del conjunto de datos agrupados

n = número total de datos

Mi = Marca de clase

fi = Frecuencia del i - ésimo intervalo

Intervalos de clase	Frecuencia
1 – 5	3
6 – 10	8
11 – 15	7
16 – 20	2
21 – 25	9
26 – 30	5
Sumas	34

Las marcas de clase de cada intervalo se calculan mediante la expresión:

Mi = Marca de clase

LI = Límite inferior

LS = Límite superior

$$Mi = \frac{LI + LS}{2}$$

Intervalos de clase	Frecuencia	Mi
1 – 5	3	3
6 – 10	8	8
11 – 15	7	13
16 – 20	2	18
21 – 25	9	23
26 – 30	5	28
Suma	34	



Se determina el producto de la frecuencia por la marca de clase

Intervalos de clase	Frecuencia	Mi	$fiMi$
1 – 5	3	3	9
6 – 10	8	8	64
11 – 15	7	13	91
16 – 20	2	18	36
21 – 25	9	23	207
26 – 30	5	28	140
Sumas	34		547

Se sustituyen valores en la ecuación:

$$\bar{x} = \frac{\sum fiMi}{n}$$

$$\bar{x} = \frac{547}{34}$$

$$\bar{x} = 16.09$$



Una empresa realizó un análisis sobre la cantidad de dinero que gastan sus empleados para trasladarse diariamente del trabajo a su casa y viceversa; los resultados fueron los siguientes. Calcula la media para datos agrupados.

Intervalos de clase	Frecuencia (empleados)	Mi	$fiMi$
30 – 35	5		
40 – 45	7		
50 – 55	12		
60 – 65	9		
70 – 75	2		
Suma	35		

2.1.2. Mediana

Datos no agrupados

Es el valor que se encuentra exactamente a la mitad de un conjunto de datos ordenados, es decir, de menor a mayor o viceversa; se representa mediante la variable $\tilde{x}$. Para el cálculo de la mediana es necesario tomar en cuenta dos factores:

1. Si el número total de datos (n) es par, la mediana es el promedio de los dos valores centrales.
2. Si el número total de datos (n) es impar, la mediana es el valor central.

Muestra núm. 1	3	4	6	3	5	7	par	
Muestra núm. 2	3	4	6	3	5	7	12	impar

Se ordena el conjunto de datos de la muestra número 1 de menor a mayor:

Muestra núm. 1	3	3	4	5	6	7	par	
----------------	---	---	---	---	---	---	-----	--

$$\tilde{x} = \frac{4+5}{2}$$

$$\tilde{x} = \frac{9}{2}$$

$$\tilde{x} = 4.5$$

También se ordenan de menor a mayor los datos de la muestra número 2.

Muestra núm. 2	3	3	4	5	6	7	12	impar
----------------	---	---	---	---	---	---	----	-------

Se localiza el valor central: $\tilde{x} = 5$

Calcula la mediana para datos no agrupados del peso de 6 de tus compañeros de clase.
Calcula la mediana para datos no agrupados de la estatura de 5 de tus compañeros.



Datos agrupados

Conociendo la tabla de distribución de frecuencia

$$\tilde{x} = L + \left(\frac{\frac{n}{2} - fa}{fx} \right) c$$

Donde:

L = límite real inferior del intervalo que contiene a la mediana

n = número total de datos

fa = frecuencia acumulada del intervalo anterior al que contiene a la mediana

fx = frecuencia del intervalo que contiene a la mediana

c = tamaño del intervalo



Intervalos de clase	Límites reales	Frecuencia
1 – 5	0.5 - 5.5	3
6 – 10	5.5 - 10.5	8
11 – 15	10.5 - 15.5	7
16 – 20	15.5 - 20.5	2
21 – 25	20.5 - 25.5	9
26 – 30	25.5 - 30.5	5
Suma		34

Se anexa una columna para calcular la frecuencia acumulada y determinar el intervalo que contiene a la mediana.

Núm.	Intervalos de clase	Límites reales	Frecuencia	Frecuencia acumulada
1	1 – 5	0.5 – 5.5	3	3
2	6 – 10	5.5 – 10.5	8	11
3	11 – 15	10.5 – 15.5	7	18
4	16 – 20	15.5 – 20.5	2	20
5	21 – 25	20.5 – 25.5	9	29
6	26 – 30	25.5 – 30.5	5	34
Sumas			34	

- El número total de datos es 34 ($n = 34$)
- Se determina el intervalo que contiene a la mediana $n = \frac{34}{2} = 17$
- Se localiza en la frecuencia acumulada la cifra más cercana; en este caso es 18
- El intervalo número 3 contiene a la mediana

3	11 - 15	10.5 - 15.5	7	18
---	---------	-------------	---	----

Para calcular el tamaño del intervalo (c), existen las siguientes ecuaciones:

$$\text{Tamaño del intervalo} = \text{límite real superior} - \text{límite real inferior}$$

$$\text{Tamaño del intervalo} = \text{límite superior} - \text{límite inferior} + \text{variación}$$

$$\text{Tamaño del intervalo} = 15.5 - 10.5 = 5$$

$$\text{Tamaño del intervalo} = 15 - 11 + 1 = 5$$

Se sustituyen los valores en la fórmula:

$$\widetilde{X} = L + \left(\frac{\frac{n}{2} - fa}{fx} \right) c$$

$$\widetilde{x} = 10.5 + \left(\frac{\frac{34}{2} - 11}{7} \right) 5$$

$$\widetilde{x} = 10.5 + \left(\frac{17 - 11}{7} \right) 5$$

$$\widetilde{x} = 10.5 + \left(\frac{6}{7} \right) 5$$

$$\widetilde{x} = 10.5 + (0.857) 5$$

$$\widetilde{x} = 10.5 + 4.285$$

$$\widetilde{x} = 14.785$$

Los resultados del examen de admisión a la Facultad de Ingeniería de la UAEM muestran los siguientes datos. Calcula la mediana para datos agrupados.

Intervalos de clase (puntos en examen)	Frecuencia (estudiantes)	Límites reales	Frecuencia acumulada
96 – 100	175		
101 – 105	143		
106 – 110	121		
111 – 115	101		
116 – 120	74		
Suma	614		



2.1.3. Moda

Es el valor que más se repite en una lista de datos; es decir, es el que se presenta con mayor frecuencia en un conjunto de datos. Si existen dos modas en el conjunto, se le llama bimodal; se representa mediante la variable $\widehat{x}$.

Datos no agrupados

3 4 6 3 5 5



El valor que más se repite

$$\hat{x} = 5$$

Datos agrupados

Conociendo la tabla de distribución de frecuencias:

$$\hat{X} = L + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2} \right] c$$

Donde:

$$\hat{x} = \text{moda}$$

L = límite real inferior del intervalo que contiene a la moda

Δ_1 = diferencia entre la frecuencia que contiene a la moda y el intervalo anterior

Δ_2 = diferencia entre la frecuencia que contiene a la moda y el intervalo siguiente

c = tamaño del intervalo

Núm.	Intervalos de clase	Límites reales	Frecuencia	Frecuencia acumulada
1	1 – 5	0.5 – 5.5	3	3
2	6 – 10	5.5 – 10.5	8	11
3	11 – 15	10.5 – 15.5	7	18
4	16 – 20	15.5 – 20.5	2	20
5	21 – 25	20.5 – 25.5	9	29
6	26 – 30	25.5 – 30.5	5	34
Suma			34	

- Se identifica la frecuencia más alta en la tabla; en este caso se encuentra en el intervalo número 5.
- Se calcula la diferencia de frecuencias:

$$\Delta_1 = 9 - 2 = 7$$

$$\Delta_2 = 9 - 5 = 4$$

- Se sustituyen los valores en la ecuación:

$$\hat{X} = L + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2} \right] c$$

$$\hat{x} = 20.5 + \left(\frac{7}{7 + 4} \right) 5$$

$$\hat{x} = 20.5 + \left(\frac{7}{11} \right) 5$$

$$\hat{x} = 20.5 + (0.636) 5$$

$$\hat{x} = 20.5 + 3.18$$

$$\hat{x} = 23.68$$

Una fábrica de zapatos ubicada en el municipio de San Mateo Atenco necesita identificar cuál es el número de calzado que más compran sus clientes. Para contar con el inventario necesario en el almacén y en sus tres sucursales.



Intervalos de clase (número de calzado)	Frecuencia (clientes)	Límites reales	Frecuencia acumulada
12 – 15	65		
15 – 18	78		
18 – 21	129		
21 – 24	154		
24 – 27	34		
Suma	460		

2.2. Medidas de posición

Las medidas de posición permiten identificar los valores ubicados en diferentes posiciones de un grupo de datos; las más comunes son cuartiles, deciles y percentiles.

2.2.1. Cuartiles

Es la medida que divide al conjunto de pares ordenados en dos mitades, los cuales se encuentran justo en medio de cada mitad; es decir que el conjunto de datos queda dividido en cuatro partes iguales; se representa mediante la variable Q_i .

Datos no agrupados (serie ordenada)

Ecuación para determinar la posición:

$$Q_i = 1 + \frac{i(n-1)}{4}$$

$$Q_i = \text{posición del cuartil deseado (1, 2 o 3)}$$

n = número total de datos

Ejemplo

12, 37, 41, 47, 54, 55, 65, 72, 80, 96, 98, 102, 105, 111, 119, 121, 130, 132, 141, 143, 147, 150, 151, 153 y 158.



Para determinar el valor del cuartil número 3 (Q_3), se sustituyen los datos en la ecuación para determinar la posición del cuartil Q_3 .

$$Q_3 = 1 + \frac{3(25 - 1)}{4}$$

$$Q_3 = 1 + \frac{3(24)}{4} = 1 + \frac{72}{4} = 1 + 18 = 19$$

La posición del Q_3 es 19, lo que significa que el valor del cuartil es 141.

En caso de que el resultado de la posición fuera un número decimal por ejemplo 19.7 se procede a restar los los números de la posición 20 y 19, es decir, $143 - 141 = 2$ y multiplicarlo por el decimal excedente, en este caso (.7) dando como resultado $= 1.4$ este resultado se suma al número de la posición 19 para obtener el valor del cuartil tres $141 + 1.4 = 142.4$ que corresponde a la posición 19.7.

Este procedimiento aplica también para los deciles y percentiles si el número obtenido en la posición es con decimales para datos no agrupados.

Datos agrupados

Conociendo la tabla de distribución de frecuencias

Núm.	Intervalos de clase	Límites reales	Frecuencia	Frecuencia acumulada
1	1 – 5	0.5 – 5.5	3	3
2	6 – 10	5.5 – 10.5	8	11
3	11 – 15	10.5 – 15.5	7	18
4	16 – 20	15.5 – 20.5	2	20
5	21 – 25	20.5 – 25.5	9	29
6	26 – 30	25.5 – 30.5	5	34
Suma			34	

Ecuación para calcular la posición del cuartil deseado:

$$Q_i = 1 + \frac{i(n - 1)}{4}$$

Q_i = Posición del cuartil deseado (1, 2 o 3)

n = número total de datos

Ecuación para calcular el valor del cuartil deseado:

$$Q_i = L + \left[\frac{i\left(\frac{n}{4}\right) - fa}{fx} \right] c$$

Q_i = cuartil deseado (1, 2 o 3)

n = número total de datos

fa = frecuencia acumulada del intervalo anterior al que contiene el cuartil deseado

fx = frecuencia del intervalo que contiene el cuartil deseado

c = tamaño del intervalo

L = límite real inferior del cuartil deseado

Se sustituyen los valores en las ecuaciones:

Calcular posición de Q_2

$$Q_2 = 1 + \left\lceil \frac{2(n-1)}{4} \right\rceil$$

$$Q_2 = 1 + \frac{2(34-1)}{4} = 1 + \frac{2(33)}{4} = 1 + \frac{66}{4} = 1 + 16.5 = 17.5$$

Se ubica el intervalo que contiene al cuartil deseado; en este caso el Q_2 se encuentra en el intervalo número 3.

Calcular valor de Q_2

$$Q_2 = 10.5 + \left\lceil \frac{2\left(\frac{34}{4}\right) - 11}{7} \right\rceil 5$$

$$Q_2 = 10.5 + \left\lceil \frac{2(8.5) - 11}{7} \right\rceil 5$$

$$Q_2 = 10.5 + \left\lceil \frac{17 - 11}{7} \right\rceil 5$$

$$Q_2 = 10.5 + \left\lceil \frac{6}{7} \right\rceil 5$$

$$Q_2 = 10.5 + 4.28$$

$$Q_2 = 14.785$$

Nota: El cuartil número 2 es igual a la media ($\bar{X} = Q_2$).

Calcula el valor de Q_3 de la tabla anterior.



2.2.2. Deciles

Es la medida que divide al conjunto de pares ordenados en diez partes iguales; se representa mediante la variable D_i .

Datos no agrupados

12, 37, 41, 47, 54, 55, 65, 72, 80, 96, 98, 102, 105, 111, 119, 121, 130, 132, 141, 143, 147, 150, 151, 153 y 158.



Ecuación para calcular la posición del decil deseado:

$$D_i = 1 + \frac{i(n-1)}{10}$$

D_i = posición del decil deseado (1, 2, 3, 4, 5, 6, 7, 8, 9)

n = número total de datos

Para calcular la posición de D_5 , se sustituyen los datos en la ecuación:

$$D_5 = 1 + \frac{5(n-1)}{10}$$

$$D_5 = 1 + \frac{5(25-1)}{10}$$

$$D_5 = 1 + \frac{5(24)}{10}$$

$$D_5 = 1 + \frac{120}{10}$$

$$D_5 = 1 + 12$$

$$D_5 = 13$$

La posición de D_5 es 13, lo que significa que el valor del decil es 105.

Deciles para datos no agrupados.

Cuando se obtiene un resultado con decimal.

$$D_4 = 1 + \frac{i(n-1)}{10}$$



Actividad
Autónoma

Determina el valor del cuarto decil en los siguientes 25 valores ordenados:

3,5,6,11,14,18,18,20,24,25,27,27,28,31,33,34,36,44,45,47,48,48,50,50 y 52

$$D_4 = 1 + \frac{4(25-1)}{10}$$

$$D_4 = 1 + 9.6$$

$$D_4 = 1 + \frac{96}{10}$$

$$D_4 = 10.6$$

El valor que corresponde a la posición 10.6 es entre 25 y 27, por lo tanto:

$$27 - 25 = 2$$

$$(2)(.6) = 1.2 \text{ por lo tanto } 25 + 1.2 = 26.2$$

El valor del cuarto decil es 26.2

Determina el D_7



Datos agrupados

Conociendo la tabla de distribución de frecuencias

Núm.	Intervalos de clase	Límites reales	Frecuencia	Frecuencia acumulada
1	1 – 5	0.5 – 5.5	3	3
2	6 – 10	5.5 – 10.5	8	11
3	11 – 15	10.5 – 15.5	7	18
4	16 – 20	15.5 – 20.5	2	20
5	21 – 25	20.5 – 25.5	9	29
6	26 – 30	25.5 – 30.5	5	34
Suma			34	

Ecuación para calcular la posición del decil deseado:

$$D_i = 1 + \frac{i(n-1)}{10}$$

Calcular sustituyendo los datos en la ecuación:

$$D_5 = 1 + \frac{2(34-1)}{10} = 1 + \frac{5(33)}{10} = 1 + \frac{165}{10} = 1 + 16.5 = 17.5$$

Se ubica el intervalo que contiene al decil deseado; en este caso D_5 se encuentra en el intervalo número 3.

Ecuación para calcular la posición del decil deseado:

$$D_i = L + \left[\frac{i\left(\frac{n}{10}\right) - fa}{fx} \right] c$$

D_i = decil deseado (1, 2, 3, 4, 5, 6, 7, 8, 9)

n = número total de datos

fa = frecuencia acumulada del intervalo anterior al que contiene el decil deseado

fx = frecuencia del intervalo que contiene el decil deseado

c = tamaño del intervalo

L = Límite real inferior del decil deseado





Calcular D_5 sustituyendo los datos en la ecuación:

$$D_5 = 10.5 + \left[\frac{5 \left(\frac{34}{10} \right) - 11}{7} \right] 5$$

$$D_5 = 10.5 + \left[\frac{5(3.4) - 11}{7} \right] 5$$

$$D_5 = 10.5 + \left[\frac{17 - 11}{7} \right] 5$$

$$D_5 = 10.5 + \left[\frac{6}{7} \right] 5$$

$$D_5 = 10.5 + (0.857)5$$

$$D_5 = 10.5 + 4.285$$

$$D_5 = 14.78$$

El valor de D_5 es **14.78**.

Calcula D_7

2.2.3 Percentiles

Son los 99 valores que dividen la serie de datos en 100 partes iguales; se representa mediante la variable P_i .

Datos no agrupados

12, 37, 41, 47, 54, 55, 65, 72, 80, 96, 98, 102, 105, 111, 119, 121, 130, 132, 141, 143, 147, 150, 151, 153 y 158.

Ecuación para calcular la posición del percentil deseado:

$$P_i = 1 + \frac{i(n-1)}{100}$$

P_i = posición del percentil deseado (1 – 99)

n = número total de datos

Calcular la posición de P_{50}

Sustituyendo los datos en la ecuación:

$$P_{50} = 1 + \frac{50(n-1)}{10}$$

$$P_{50} = 1 + \frac{50(n-1)}{100}$$

$$P_{50} = 1 + \frac{50(25-1)}{100} = 1 + \frac{50(24)}{100} = 1 + \frac{1200}{100} = 1 + 12 = 13$$

La posición de P_{50} es 13, lo que significa que el valor del percentil es 105.

Datos agrupados

Conociendo la tabla de distribución de frecuencias

Núm.	Intervalos de clase	Límites reales	Frecuencia	Frecuencia acumulada
1	1 – 5	0.5 – 5.5	3	3
2	6 – 10	5.5 – 10.5	8	11
3	11 – 15	10.5 – 15.5	7	18
4	16 – 20	15.5 – 20.5	2	20
5	21 – 25	20.5 – 25.5	9	29
6	26 – 30	25.5 – 30.5	5	34
Suma			34	

Ecuación para calcular la posición del percentil deseado:

$$P_i = 1 + \frac{i(n-1)}{10}$$

Calcular P_{50} sustituyendo los datos en la ecuación:

$$P_{50} = 1 + \frac{50(34-1)}{100} = 1 + \frac{50(33)}{100} = 1 + \frac{1650}{100} = 1 + 16.5 = 17.5$$

Se ubica el intervalo que contiene al percentil deseado; en este caso P_{50} se encuentra en el intervalo número 3.

Ecuación para calcular la posición del percentil deseado:

$$P_i = L + \left[\frac{i \left(\frac{n}{100} \right) - fa}{fx} \right] c$$



$P_1 = \text{percentil deseado } (1 - 99)$

$n = \text{número total de datos}$

$f_i = \text{frecuencia acumulada del intervalo anterior al que contiene el percentil deseado}$

$f_x = \text{frecuencia del intervalo que contiene el percentil deseado}$

$c = \text{tamaño del intervalo}$

$L = \text{límite real inferior del percentil deseado}$



Calcular P_{50} sustituyendo los datos en la ecuación:

$$P_{50} = 10.5 + \left[\frac{50 \left(\frac{34}{100} \right) - 11}{7} \right] 5$$

$$P_{50} = 10.5 + \left[\frac{50(0.34) - 11}{7} \right] 5$$

$$P_{50} = 10.5 + \left[\frac{17 - 11}{7} \right] 5$$

$$P_{50} = 10.5 + \left[\frac{6}{7} \right] 5$$

$$P_{50} = 10.5 + (0.857) 5$$

$$P_{50} = 10.5 + 4.285$$

$$P_{50} = 14.785$$

El valor de P_{50} es **14.785**

Calcula P_{75}

Ejercicios

1. Las calificaciones del grupo 512 del Plantel "Dr. Pablo González Casanova" en la asignatura de Estadística en el semestre agosto 201_-enero 201_ se presentan a continuación:

6.9	7.6	8.4	6.0
8.7	8.0	7.0	5.5
6.9	8.0	7.1	8.0

8.0	7.9	8.2	8.6
4.0	9.0	5.0	5.0
10.0	8.7	10.0	8.0
8.0	9.5	8.5	7.3
10.0	8.2	8.1	

Determina las medidas de tendencia central:

- ✓ Media aritmética
- ✓ Mediana
- ✓ Moda

2. Determina la moda para la siguiente tabla de frecuencias.

Núm.	Intervalos de clase	Frecuencia
1	1000 - 1900	1500
2	2000 - 2900	2300
3	3000 - 3900	1200
4	4000 - 4900	2200
5	5000 - 5900	2800
6	6000 - 6900	2900
7	7000 - 7900	3500
8	8000 - 8900	4200
9	9000 - 9900	2500

3. En la siguiente tabla se muestran los reportes de automovilistas que excedieron el límite de velocidad establecido, que es de 80 km/h.

Núm.	Intervalos de clase (velocidad)	Frecuencia (automóviles multados)
1	81 - 85	15
2	86 - 90	40
3	91 - 95	25
4	96 - 100	12
5	101 - 105	8





Módulo 2

Determina las medidas de tendencia central:

- ✓ Media aritmética
- ✓ Mediana
- ✓ Moda

Determina las medidas de posición:

- ✓ Cuartil dos (Q_2)
- ✓ Decil tres (D_3)
- ✓ Percentil setenta y cinco (P_{75})

Autoevaluación

Marca en el recuadro correspondiente la respuesta sí o no a los siguientes indicadores.

	Indicadores	Sí	No
Módulo 2			
1	Identificas la aplicación de la media aritmética y la calculas de forma correcta en los datos no agrupados y agrupados		
2	Identificas la aplicación de la mediana y la calculas de forma correcta en los datos no agrupados y agrupados		
3	Identificas la aplicación de la moda y la calculas de forma correcta en los datos no agrupados y agrupados		
4	Calculas de forma correcta los cuartiles para datos no agrupados y agrupados		
5	Calculas de forma correcta los deciles para datos no agrupados y agrupados		
6	Calculas de forma correcta los percentiles para datos no agrupados y agrupados		

Para las respuestas negativas, escribe cuáles son las actividades o acciones que debes reforzar.

1	
2	
3	
4	
5	
6	

Escribe cuál es la razón por la que se te facilitó o dificultó el proceso de aprendizaje del módulo dos.

Fácil	Difícil

Formulario del módulo dos

Datos no agrupados

Media aritmética:

$$\overline{X} = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n}$$

Mediana:

a) Si el número total de datos (n) es par, la mediana es el promedio de los dos valores centrales.

b) Si el número total de datos (n) es impar, la mediana es el valor central.

Moda:

El valor que más se repite

Datos agrupados

$$\overline{X} = \frac{\sum f_i M_i}{n}$$

$$\widetilde{X} = L + \left(\frac{\frac{n}{2} - fa}{fx} \right) c$$

$$\widehat{X} = L + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2} \right] c$$



Cuartiles:

$$Q_1 = 1 + \left\lceil \frac{i(n-1)}{4} \right\rceil$$

$$Q_1 = 1 + \left\lceil \frac{i(n-1)}{4} \right\rceil$$

$$Q_i = L + \left\lceil \frac{i\left(\frac{n}{4}\right) - fa}{fx} \right\rceil c$$

Deciles:

$$D_1 = 1 + \frac{i(n-1)}{10}$$

$$D_1 = 1 + \frac{i(n-1)}{10}$$

$$D_i = L + \left\lceil \frac{i\left(\frac{n}{10}\right) - fa}{fx} \right\rceil c$$

Percentiles:

$$P_1 = 1 + \frac{i(n-1)}{100}$$

$$P_1 = 1 + \frac{i(n-1)}{100}$$

$$P_i = L + \left\lceil \frac{i\left(\frac{n}{100}\right) - fa}{fx} \right\rceil c$$

MÓDULO TRES

Las medidas de dispersión en la estadística descriptiva y recta de regresión



Competencias disciplinares

Matemáticas básicas y extendidas

3. Explica e interpreta los resultados obtenidos mediante procedimientos matemáticos y los contrasta con modelos establecidos en situaciones reales.
4. Argumenta la solución obtenida de un problema, con métodos numéricos, gráficos, analíticos o variacionales, mediante el lenguaje verbal, matemático y el uso de las tecnologías de la información y la comunicación.
5. Analiza las relaciones entre dos o más variables de un proceso social o natural para determinar o estimar su comportamiento.
8. Interpreta tablas, gráficas, mapas, diagramas y textos con símbolos matemáticos y científicos.

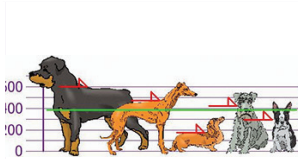
Competencias genéricas

1. Se conoce y valora a sí mismo y aborda problemas y retos teniendo en cuenta los objetivos que persigue.
 - 1.1 Enfrenta las dificultades que se le presentan y es consciente de sus valores, fortalezas y debilidades.
4. Escucha, interpreta y emite mensajes pertinentes en distintos contextos mediante la utilización de medios, códigos y herramientas apropiados.
 - 4.1 Expresa ideas y conceptos mediante representaciones lingüísticas, matemáticas o gráficas.
5. Desarrolla innovaciones y propone soluciones a problemas a partir de métodos establecidos.
 - 5.1 Sigue instrucciones y procedimientos de manera reflexiva, comprendiendo cómo cada uno de sus pasos contribuye al alcance de un objetivo.
 - 5.2 Ordena información de acuerdo a categorías, jerarquías y relaciones.
 - 5.6 Utiliza las tecnologías de la información y comunicación para procesar e interpretar información.
8. Participa y colabora de manera efectiva en equipos diversos.
 - 8.1 Propone maneras de solucionar un problema o desarrollar un proyecto en equipo, definiendo un curso de acción con pasos específicos.
 - 8.2 Aporta puntos de vista con apertura y considera los de otras personas de manera reflexiva.
 - 8.3 Asume una actitud constructiva, congruente con los conocimientos y habilidades con los que cuenta dentro de distintos equipos de trabajo.

Propósito

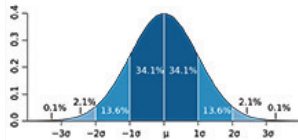
Comprende el uso y aplicación de la desviación media, varianza, desviación estándar y recta de regresión para el cálculo y apreciación de la dispersión de los datos para su análisis.

Temas módulo tres



3.1 Desviación media

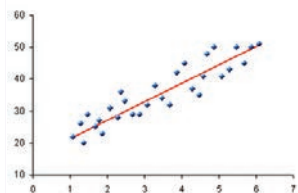
3.1.1. Interpretación de la desviación media



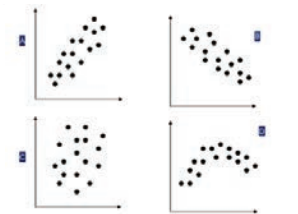
3.2. Varianza y desviación estándar

3.2.1. Interpretación de la varianza y sesgo

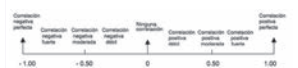
3.2.2. Interpretación de la desviación estándar



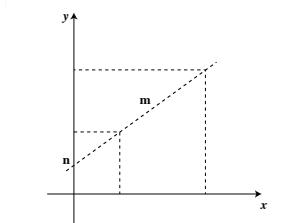
3.3. Diagrama de dispersión y centro de gravedad



3.4. Covarianza



3.5. Coeficiente de correlación



3.6. Ecuación de la recta de regresión y aplicación a series de tiempo



Introducción

Las medidas de dispersión, también llamadas medidas de variación, indican qué tan alejados o dispersos se encuentran los datos entre sí o respecto a la media. Las medidas de dispersión que analizaremos en el presente módulo son desviación media, varianza, desviación estándar; asimismo, abordaremos otros temas como diagrama de dispersión, centro de gravedad, covarianza, coeficiente de correlación y ecuación de la recta de regresión.

Evaluación diagnóstica

1. ¿Qué son las medidas de dispersión?
2. ¿Para qué se utilizan las medidas de dispersión?
3. ¿Qué es la covarianza?
4. ¿Escribe en qué se aplica la regresión lineal?

3.1. Desviación media

Es el promedio de los valores absolutos de las desviaciones de cada dato con respecto a la media; representa la lejanía de los datos con relación a la media o promedio.

Datos no agrupados

Ecuación para calcular la desviación media:

$$DM = \frac{\sum |X_i - \bar{X}|}{n}$$

Donde:

DM = desviación media

X_i = valor del i – ésimo dato

$\bar{X}$ = media del conjunto de datos

n = número total de datos

Se determina la media del conjunto de datos:

Muestra núm. 1	3	4	6	3	5	7
----------------	---	---	---	---	---	---

$$\overline{X} = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n}$$

$$\overline{X} = \frac{3 + 3 + 4 + 5 + 6 + 7}{6}$$

$$\overline{X} = \frac{28}{6}$$

$$\overline{X} = 4.667$$

Se calcula la desviación media:

$$DM = \frac{|3 - 4.66| + |3 - 4.66| + |4 - 4.66| + |5 - 4.66| + |6 - 4.66| + |7 - 4.66|}{6}$$

$$DM = \frac{|1.66| + |1.66| + |0.66| + |0.33| + |1.33| + |2.33|}{6}$$

$$DM = \frac{8}{6}$$

$$DM = 1.33$$

Datos agrupados

Intervalos de clase	Frecuencia	M_i	$f_i M_i$
1 - 5	3	3	9
6 - 10	8	8	64
11 - 15	7	13	91
16 - 20	2	18	36
21 - 25	9	23	207
26 - 30	5	28	140
Sumas	34		547

Ecuación para calcular la desviación media:

$$DM = \frac{\sum f_i |M_i - \overline{X}|}{n}$$

Donde:

DM = desviación media

f_i = valor de la frecuencia del i – ésimo dato

M_i = valor de la marca de clase del i – ésimo dato



n = número total de datos

$\bar{X}$ = media del conjunto de datos

Se calcula la $\bar{X}$ mediante la ecuación:

$$\bar{X} = \frac{\sum f_i M_i}{n} = \frac{547}{34} = 16.09$$

Se agrega la columna $|M_i - \bar{X}|$ y $f_i |M_i - \bar{X}|$ a la tabla con los respectivos valores en cada uno de los intervalos.

Ejemplo

La siguiente tabla muestra los datos correspondientes al número de libros que un grupo de estudiantes consulta de forma mensual.

Núm.	Intervalos de clase	Límites reales	Frecuencia	Frecuencia acumulada	M_i	$f_i M_i$	$ M_i - \bar{x} $	$f_i M_i - \bar{x} $
1	1 - 5	0.5 - 5.5	3	3	3	9	13.09	39.27
2	6 - 10	5.5 - 10.5	8	11	8	64	8.09	64.72
3	11 - 15	10.5 - 15.5	7	18	13	91	3.09	21.63
4	16 - 20	15.5 - 20.5	2	20	18	36	1.91	3.82
5	21 - 25	20.5 - 25.5	9	29	23	207	6.91	62.19
6	26 - 30	25.5 - 30.5	5	34	28	140	11.91	59.55
Sumas			34			547		251.18

Se sustituyen los valores en la ecuación:

$$DM = \frac{\sum f_i |M_i - \bar{X}|}{n}$$

$$DM = \frac{251.18}{34}$$

$$DM = 7.38$$

3.1.1. Interpretación de la desviación media

La desviación media indica en promedio el número de unidades en que los datos se encuentran alejados de la media.

Ejemplo

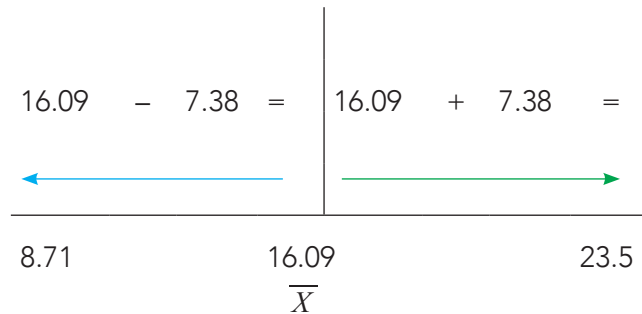
En la tabla de estudiantes que consultan mensualmente libros la biblioteca de su plantel, se observa que la media corresponde a 16.09; es decir, en promedio son 16 libros los que consultan los 34 estudiantes.

$$\bar{X} = \frac{\sum f_i M_i}{n} = \frac{547}{34} = 16.09$$

La desviación representa la distancia a la que se ubican el préstamo de los libros de forma mensual con respecto a los 16 que es la media; con base en el resultado obtenido se tiene que es 7.

$$DM = 7.38$$

La desviación media se localiza entre 8.71 y 23.5 (la lejanía de los datos con relación al promedio)

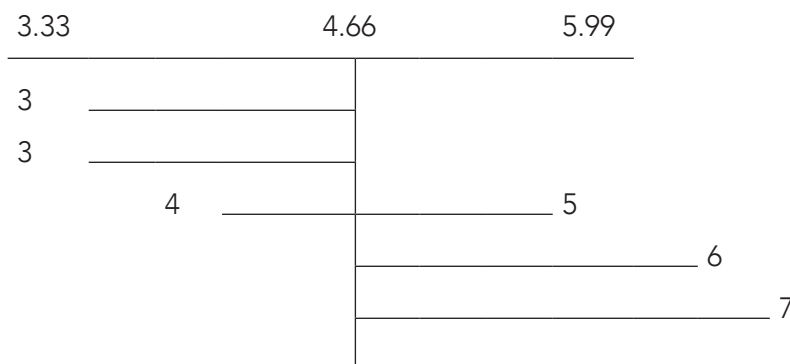


La media para la consulta de libros de forma mensual se ubica entre 8.71 y 23.5 con respecto a la media de 16.09

Datos no agrupados

Muestra núm. 1	3	3	4	5	6	7
----------------	---	---	---	---	---	---

La desviación media para el conjunto de datos no agrupados integrado por 6 números es 4.667 y su desviación media de 1.33; es decir que los datos se encuentran alejados de la media a 1.33.



La desviación media para 6 datos no agrupados se ubica entre 3.33 y 5.99 con respecto a la media de 4.66.

3.2 Varianza y desviación estándar

Varianza

Es una de las medidas de dispersión más utilizadas; indica la media de las diferencias (elevadas al cuadrado) de cada valor con respecto a la media; se representa mediante la variable σ^2 .



Datos no agrupados

Muestra núm. 1

3

3

4

5

6

7

Ecuación para calcular σ^2 :

$$\sigma^2 = \frac{\sum (X_i - \bar{X})^2}{n}$$

Donde:

$$\sigma^2 = \text{varianza}$$

X_i = valor del i - ésimo dato

$\bar{X}$ = media del conjunto de datos

n = número total de datos

Se determina el valor de la $\bar{X}$ mediante la ecuación:

$$\bar{X} = \frac{3 + 3 + 4 + 5 + 6 + 7}{6} = \frac{28}{6} = 4.66$$

Se calcula la varianza:

$$\sigma^2 = \frac{(3 - 4.66)^2 + (3 - 4.66)^2 + (4 - 4.66)^2 + (5 - 4.66)^2 + (6 - 4.66)^2 + (7 - 4.66)^2}{6}$$

$$\sigma^2 = \frac{(-1.77)^2 + (-1.77)^2 + (0.7)^2 + (0.33)^2 + (1.33)^2 + (2.33)^2}{6}$$

$$\sigma^2 = \frac{(2.78) + (0.44) + (0.11) + (1.78) + (5.44)}{6}$$

$$\sigma^2 = \frac{13.333}{6}$$

$$\sigma^2 = 2.22$$

Datos agrupados

Núm.	Intervalos de clase	Límites reales	Frecuencia	Frecuencia acumulada	Mi	$fiMi$
1	1 - 5	0.5 - 5.5	3	3	3	9
2	6 - 10	5.5 - 10.5	8	11	8	64
3	11 - 15	10.5 - 15.5	7	18	13	91
4	16 - 20	15.5 - 20.5	2	20	18	36
5	21 - 25	20.5 - 25.5	9	29	23	207
6	26 - 30	25.5 - 30.5	5	34	28	140
Sumas			34			547

Ecuación para calcular la σ^2 :

$$\sigma^2 = \frac{\sum f_i (M_i - \bar{x})^2}{n}$$

Donde:

σ^2 = varianza

$\bar{X}$ = media del conjunto de datos

n = número total de datos

f_i = frecuencia del i – ésimo intervalo

M_i = marca de clase

Se procede al cálculo de la $\bar{X}$ mediante la ecuación:

$$\bar{X} = \frac{\sum f_i M_i}{n} = \frac{547}{34} = 16.09$$

Una vez calculada la media se agregan las columnas $M_i - \bar{X}$ así como $(M_i - \bar{X})^2$ y $f_i(M_i - \bar{X})^2$ con sus respectivos valores.

Núm.	Intervalos de clase	Límites reales	Frecuencia	Frecuencia acumulada	M_i	$f_i M_i$	$ M_i - \bar{X} $	$(M_i - \bar{X})^2$	$f_i(M_i - \bar{X})^2$
1	1 - 5	0.5 - 5.5	3	3	3	9	13.09	171.35	514.04
2	6 - 10	5.5 - 10.5	8	11	8	64	8.09	65.45	523.58
3	11 - 15	10.5 - 15.5	7	18	13	91	3.09	9.55	66.84
4	16 - 20	15.5 - 20.5	2	20	18	36	1.91	3.65	7.30
5	21 - 25	20.5 - 25.5	9	29	23	207	6.91	47.75	429.73
6	26 - 30	25.5 - 30.5	5	34	28	140	11.91	141.85	709.24
Sumas			34			547		439.59	2250.74

Se sustituyen valores en la ecuación:

$$\sigma^2 = \frac{\sum f_i (M_i - \bar{x})^2}{n}$$

$$\sigma^2 = \frac{2250.74}{34}$$

$$\sigma^2 = 66.19$$

Desviación estándar

Es la raíz cuadrada de la varianza; se representa mediante la variable σ .



Datos no agrupados

Muestra núm. 1

3

3

4

5

6

7

Se determina mediante la ecuación:

$$\sigma = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n}}$$

Donde:

σ = desviación estándar

X_i = valor del i - ésimo dato

$\bar{X}$ = media del conjunto de datos

n = número total de datos

Se calcula la media

$$\bar{X} = \frac{3 + 3 + 4 + 5 + 6 + 7}{6} = \frac{28}{6} = 4.66$$

Una vez determinada la media se procede con el cálculo de la **desviación estándar**:

$$\sigma = \sqrt{\frac{(3 - 4.66)^2 + (3 - 4.66)^2 + (4 - 4.66)^2 + (5 - 4.66)^2 + (6 - 4.66)^2 + (7 - 4.66)^2}{6}}$$

$$\sigma = \sqrt{\frac{(-1.77)^2 + (-1.77)^2 + (0.7)^2 + (0.33)^2 + (1.33)^2 + (2.33)^2}{6}}$$

$$\sigma = \sqrt{\frac{(2.78) + (0.44) + (0.11) + (1.78) + (5.44)}{6}}$$

$$\sigma = \sqrt{\frac{13.333}{6}}$$

$$\sigma = \sqrt{2.22}$$

$$\sigma = 1.49$$

Datos agrupados

Núm.	Intervalos de clase	Límites reales	Frecuencia	Frecuencia acumulada	Mi	fMi	$Mi - \bar{X}$	$(Mi - \bar{X})^2$	$f_i(Mi - \bar{X})^2$
1	1 - 5	0.5 - 5.5	3	3	3	9	13.09	171.35	514.04
2	6 - 10	5.5 - 10.5	8	11	8	64	8.09	65.45	523.58
3	11 - 15	10.5 - 15.5	7	18	13	91	3.09	9.55	66.84
4	16 - 20	15.5 - 20.5	2	20	18	36	1.91	3.65	7.30

Núm.	Intervalos de clase	Límites reales	Frecuencia	Frecuencia acumulada	M_i	$f_i M_i$	$M_i - \bar{X}$	$(M_i - \bar{X})^2$	$f_i (M_i - \bar{X})^2$
5	21 - 25	20.5 - 25.5	9	29	23	207	6.91	47.75	429.73
6	26 - 30	25.5 - 30.5	5	34	28	140	11.91	141.85	709.24
Sumas			34			547		439.59	2250.74

Se determina mediante la ecuación:

$$\sigma = \sqrt{\frac{\sum f_i (M_i - \bar{X})^2}{n}}$$

Donde:

σ = desviación estándar

M_i = marca de clase del i – ésimo intervalo

$\bar{X}$ = media del conjunto de datos

n = número total de datos

f_i = frecuencia del i – ésimo intervalo

Se calcula la desviación media:

$$\sigma = \sqrt{\frac{\sum f_i (M_i - \bar{X})^2}{n}} = \frac{2250.74}{34} = 66.19$$

Se calcula la desviación estándar:

$$\sigma = \sqrt{\frac{\sum f_i (M_i - \bar{X})^2}{n}}$$

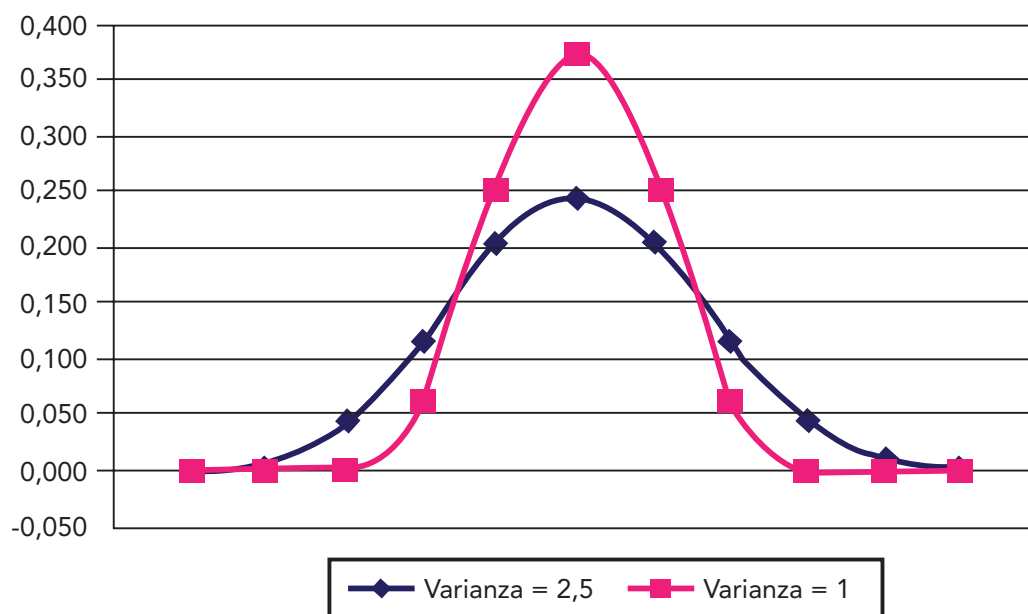
$$\sigma = \sqrt{\frac{2250.74}{34}}$$

$$\sigma = \sqrt{66.19}$$

$$\sigma = 8.13$$

3.2.1. Interpretación de la varianza y sesgo

La varianza es la medida de dispersión más importante; tiene como base el promedio aritmético de las desviaciones (distintas de un valor con respecto a la media) elevado al cuadrado.



Representa la base para determinar la desviación estándar al eliminar las cantidades elevadas al cuadrado por medio de la raíz.

Para determinar el sesgo, es necesario contar con la desviación estándar que ya fue desarrollada en el tema 3.2.

Sesgo

Representa el grado de simetría o asimetría en la distribución de frecuencias; es decir, representa la distribución de datos con respecto a la media. Las medidas más comunes para obtener el sesgo son la media, la mediana y la moda, y éste va a estar representado por curvas de sesgo positivo, negativo o cero.

En este momento vamos a calcular el primer y segundo sesgo para datos no agrupados y agrupados.

Ecuaciones:

Primer coeficiente de sesgo

$$1^{\circ} \text{ Sesgo} = \frac{\overline{X} - \widehat{X}}{\sigma}$$

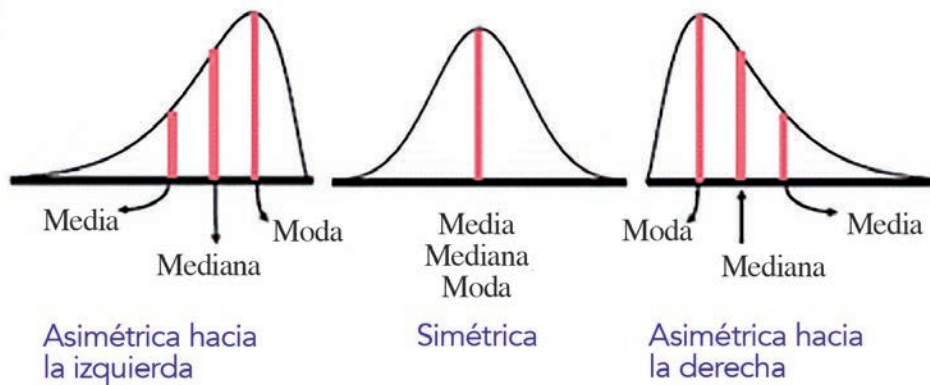
Segundo coeficiente de sesgo

$$2^{\circ} \text{ Sesgo} = \frac{3\overline{X} - 3\widetilde{X}}{\sigma}$$

Interpretación

- El sesgo es cero cuando media, mediana y moda son iguales; su curva es normal y se denomina campana de Gauss.

- El sesgo es positivo o sesgado a la derecha cuando la media es mayor a la mediana y a la moda.
- El sesgo es negativo o sesgado a la izquierda cuando la media es menor a la mediana y a la moda.



Ejemplo

La siguiente tabla muestra la cantidad de revisiones médicas de los derechohabientes entre las edades de 1 a 30 años, durante el año 2017, en la clínica del ISSEMYM del municipio de Tenancingo. Calcula el primer y segundo sesgo de Pearson.

Intervalos de clase	Frecuencia
1 – 5	3
6 – 10	8
11 – 15	7
16 – 20	2
21 – 25	9
26 – 30	5
Suma	34

Núm.	Intervalos de clase	Límites reales	Frecuencia	Frecuencia acumulada	M_i	$f_i M_i$	$ M_i - \bar{X} $	$(M_i - \bar{X})^2$	$f_i (M_i - \bar{X})^2$
1	1 - 5	0.5 - 5.5	3	3	3	9	13.09	171.35	514.04
2	6 - 10	5.5 - 10.5	8	11	8	64	8.09	65.45	523.58
3	11 - 15	10.5 - 15.5	7	18	13	91	3.09	9.55	66.84
4	16 - 20	15.5 - 20.5	2	20	18	36	1.91	3.65	7.30
5	21 - 25	20.5 - 25.5	9	29	23	207	6.91	47.75	429.73
6	26 - 30	25.5 - 30.5	5	34	28	140	11.91	141.85	709.24
Sumas			34			547		439.59	2250.74



$$\bar{X} = 16.09$$

$$\widetilde{X} = 14.78$$

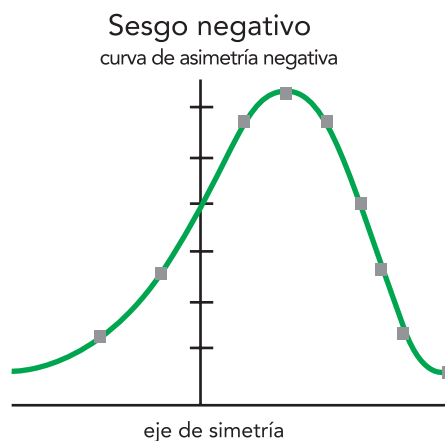
$$\widehat{X} = 23.68$$

$$\sigma = 8.13$$

Primer coeficiente de sesgo

$$1^\circ \text{ Sesgo} = \frac{\bar{X} - \widetilde{X}}{\sigma} = \frac{16.09 - 23.68}{8.13} = -.093$$

El sesgo es negativo, lo que significa que la distribución de frecuencias se encuentra sesgada a la izquierda.



Segundo coeficiente de sesgo

$$2^\circ \text{ Sesgo} = \frac{3\bar{X} - 3\widehat{X}}{\sigma} = \frac{3(16.09) - 3(23.68)}{8.13} = \frac{48.27 - 71.04}{8.13} = -0.28$$

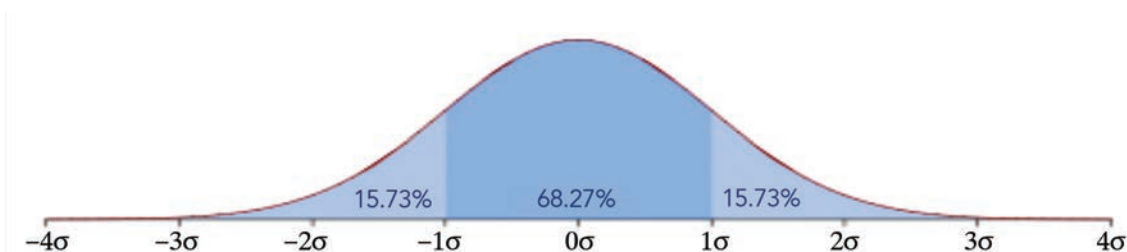
El sesgo es positivo, lo que significa que la distribución de frecuencias se encuentra sesgada a la derecha.



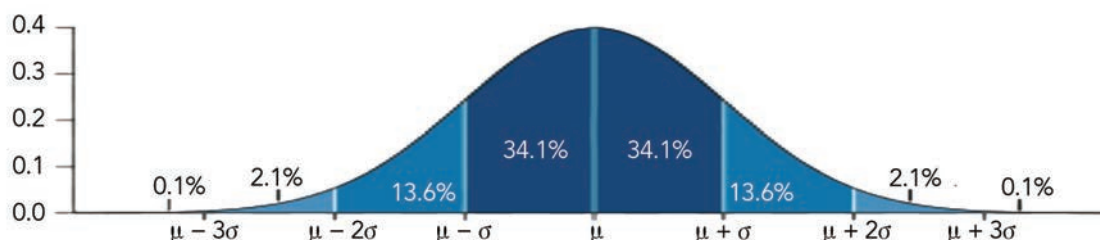
3.2.2 Interpretación de la desviación estándar

La desviación estándar es la medida de dispersión más utilizada, ya que el resultado se expresa en las mismas unidades que los datos originales.

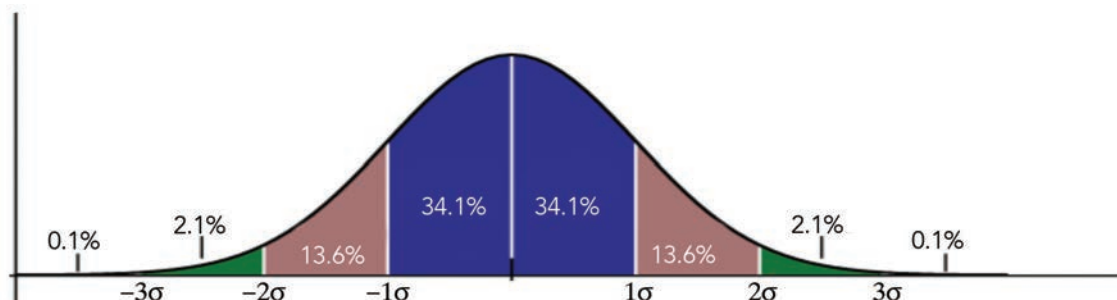
1. **Para distribuciones de frecuencia aproximadamente simétricas, el intervalo que tiene por límites $\bar{X} - \sigma$ y $\bar{X} + \sigma$ contiene aproximadamente 68.27% del valor total de los datos, o bien 0.6827 del área bajo la curva normal; esto es:**



2. Para distribuciones de frecuencia aproximadamente simétricas, el intervalo que tiene por límites $\bar{X} - 2\sigma$ y $\bar{X} + 2\sigma$ contiene aproximadamente 95.45% del valor total de los datos, o bien 0.9545 del área bajo la curva normal; esto es



3. Para distribuciones de frecuencia aproximadamente simétricas, el intervalo que tiene por límites $\bar{X} - 3\sigma$ y $\bar{X} + 3\sigma$ contiene aproximadamente 99.73% del valor total de los datos, o bien 0.9973 del área bajo la curva normal; esto es:



Con los siguientes ejemplos se va a realizar la interpretación de la desviación estándar.

La siguiente tabla muestra la cantidad de revisiones médicas de los derechohabientes entre las edades de 1 a 30 años, durante el año 2017, en la clínica del ISSEMYM del municipio de Tenancingo. Calcula la interpretación de la desviación estándar.

Núm.	Intervalos de clase	Límites reales	Frecuencia	Frecuencia acumulada	Mi	$fiMi$	$ Mi - \bar{X} $	$(Mi - \bar{X})^2$	$fi(Mi - \bar{X})^2$
1	1 - 5	0.5 - 5.5	3	3	3	9	13.09	171.35	514.04
2	6 - 10	5.5 - 10.5	8	11	8	64	8.09	65.45	523.58
3	11 - 15	10.5 - 15.5	7	18	13	91	3.09	9.55	66.84



Núm.	Intervalos de clase	Límites reales	Frecuencia	Frecuencia acumulada	M_i	$f_i M_i$	$ M_i - \bar{X} $	$(M_i - \bar{X})^2$	$f_i(M_i - \bar{X})^2$
4	16 - 20	15.5 - 20.5	2	20	18	36	1.91	3.65	7.30
5	21 - 25	20.5 - 25.5	9	29	23	207	6.91	47.75	429.73
6	26 - 30	25.5 - 30.5	5	34	28	140	11.91	141.85	709.24
Sumas			34			547		439.59	2250.74

$$\bar{X} = 16.09$$

$$\widetilde{X} = 14.78$$

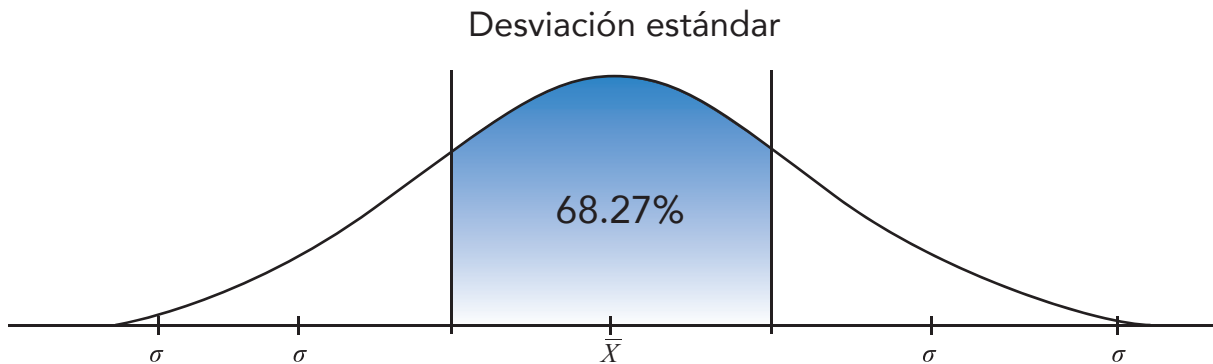
$$\widehat{X} = 23.68$$

$$\sigma = 8.13$$

Caso 1: Con área bajo la curva de 68.27%
 $\bar{X} - \sigma$ y $\bar{X} + \sigma$

$$16.09 - 8.13 = 7.96 \quad \text{y} \quad 16.09 + 8.13 = 24.42$$

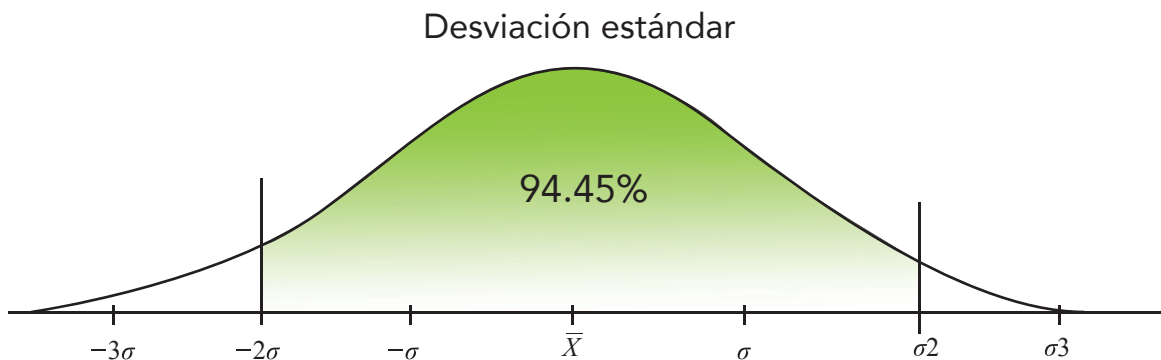
El 68.27% de los datos bajo la curva se localizan entre los años de los derechohabientes de 7.96 a 24.42.



Caso 2: Con área bajo la curva de 94.45%
 $\bar{X} - 2\sigma$ y $\bar{X} + 2\sigma$

$$16.09 - (2)8.13 = -0.17 \quad \text{y} \quad 16.09 + (2)8.13 = 32.35$$

El 94.45% de los datos bajo la curva se localizan entre los años de los derechohabientes de -0.17 a 32.35.

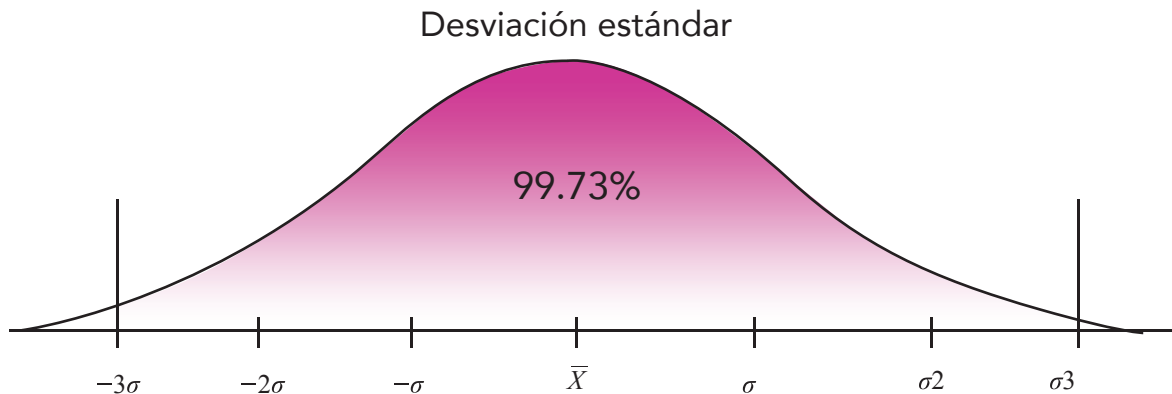


Caso 3. Con área bajo la curva de 99.73%

$X^- - 3\sigma$ y $X^+ + 3\sigma$

$$16.09 - (3)8.13 = -8.3 \quad y \quad 16.09 + (3)8.13 = 40.48$$

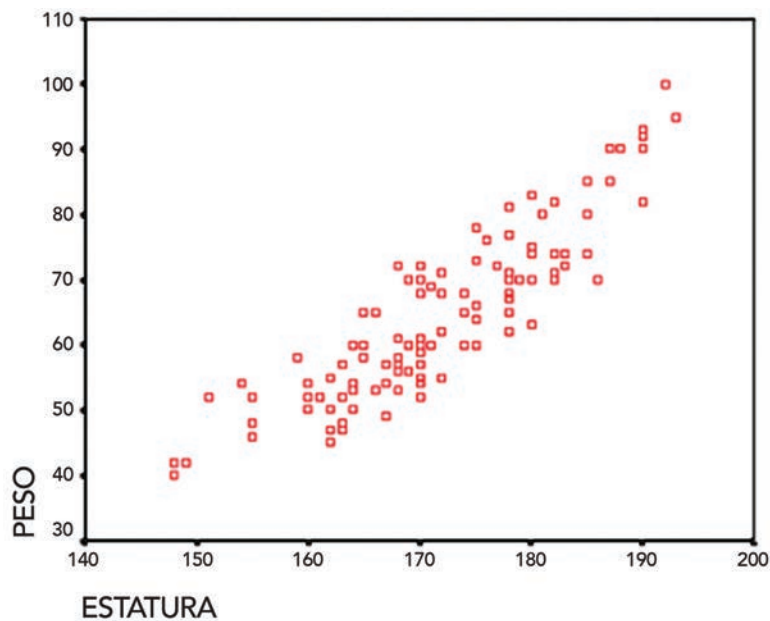
El 99.73% de los datos bajo la curva se localizan entre los años de los derechohabientes de -8.3 a 40.48.



3.3. Diagrama de dispersión y centro de gravedad

a. Diagrama de dispersión

El diagrama de dispersión, también llamado centroe, es una herramienta gráfica que ayuda a identificar la posible relación entre dos variables, lo que hace más fácil visualizar e interpretar los datos.



Ejemplo

En la siguiente tabla se muestra el peso y la estatura de 10 alumnos del sexto semestre del Plantel "Dr. Pablo González Casanova".



Estudiante	Estatura (metros)	Peso (kg)
1	1.62	55
2	1.54	52
3	1.89	68
4	1.68	75
5	1.82	82
6	1.87	89
7	1.58	45
8	1.75	86
9	1.54	58
10	1.90	90



b. Centro de gravedad

Es el punto de equilibrio de un conjunto de datos en un diagrama de dispersión, cuyas coordenadas son $(\bar{X}, \bar{Y})$.

Donde:

$\bar{x}$ = media aritmética de los primeros elementos de los pares ordenados

$\bar{y}$ = media aritmética de los segundos elementos de los pares ordenados

Determina el centro de gravedad del siguiente conjunto de datos:



Estudiante	Estatura (metros)	Peso (kg)
1	1.62	55
2	1.54	52
3	1.89	68
4	1.68	75
5	1.82	82
6	1.87	89
7	1.58	45
8	1.75	86
9	1.54	58
10	1.90	90
Sumas	17.19	700

Se determina la $\bar{X}$

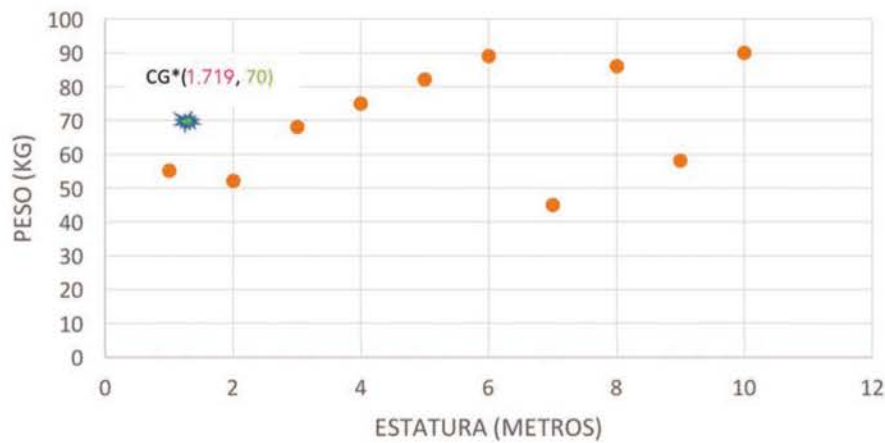
$$\bar{X} = \frac{17.19}{10} = 1.719$$

Se determina la $\bar{Y}$

$$\bar{Y} = \frac{700}{10} = 70$$

Las coordenadas del centro de gravedad son $CG^*(1.719, 70)$.

Diagrama de dispersión

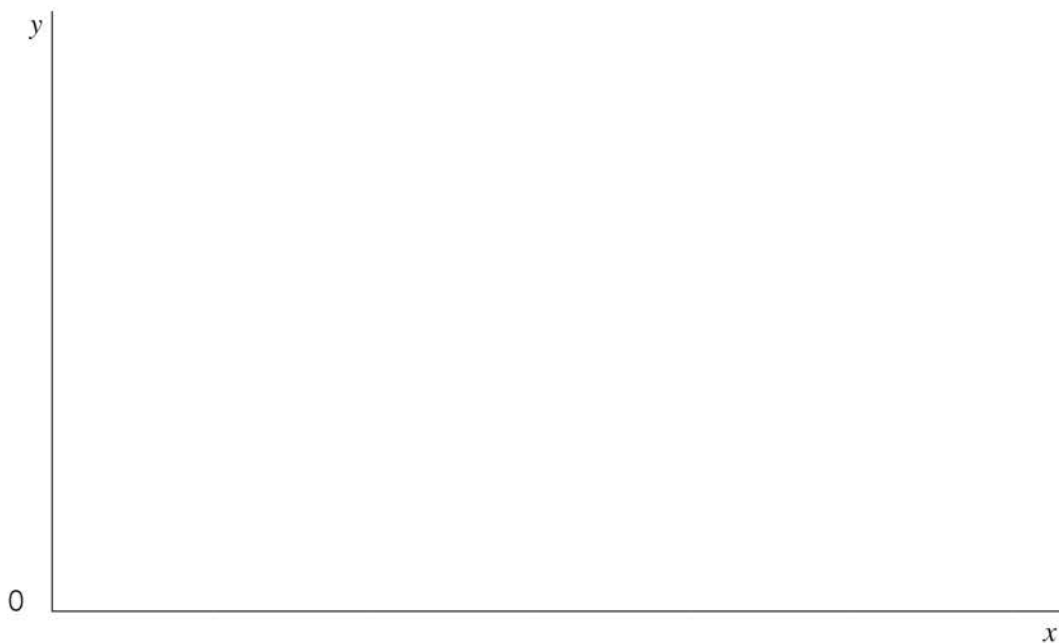


Calcula el centro de gravedad y representa el diagrama de dispersión con el centro de gravedad de los siguientes datos.

Una empresa de cosméticos realizó un análisis de la adquisición de su maquillaje estrella al año con relación a las ventas netas anuales en la zona centro de la república mexicana. Los resultados fueron los siguientes.



Venta de producto estrella	5	7	10	12	15	16	19	20	21
Ventas netas (en miles)	7	10	22	24	17	18	20	32	29





3.4 Covarianza

Es una medida descriptiva que permite el tipo de asociación lineal entre dos variables; se obtiene mediante la ecuación:

$$\text{Cov}(x,y) = \frac{1}{n} \sum (X_i - \bar{X})(Y_i - \bar{Y})$$

Donde:

$\text{Cov}(x, y) = \text{covarianza}$

$X_i = \text{valor } i - \text{ésimo de la variable } X$

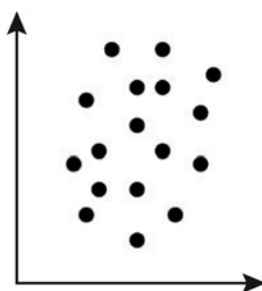
$Y_i = \text{valor } i - \text{ésimo de la variable } Y$

$\bar{Y} = \text{media de la variable } Y$

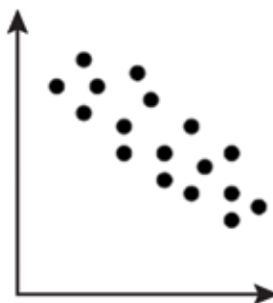
$n = \text{número de datos}$

Interpretación del valor de la covarianza:

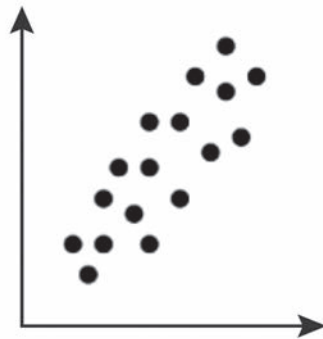
1. Si la covarianza es cero, no existe relación entre las variables.



2. Si la covarianza es negativa, hay una relación inversa entre las variables; esto indica que cuando X aumenta de valor de Y disminuye, y viceversa.



3. Si la covarianza es positiva, hay una relación directa entre las variables; esto indica que cuando el valor de X aumenta, el de Y también lo hace, y viceversa.



Retomando el ejercicio anterior, determinar la covarianza.

Estudiante	Estatura (metros)	Peso (kg)
1	1.62	55
2	1.54	52
3	1.89	68
4	1.68	75
5	1.82	82
6	1.87	89
7	1.58	45
8	1.75	86
9	1.54	58
10	1.90	90
Sumas	17.19	700

Se determina la $\bar{X}$

$$\bar{X} = \frac{17.19}{10} = 1.719$$

Se determina la $\bar{Y}$

$$\bar{Y} = \frac{700}{10} = 70$$

Se agregan las columnas $(X - \bar{X})$ y $(Y - \bar{Y})$ y se determinan sus respectivos valores.

Estudiante	Estatura (metros)	Peso (kg)	$(X_i - \bar{X})$	$(Y_i - \bar{Y})$	$(X_i - \bar{X})(Y_i - \bar{Y})$
1	1.62	55	-0.099	-15	1.485
2	1.54	52	-0.179	-18	3.222
3	1.89	68	-0.171	-2	-0.342
4	1.68	75	-0.039	5	-0.195



5	1.82	82	-0.101	12	1.212
6	1.87	89	-0.151	19	2.869
7	1.58	45	-0.139	-25	3.475
8	1.75	86	-0.031	16	0.496
9	1.54	58	-0.179	-12	2.148
10	1.90	90	-0.181	20	3.62
Sumas	17.19	700			17.99

El valor de la covarianza es:

$$Cov(X, Y) = \frac{1}{n} \sum (X_i - \bar{X})(Y_i - \bar{Y})$$

$$Cov(X, Y) = \frac{1}{10} (17.99)$$

$$Cov(X, Y) = 0.1 (17.99)$$

$$Cov(X, Y) = 1.799$$

Por lo tanto, si la covarianza es positiva, hay una relación directa entre las variables; esto indica que cuando el valor de X aumenta, el de Y también lo hace, y viceversa.



Una farmacia desea conocer la relación que existe entre la edad y el número de veces que asisten a consulta sus clientes, para lo cual realiza preguntas a 10 personas de forma aleatoria, quienes oscilan entre la edad de 40 a 80 años. Calcula la covarianza y escribe su interpretación.

Edad	Consultas
40	1
41	1
43	3
45	2
50	3
57	4
70	3
77	6
79	7
80	11

3.5 Coeficiente de correlación

El coeficiente de correlación se utiliza para medir la magnitud de la relación lineal entre dos variables; es decir, indica cuán fuerte o débil es una relación lineal; se representa mediante la variable r , también llamada r de Pearson, y se obtiene mediante la ecuación:

$$r = \frac{Cov(x, y)}{\sigma_x \sigma_y}$$

Donde

r = coeficiente de correlación

$Cov(x, y)$ = covarianza del conjunto de datos

σ_x = desviación estándar de X

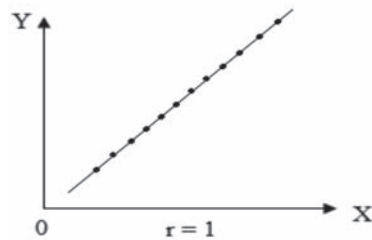
σ_y = desviación estándar de Y

El coeficiente de correlación muestra la relación lineal que existe entre las variables, así como qué tan estrecha es la dependencia o relación.

El rango de valores oscila de -1 a $+1$; los valores intermedios se interpretan intuitivamente.

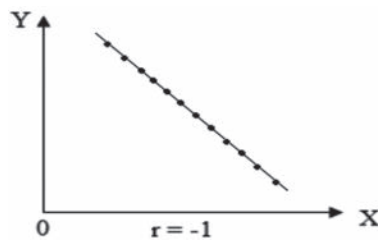
✓ Observación gráfica

✓ Correlación positiva perfecta



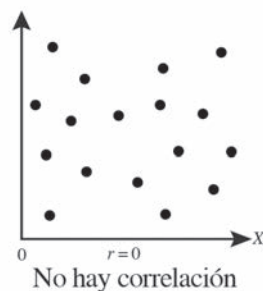
Si r es positivo, indica que la variable Y aumenta al aumentar la variable X .

✓ Correlación negativa perfecta



Si r es negativa, indica que la variable Y decrece si la variable X aumenta.

✓ Correlación cero





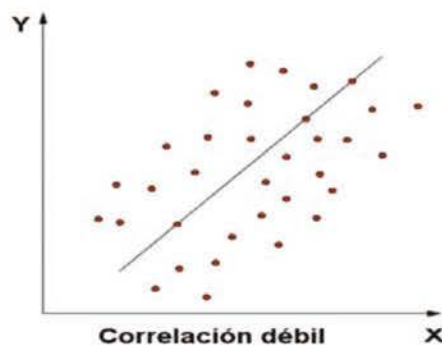
Si $r = 0$, no existe ninguna relación entre las variables.

✓ Correlación fuerte

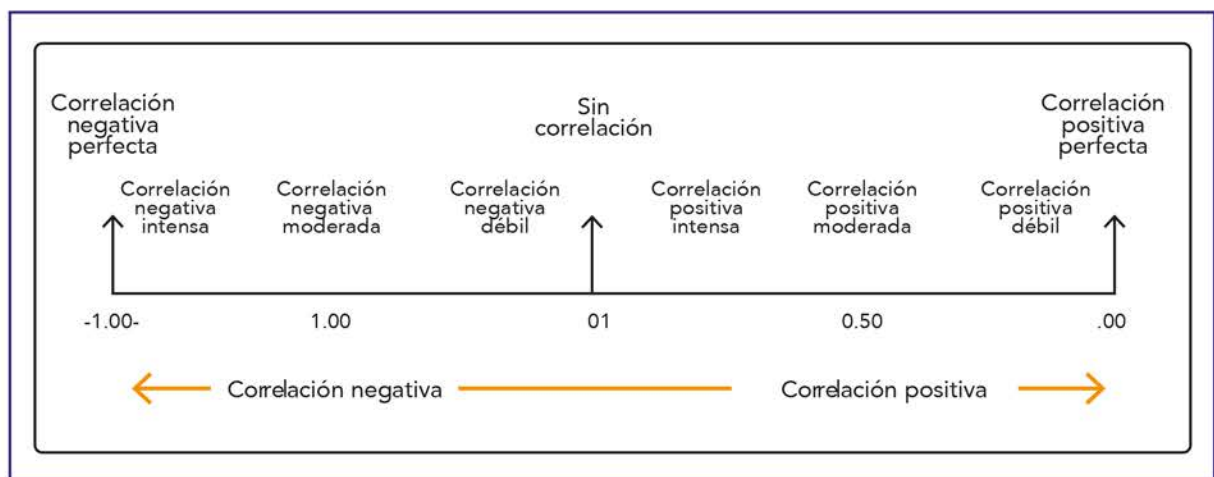


Si r está próximo a 1 o -1

✓ Correlación débil



Si r está próxima a cero



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

Ejemplo

Obtener el coeficiente de correlación para el siguiente conjunto de datos

Estudiante	Estatura (metros)	Peso (kg)
1	1.62	55
2	1.54	52
3	1.89	68
4	1.68	75
5	1.82	82
6	1.87	89
7	1.58	45
8	1.75	86
9	1.54	58
10	1.90	90
Sumas	17.19	700

Ecuación del coeficiente de correlación:

$$r = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

Se obtiene el valor de la media aritmética para las variables X y Y .

Se determina la $\bar{X}$

$$\bar{X} = \frac{17.19}{10} = 1.719$$

Se determina la $\bar{Y}$

$$\bar{Y} = \frac{700}{10} = 70$$

Se agregan columnas a la tabla para el cálculo de la varianza y de la desviación estándar de X y de Y .

Estudiante	Estatura (metros)	Peso (kg)	$(X_i - \bar{X})$	$(Y_i - \bar{Y})$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$
1	1.62	55	-0.099	-15	1.485	0.009801	225
2	1.54	52	-0.179	-18	3.222	0.032041	324
3	1.89	68	-0.171	-2	-0.342	0.029241	4
4	1.68	75	-0.039	5	-0.195	0.001521	25
5	1.82	82	-0.101	12	1.212	0.010201	144
6	1.87	89	-0.151	19	2.869	0.022801	361
7	1.58	45	-0.139	-25	3.475	0.019231	625
8	1.75	86	-0.031	16	0.496	0.000961	256
9	1.54	58	-0.179	-12	2.148	0.032041	144
10	1.90	90	-0.181	20	3.62	0.032761	400
Sumas	17.19	700			17.99	0.19069	2508



El valor de la covarianza es:

$$Cov(X, Y) = \frac{1}{n} \Sigma (X_i - \bar{X})(Y_i - \bar{Y})$$

$$Cov(X, Y) = \frac{1}{10} (17.99)$$

$$Cov(X, Y) = 0.1 (17.99)$$

$$Cov(X, Y) = 1.799$$

El valor de la desviación estándar de X es:

$$\sigma_X = \sqrt{\frac{\Sigma (X_i - \bar{X})^2}{n}}$$

$$\sigma_X = \sqrt{\frac{0.19169}{10}}$$

$$\sigma_X = \sqrt{0.019069}$$

$$\sigma_X = 0.1381$$

El valor de la desviación estándar de Y es:

$$\sigma_Y = \sqrt{\frac{\Sigma (Y_i - \bar{Y})^2}{n}}$$

$$\sigma_Y = \sqrt{\frac{2508}{10}}$$

$$\sigma_Y = \sqrt{250.8}$$

$$\sigma_Y = 15.83$$

Se sustituye en ecuación:

$$r = \frac{Cov(x, y)}{\sigma_X \sigma_Y}$$

$$r = \frac{17.99}{(0.1381)(15.83)}$$

$$r = \frac{1.799}{2.1861}$$

$$r = 0.8229$$

Interpretación: como el coeficiente de correlación es muy próximo a 1, existe una marcada dependencia lineal entre las variables.

Calcula el coeficiente de correlación de la siguiente información y realiza su interpretación.

En la carretera México-Toluca se realizó un estudio de la cantidad de accidentes en los cuales se encuentran implicadas personas de entre 20 y 25 años de edad en un año. Los resultados son los siguientes.

Accidentes	Frecuencias
0	129
1	156
2	141
3	75
4	24
5	30
6	21

3.6 Ecuación de la recta de regresión y aplicación a series de tiempo

Ecuación de la recta de regresión

La ecuación de la recta de regresión permite pronosticar la puntuación que alcanzará cada supuesto en una variable Y , conociendo su puntuación en la variable X .

La ecuación de una recta se determina de acuerdo con las condiciones que cumple el conjunto de puntos que la forman.

Conociendo la pendiente (ángulo de inclinación) de una recta, así como la ordenada al origen, se dice que la recta es de la forma "pendiente ordenada al origen".

Ecuación de la recta:

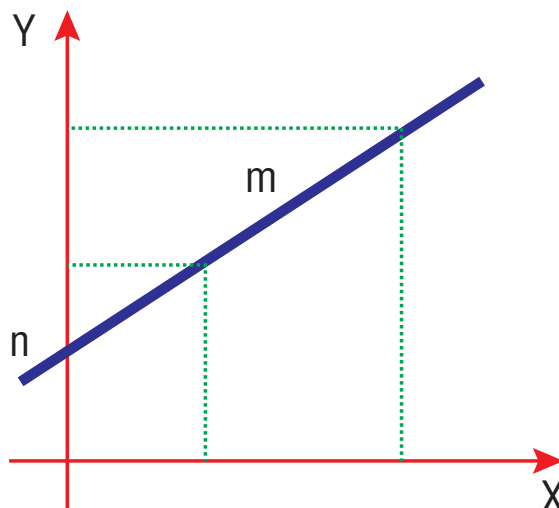
$$Y = mx + b$$

Donde:

$m =$ pendiente de la recta

$b =$ ordenada del origen

Su representación gráfica es:





La pendiente de una recta se determina como la tangente de su ángulo de inclinación y representa la inclinación de la recta con respecto al eje X .

La ecuación de la pendiente es:

$$m = \frac{\text{cov}(x, y)}{\text{var}(x)}$$

Ejemplo

Traza la gráfica de la recta indicando la pendiente. Considera la ordenada al origen con un valor de tres ($b = 3$)

Pendiente: $m = 1$

Ecuación $x + 2$

$$y = mx + b$$

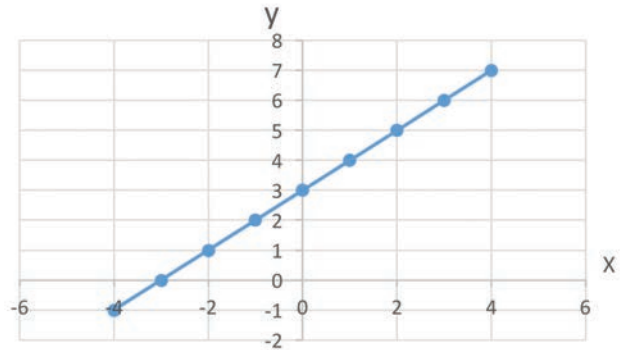
Se sustituye el valor de la pendiente y de la ordenada en la ecuación anterior:

$$y = (1)(x) + 3$$

$$y = X + 3$$

Se asignan valores $[4, -4]$

x	$y = x + 3$	Coordenadas (X, Y)
-4	$(-4 + 3) = -1$	$(-4, -1)$
-3	$(-3 + 3) = 0$	$(-3, 0)$
-2	$(-2 + 3) = 1$	$(-2, 1)$
-1	$(-1 + 3) = 2$	$(-1, 2)$
0	$(-0 + 3) = 3$	$(0, 3)$
1	$(1 + 3) = 4$	$(1, 4)$
2	$(2 + 3) = 5$	$(2, 5)$
3	$(3 + 3) = 6$	$(3, 6)$
4	$(4 + 3) = 7$	$(4, 7)$



Traza la gráfica de la siguiente recta indicando su pendiente y su ordenada al origen:

$$y = x - 2$$

Series de tiempo

En estadística se le llama así a un conjunto de valores observados durante una serie de periodos temporales (semanales, mensuales, trimestrales o anuales) secuencialmente

ordenada. Se representa a través de una gráfica de líneas sobre cuyo eje X se registran los periodos, y en cuyo eje Y los valores de la serie de tiempo.

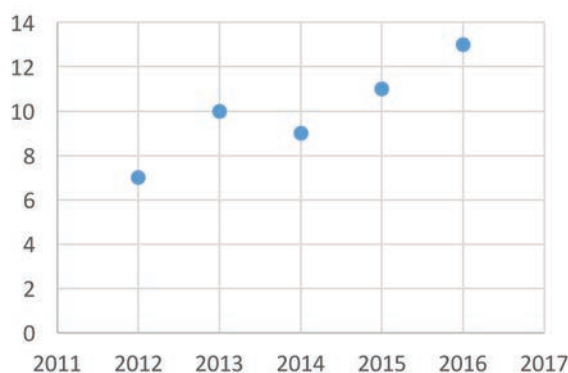
El análisis de series de tiempos es el procedimiento por el cual se identifican y aíslan los factores relacionados con el tiempo, que influyen en los valores observados en las series de tiempo, para que una vez identificados puedan contribuir a la interpretación de valores históricos y así predecir valores futuros.

Algunos ejemplos de series de tiempo son el registro periódico de las ventas anuales de una empresa, el índice de precios al consumidor (INPC), la calificación otorgada a los profesores de la UAEM en la apreciación estudiantil, entre otros.

Ejemplo

En la siguiente tabla se muestran las ventas de los últimos 5 años de una empresa del ramo alimentaria llamada La Garza S. A. de C. V.; analiza el proceso para obtener la recta de ajuste, la gráfica y las ventas que se esperan para el año 2017 y 2018.

Año	Ventas (millones de \$)
2012	7
2013	10
2014	9
2015	11
2016	13
Σ 10070	50



Se determinan las $\bar{X}$ y $\bar{Y}$.

$$\bar{X} = \frac{10070}{5} = 2014$$

$$\bar{Y} = \frac{50}{5} = 10$$

Determinar la $Cov(x, y)$ y la $var(x)$, construir tabla

Año	Ventas (millones de \$)	$(X_i - \bar{X})$	$(Y_i - \bar{Y})$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$
2012	7	-2	-3	6	4
2013	10	-1	0	0	1
2014	9	0	-1	0	0
2015	11	1	1	1	1
2016	13	2	3	6	4
10070	50			13	10



Ecuación para determinar la varianza de x

$$var(x) = \frac{\sum (x_i - \bar{x})^2}{n}$$

$$var(x) = \frac{(10)}{5}$$

$$var(x) = 2$$

Ecuación para el cálculo de la $Cov(x, y)$

$$Cov(X, Y) = \frac{1}{n} \sum (X_i - \bar{X})(Y_i - \bar{Y})$$

$$Cov(X, Y) = \frac{1}{5}(13)$$

$$Cov(X, Y) = (0.2)(13)$$

$$Cov(X, Y) = 2.6$$

Se obtiene el valor de la pendiente mediante la ecuación.

$$m = \frac{cov(x, y)}{var(x)}$$

$$m = \frac{2.6}{2}$$

$$m = 1.3$$

El valor de b es:

$$b = \bar{y} - m\bar{x}$$

$$b = 10 - (1.3)(2014)$$

$$b = 10 - (2618.20)$$

$$b = -2608.20$$

Se sustituyen los valores en la ecuación de la recta

$$Y = mx + b$$

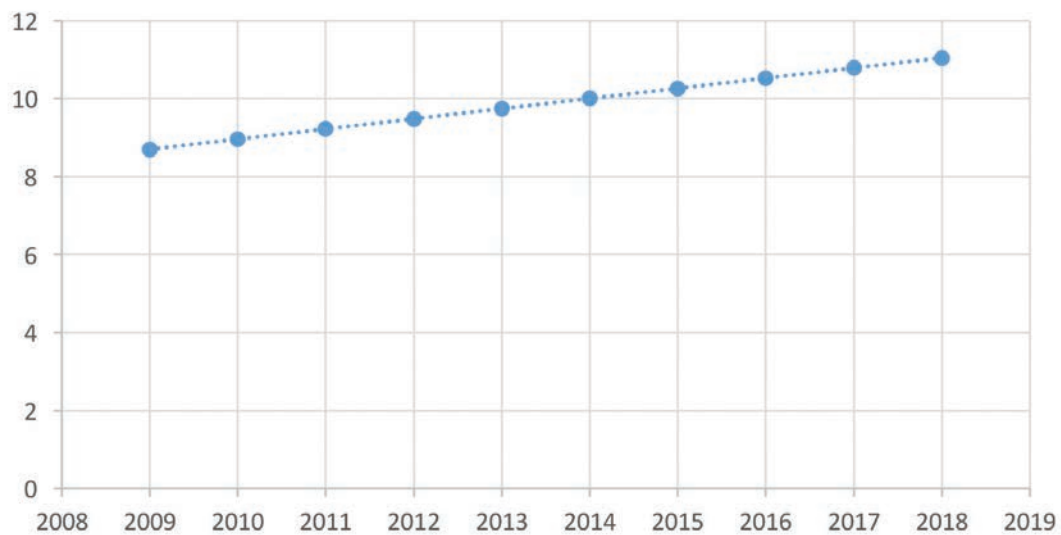
$$Y = 1.3x + (-2608.20)$$

$$Y = 1.3x - 2608.20$$

Por lo tanto, la gráfica de la recta de ajuste es:

x	$Y = 1.3X - 2608.20$	Coordenadas (X, Y)
2009	$(1.3)(2009) - 2608.20 = 3.5$	(2009, 3.5)
2010	$(1.3)(2010) - 2608.20 = 4.8$	(2010, 4.8)
2011	$(1.3)(2011) - 2608.20 = 6.1$	(2011, 6.1)

x	$Y = 1.3X - 2608.20$	Coordenadas (X, Y)
2012	$(1.3)(2012) - 2608.20 = 7.4$	(2012, 7.4)
2013	$(1.3)(2013) - 2608.20 = 8.7$	(2013, 8.7)
2014	$(1.3)(2014) - 2608.20 = 10$	(2014, 10)
2015	$(1.3)(2015) - 2608.20 = 11.3$	(2015, 11.3)
2016	$(1.3)(2016) - 2608.20 = 12.6$	(2016, 12.6)
2017	$(1.3)(2017) - 2608.20 = 13.9$	(2017, 13.9)
2018	$(1.3)(2018) - 2608.20 = 15.2$	(2018, 15.2)



✓ Las ventas esperadas para el año 2017

$$Y = 1.3x - 2608.20$$

$$Y = 1.3(2017) - 2608.20$$

$$Y = 2622.10 - 2608.20$$

$$Y = 13.9$$

✓ Las ventas esperadas para el año 2018

$$Y = 1.3x - 2608.20$$

$$Y = 1.3(2018) - 2608.20$$

$$Y = 2623.4 - 2608.20$$

$$Y = 15.2$$

Se espera que para el año 2017 las ventas proyectadas sean de 13.9 y para el 2018 de 15.2 (cifras en millones de pesos).



Ejercicios

1. Busca en la sopa de letras la respuesta que completa las definiciones que se presentan a continuación.

- Las medidas de _____ o _____ indican qué tan alejados o dispersos se encuentran los datos entre sí o con respecto a la media.
- El _____ es una medida de dispersión y considera los valores extremos de un conjunto de datos.
- _____ es el promedio de los valores absolutos de las desviaciones de cada dato con respecto a la _____.
- La _____ es el promedio de los cuadrados de las desviaciones de los datos con respecto a la media y a la desviación.
- _____ es la raíz cuadrada de la varianza.
- El _____ es el grado de asimetría de una distribución de frecuencias.

E	R	T	G	V	S	O	Z	X	C	V	B	V	R	T
G	F	U	I	O	G	Q	W	E	R	T	A	T	Y	U
D	R	Y	U	S	I	O	P	O	Y	R	Y	M	Q	M
A	E	S	E	D	F	G	H	J	I	K	L	Ñ	Q	E
W	E	S	R	T	Y	U	I	A	O	P	Q	A	S	D
Z	X	C	V	B	N	J	N	K	L	I	O	P	W	I
M	N	B	V	I	F	Z	R	T	Y	U	I	O	P	A
A	R	Y	U	Q	A	Q	W	E	R	T	Y	U	I	O
S	A	S	D	F	G	C	R	A	D	N	A	T	S	E
D	A	S	F	G	V	B	I	T	Y	F	G	H	I	Ñ
F	N	Q	R	E	T	Y	I	O	P	O	I	U	Y	O
G	H	O	P	I	T	R	E	T	N	D	G	Y	I	P
H	E	T	I	Q	W	E	R	T	Y	M	U	I	O	U
J	Y	E	R	S	R	Y	U	I	O	P	E	I	P	Y
K	O	G	N	A	R	P	U	Y	T	R	E	D	8	T
L	A	E	I	O	U	E	A	E	I	O	U	Ñ	I	O
Ñ	P	P	O	I	U	Y	P	R	E	W	Q	O	P	A
O	P	P	O	I	U	Y	P	S	M	O	P	P	O	O
P	P	P	O	I	U	Y	P	I	I	O	P	U	I	U
U	O	E	E	R	T	G	V	S	I	D	Q	Y	Y	Y
Y	B	M	E	R	T	G	V	S	I	R	W	T	T	T
T	X	V	V	A	R	I	A	C	I	O	N	N	R	R

2. Una empresa que vende productos por catálogo realizó un estudio a fin de determinar la relación que hay entre el número de agentes de ventas y las ventas netas en cierta localidad. El registro de las ventas es en miles de pesos y se obtuvieron los siguientes resultados.

Realiza un diagrama de dispersión y localiza el centro de gravedad.

Agentes X	Ventas Y
10	20
11	25
12	32
13	30
14	33
15	34
16	34
17	31
18	28
19	24
20	20

3. Determina la covarianza del siguiente conjunto de datos.

X	1600	1200	1100	1000	9000	7000	6000	5000	4000	2000
Y	5000	8000	6000	3500	8600	5000	4000	10000	5650	1800

4. Determina el coeficiente de correlación del siguiente conjunto de datos.

X	Y	$(X_i - \bar{X})$	$(Y_i - \bar{Y})$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$
800	1300					
10000	1700					
12500	1200					
13000	1800					
14000	1700					
15000	2100					



X	Y	$(X_i - \bar{X})$	$(Y_i - \bar{Y})$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$
17500	1900					
20000	1100					
Σ						

5. En la siguiente tabla se muestra la cantidad de playeras elaboradas durante el periodo 2010-2016 en una empresa textil.

Analiza el proceso para obtener la ecuación de la recta de ajuste, la gráfica y el número de productos que se espera que la empresa produzca para los años 2017, 2018, 2019.

Año	Producción
2010	1 000
2011	1 250
2012	1 280
2013	1 350
2014	1 400
2015	1 500
2016	1 946

6. Traza la gráfica de las siguientes rectas indicando su pendiente y su ordenada al origen:

$$Y = x - 2$$

$$Y = x + 2$$

7. Determinar desviación media, varianza y desviación estándar del siguiente ejercicio:

Intervalo	Frecuencia
15 - 25	8
26 - 36	6
37 - 47	15
48 - 58	13
59 - 69	2

Autoevaluación



Módulo 3

Marca en el recuadro correspondiente la respuesta sí o no a los siguientes indicadores.

Núm.	Indicadores	Sí	No
1	Calculas de forma correcta la desviación media para datos no agrupados y agrupados		
2	Identificas la diferencia entre la varianza y la desviación estándar		
3	Calculas de forma correcta la varianza y la desviación estándar		
4	Reconoces la relación entre el diagrama de dispersión y centro de gravedad		
5	Sabes determinar la covarianza		
6	Realizas de forma correcta los ejercicios de coeficiente de correlación		
7	Sabes aplicar la ecuación de la recta de regresión a series de tiempo		

Para las respuestas negativas, escribe cuáles son las actividades o acciones que debes reforzar.

1	
2	
3	
4	
5	
6	



M
3



Módulo 3

Escribe cuál es la razón por la que se te facilitó o dificultó el proceso de aprendizaje del módulo tres.

Fácil	Difícil

Formulario del módulo tres

Datos no agrupados

Desviación media:

$$DM = \frac{\sum |X_i - \bar{X}|}{n}$$

Varianza:

$$\sigma^2 = \frac{\sum (X_i - \bar{X})^2}{n}$$

Desviación estándar:

$$\sigma = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n}}$$

Datos agrupados

$$DM = \frac{\sum f_i |M_i - \bar{X}|}{n}$$

$$\sigma^2 = \frac{\sum f_i (M_i - \bar{X})^2}{n}$$

$$\sigma = \sqrt{\frac{\sum f_i (M_i - \bar{X})^2}{n}}$$

Sesgo:

Primer coeficiente de sesgo:

$$1^\circ \text{ sesgo} = \frac{\bar{X} - \widehat{X}}{\sigma}$$

Segundo coeficiente de sesgo:

$$2^\circ \text{ sesgo} = \frac{3\bar{X} - 3\widetilde{X}}{\sigma}$$

**Centro de gravedad:**

$\bar{x}$ = media aritmética de los primeros elementos de los pares ordenados

$\bar{y}$ = media aritmética de los segundos elementos de los pares ordenados

Covarianza:

$$\text{Cov}(X, Y) = \frac{1}{n} \sum (X_i - \bar{X})(Y_i - \bar{Y})$$

Coeficiente de correlación:

$$r = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y}$$

Ecuación de la recta de regresión:

$$y = mx + b$$

Series de tiempo:

$$\sigma^2 = \frac{\sum (X_i - \bar{X})^2}{n}$$

$$\text{Cov}(X, Y) = \frac{1}{n} \sum (X_i - \bar{X})(Y_i - \bar{Y})$$

$$m = \frac{\text{cov}(x, y)}{\text{var}(x)}$$

The background is a solid dark blue. It features several faint, semi-transparent illustrations: four playing cards (Ace of Hearts, Ace of Spades, Ace of Diamonds, and Ace of Clubs) are scattered across the upper half. Interspersed among the cards are several interlocking mechanical gears of different sizes. In the lower right corner, there are three dice, some showing different faces with black pips.

MÓ DU LO CUATRO

Introducción a la probabilidad



Competencias disciplinares

Matemáticas básicas y extendidas

3. Explica e interpreta los resultados obtenidos mediante procedimientos matemáticos y los contrasta con modelos establecidos en situaciones reales.
4. Argumenta la solución obtenida de un problema, con métodos numéricos, gráficos, analíticos o variacionales, mediante el lenguaje verbal, matemático y el uso de las tecnologías de la información y la comunicación.
5. Analiza las relaciones entre dos o más variables de un proceso social o natural para determinar o estimar su comportamiento.
8. Interpreta tablas, gráficas, mapas, diagramas y textos con símbolos matemáticos y científicos.

Competencias genéricas

1. Se conoce y valora a sí mismo y aborda problemas y retos teniendo en cuenta los objetivos que persigue.
 - 1.1 Enfrenta las dificultades que se le presentan y es consciente de sus valores, fortalezas y debilidades.
4. Escucha, interpreta y emite mensajes pertinentes en distintos contextos mediante la utilización de medios, códigos y herramientas apropiados.
 - 4.1 Expresa ideas y conceptos mediante representaciones lingüísticas, matemáticas o gráficas.
5. Desarrolla innovaciones y propone soluciones a problemas a partir de métodos establecidos.
 - 5.1 Sigue instrucciones y procedimientos de manera reflexiva, comprendiendo cómo cada uno de sus pasos contribuye al alcance de un objetivo.
 - 5.2 Ordena información de acuerdo con categorías, jerarquías y relaciones.
 - 5.6 Utiliza las tecnologías de la información y comunicación para procesar e interpretar información.
8. Participa y colabora de manera efectiva en equipos diversos.
 - 8.1 Propone maneras de solucionar un problema o desarrollar un proyecto en equipo, definiendo un curso de acción con pasos específicos.
 - 8.2 Aporta puntos de vista con apertura y considera los de otras personas de manera reflexiva.
 - 8.3 Asume una actitud constructiva, congruente con los conocimientos y habilidades con los que cuenta dentro de distintos equipos de trabajo

Propósito

Reconoce los elementos de probabilidad básica y los aplica en eventos de la vida diaria.

Temas del módulo cuatro



4.1. El papel de la probabilidad

4.1.1 Evento y espacio muestral



4.2. Problemas de conteo

4.2.1 Factorial, permutaciones y combinaciones

4.2.1.1 Diagrama de árbol



4.3. Clasificación de la probabilidad

4.3.1 Eventos mutuamente excluyentes o disjuntos

4.3.2 Eventos independientes

4.3.3 Eventos dependientes

4.3.4 Eventos no excluyentes entre sí

Introducción

La probabilidad es una herramienta necesaria en la estadística para evaluar la confiabilidad de los sucesos en un experimento. Actualmente, observamos su uso en la política, economía, medicina, biología, finanzas, entre otras, se ha convertido en una herramienta necesaria para la toma de decisiones basadas en predicciones o análisis de viabilidad.

Es este módulo cuatro se revisan los conceptos básicos de probabilidad y sus conceptos y aplicaciones, como espacios muestrales, combinaciones, permutaciones, diagramas de árbol, además de eventos dependientes e independientes, entre otros.

Evaluación diagnóstica

1. ¿Qué es la probabilidad?



2. ¿En dónde has visto la aplicación de la probabilidad?

3. ¿Qué es un espacio muestral?

4.1 El papel de la probabilidad

La probabilidad es la base de la estadística inferencial y apoya la interpretación de los resultados obtenidos en un suceso; es decir, analiza la posibilidad de que un hecho se realice; además, se encarga de medir la frecuencia de los eventos presentes en los experimentos, denominados resultados posibles o probabilidades de ocurrencia, según sus condiciones.

Supongamos que se organizan varias rifas para recaudar fondos para construir una nueva cancha de fútbol y se observa que un estudiante gana dos veces el primer premio; la primera impresión es que hizo trampa; sin embargo, la posibilidad es muy baja en relación con improbabilidad de que suceda este hecho. La justificación es lo que se pretende explicar en este módulo: el uso y aplicación de la probabilidad en diversas actividades.

4.1.1 Evento y espacio muestral

Un **evento** es el conjunto de todos los resultados o sucesos posibles de una actividad y representa las observaciones realizadas a datos estadísticos; un **evento** o suceso **simple** es aquel que no puede desglosarse en un evento menor. Los eventos se clasifican en controlados o no controlados.

Los eventos no controlados se localizan en la naturaleza, como los huracanes, que son observados y analizados para predecir su trayectoria y velocidad. Por otra parte, un evento controlado es aquel en el que puede predecirse los resultados.

Un espacio muestral, es el conjunto de todos los resultados posibles de un evento.

Ejemplo

Si se lanza un dado una sola vez, se sabe que se existen seis posibilidades de que **caiga** un número del 1 al 6; éstos representan los eventos o el conjunto de resultados posibles. Las caras posibles al lanzar un dado son 6 (1, 2, 3, 4, 5 o 6); esto representa un espacio muestral.

Si al lanzar el dado aparece el número 3, se trata de un evento simple, debido a que no existe en esa cara del dado otra opción como resultado.

Espacio muestral:

1, 2, 3, 4, 5, 6



De forma cotidiana observamos diversos eventos, como el lanzamiento de una moneda, con la opción de que caiga cara o cruz, que se considera un evento simple y se representa de la siguiente forma E_x :

E_1 : cara

E_2 : cruz

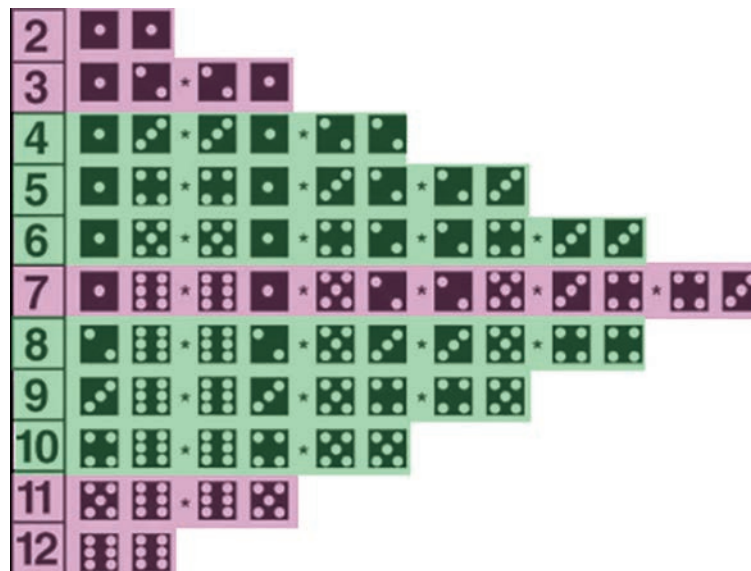
A continuación se muestran algunos eventos:

1. Calificación de los estudiantes en el primer parcial
2. Cantidad de gas que gastan los inquilinos por día
3. Opinión del uso de áreas verdes en la colonia donde vives
4. Revisión de artículos defectuosos en las empresas
5. Cantidad de calorías consumidas

Escribe tres eventos en los que se aplica la probabilidad en tu entorno.

- 1.
- 2.
- 3.

Un espacio muestral representa el conjunto de posibles resultados de un experimento o evento.



Espacio muestral
Fuente: www.eyeintheskygroup.com



Se lanzan dos dados y se solicita sumar los puntos obtenidos en cada lanzamiento; el espacio muestral es 2, 3, 4, 5... 12; es decir, se tienen 11 eventos en el espacio muestral.

Ejemplo

Se tiran dos dados, uno verde y uno rojo, y se desea conocer la probabilidad de que el verde caiga en 3 y el rojo en 4, el espacio muestral es el siguiente:

1, 6	1, 5	1, 4	1, 3	1, 2	1, 1
2, 6	2, 5	2, 4	2, 3	2, 2	2, 1
3, 6	3, 5	3, 4	3, 3	3, 2	3, 1
4, 6	4, 5	4, 4	4, 3	4, 2	4, 1
5, 6	5, 5	5, 4	5, 3	5, 2	5, 1
6, 6	6, 5	6, 4	6, 3	6, 2	6, 1

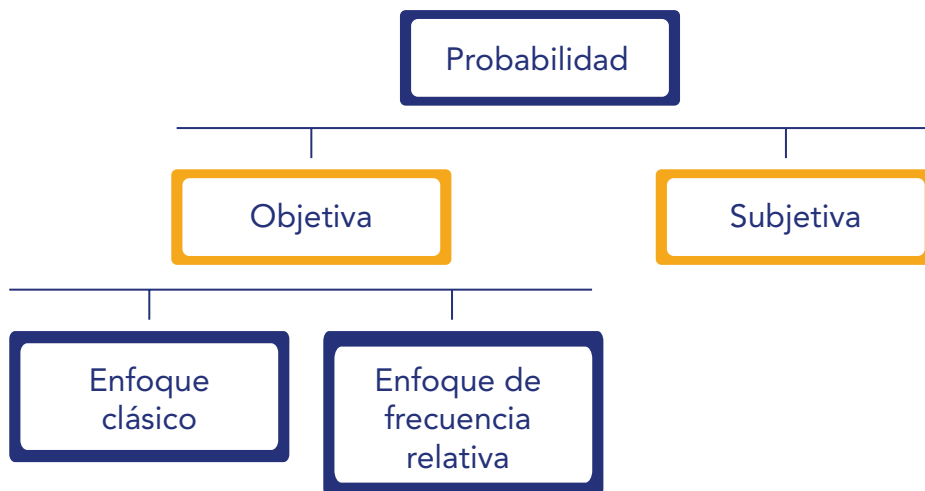
Contienen 36 posibilidades el espacio muestral y sólo una de ellas representa la probabilidad de que caiga en verde 3 y en rojo 4.



Escribe el ejemplo de un espacio muestral.

Clasificación de la probabilidad

Para aplicar la probabilidad a los diversos eventos cotidianos, es importante identificar los conceptos, usos y aplicaciones de la probabilidad, que se clasifica en objetiva y subjetiva (Sánchez, 2004).



La **probabilidad subjetiva** va relacionada a la experiencia de las observaciones realizadas; corresponde a las predicciones de un evento, con base en los eventos relevantes que se han venido observando. Un ejemplo de la probabilidad subjetiva son las predicciones meteorológicas, que estiman la probabilidad de lluvia, la cantidad de días soleados, entre otros.

Escribe un evento relacionado con la probabilidad subjetiva.



La probabilidad objetiva es la relación entre el número de veces que se realiza u observa un suceso en relación con el número de veces que sucede.

La probabilidad de un evento se obtiene dividiendo el número de resultados exitosos entre el total de eventos. Laplace define la probabilidad de un evento mediante el análisis de casos favorables (Possani y Barreiro, 2008).

Enfoque clásico

En este enfoque se obtiene un resultado exacto; para este caso, la probabilidad de que ocurra un evento se encuentra representada por el número de sucesos simples diversos, cada uno de los cuales tiene la misma posibilidad de suceder, y se representa por la siguiente ecuación:

$$P(A) = \frac{\text{número de formas en que puede ocurrir } A}{\text{número de sucesos simples diferentes}} = \frac{n}{N}$$

Donde:

P = probabilidad

A, B, C representan los sucesos o eventos

$P(A)$ = probabilidad de que ocurra el suceso A

Ejemplo

Si tenemos un recipiente de cristal con 50 canicas rojas y 200 canicas amarillas, ¿cuál es la probabilidad de sacar una canica roja?

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

Donde:

Las formas en que se puede obtener una canica roja es $n = 50$

El total de sucesos simples está representado por el total de eventos $N = 200 + 50$



$$P(A) = \frac{50}{50 + 200}$$

$$P(A) = \frac{1}{5}$$

$$P(A) = 0.2$$

Continuando con el ejemplo de tirar dos dados de colores, uno verde y uno rojo, calcula la probabilidad de que caiga **verde 3** y **rojo 4**.

Existen 36 posibilidades, que corresponden al espacio muestral, y sólo una de ellas representa la probabilidad de que caiga en **verde 3** y en **rojo 4**.

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

$$P(3, 4) = \frac{1}{36}$$

La probabilidad de que se obtenga **3 verde** y **4 rojo** es de 0.027

Enfoque de frecuencia relativa

Es la relación del número de eventos favorables obtenidos con relación al número total de intentos. En este enfoque se trabaja con estimaciones de sucesos o eventos; se observa un evento varias veces y se registran las ocasiones en que un evento A ocurre; se calcula mediante la siguiente ecuación:

$$P(A) = \frac{\text{número de veces que ocurre } A}{\text{número de veces que se repitió el ensayo}}$$

Ejemplo

En el caso de un recipiente de cristal que contiene 100 canicas rojas y amarillas, se han extraído 80 piezas, de las cuales 15 han sido de color rojo: ¿cuál es la probabilidad en estos eventos?

Para este caso, no interesa el total de canicas rojas, sino la frecuencia relativa, que se va a determinar con el total de eventos que se han suscitado en la extracción de la canica de color rojo.

$$P(A) = \frac{\text{número de veces que ocurre } A}{\text{número de veces que se repitió el ensayo}}$$

$$\text{Probabilidad} = \frac{15}{80}$$

$$\text{Probabilidad} = 0.1875$$

Es decir, existe una probabilidad de 18.75%.

Para obtener la probabilidad de fracaso en un evento, la ecuación es la siguiente:

$$q = 1 - p$$

Donde:

p = probabilidad de éxito

q = probabilidad de fracaso

Para calcular la probabilidad de fracaso del ejercicio anterior se realiza la siguiente operación:

$$q = 1 - 0.1875 = 0.8125$$

$$q = 81.25\%$$

Por lo tanto, la probabilidad de fracaso al obtener una canica roja es de 81.25% y la probabilidad de éxito de 18.75%.

Ejemplo

Se cuenta con una muestra al azar de 600 personas en el aeropuerto de la Ciudad de México; se identificó que 400 han realizado vuelos internacionales: calcula la probabilidad de que las personas hayan viajado al extranjero.

$$P(A) = \frac{\text{número de veces que ocurre } A}{\text{número de veces que se repitió el ensayo}}$$

$$P(A) = \frac{400}{600}$$

$$P(A) = 0.666$$

Es decir, la probabilidad de éxito es de 66.66%.

Para calcular la probabilidad de fracaso:

$$q = 1 - 0.666 = 0.3334$$

$$q = 33.34\%$$

Ejemplo

En un seguimiento del número de nacimientos en el Hospital del Niño, se observó que, de 205 nacidos en una semana, 105 son niños. Si un bebé es seleccionado del grupo, ¿cuál es la probabilidad de que no sea niño?

$$P(A) = \frac{\text{número de veces que ocurre } A}{\text{número de veces que se repitió el ensayo}}$$

$$P(A) = \frac{105}{205}$$

$$P(A) = 0.5121$$

Esta cantidad representa la probabilidad de que sea niña, pero lo que nos interesa identificar es la probabilidad de que no sea niño.

Por lo que, aplicando la ecuación de la probabilidad de fracaso, se tiene:

$$q = 1 - 0.5121 = 0.4878$$

$q = 48.78\%$ representa la probabilidad de no seleccionar un niño



En resumen, en el enfoque clásico es necesario contar con resultados igualmente probables, mientras que en el de frecuencias relativas se obtiene un estimado en lugar de un número exacto. En la medida en que el número de observaciones se incrementa tiende a acercarse a la probabilidad real (Triola, 2004).

4.2 Problemas de conteo

El proceso de conteo es la abstracción de realizar una operación aritmética con números cardinales, ya sea suma, resta, multiplicación y división.

Ejemplo

1. Supongamos que se cuenta con 40 libros de biología y 30 de anatomía humana en la biblioteca de tu plantel; en total hay $(40 + 30)$ 70 opciones para estudiar.
2. Se quiere conocer la cantidad de placas que pueden integrarse con 2 letras y 4 números; este problema se puede resolver considerando tres situaciones:
 - a. No se pueden repetir las letras ni los números.
 - b. Conociendo que el alfabeto lo conforman 27 letras, la primera opción es 27 y la segunda 26, y así sucesivamente.
 - c. Para el caso de los números, se cuenta con 10 dígitos (0, 1, 2, 3, 4, 5, 6, 7, 8, 9), por lo que la primera opción será 10, posteriormente 9, 8 y 7... 0.

$$(27)(26)(10)(9)(8)(7) = 3\,538\,080 \text{ opciones de placas}$$

Si se pueden repetir letras:

$$(27)(27)(10)(10)(10)(10) = 7\,290\,000 \text{ opciones de placas.}$$

Los problemas de conteo proporcionan las herramientas para realizar las operaciones aritméticas necesarias. A continuación se expondrá la forma de trabajar la factorial y sus diversas aplicaciones en la probabilidad.

4.2.1 Factorial, permutaciones y combinaciones

a. Factorial

La factorial representa la cantidad de veces que puedes ordenar un evento; se identifica con el signo $!$, y se usa en diversas áreas de las matemáticas; en el caso de la estadística es en las permutaciones y combinaciones. Una de sus características es que pertenece a los números naturales (conjunto de números enteros y positivos); se obtiene multiplicando todos los números menores naturales, incluyendo el cero, al evento que se quiere determinar.

Ecuación:

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

Ejemplo

Supongamos que cuentas con seis libros de estadística y quieres conocer las diferentes formas que pueden ser ordenados en un librero; la factorial te ayuda a identificar esa cantidad de eventos.



Es decir, $6!$ (seis factorial) representa las formas de ordenar los libros.

Donde:

$$6! = 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$$

Se cuenta con 720 formas diferentes de ordenar los libros

El cálculo de la factorial $3!$ y $7!$ se realiza de la siguiente manera con base a su ecuación:

$$3! = 3 \times 2 \times 1 = 6$$

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

$$3! = 3(3-1)! = 6$$

$$7! = 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 5\,040$$

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

$$7! = 7(7-1)! = 5\,040 \longrightarrow (7-1)! = 6!$$

$$7! = 7(720) = 5\,040 \longleftarrow 6! = 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$$

Esto quiere decir que la factorial de un número " x ", es igual al número " x " por la factorial del número menor que se determinará. A continuación se muestra el cálculo de la factorial $10!$

$$10! = 10 \times 9!$$

$$10! = 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$$



Escribe el resultado de la factorial diez ($10!$).



Para el caso de 0, la aplicación es $0! = 1$
y, por lo tanto: $1! = 1$

A continuación, vas a localizar la factorial del número 0 al 25.

N	$n!$
0	1
1	1
2	2
3	6
4	24
5	120
6	720
7	5040
8	40 320
9	362 880
10	3 628 800
11	39 916 800
12	479 001 600
13	6 227 020 800
14	87 178 291 200
15	1 307 674 368 000
16	20 922 789 888 000
17	355 687 428 096 000
18	6 402 373 705 728 000
19	121 645 100 408 832 000
20	2 432 902 008 176 640 000
21	51 090 942 171 709 400 000
22	1 124 000 727 777 610 000 000
23	25 852 016 738 885 000 000 000
24	620 448 401 733 239 000 000 000
25	15 511 210 043 331 000 000 000 000

Como se observa, la factorial crece rápido.

Con las factoriales se pueden realizar las operaciones básicas, como suma, resta, multiplicación y división.

Suma

$$\begin{array}{rcl} 7! & + & 3! \\ 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 & + & 3 \times 2 \times 1 \\ 5\,040 & + & 6 \\ & = & 5\,046 \end{array}$$

Resta

$$\begin{array}{rcl} 5! & - & 6! \\ 5 \times 4 \times 3 \times 2 \times 1 & - & 1 \times 2 \times 3 \times 4 \times 5 \times 6 \\ 120 & - & 720 \\ & = & 600 \end{array}$$

Multiplicación

$$\begin{array}{rcl} 2! & \times & 4! \\ (1 \times 2) & \times & (1 \times 2 \times 3 \times 4) \\ 2 & \times & 24 \\ & = & 48 \end{array}$$

División

$$\frac{9!}{5} = \frac{1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8 \times 9}{5} = \frac{362\,880}{5} = 72\,576$$

Simplificación de expresiones factoriales

$$\frac{X!Z!}{Z!} = X! \qquad \frac{X!\cancel{Z!}}{\cancel{Z!}} = X!$$

Para el caso de las expresiones factoriales, recuerda que factores iguales al dividirse el resultado es 1. Todo número multiplicado por la unidad es igual al mismo número.

$$\frac{8!4!}{4!} = 8:$$

$$\frac{8!(\cancel{1 \times 2 \times 3 \times 4})}{(\cancel{1 \times 2 \times 4 \times 4})} = 8!$$

$$\frac{(\cancel{8})4!}{\cancel{4!}4!} = \frac{8}{4!}$$

$$\frac{8}{4!} = \frac{8}{4 \times 3 \times 2 \times 1} = \frac{8}{24} = 0.33$$

Ejemplo

Se cuenta con la siguiente operación, calcula x .

$$\frac{(x+2)!}{(x+1)!} = 10$$



Para desglosar la factorial, se realiza una multiplicación de números; para el caso del número 2, se tiene que $2! = 2 \times 1$

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

Es decir: $(x+2)(x-1)!$

$$\frac{(x+2)\cancel{(x+1)!}}{\cancel{(x+1)!}} = 10$$

$$x+2 = 10$$

$$x = 10 - 2$$

$$x = 8$$

Ejemplo

El autobús escolar cuenta con 15 asientos para el mismo número de personas, ¿de cuántas maneras pueden sentarse los estudiantes en el autobús?

Para identificar el número de eventos en los que se pueden acomodar 15 personas en 15 lugares se representa por la factorial 15!

Buscando el número de la factorial en la pirámide se sabe que hay 1 307 674 368 000 formas para sentarse en los 15 asientos.

Ejemplo

Se tienen 40 cartas de una baraja; por lo tanto, se tienen 40! formas posibles para repartir las cartas. Supongamos que las primeras 4 cartas te tocan a ti, teniendo 4! formas de que sean repartidas estas cartas, quedando 36! posibles formas de repartir, y se desea conocer de cuántas formas es posible repartir las cartas, considerando que siempre van a ser repartidas a ti en primer lugar.

La operación es la siguiente:

$$\frac{40!}{36!4!}$$



Calcula el resultado del ejercicio (recuerda que factores iguales se cancelan).

$$\frac{40!}{36!4!} =$$

Calcula las siguientes factoriales.

$$15! =$$

$$21! =$$

$$4! + 11! =$$

$$8! \times 1! =$$

b. Permutaciones

Dentro de la probabilidad existen formas de presentar los resultados de un proceso, lo que ayuda a conocer la cantidad de eventos posibles en un espacio muestral (Sánchez, 2004). A continuación, se van explicar las permutaciones y las combinaciones.

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 un evento de 3 dígitos 2, 3, 4, existen diversas formas de presentarlo: 234, 342, 423, 432, 324; cada una de estas cifras representa una permutación de los dígitos 2, 3, 4 y, como se observa, muestra valores diferentes cada una de las permutaciones.

Con las letras de la palabra *silbar*, se pueden formar las palabras libras o Brasil, y podemos integrar otras palabras cuyos vocablos no son usuales, como 'rasilb' o 'labris'.

Por lo tanto, contamos con un conjunto de eventos, integrado por seis letras para seleccionar las permutaciones posibles, respetando el orden de aparición de las letras {s, i, l, b, a, r}.

— — — — —

En este caso, se puede llenar el primer espacio de seis maneras, con alguna de las letras del conjunto {s, i, l, b, a, r}: una vez elegida la primera letra se cuenta con 5 opciones, respetando el orden de aparición. Con base en esta información, tenemos que para el tercer espacio se cuenta con 4 formas y así sucesivamente.

Aplicando el principio de multiplicación factorial tenemos que:
 $5 \times 4 \times 3 \times 2 \times 1 = 6! = 720$

Tenemos 720 permutaciones simples o formas de reagrupar las letras del vocablo *silbar*.

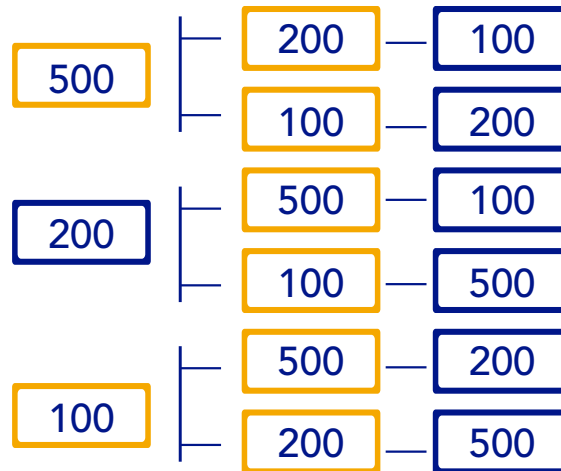
Nota: para las permutaciones es importante respetar el orden de aparición para dar paso a la probabilidad del siguiente evento.

Ejemplo

En un recipiente se coloca un billete de 500, uno de 200 y uno de 100; tres personas toman un billete sin ver; determina las formas posibles en que pueden permutarse o distribuirse los billetes.

Tenemos que la primera persona cuenta con tres opciones, la segunda con dos opciones y la última toma el billete restante.

Equivale a tener $3 \times 2 \times 1 = 3! = 6$



Por lo tanto, existen 6 formas de permutar los tres billetes.

- En una caja colocamos cuatro pelotas (morada, rosa, azul y verde). Al ir extrayendo una por una, ¿cuál es el orden en el que pueden aparecer?
- Con las letras de la palabra *lápiz*, ¿cuántas palabras distintas se pueden formar? (recuerda que las palabras pueden tener sentido o no).

La ecuación para determinar el número de permutaciones es:

Cuando se cuenta con n cantidad de objetos y sólo se considera una r cantidad para ordenar.

$$P(n, r) = \text{Permutaciones de } n \text{ objetos } P = \frac{n!}{(n-r)!}$$

Donde:

n = número total de eventos u objetos

r = número de objetos que se desea considerar

Ejemplo

En un recipiente colocamos 4 listones (rojo, verde, amarillo y azul); si se toman dos listones del recipiente, ¿cuál es el orden en el que pueden aparecer?

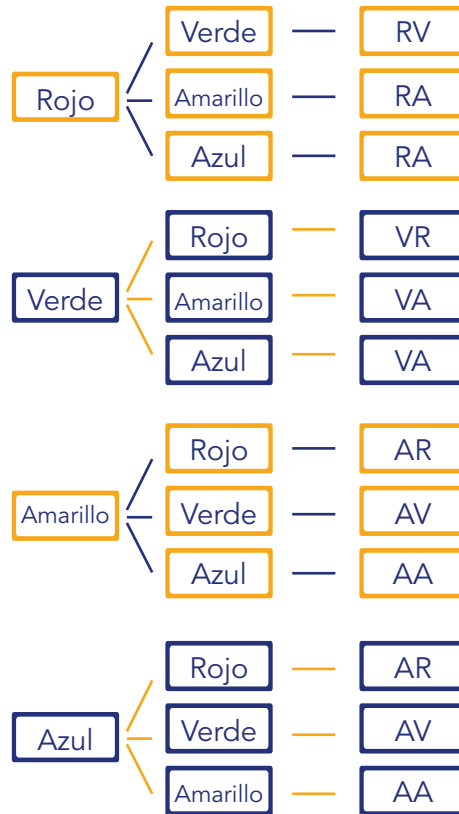
$$P(n, r) = \frac{4!}{(4-2)!}$$

$$P(n, r) = \frac{24}{2}$$

$$P(n, r) = 12$$

Otra forma de calcular es por medio de la factorial $4 \times 3 = 12$

Esto se puede corroborar en la siguiente representación



Comprobamos que existen 12 permutaciones diferentes.



Actividad
Autónoma

- Derivado del cambio de administración en un plantel (director, subdirector académico y subdirector administrativo), se cuenta con 2 candidatos para ocupar los puestos. Bajo el supuesto de que cualquiera de ellos cumple con los requisitos y las aptitudes para cualquiera de los tres puestos, determina el número de formas en que puede quedar integrada la dirección.
- El docente de educación física del plantel desea integrar equipos de 5 mujeres; el grupo se compone de 39 estudiantes, y de éstos 19 son mujeres: calcula el número de permutaciones posibles para equipos de 4 personas.



Permutaciones con remplazo

Cuando el número de objetos n es limitado, pero el número de veces m es indefinido, se le denomina remplazo y puede ser nuevamente elegido.

Ecuación de permutaciones con remplazo:

$$\text{permutaciones con remplazo} = n^m$$

Donde:

n = son las opciones u objetos diferentes

m = es el número de intentos con remplazo

Ejemplo

Un equipo de amigos ha ingresado al torneo interno de la escuela. Se cuenta con 4 juegos con la opción de ganar o perder, determina la cantidad de resultados posibles en este torneo.

n = corresponde a las opciones de ganar o perder (2)

m = es la cantidad de juegos con opción de ser elegido nuevamente (4)

Aplicando la ecuación:

$$\text{permutaciones con remplazo} = n^m$$

$$\text{permutaciones con remplazo} = 2^4$$

$$\text{permutaciones con remplazo} = 16$$

Cuando se cuenta con un solo evento en la permutación con remplazo, tenemos 16 formas diferentes de obtener un resultado de ganar o perder en 4 juegos.

Ejemplo

Fabián desea conocer el número de códigos posibles que pueden ingresarse en un dispositivo electrónico de 4 dígitos, considerando que el número de permutaciones con remplazo es para 10 dígitos, es decir, el evento muestral corresponde a 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. Calcula cuántos códigos son posibles permutar para un código de 4 dígitos.

$$\text{permutaciones con remplazo} = 10^4$$

$$\text{permutaciones con remplazo} = 10\,000$$

Permutaciones con más de un evento de remplazo

Con los siguientes 9 números: 2, 2, 2, 3, 3, 3, 3, 4, 4, se desea identificar todas las cifras que pueden integrarse.

1. Se observa que se cuenta con números repetidos:

2 = 3 veces repetido
 3 = 4 veces repetido
 4 = 2 veces repetido

Para las permutaciones con repetición de n elementos, se toma en cuenta que el primer elemento se repite a veces, el segundo b veces, el tercero c veces, y así sucesivamente.

$n = a + b + c + x$ veces

Ecuación:

$$P(n, r) = \binom{a, b, c}{n} = \left(\frac{P_n}{a! b! c! \dots} \right)$$

$$P = \binom{3, 4, 4}{9} = \left(\frac{9!}{3! 4! 2! \dots} \right)$$

$$P = \binom{3, 4, 4}{9} = \frac{9 \times 8 \times 7 \times 6 \times \cancel{5 \times 4 \times 3 \times 2 \times 1}}{(3 \times 2 \times 1)(\cancel{4 \times 3 \times 2 \times 1})(2 \times 1)} = 1\,260$$

Con 9 dígitos se pueden realizar 1 260 permutaciones con repetición de eventos.

Ejemplo

Un pequeño está jugando con 5 caballos, 3 automóviles y 6 soldados, en ¿cuántas formas diferentes pueden ordenarse sus juguetes?

$$P = \binom{a, b, c}{n} = \left(\frac{P_n}{a! b! c! \dots} \right)$$

$$P = \binom{3, 4, 6}{14} = \left(\frac{14!}{5! 3! 6! \dots} \right)$$

$$P = \binom{5, 3, 6}{14} = \left(\frac{14 \times 13 \times \cancel{12 \times 11 \times 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}}{(\cancel{5 \times 4 \times 3 \times 2 \times 1})(3 \times 2 \times 1)(\cancel{6 \times 5 \times 4 \times 3 \times 2 \times 1})} \right) = \left(\frac{14 \times 13 \times 11 \times 9 \times 8 \times 7}{3 \times 2 \times 1} \right)$$

$$P = \binom{5, 3, 6}{14} = \frac{1\,009\,008}{6} = 168\,168 \text{ son las formas de ordenar los juguetes.}$$

Se sustituye el $12 \times 10 = 120$, debido a que $5 \times 4 \times 3 \times 2 \times 1 = 120$.

Asimismo, si se desea identificar las formas de ordenar los juguetes considerando que los 5 caballos quedan al frente, entonces:

$$P = \binom{3, 6}{9} = \left(\frac{9 \times 8 \times 7 \times \cancel{6 \times 5 \times 4 \times 3 \times 2 \times 1}}{(3 \times 2 \times 1)(\cancel{6 \times 5 \times 4 \times 3 \times 2 \times 1})} \right) = \left(\frac{9 \times 8 \times 7}{3 \times 2 \times 1} \right)$$

$$P = \binom{5, 3, 6}{14} = \frac{504}{6} = 84 \text{ son las formas de ordenar los juguetes si los 5 caballos quedan al frente.}$$



¿Cuál es la probabilidad de que en los cinco primeros lugares queden los caballos?

$$P(A) = \frac{\text{número de formas en que puede ocurrir } A}{\text{números de sucesos simples diferentes}} = \frac{n}{N}$$

$$P(A) = \frac{84}{168 \cdot 168} = 0.00049$$



El equipo de fútbol femenino de Tenancingo va a participar en 12 juegos en la temporada de primavera. Considerando que en los 12 partidos obtengan 7 victorias, 3 empates y 2 juegos perdidos, ¿calcula de cuántas formas diferentes pueden quedar los resultados de estos partidos?

c. Combinaciones

Es la manera en la cual pueden presentarse los eventos, y para este caso el orden de aparición no importa.

Ejemplo

El club de ajedrez del plantel cuenta con 10 integrantes; la próxima competencia es en dos semanas y sólo existe espacio para tres integrantes. ¿De cuántas formas diferentes pueden elegirse de los 10 integrantes de la selección a los representantes de la competencia?

Tenemos 3 posibles representantes de 10, es decir $10 \times 9 \times 8$ de 10 variaciones del conjunto de elementos que van de tres en tres.

Existen 3! formas diferentes de ordenar a los integrantes de un conjunto de 10 integrantes. Ya conocemos que existen $n = 10$ eventos y sólo se desea considerar a $r = 3$ elementos para la combinación.

La ecuación para determinar el número de combinaciones es:

$$\text{Combinación de } n \text{ objetos tomados de } r \text{ en } r = Cr = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Donde:

n = número de eventos u objetos

r = número de objetos que se desea considerar

Son valores enteros positivos.

Esto quiere decir que 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).

Ejemplo

Un equipo de fútbol se encuentra integrado por 25 jugadores, ¿cuántas combinaciones de jugadores puede realizar el entrenador, considerando que sólo 11 jugadores ingresan a la cancha?

$$C_r = \binom{25}{11} = \frac{25!}{11!(25-11)!}$$

$$C_r = \binom{25}{11} = \frac{1,551121 \times 10^{22}}{(39\,916\,800)(8.717829 \times 10^{10})}$$

$$C_r = \binom{25}{11} = 4\,457\,400$$

Las combinaciones posibles para 25 jugadores considerando 11 jugadores en la cancha es 4 457 400.

Ejemplo

Se organiza un comité de 5 personas para la revisión de desastres en el municipio de Metepec, integrado por 2 arquitectos de los 7 que existen en el departamento y 3 ingenieros de los 10 que trabajan en el municipio, ¿de cuántas formas diferentes puede organizarse el comité?

En primer lugar, se calculan las diferentes combinaciones para elegir a los arquitectos.

$$C_r = \binom{7}{2} = \frac{7!}{2!(7-2)!}$$

$$C_r = \binom{7}{2} = \frac{7!}{2!5!} = \frac{7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{(2 \times 1)(5 \times 4 \times 3 \times 2 \times 1)} = \frac{7 \times 6}{2 \times 1}$$

$$C_r = \binom{7}{2} = 21$$

Se calcula para el caso de los ingenieros.

$$C_r = \binom{10}{3} = \frac{10!}{3!(10-3)!}$$

$$C_r = \binom{10}{3} = \frac{10!}{3!7!} = \frac{10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{(3 \times 2 \times 1)(7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1)} = \frac{10 \times 9 \times 8}{3 \times 2 \times 1}$$

$$C_r = \binom{10}{3} = 120$$

Aplicando el principio de multiplicación de los eventos, se tiene que las diversas combinaciones para integrar el comité de revisión corresponden a:

(21) (120) = 2 520 formas diferentes.



Se colocan en un recipiente seis pelotas de golf numeradas del 1 al 6; de éstas se van a tomar cuatro al azar, ¿de cuántas formas diferentes se pueden tomar las pelotas?

4.2.1.1. Diagrama de árbol

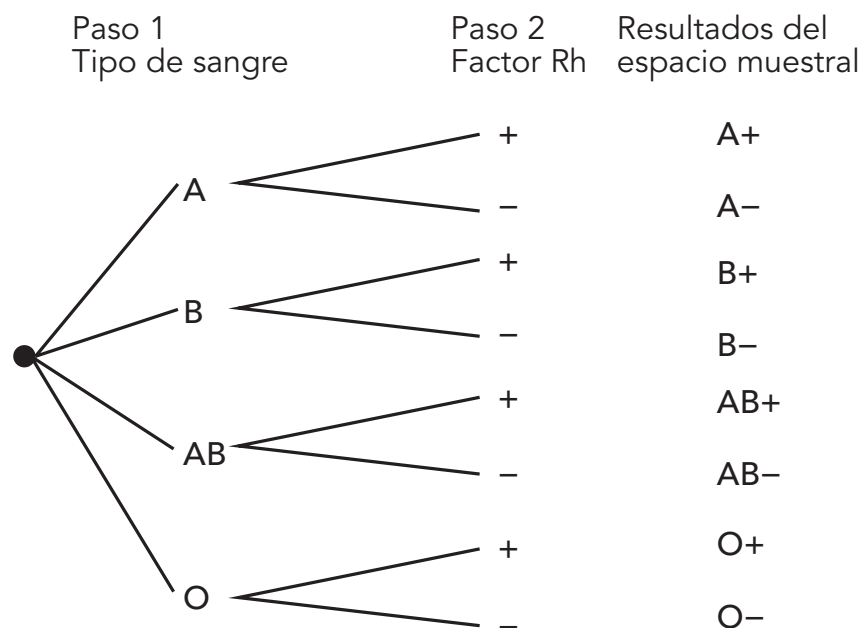
El diagrama de árbol o árbol de probabilidad es una representación gráfica de los posibles resultados de un evento, con sus diversas probabilidades de que suceda. Se construye mediante ramificaciones, las cuales representan la probabilidad de los eventos. Al realizar un análisis de los resultados posibles, la suma de la probabilidad es igual a 1 o 100% en las ramas del árbol, las cuales representan los diversos niveles que se deben de avanzar para alcanzar el resultado.

Ejemplo

La siguiente tabla muestra información de un evento simple que corresponde al tipo de sangre

Tipo de sangre / Factor RH	A	B	AB	O
Positivo	A–	B–	AB–	O–
Negativo	A+	B+	AB+	O+

Con base en esta información elaboramos el diagrama de árbol para obtener los resultados.

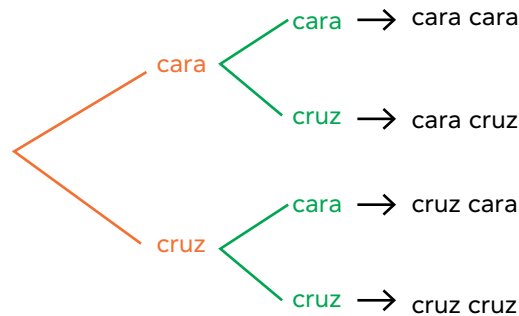


Por lo tanto, la probabilidad de tener alguno de los tipos de sangre para una persona es igual a $1/8$.

La probabilidad de un evento A es igual a la suma de las probabilidades de los eventos simples contenidos en A ; es decir, la probabilidad de que suceda un evento simple es de $\frac{1}{8}$, como se observa en los resultados del espacio muestral.

Ejemplo

El diagrama de árbol que se presenta a continuación muestra el lanzamiento de una moneda y sus opciones de cara (c) o cruz (x) en 4 lanzamientos

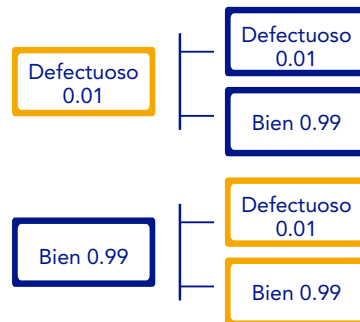


Cuatro sucesos elementales

$E = \{\text{cara cara, cara cruz, cruz cara, cruz cruz}\}$

Ejemplo

Una fábrica de ropa detectó en su producción que 1% de sus artículos tienen algún defecto; elabora un diagrama de árbol mostrando la probabilidad de elegir una prenda defectuosa, suponiendo 2 revisiones de forma aleatoria.



Elabora un diagrama de árbol para dos equipos de la liga de fútbol, considerando la posibilidad de ganar o perder.



4.3 Clasificación de la probabilidad

Un evento es un resultado posible y representa la mínima unidad de análisis para calcular una probabilidad.

Los eventos pueden clasificarse en:



Mutualmente excluyentes o disjuntos	Son aquellos que no pueden ocurrir al mismo tiempo
Independientes	No se ven afectados por otros eventos
Dependientes	Un evento afecta la probabilidad de que suceda otro
No excluyentes entre sí	La ocurrencia de un evento no impide la de otro

Fuente: Sánchez (2004).

4.3.1 Eventos mutuamente excluyentes o disjuntos

Dos eventos no pueden ocurrir al mismo tiempo: es uno u otro. Por ejemplo, si lanzamos una moneda, puede caer cara o cruz pero no las dos opciones; es decir, la ocurrencia de un evento impide automáticamente que suceda el otro.

Haciendo uso de la ecuación de la probabilidad:

$$P(A) = \frac{\text{números de eventos exitosos}}{\text{total de intentos o eventos}} = \frac{n}{N}$$

Ejemplo

En un cofre tenemos 3 pulseras azules, 6 collares verdes y un par de aretes cafés; ¿cuál es la probabilidad de extraer los aretes de las diversas piezas de joyería?

$$P(A) = \frac{n}{N} = \frac{1}{3 + 6 + 1}$$

$$P(A) = 0.1 = 10\%$$

La probabilidad de elegir los aretes del recipiente corresponde a un 10%.



En una baraja de 52 cartas se cuenta con 4 reyes, ¿qué probabilidad existe que una carta sea rey? ¿qué probabilidad existe que una carta sea reina y rey?

4.3.2 Eventos independientes

Este tipo de eventos no se ven afectados por otros eventos. Para los eventos independientes, la probabilidad total es la multiplicación de las probabilidades individuales.

Ejemplo

Continuando con el ejercicio de las pulseras, collares y aretes, qué sucede cuando deseamos determinar la probabilidad de sustraer una pulsera y los aretes.

$$P(A) = \left(\frac{\text{evento } A}{\text{total de eventos}} \right) \left(\frac{\text{número de eventos } B}{\text{total de eventos}} \right)$$

$$P(A) = \left(\frac{\text{número de pulseras}}{\text{total de eventos}} \right) \left(\frac{\text{número de aretes}}{\text{total de eventos}} \right)$$

$$P(A) = \left(\frac{3}{3+6+1} \right) \left(\frac{1}{3+6+1} \right)$$

$$P(A) = (0.3)(0.1)$$

$$P(A) = 0.3 = 3\%$$

Describe la conclusión:



4.3.3 Eventos dependientes

Se presentan cuando un evento afecta la probabilidad de que suceda otro; su dependencia influye en el resultado. La probabilidad de que 2 eventos se presenten de forma simultánea se determina multiplicando la probabilidad de ambos eventos.

Ejemplo

Continuando con el caso de la joyería (6 collares, 3 pulseras y 1 par de aretes), se desea determinar la probabilidad de extraer al azar dos collares.

Para que se cumpla el primer evento, la primera pieza que se extraiga del recipiente debe ser un collar; la probabilidad es la siguiente:

$$P(A) = \frac{\text{número de collares}}{\text{total de joyería}}$$

$$P(A) = \frac{6}{3+6+1}$$



$$P(A) = \frac{6}{10}$$

$$P(A) = 0.6$$

$$P(A) = 60\%$$

La segunda condición es que la segunda pieza que se extraiga sea también un collar; se determina la probabilidad mediante:

$$P(A) = \frac{5}{3 + 5 + 1}$$

$$P(A) = \frac{5}{9}$$

$$P(A) = 0.5556$$

$$P(A) = 55.56\%$$

Para el segundo evento, se considera que existe un collar menos dentro del cofre.

Para determinar la probabilidad de que dos eventos sean simultáneos se realiza una multiplicación:

$$P(A) = (\text{resultado del evento } A)(\text{resultado del evento } B)$$

$$P(A) = (0.6)(0.5556)$$

$$P(A) = 0.33$$

La probabilidad de que se extraigan dos collares es de 33%.



En una fiesta de disfraces se reunieron 20 personas, de las cuales 11 son hombres y 9 son mujeres; de los disfraces utilizados 5 hombres usan capa y 7 mujeres también. Calcula la probabilidad de que la segunda mujer use capa también.

4.3.4 Eventos no excluyentes entre sí

En este tipo de eventos la característica principal es que la posibilidad de que se presente un evento, no impide que suceda también otro.

La ecuación que se aplica es:

$$P(A) = \frac{\text{número de eventos exitosos}}{\text{total de intentos o eventos}} = \frac{n}{N}$$

Ejemplo

Se lanzan dos dados al aire, considerando cada suceso como el evento A y el evento B . En el evento A , uno de los dados cae en el número seis y corresponde al primer suceso. Esto representa una probabilidad de $6/36$ de acuerdo con el siguiente espacio muestral (1-6, 6-1, 2-6, 6-2, 3-6, 6-3, 4-6, 6-4, 5-6, 6-5, 6-6).

Y en el caso del segundo suceso, denominado B , se obtenga una suma de puntos que sea 8. Corresponde una probabilidad de $1/36$, representan la cantidad de eventos que cumplen con esta característica, y corresponde a (6, 2).

Como se observa se cumple la existencia de un evento que es la obtención de un seis en el primer evento y la suma de ocho con el lanzamiento del segundo evento.

El espacio muestral es 36 y representa la cantidad de posibles resultados al lanzar dos dados y las diferentes opciones que pueden ocurrir.

+						
	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

Escribe un ejemplo de un evento no excluyente entre sí.



Ejemplo

En el plantel de la escuela preparatoria se realizó una encuesta a 45 estudiantes del grupo 506 con relación a las lenguas extranjeras que estudian. Obteniendo los siguientes datos: 28 estudiantes cursan inglés, 16 francés, 12 no estudian ninguna de las dos lenguas extranjeras, los restantes cursan inglés y francés.

¿Qué probabilidad existe de que al entrevistar a un estudiante, éste se encuentre en las clases de inglés y francés?

$$\begin{array}{rccccccc} \text{Inglés} & + & \text{francés} & + & \text{ninguna lengua} & - & \text{total de estudiantes} & = \\ 28 & + & 16 & + & 12 & - & 45 & = 11 \end{array}$$

Por lo tanto, 11 jóvenes estudian inglés y francés

La operación para determinar la probabilidad de que 1 estudiante curse dos lenguas extranjeras es:

$$P(A) = \frac{11}{45}$$

$$P(A) = 0.244 = 24.4\%$$

Ejemplo

¿Cuál es la probabilidad de que en 15 tarjetas numeradas del 1 al 15 se tome una que sea par y menor o igual a 10?

De 1 al 15 hay (2, 4, 6, 8, 10, 12, 14) siete números pares, y de éstos 5 cumplen la regla de menor o igual a 10.

$$P(A) = \frac{5}{15}$$

$$P(A) = 0.3333 = 33.33\%$$

Es un evento no excluyente entre sí al contar con números pares que son menores o iguales a 10; por tanto, es posible que sucedan los dos eventos simples.

Ejemplo

Localiza la probabilidad de que al levantar una ficha de dominó se obtenga un número de puntos mayor a 9 y que cumplan la característica de ser múltiplos de 4, tomando en cuenta que el dominó consta de 28 fichas.

Evento A: (4, 6), (5, 5), (5, 6), (6, 6) por lo tanto corresponden a 4 eventos

No se considera (3, 6) o (5, 4) debido a que el resultado es 9 y no mayor a 9

Evento B. (0, 4), (1, 3), (2, 2), (2, 6), (3, 5), (4, 4), (6, 6), por lo tanto corresponden a 5 debido a que la combinación de la ficha de ajedrez (6, 6) ya se localiza en el evento A.

$$P(A) = \frac{10}{28} = \frac{5}{14}$$

$$P(A) = 0.3571 = 35.71\%$$

Ejercicios

1. Se realiza el lanzamiento de dos dados de forma simultánea. Determina la probabilidad de que la suma del lanzamiento sea 7.
2. El docente de Álgebra cuenta con una amplia asistencia de estudiantes entre hombres y mujeres a las asesorías semanales. Se solicitó que se registraran para asesorías del lunes 100 estudiantes, de los cuales se han inscrito 30 hombres. Se desea conocer qué probabilidad existe de que el siguiente alumno sea hombre.
3. En un examen de estadística y probabilidad se solicita a los estudiantes que respondan 6 preguntas de 10 reactivos, no importa el orden de las respuestas. Calcula de cuantas formas pueden responder el examen.
4. En una caja encontramos 40 diversas flores para el día 14 de febrero. Si se toman al azar 7 flores, ¿de cuántas formas diferentes se pueden combinar?
 - a) Considera el orden de aparición
 - b) No consideres el orden de aparición

5. Calcula las siguientes factoriales.

$$5! + 9! =$$

$$3! - 8! =$$

$$4! \times 4! =$$

$$\frac{5!}{2!} =$$



6. La familia Rodríguez ha reunido 5 bicicletas de los siguientes colores: rosa, roja, negra, azul y plateada; si se toman 2 bicicletas de la cochera, ¿cuál es el orden en el que pueden aparecer?
7. El fin de semana se rentó un autobús con 40 lugares para asistir a un concierto en la CDMX; si sólo asistieron 30 personas, ¿de cuántas formas diferentes se pueden sentar en el autobús?
8. Elabora un diagrama de árbol con la opción de aprobar o reprobar tres asignaturas en el primer y segundo parcial.
9. Una asociación civil ha reunido ropa para diversos albergues de los municipios cercanos a Toluca, recaudó 10 pantalones, 8 sudaderas y 3 pares de zapatos.
 - a) ¿Cuál es la probabilidad de que al elegir una prenda al azar, ésta sea un pantalón?
 - b) ¿Cuál es la probabilidad de elegir un par de zapatos y una sudadera?
 - c) ¿Cuál es la probabilidad de elegir dos pantalones de forma continua?
10. El grupo 307 cuenta con 40 alumnos, con los cuales se conformarán equipos de 5 integrantes, ¿cuántas combinaciones de equipos pueden organizarse?

11. El año pasado se realizó una encuesta para identificar de un grupo de estudiantes la cantidad de ellos que fuma y la que no fuma.

Sexo	Fuma	No fuma
Femenino	25	35
Masculino	40	15

Se elige al azar un estudiante. Complementa las oraciones identificando si se trata de un evento mutuamente excluyente o mutuamente no excluyente entre sí.

1. Femenino y fuma _____ se tienen a 25 estudiantes que cuentan con estas características.
2. Femenino y masculino _____
3. Masculino y no fuma _____ existen 15 estudiantes con estas características
4. Fuma y no fuma _____

Fuente: Possani y Barreiro (2008).

Autoevaluación

Marca en el recuadro correspondiente la respuesta sí o no a los siguientes indicadores.

Núm.	Indicadores	Sí	No
1	Identificas la diferencia entre evento y espacio muestral		
2	Realizas ejercicios de forma correcta con factorial, combinaciones y permutaciones		
3	Sabes elaborar un diagrama de árbol		
4	Calculas de forma correcta ejercicios de eventos mutuamente excluyentes o disjuntos		
5	Calculas de forma correcta ejercicios de eventos independientes		
6	Calculas de forma correcta ejercicios de eventos dependientes		
7	Calculas de forma correcta ejercicios de eventos no excluyentes entre sí		



En el caso de las respuestas negativas, escribe cuáles son las actividades o acciones que debes reforzar.

1	
2	
3	
4	
5	
6	

Escribe cuál es la razón por la que se te facilitó o dificultó el proceso de aprendizaje del módulo cuatro.

Fácil	Difícil

Formulario del módulo cuatro

Enfoque clásico:

$$P(A) = \frac{\text{número de formas en que puede ocurrir } A}{\text{número de sucesos simples diferentes}} = \frac{n}{N}$$

Enfoque de frecuencia relativa:

$$P(A) = \frac{\text{número de veces que ocurre } A}{\text{número de veces que se repitió el ensayo}}$$

Probabilidad de fracaso:

$$q = 1 - p$$

Factorial:

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

Permutaciones:

Cuando se cuenta con n cantidad de objetos y sólo se considera una r cantidad para ordenar

$$\text{Permutaciones de } n \text{ objetos } r = \frac{n!}{(n - r)!}$$

Permutaciones con remplazo:

Cuando el número de objetos n es limitado, pero el número de veces m es indefinido,

$$\text{permutaciones con remplazo} = n^m$$

Permutaciones con más de un evento de remplazo:

$$(a, b, c, \text{ y } n) = \frac{a, b, c}{n} = \frac{P_n}{a! b! c! \dots}$$

Combinaciones:

$$\text{Combinación de } n \text{ objetos tomados de } r \text{ en } r = Cr = \binom{n}{r} = \frac{n!}{r!(n - r)!}$$

Eventos mutuamente excluyentes o disjuntos:

$$\text{probabilidad} = \frac{\text{número de eventos exitosos}}{\text{total de intentos o eventos}} = \frac{n}{N}$$

Eventos independientes:

$$\text{probabilidad} = \left(\frac{\text{número de eventos } A}{\text{total de eventos}} \right) \left(\frac{\text{número de eventos } B}{\text{total de eventos}} \right)$$

Eventos dependientes:

$$\text{probabilidad} = (\text{resultado del evento } A)(\text{resultado del evento } B)$$

Eventos no excluyentes entre sí:

$$\text{probabilidad} = \frac{\text{número de eventos exitosos}}{\text{total de intentos o eventos}} = \frac{n}{N}$$

GLOSARIO

Centro de gravedad: es el punto de equilibrio de un conjunto de datos en un diagrama de dispersión.

Combinaciones: manera en la cual pueden presentarse los eventos y para este caso el orden de aparición no importa.

Covarianza: es una medida descriptiva que permite el tipo de asociación lineal entre dos variables.

Dato estadístico: es un valor, medida o característica que permite ser observada o contada para su posterior análisis.

Desviación estándar: es la raíz cuadrada de la varianza.

Desviación media: es el promedio de los valores absolutos de las desviaciones de cada dato con respecto a la media.

Diagrama de árbol: representación gráfica de los posibles resultados de un evento, con sus diversas probabilidades de que suceda.

Diagrama de dispersión: es una herramienta gráfica que ayuda identificar la posible relación entre dos variables.

Estadística: es aquella técnica que se encarga de la recolección, organización, tratamiento, análisis para extraer conclusiones o tomar decisiones.

Estadística descriptiva: es la técnica de recolección, organización y tratamiento de los datos. Es la responsable de recabar, organizar y mostrar en graficas la información seleccionada.

Estadística inferencial: analiza la información para elaborar las conclusiones correspondientes y tomar decisiones.

Frecuencia: representa la cantidad o número de veces que se repite un dato estadístico en un intervalo.

Frecuencia acumulada: es la sumatoria o acumulación de la frecuencia normal en cada uno de los intervalos subsecuentes.

Frecuencia relativa: es la proporción que representa para cada intervalo la cantidad de datos que se ubican en él.

Gráfica de barras: es una forma de representar gráficamente un conjunto de datos o valores, y está conformada por barras rectangulares cuya longitud es proporcional a los valores o las cantidades representados.

Histograma de frecuencias: representación gráfica de una variable en forma de barras, donde la superficie de cada barra es proporcional a la frecuencia de los valores representados.

Marca de clase: es el punto central o medio de un intervalo de clase.

Media aritmética: es una medida de tendencia central también conocida como promedio.

Mediana: es el valor que se encuentra exactamente a la mitad de un conjunto ordenado de datos.

Moda: es el valor que más se repite en una lista de datos; es decir, el que se presenta con mayor frecuencia en un conjunto de datos.

Muestra: es la parte proporcional de la población que será analizada y que es representativa de la información que se necesita.

Ojiva: es una representación gráfica de líneas.

Permutación: forma en la cual se presentan los eventos, respetando el orden de aparición.

Población: conjunto de personas que habitan un determinado lugar o aquellos individuos de la misma especie que ocupan un espacio geográfico.

Probabilidad: herramienta necesaria en la estadística para evaluar la confiabilidad de los sucesos que se presentan, a partir de las conclusiones que se tienen de una muestra.

Rango: es la diferencia entre el valor mayor menos el menor de un conjunto de datos estadísticos.

Serie de tiempo: un conjunto de valores observados durante una serie de periodos temporales (semanales, mensuales, trimestrales o anuales) secuencialmente ordenada.

Tabla de frecuencias: cuadro en el que se registran de forma organizada los datos: en clases, categorías o intervalos, de los datos estadísticos obtenidos, y pueden ser cualitativos o cuantitativos.

Variable: es la letra o número que representa un conjunto de datos integrados por la misma característica.

Variable discreta: es aquella que toma un número finito o valor de conteo en el cual intervienen exclusivamente números enteros.

Variación: representa la diferencia mínima entre los datos estadísticos.

Varianza: es una de las medidas de dispersión más utilizadas, indica la media de las diferencias (elevadas al cuadrado) de cada valor con respecto a la media.

REFERENCIAS

- Chong, M. J. (2016). *Metodología de la investigación*. Libro de texto. México: UAEM.
- Gutiérrez, A. (2012). *Probabilidad y estadística un enfoque por competencias*. México: McGraw Hill.
- Mendenhall, W. Beaver, R. y Beaver, B. (2008). *Introducción a la probabilidad y estadística*. México: Cengage Learning.
- Mora, R. (2000). *Probabilidad e inferencia estadística*. Lima: San Marcos.
- Portus, L. (1998). *Introducción a la estadística*. México: McGraw Hill.
- Possani, E. y Barreiro, L. (2008). *Estadística y probabilidad*. México: Santillana.
- Real Academia Española (2017). *Diccionario de la Lengua Española*. Recuperado de <http://www.rae.es/>, consulta: 25 de septiembre del 2017.
- Sánchez, O. (2004). *Probabilidad y estadística*. México: McGraw Hill.
- Spiegel, M. (2009). *Estadística*. México: McGraw Hill.
- Toca, A. (17 mayo 2009). *El cuento de las cuentas o el valor real de la estadística*. Recuperado de <https://www.sportyou.es/noticias/author/antonio-toca/page/2>, consulta: 26 de septiembre del 2017.
- Triola, M. (2004) *Probabilidad y estadística*. México: Pearson.

DIRECTORIO

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
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 Lujá

Secretaria de Difusión Cultural

Doctor en Ciencias del Agua
Francisco Zepeda Mondragón

Secretario de Extensión y Vinculación

Doctor en Educación
Octavio Crisóforo Bernal Ramos

Secretario de Finanzas

Doctora en Ciencias Económico Administrativas
Eréndira Fierro Moreno

Secretaria de Administración

Doctora en Ciencias Administrativas
María Esther Aurora Contreras Lara Vega

Secretaria de Planeación y Desarrollo Institucional

Doctora en Derecho
Luz María Consuelo Jaimes Legorreta

Abogada General

Doctor en Ciencias Sociales
Luis Raúl Ortíz Ramírez

Secretario Técnico de la Rectoría

Licenciada en Comunicación
Ginarely Valencia Alcántara

Directora General de Comunicación Universitaria

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

PLANTELES DEPENDIENTES de la Escuela Preparatoria

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"