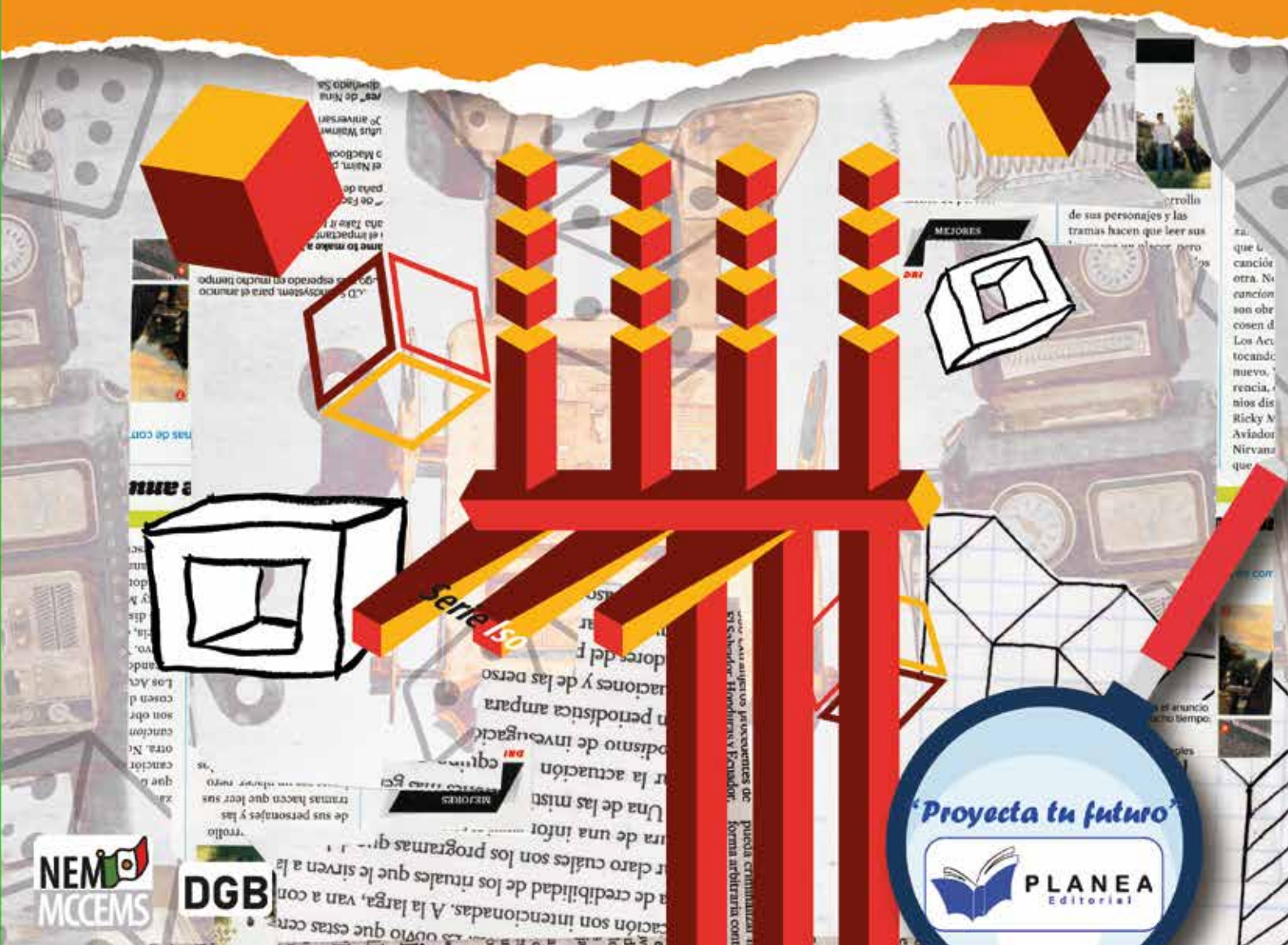


Taller de probabilidad y estadística 2

René Pérez Moreno

DGB



DGB





Primera edición 2026

Copyright © Editorial Planea

ISBN: En trámite **Clave: 20265**

Impreso en México

Contacto: 771-655-6186

Correo electrónico:

informes@editorialplanea.com.mx

Se reservan todos los derechos. Está prohibida la reproducción, almacenamiento en sistemas de recuperación o transmisión de estas publicaciones, ya sea de forma electrónica, mecánica, mediante fotocopia, grabación u otros medios, sin el consentimiento previo del editor. Esto incluye su distribución en redes, almacenamiento electrónico o transmisión para fines de aprendizaje a distancia.

Editor en jefe: Cosme Lorenzo Rodríguez

Autor: René Pérez Moreno

Correctora: Angélica María Alvarado Carreón

Diseño: Nasbbi Irazú Portes Loeza

Imágenes: Adobe Stock

Aviso de exención de responsabilidad:

Los enlaces incluidos en este libro no son propiedad de Editorial Planea, por lo que no se tiene control sobre la información proporcionada por los sitios web en un momento determinado, ni se puede garantizar la exactitud de la información proporcionada por terceros (enlaces externos). Aunque la información se recopila con cuidado y se actualiza de manera constante, no se asume responsabilidad alguna por su exactitud, integridad o actualidad.

Los artículos atribuidos a los autores reflejan sus opiniones y, a menos que se indique de forma específica, no representan las opiniones del editor. Además, la reproducción de este libro o cualquier material de los sitios web incluidos en él no está autorizada, ya que dicho material puede estar sujeto a derechos de propiedad intelectual.

Los derechos pertenecen a sus respectivos propietarios, y Editorial Planea no se hace responsable de la información mostrada en los enlaces proporcionados.



Presentación

En la Editorial Planea estamos comprometidos por ofrecer materiales didácticos de alta calidad, apegados al Nuevo Modelo Educativo de la Educación Media Superior, basado en la premisa de desarrollar en ti joven estudiante un aprendizaje situado en tu entorno, que te ayude en tu día a día, adaptándose a los cambios y brindarte un constante aprendizaje inclusivo, pluricultural, colaborativo y equitativo, basado en los principios de la Nueva Escuela Mexicana.

Este libro se encuentra apegado al 100% al programa de estudio basado en progresiones de aprendizaje del NME de la EMS, abordando las categorías y subcategorías para lograr los aprendizajes meta que propone el programa de Taller de Probabilidad y Estadística II, para la Dirección General de Bachillerato (DGB).

Estas progresiones, se encuentran organizadas en dos unidades de aprendizaje, la primera aborda los "fundamentos de la inferencia estadística", desarrollando las cuatro primeras progresiones de aprendizaje referentes a la exploración y análisis de datos, la estimación de probabilidad con la distribución normal y la distribución t student; en la segunda unidad denominada "Análisis y contrastación de relaciones estadísticas" se abordan las progresiones cinco a ocho del programa de estudios donde se analiza la contrastación de hipótesis, comparación de medias mediante la prueba t, correlación y regresión lineal y el análisis de independencia con chi cuadrada.

Este libro, está diseñado para ti, trata de proporcionar elementos que te ayuden a la comprensión y aplicación de los conceptos de variabilidad e incertidumbre inherentes en diversos fenómenos del entorno, asimismo, desarrollar estrategias de procesamiento y análisis de información desde un enfoque humanista para tomar decisiones razonadas.



La Nueva Escuela Mexicana NEM

La Nueva Escuela Mexicana (NEM) parte de un diagnóstico donde la educación se entendía como tres ciclos sin conexión, la educación básica (preescolar, primaria y secundaria), la educación media superior y la educación superior, con base en este diagnóstico se construye una propuesta donde la educación debe ser entendida para toda la vida, bajo el concepto de aprender a aprender, la actualización continua, adaptación a los cambios y el aprendizaje permanente.

La NEM propone un plan de 23 años en los diferentes niveles educativos, los cuales estén interconectados entre sí, donde se potencialice la formación integral de las niñas, niños, adolescentes y jóvenes con el objetivo de promover el aprendizaje de excelencia, inclusivo, pluricultural, colaborativo y equitativo a lo largo de su formación.

Para alcanzar el bienestar y la prosperidad incluyente, la NEM se fundamenta en los siguientes principios:



Fomento de la identidad con México. El amor a la patria, el aprecio por su cultura, el conocimiento de su historia y el compromiso de los valores plasmados en la Constitución Política, son las acciones que forman este principio.

Responsabilidad ciudadana. El principio implica la aceptación de derechos y deberes personales y comunes, el respeto por los valores cívicos por parte de los estudiantes formados en la NEM es esencial para transmitir los valores de honestidad, respeto, justicia, solidaridad, reciprocidad, lealtad, libertad, equidad y gratitud.



Honestidad. Se destaca este valor dentro de la responsabilidad social de los estudiantes, el cual permite formar una sociedad con base en la confianza y el sustento de la verdad de todas las acciones para permitir una sana relación entre los ciudadanos.

Respeto de la dignidad humana. Promover el respeto irrestricto a la dignidad y los derechos humanos de las personas, con base en la convicción de la igualdad de todos los individuos en derechos, trato y oportunidades.





Respeto por la naturaleza y cuidado del medio ambiente. La conciencia ambiental favorece la protección y conservación del medio ambiente, la prevención de la contaminación y cambio climático comienza con la educación del desarrollo sostenible.

Promoción de la interculturalidad.

El aprecio y la comprensión por la diversidad cultural y lingüística, así como, el diálogo y el intercambio cultural es una fuerza motriz para tener una vida intelectual, afectiva, moral y espiritual.



Participación en la transformación de la sociedad.

La superación de cada persona por iniciativa propia es la base de este principio, el sentido social de la educación permite construir relaciones cercanas, solidarias y fraternas que superan las indiferencias y la apatía por transformar la sociedad.



Promoción de la cultura de la paz. El objetivo de la agenda 2030 que promueve "Paz, justicia e instituciones sólidas", tiene como fundamento promover sociedades pacíficas, inclusivas, que faciliten el desarrollo sostenible, el acceso a la justicia para todos y la construcción a todos los niveles de instituciones eficaces e inclusivas que rindan cuentas.





Conoce tu libro

Dentro del libro se encuentra desarrollado el Nuevo Modelo Educativo de la Educación Media Superior, el cual se basa en un programa de estudio por progresiones de aprendizaje, las cuales se desarrollan en tres momentos que son:



Apertura. En este primer momento se busca despertar el interés y la motivación del estudiante por el tema que se va a abordar.



Cierre. En este último momento se busca consolidar los aprendizajes y hacer una evaluación del proceso.



Desarrollo. Se presenta el contenido y se realiza una explicación clara y detallada de los conceptos clave.



También se encuentran las secciones:

Evaluación diagnóstica. Se encuentra al inicio de cada unidad de aprendizaje, ayuda a identificar las fortalezas y debilidades con los temas que se van a abordar.

Aprendizaje situado en contextos:



Escuela



Aula



Comunidad



Prácticas transversales.

Donde se enlazan los aprendizajes de los recursos socio-cognitivos con las disciplinas de las áreas de conocimiento.

Prácticas socioemocionales.

El currículum ampliado se vincula con los recursos sociocognitivos, áreas de conocimiento por medio de los diferentes ámbitos de los recursos socioemocionales que están presentes en este tipo de actividades.





Prácticas de aprendizaje. La mejor manera de aplicar los conocimientos y habilidades aprendidas es a través de este tipo de prácticas, las cuales están numeradas, ubicadas en un contexto de aprendizaje y potencializando un principio de la NEM, como se muestra en el siguiente ejemplo:



Práctica de aprendizaje



Lectura NEM. Es una actividad de comprensión lectora que aborda uno de los principios de la Nueva Escuela Mexicana.



Evaluación de la unidad de aprendizaje. Son reactivos que abordan los temas de cada unidad de aprendizaje.

Categorías, subcategorías y metas de aprendizaje. Cada progresión tiene al inicio las categorías, subcategorías y metas de aprendizaje que aborda su contenido como se muestra a continuación:

Categorías de aprendizaje

Subcategoría de aprendizaje

C1 S1 S2
M1 M2

Metas de aprendizaje



Proyecto Aula - Escuela - Comunidad (PAEC). En estos códigos QR podrás realizar las actividades de las progresiones que son parte del PAEC.

Maestro Iso. Cada vez que veas al maestro Iso, él te explicará la progresión de manera dinámica, escaneando el código QR.



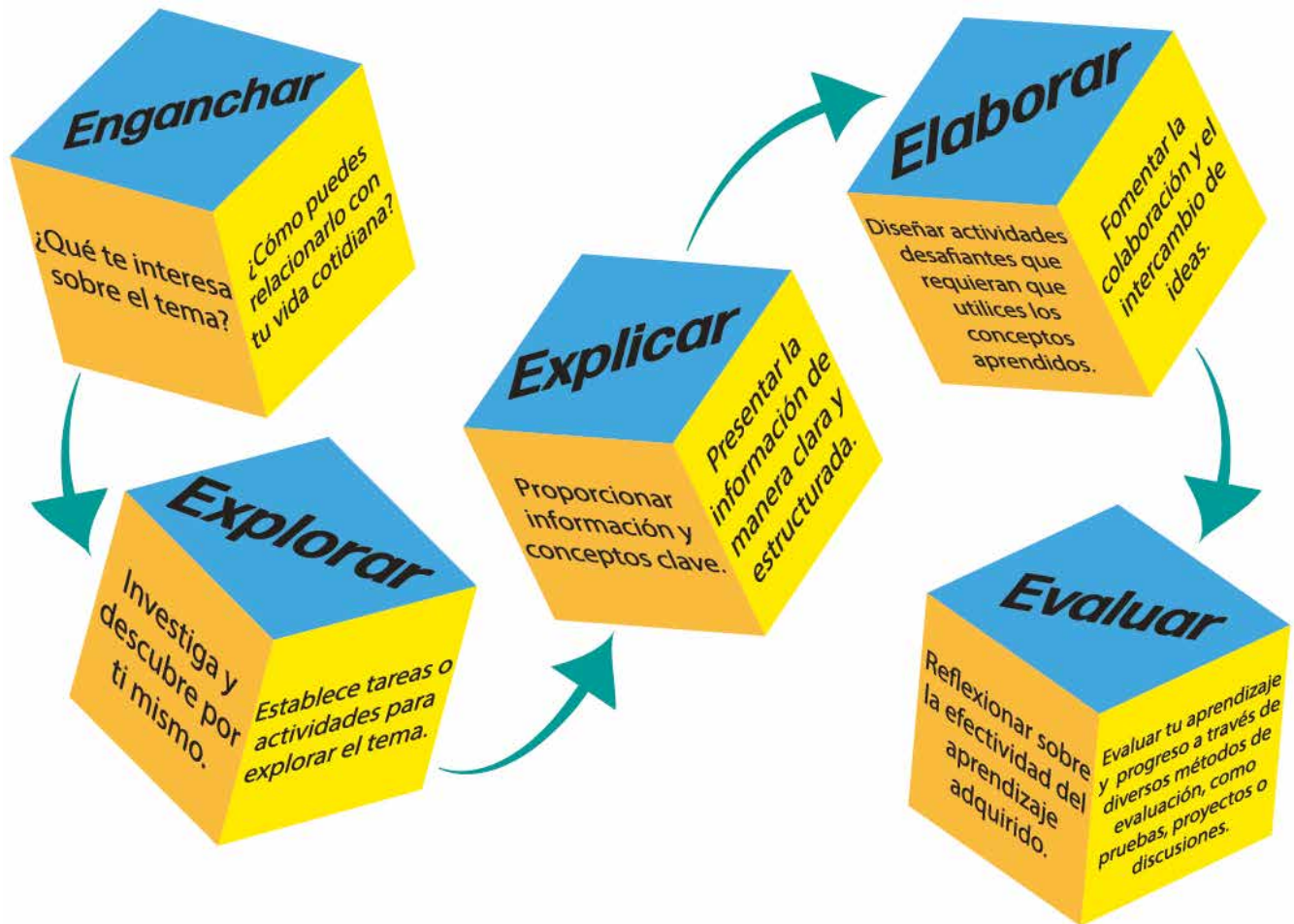
Progresiones de aprendizaje

1. Emplea un análisis exploratorio de datos no agrupados a partir del cálculo de medidas de tendencia central, de dispersión y representaciones gráficas, tales como histogramas y diagramas de caja y bigotes para efectuar comparaciones, identificar patrones y tendencias que permitan verificar si se distribuyen normalmente en problemáticas de diversos contextos, haciendo uso de los recursos tecnológicos al alcance.
2. Identifica y distingue en una situación de su entorno la población y el tipo de muestreo utilizado para analizar la distribución de las medias muestrales a medida que se incrementa el tamaño de la muestra, favoreciendo una aproximación informal al Teorema del Límite Central a través del uso de recursos tecnológicos disponibles.
3. Calcula intervalos de confianza, utilizando la distribución normal como modelo matemático, cuando se conoce la desviación típica poblacional o se tiene una muestra grande (mayor a 30) para efectuar una estimación de la media a partir de la información obtenida de una muestra en problemáticas o situaciones propias de las ciencias y su entorno.
4. Calcula intervalos de confianza, utilizando la distribución t student como modelo matemático, cuando se desconoce la desviación típica poblacional o se tiene una muestra pequeña (menor a 30) para efectuar una estimación de la media a partir de la información obtenida de una muestra en problemáticas o situaciones propias de las ciencias y su entorno.
5. Identifica los conceptos y aplica el procedimiento para contrastar hipótesis estadísticas a partir de la selección del modelo apropiado (distribución z o t, considerando sigma conocida o tamaño de muestra pequeño), haciendo uso de los recursos tecnológicos disponibles y favoreciendo el pensamiento crítico en situaciones de toma de decisiones en problemáticas contextuales.
6. Contrasta hipótesis a partir de datos estadísticos que provienen de variables cuantitativas para determinar si la media de dos muestras son estadísticamente iguales o distintas a través de la prueba t, haciendo uso de recursos tecnológicos disponibles para la toma de decisiones informadas en diversos contextos.
7. Analiza y establece la relación existente entre dos variables cuantitativas por medio de una representación gráfica y el cálculo del coeficiente de correlación de Pearson para la construcción de un modelo de regresión lineal, mediante el uso de los recursos tecnológicos disponibles, donde se explique la causalidad existente en problemas de su entorno, valorando sus alcances y limitaciones para propiciar el pensamiento crítico.
8. Contrasta hipótesis a partir de datos estadísticos que provienen de variables categóricas para determinar si existe independencia entre ellas a través de la prueba de chi cuadrada χ^2 y el uso de recursos tecnológicos disponibles, y con ello generar conciencia en la toma de decisiones en diversos contextos.

Estrategias para trabajo colaborativo

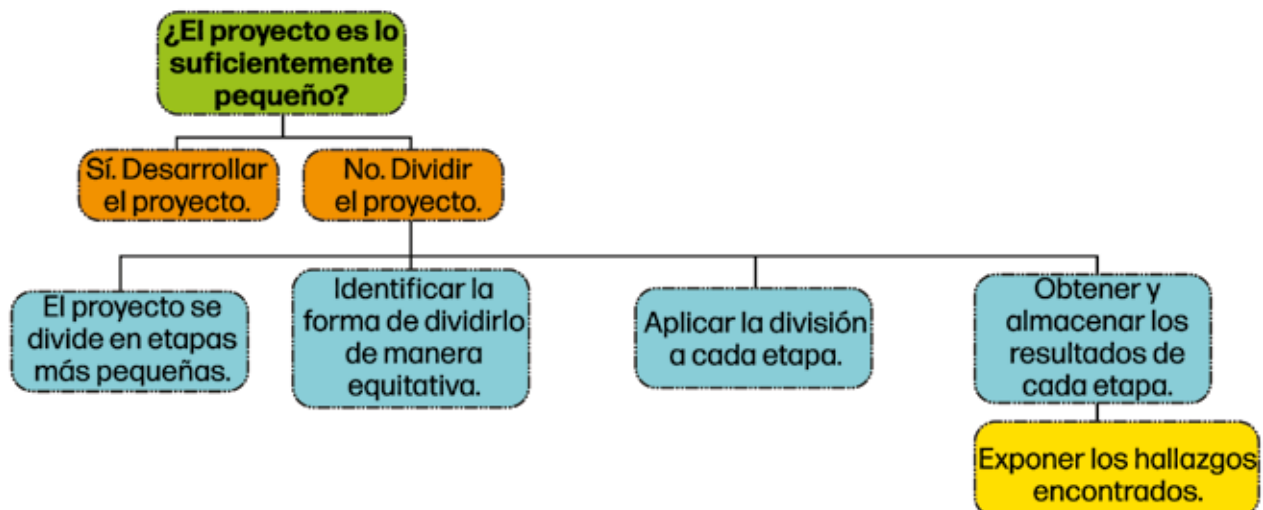
Estrategia 5E

Es una estrategia utilizada en educación para el trabajo colaborativo y diseño de proyectos, consiste en:



Divide y vencerás

Consiste en no ver un proyecto como una unidad, sino como una serie de etapas que pueden desarrollarse de manera individual para después integrar y exponer los hallazgos encontrados, a continuación se muestran los pasos a seguir.



Contenido

Unidad de aprendizaje 1. Fundamentos de la inferencia estadística

- Exploración y análisis de datos.
- Población, muestra y distribución muestral.
- Estimación con distribución normal.
- Estimación con distribución t Student.

Unidad de aprendizaje 2. Análisis y contrastación de relaciones estadísticas

- Introducción a la contrastación de hipótesis.
- Comparación de medias mediante la prueba t.
- Correlación y regresión lineal.
- Análisis de independencia con chi cuadrada.





Unidad de aprendizaje 1

Fundamentos de la inferencia estadística

Categorías de aprendizaje:

■ C1. Procedural.

Subcategorías:

- S4. Manejo de datos e incertidumbre.

■ C2. Procesos de intuición y razonamiento

Subcategorías:

- S1. Capacidad para observar y conjeturar.
- S2. Pensamiento intuitivo.

■ C3. Solución de problemas y modelación

Subcategorías:

- S1. Uso de modelos.

Meta de aprendizaje:

- CIM1. Ejecuta cálculos y algoritmos para resolver problemas matemáticos de las ciencias y de su entorno.
- CIM2. Analiza los resultados obtenidos al aplicar procedimientos algorítmicos propios del pensamiento matemático en la resolución de problemáticas teóricas y de su contexto.
- CIM3. Comprueba los procedimientos usados en la resolución de problemas utilizando diversos métodos, empleando recursos tecnológicos o la interacción con sus pares
- C2M1. Observa y obtiene información de una situación o fenómeno para establecer estrategias o formas de visualización que ayuden a entenderlo.

- **C2M3.** Compara hechos, opiniones o afirmaciones para organizarlos en formas lógicas útiles en la solución de problemas y explicación de situaciones y fenómenos.
- **C2M4.** Argumenta a favor o en contra de afirmaciones acerca de situaciones, fenómenos o problemas propios de la matemática, de las ciencias o de su contexto.
- **C3M1.** Selecciona un modelo matemático por la pertinencia de sus variables y relaciones para explicar una situación, fenómeno o resolver un problema tanto teórico como de su contexto.
- **C3M3.** Aplica procedimientos, técnicas y lenguaje matemático para la solución de problemas propios del pensamiento matemático, de áreas de conocimiento, recursos sociocognitivos, recursos socioemocionales y de su entorno.
- **C3M4.** Construye y plantea posibles soluciones a problemas de áreas de conocimiento, recursos sociocognitivos, recursos socioemocionales y de su entorno, empleando técnicas y lenguaje matemático.

Aprendizaje de trayectoria:

- Valora la aplicación de procedimientos automáticos y algorítmicos, así como la interpretación de sus resultados para anticipar, encontrar y validar soluciones a problemas matemáticos, de áreas del conocimiento y de su vida personal.
- Adopta procesos de razonamiento matemático tanto intuitivos como formales tales como observar, intuir, conjeturar y argumentar, para relacionar información y obtener conclusiones de problemas (matemáticos, de las ciencias naturales, experimentales y tecnología, sociales, humanidades y de la vida cotidiana).
- Modela y propone soluciones a problemas tanto teóricos como de su entorno, empleando lenguaje y técnicas matemáticas.
- Explica el planteamiento de posibles soluciones a problemas y la descripción de situaciones en el contexto que les dio origen empleando lenguaje matemático y lo comunica a sus pares para analizar su pertinencia.

Progresiones:

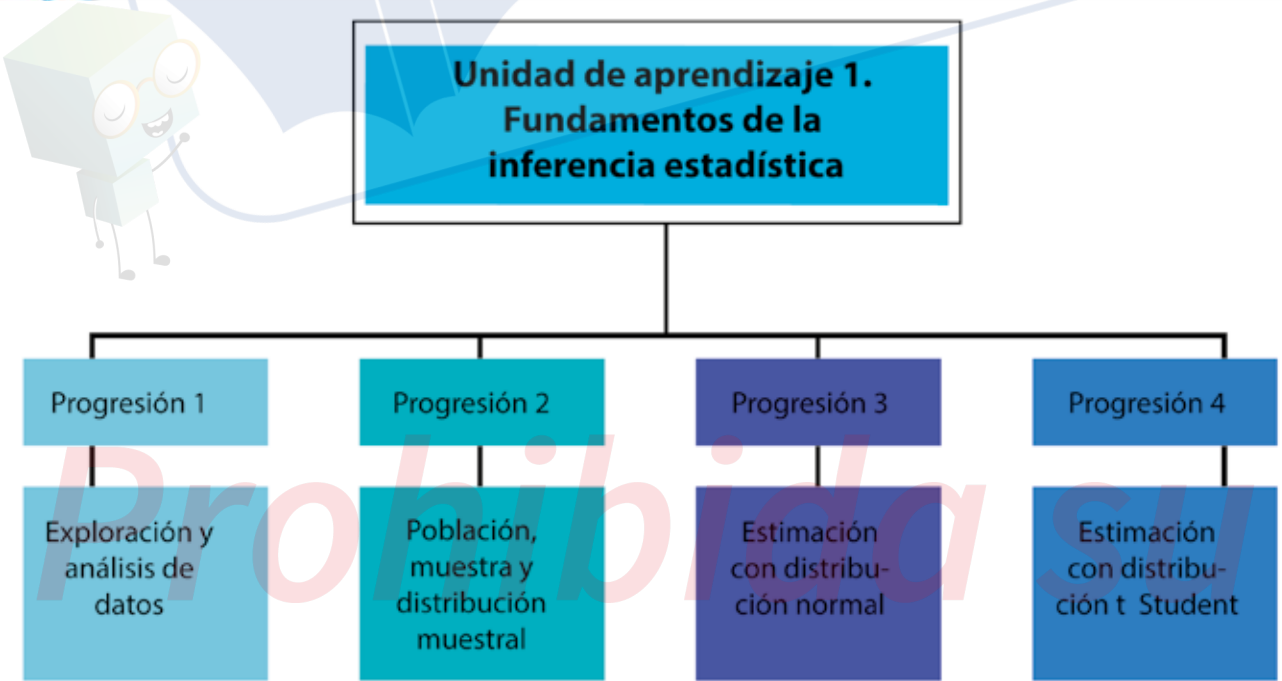
1. Emplea un análisis exploratorio de datos no agrupados a partir del cálculo de medidas de tendencia central, de dispersión y representaciones gráficas, tales como histogramas y diagramas de caja y bigotes para efectuar comparaciones, identificar patrones y tendencias que permitan verificar si se distribuyen normalmente en problemáticas de diversos contextos, haciendo uso de los recursos tecnológicos al alcance.
2. Identifica y distingue en una situación de su entorno la población y el tipo de muestreo utilizado para analizar la distribución de las medias muestrales a medida que se incrementa el tamaño de la muestra, favoreciendo una aproximación informal al Teorema del Límite Central a través del uso de recursos tecnológicos disponibles.
3. Calcula intervalos de confianza, utilizando la distribución normal como modelo matemático, cuando se conoce la desviación típica poblacional o se tiene una muestra grande (mayor a 30) para efectuar una estimación de la media a partir de la información obtenida de una muestra en problemáticas o situaciones propias de las ciencias y su entorno.
4. Calcula intervalos de confianza, utilizando la distribución t student como modelo matemático, cuando se desconoce la desviación típica poblacional o se tiene una muestra pequeña (menor a 30) para efectuar una estimación de la media a partir de la información obtenida de una muestra en problemáticas o situaciones propias de las ciencias y su entorno.

Presentación

La primera unidad de aprendizaje denominada “Fundamentos de la inferencia estadística” se introduce en el análisis estadístico como herramienta para comprender fenómenos presentes en distintos contextos. Parte del estudio de datos no agrupados mediante medidas de tendencia central, de dispersión y representaciones gráficas que permiten comparar información, reconocer patrones y valorar la pertinencia de modelos de distribución, con apoyo de recursos tecnológicos accesibles.

El recorrido formativo incorpora la identificación de poblaciones y tipos de muestreo en situaciones cercanas al entorno. Este enfoque facilita la comprensión de la variabilidad de las medias muestrales y su comportamiento al modificarse el tamaño de la muestra, lo que abre paso a una aproximación intuitiva al Teorema del Límite Central mediante herramientas digitales. También se desarrolla la capacidad de estimar parámetros poblacionales mediante intervalos de confianza basados en la distribución normal, en escenarios donde se conoce la desviación típica poblacional o se dispone de muestras grandes.

Por último, se profundiza en la construcción de intervalos de confianza sustentados en la distribución t de Student, adecuados para situaciones con muestras pequeñas o con desviación típica poblacional desconocida. Los temas específicos se muestran en el siguiente diagrama.



reproducción



Evaluación diagnóstica

1. ¿Cuál es el propósito principal del Análisis Exploratorio de Datos (AED) en un conjunto de datos no agrupados?
 - a) Predecir resultados futuros utilizando regresión lineal
 - b) Calcular la media poblacional exacta
 - c) Seleccionar el tipo de muestreo más adecuado
 - d) Identificar patrones, tendencias y posibles anomalías mediante medidas de tendencia central, dispersión y representaciones gráficas
2. ¿Cuál de las siguientes opciones describe mejor la diferencia entre población y muestra en un estudio estadístico?
 - a) La población incluye todos los datos posibles, mientras que la muestra es un subconjunto representativo de la población
 - b) La muestra siempre es más grande que la población
 - c) La muestra contiene menos errores que la población
 - d) Población y muestra significan lo mismo en estadística
3. ¿Qué sucede con la distribución de las medias muestrales a medida que aumenta el tamaño de la muestra, según el Teorema del Límite Central?
 - a) Se vuelve más dispersa y menos predecible
 - b) Tiende a una distribución normal, sin importar la forma de la población original
 - c) Permanece igual que la población original
 - d) Desaparece cualquier variabilidad
4. ¿En qué situación es apropiado calcular un intervalo de confianza utilizando la distribución normal?
 - a) Cuando se desconoce la desviación típica poblacional y la muestra es pequeña
 - b) Cuando se analiza una variable cualitativa
 - c) Cuando se conoce la desviación típica poblacional o la muestra es grande (mayor a 30)
 - d) Cuando no se requiere estimar parámetros poblacionales

Exploración y análisis de datos

Enganchar 1



Apertura

Unidad 1

Vivimos en un mundo repleto de información. Cada interacción, cada medición científica, cada transacción comercial genera datos. Pero estos números, por sí solos, suelen ser solo ruido. ¿Cómo transformamos ese ruido en conocimiento? La respuesta comienza con el Análisis Exploratorio de Datos (AED).

A lo largo de la progresión de aprendizaje se desarrollan las habilidades para usar herramientas para describir, resumir y visualizar conjuntos de datos no agrupados, con el fin de revelar las historias que esconden. A partir de las medidas de tendencia central, dispersión y representaciones gráficas.



Práctica de aprendizaje



Explorar 2

El dilema del gestor deportivo

Eres el gestor de un club de atletismo que debe decidir en qué especialidad, 100 metros planos (velocidad) o lanzamiento de bala (fuerza), invertir recursos para formar un nuevo equipo juvenil. Tienes dos conjuntos de datos con los mejores registros personales (en segundos para 100 m y en metros para el lanzamiento de bala de 15 atletas juveniles prometedores en cada disciplina:

■ 100 metros planos:

11.2, 11.5, 11.8, 11.9, 12.0, 12.1, 12.1, 12.2, 12.3, 12.4, 12.5, 12.6, 12.7, 12.8, 13.

■ Lanzamiento de bala:

10.1, 11.3, 11.5, 12.0, 12.2, 12.5, 12.7, 13.0, 13.4, 13.8, 14.2, 14.5, 15.0, 15.9, 17.5

A partir de los datos se plantean las siguientes preguntas:

a) ¿En cuál disciplina el rendimiento del grupo es más parejo (consistente)?

b) ¿Qué grupo tiene un rendimiento general más alto en comparación con sus propios registros?

c) ¿Hay atletas con registros excepcionalmente buenos o malos (posibles talentos excepcionales o casos que necesiten apoyo específico)?

d) Basándote en la forma de la distribución, ¿cuál conjunto de datos se aproxima más a una distribución normal?



Medidas de tendencia central

Las medidas de tendencia central son valores únicos que intentan representar o describir un conjunto completo de datos. Son el primer acercamiento para resumir la información.

- **Media (Promedio Aritmético).** Se calcula sumando todos los valores y dividiendo entre el número total de datos.
- **Media (\bar{x}) = (Suma de todos los valores) / (Número de datos)**
- **Mediana:** Es el valor que ocupa la posición central cuando los datos están ordenados de menor a mayor. Separa el 50% inferior del 50% superior.
- **Moda:** Es el valor que aparece con mayor frecuencia en el conjunto de datos. Puede haber más de una moda o ninguna.

Para comprender el uso de las medidas de tendencia central analiza el siguiente ejemplo.

Se realiza un análisis sobre las notas finales (sobre 10) de un grupo de 7 estudiantes en Matemáticas, los resultados son:

7.5, 8.0, 6.0, 9.5, 8.0, 5.5, 8.0

1. Ordenar los datos:

5.5, 6.0, 7.5, 8.0, 8.0, 8.0, 9.5

2. Calcular la media:

- $\text{Suma} = 5.5 + 6.0 + 7.5 + 8.0 + 8.0 + 8.0 + 9.5 = 52.5$
- $\text{Número de datos } (n) = 7$
- $\text{Media } (\bar{x}) = 52.5 / 7 \approx 7.5$

3. Calcular la mediana:

- **Datos ordenados:** 5.5, 6.0, 7.5, 8.0, 8.0, 8.0, 9.5
- La posición central es la 4ª (7 datos). El valor en esa posición es 8.0.
- **Mediana = 8.0**

4. Identificar la moda:

- El valor 8.0 aparece 3 veces, más que cualquier otro.
- **Moda = 8.0**

Con los resultados obtenidos se puede establecer que la nota típica del grupo ronda entre 7.5 y 8.0. La media (7.5) es ligeramente afectada por la nota baja (5.5), mientras la mediana y moda (8.0) muestran una tendencia central ligeramente mayor.



■ Visualización de las medidas de tendencia central. Los puntos representan cada dato. La media (rojo) es el equilibrio numérico, la mediana (verde) es el valor central físico y la moda (amarillo) es el valor más frecuente



Práctica de aprendizaje



1. Calcula la media, mediana y moda del siguiente conjunto de datos que representa el número de horas de sueño de 8 adolescentes: {6, 7, 8, 9, 7, 6.5, 10, 7}.

2. Si al conjunto anterior se añade un nuevo dato, 12 horas, ¿cuál medida de tendencia central cambia más: la media o la mediana? Justifica.

3. Inventa un conjunto pequeño de datos donde la media sea mayor que la mediana. ¿Qué implica esto sobre la distribución de los datos?

Medidas de dispersión

Conocer el centro no es suficiente. ¿Los datos están todos agrupados cerca de la media o muy esparcidos? Dos equipos pueden tener la misma media de puntos por juego, pero uno puede ser muy consistente y el otro muy irregular. Las medidas de dispersión son:

■ Rango.

Es la diferencia entre el valor máximo y el mínimo.

$$\text{Rango} = \text{Máximo} - \text{Mínimo}$$

Simple pero muy sensible a valores extremos.

■ Varianza (s^2).

Mide el promedio de las distancias al cuadrado de cada dato respecto a la media. Indica la dispersión en las unidades originales al cuadrado.

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}, \text{ para una muestra.}$$

■ Desviación estándar (s).

Es la raíz cuadrada de la varianza. Regresa a las unidades originales de los datos y es la medida de dispersión más útil. Mide, en promedio, cuánto se desvía un dato de la media.

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

Para comprender el cálculo de las medidas de dispersión observa el siguiente ejemplo.

Se cuentan con los siguientes datos: 5.5, 6.0, 7.5, 8.0, 8.0, 8.0, 9.5, calcular las medidas de dispersión.

Antes de comenzar con el cálculo de las medidas de dispersión es necesario calcular la media:

$$\bar{x} = \frac{5.5 + 6.0 + 7.5 + 8.0 + 8.0 + 8.0 + 9.5}{7} = \frac{52.5}{7} = 7.5$$

1. Rango = $9.5 - 5.5 = 4.0$

2. Varianza (s^2):

■ Calcular diferencias con la media y elevarlas al cuadrado:

$$(5.5-7.5)^2=4, (6-7.5)^2=2.25, (7.5-7.5)^2=0, (8-7.5)^2=0.25, (8-7.5)^2=0.25, (8-7.5)^2=0.25, (9.5-7.5)^2=4.$$

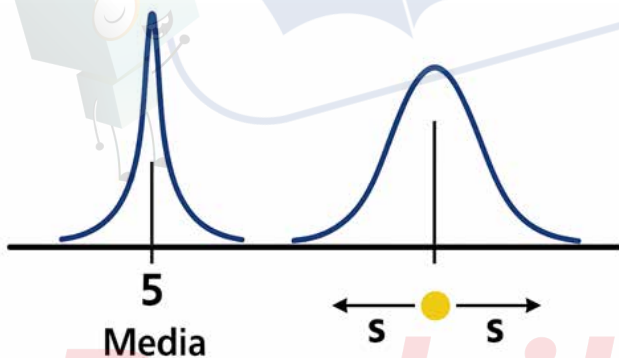
■ Suma de cuadrados = $4+2.25+0+0.25+0.25+0.25+4 = 11.0$

$$s^2 = 11.0 / (7-1) = 11.0 / 6 \approx 1.83$$

3. Desviación estándar (s):

$$s = \sqrt{1.83} \approx 1.35$$

Las notas tienen un rango de 4 puntos. La desviación estándar de ≈ 1.35 indica que, en promedio, la nota de un estudiante se desvía de la media (7.5) en aproximadamente 1.35 puntos.



■ El rol de la desviación estándar. Mientras mayor sea la desviación estándar (s), más ancha es la curva y más dispersos están los datos alrededor de la media. Una ' s ' pequeña indica datos muy agrupados



Refuerza tus conocimientos en las medidas de tendencia central y dispersión a través del video del código QR.





Práctica de aprendizaje



1. Calcula el rango, varianza y desviación estándar de las horas de sueño de la práctica de aprendizaje 2 (sin el dato 12): {6, 7, 8, 9, 7, 6.5, 10, 7}. Media ≈ 7.56 .

2. Compara la desviación estándar de ese conjunto con la del conjunto {6, 7, 7, 7.5, 8, 8, 8.5, 9} (Media=7.63). ¿Cuál es más homogéneo?

3. ¿Por qué en el cálculo de la varianza muestral se divide entre $(n-1)$ y no entre (n) ?

Representaciones Gráficas

Histograma

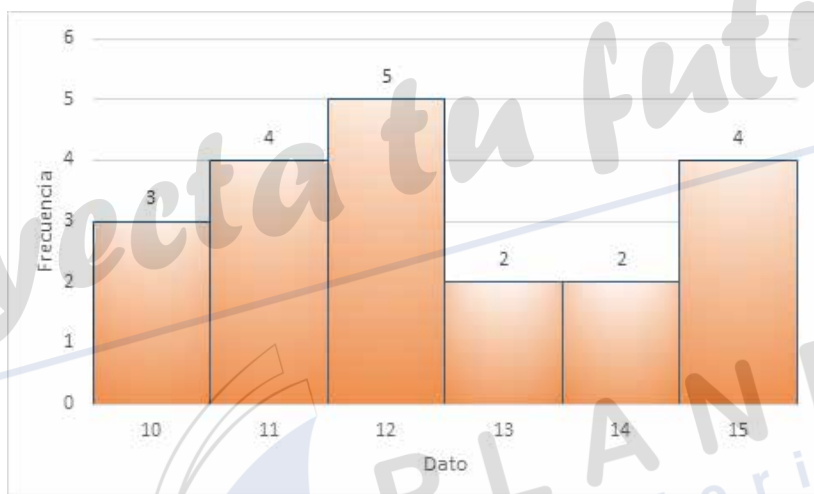
Un histograma es una gráfica de barras apiladas para representar variables cuantitativas, esta es la principal diferencia con respecto a las gráficas de barras del tema anterior, para realizar la gráfica, se coloca en el eje horizontal los datos (x) y en el eje vertical la frecuencia F , para ejemplificar el proceso para su creación, retomemos el ejemplo de la rúbrica del profesor de química que tiene las siguientes dos columnas de la tabla de distribución de frecuencia

- Se colocan los rótulos sobre los ejes

x	F
10	3
11	4
12	5
13	2
14	2
15	4
n	20



- Se dibujan las barras apiladas (sin dejar espacio entre ellas), en cada uno de los datos y de esta manera se obtiene el histograma.



Observa el video del código QR, donde se muestra paso a paso como se puede construir un histograma con una hoja de cálculo



Diagrama de cajas y bigotes

El diagrama de caja (boxplot) es un resumen gráfico de 5 números que ofrece una vista clara de la dispersión y simetría de los datos, destacando valores extremos.

Los 5 números son:

1. Mínimo. Valor más bajo (excluyendo outliers).
2. Primer Cuartil (Q1). Mediana del 50% inferior de los datos.
3. Mediana (Q2). Valor central.
4. Tercer Cuartil (Q3). Mediana del 50% superior.
5. Máximo. Valor más alto (excluyendo outliers).

- Rango Intercuartílico (IQR):

$$\text{IQR} = \text{Q3} - \text{Q1}.$$

Es el ancho de la caja.

- Bigotes:

Se extienden desde los bordes de la caja hasta el valor mínimo y máximo dentro del límite.

$$\text{Límite inferior} = \text{Q1} - 1.5 \times \text{IQR}$$

$$\text{Límite superior} = \text{Q3} + 1.5 \times \text{IQR}$$

- Valores Atípicos (Outliers):

Puntos individuales que caen fuera de los límites de los bigotes.

Para comprender la construcción de diagrama de caja analiza el siguiente ejemplo.

Un profesor de educación física quiere analizar el rendimiento de dos grupos de estudiantes en una prueba de flexiones en 1 minuto. Los datos obtenidos son:

- Grupo A (10 estudiantes): 22, 24, 25, 26, 27, 28, 29, 30, 31, 35
- Grupo B (12 estudiantes): 18, 20, 22, 24, 25, 26, 26, 27, 28, 30, 33, 40

Paso 1: ordenar los datos de menor a mayor

- Grupo A (ya ordenados):

22, 24, 25, 26, 27, 28, 29, 30, 31, 35

- Grupo B (ordenados):

18, 20, 22, 24, 25, 26, 26, 27, 28, 30, 33, 40

Paso 2: calcular los 5 números resumen para cada grupo

Para el Grupo A (n=10):

- Mínimo: 22
- Primer Cuartil (Q1): Posición de Q1 = $0.25 \times (n+1) = 0.25 \times 11 = 2.75$

Esto significa que Q1 está al 75% del camino entre el 2º y 3º dato: 2º dato = 24, 3º dato = 25

$$Q1 = 24 + 0.75(25-24) = 24 + 0.75 = 24.75$$

- Mediana (Q2):

Posición de Q2 = $0.50 \times (n+1) = 0.50 \times 11 = 5.5$

Entre el 5º y 6º, dato: 27 y 28

$$Q2 = \frac{27 + 28}{2} = 27.5$$

- Tercer Cuartil (Q3):

Posición de Q3 = $0.75 \times (n+1) = 0.75 \times 11 = 8.25$

Entre el 8º y 9º, dato: 30 y 31

$$Q3 = 30 + 0.25(31-30) = 30 + 0.25 = 30.25$$

- Máximo: 35

5 números del grupo A: {22, 24.75, 27.5, 30.25, 35}

Para el Grupo B (n=12):

- Mínimo: 18
- Primer Cuartil (Q1): Posición de Q1 = $0.25 \times (n+1) = 0.25 \times 13 = 3.25$

Entre el 3º y 4º dato: 22 y 24

$$Q1 = 22 + 0.25(24-22) = 22 + 0.5 = 22.5$$

- Mediana (Q2):

$$\text{Posición de Q2} = 0.50 \times (n+1) = 0.50 \times 13 = 6.5$$

Entre el 6º y 7º, dato: 26 y 26

$$Q2 = \frac{26 + 26}{2} = 26$$

- Tercer Cuartil (Q3):

$$\text{Posición de Q3} = 0.75 \times (n+1) = 0.75 \times 13 = 9.75$$

Entre el 9º y 10º dato: 28 y 30

$$Q3 = 28 + 0.75(30-28) = 28 + 1.5 = 29.5$$

- Máximo: 40

5 números del grupo B: {18, 22.5, 26, 29.5, 40}

Paso 3: calcular el rango intercuartílico (IQR) y límites para valores atípicos

Grupo A:

$$IQR = Q3 - Q1 = 30.25 - 24.75 = 5.5$$

- Límite inferior = $Q1 - 1.5 \times IQR = 24.75 - 1.5 \times 5.5 = 24.75 - 8.25 = 16.5$
- Límite superior = $Q3 + 1.5 \times IQR = 30.25 + 1.5 \times 5.5 = 30.25 + 8.25 = 38.5$

No hay valores atípicos (todos los datos están entre 16.5 y 38.5)

Grupo B:

$$IQR = Q3 - Q1 = 29.5 - 22.5 = 7$$

- Límite inferior = $22.5 - 1.5 \times 7 = 22.5 - 10.5 = 12$
- Límite superior = $29.5 + 1.5 \times 7 = 29.5 + 10.5 = 40$

Se verifica que no existan valores atípicos:

El valor 40 está justo en el límite superior ($40 = 40$), por lo que no se considera atípico

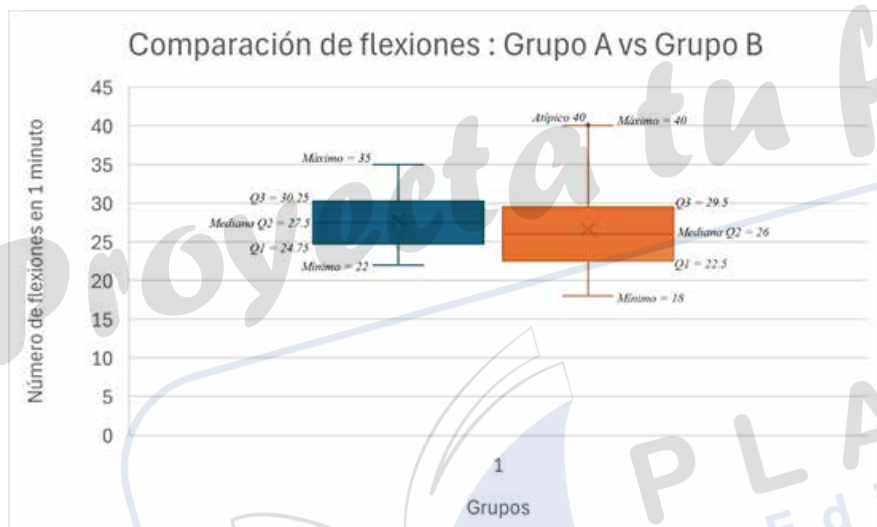
Todos los datos están entre 12 y 40

Paso 4: construir el diagrama de caja manualmente

Para la construcción del diagrama se realizan los siguientes pasos:

1. Para cada grupo, se dibuja una línea desde el mínimo hasta el máximo
2. Dibujar una caja rectangular desde Q1 hasta Q3
3. Se marca la mediana con una línea dentro de la caja

4. Los bigotes van desde los bordes de la caja hasta el mínimo y máximo (excluyendo outliers)
5. Si hubiera valores atípicos, los marcaríamos con puntos (•) o asteriscos (*)



■ Diagrama de cajas y bigotes

Paso 5: crear el diagrama con tecnología (Excel/Google Sheets)

Instrucciones para crear el diagrama de cajas con bigotes en la hoja de cálculo.

1. Introducir los datos:

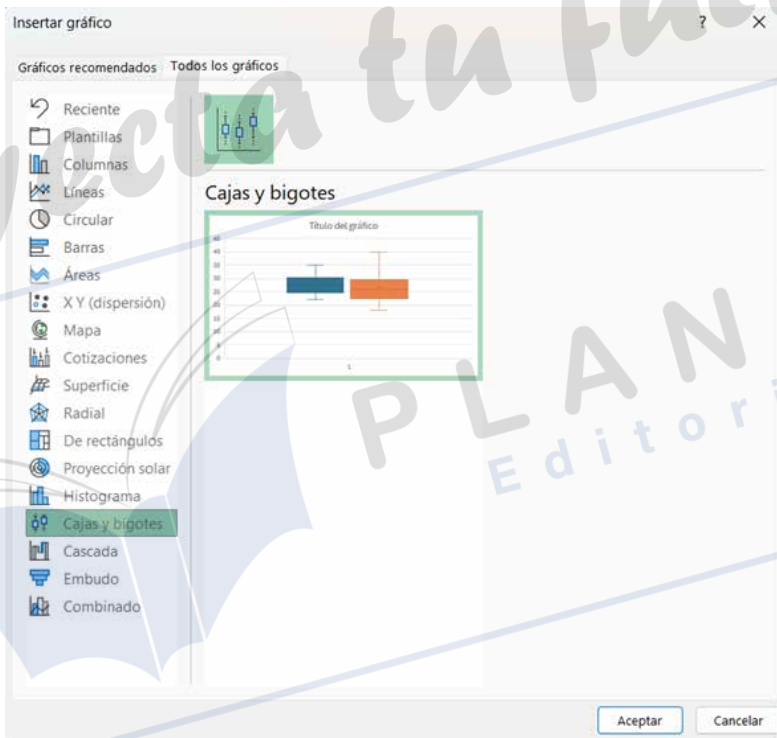
	A	B
1	Grupo A	Grupo B
2		22 18
3		24 20
4		25 22
5		26 24
6		27 25
7		28 26
8		29 26
9		30 27
10		31 28
11		35 30
12		33
13		40

2. Seleccionar todos los datos (A1:B12)

Prohibida su reproducción

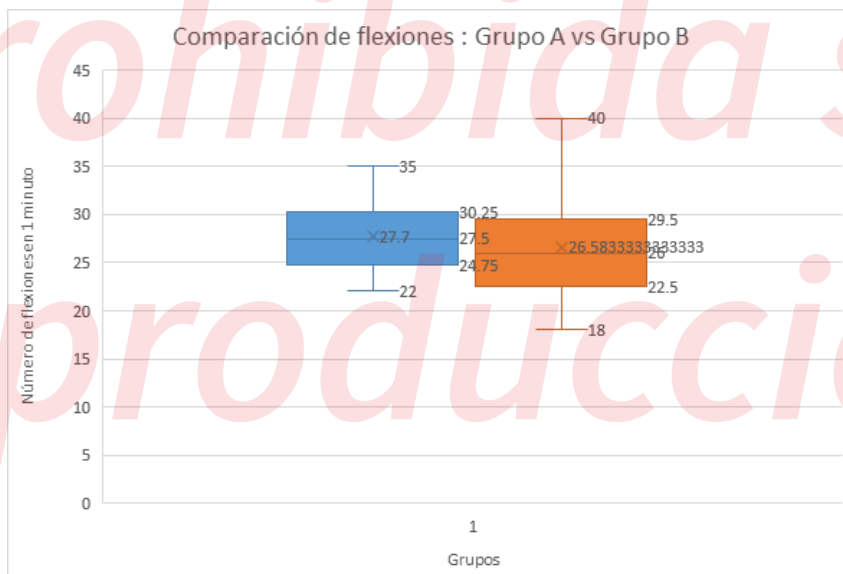
3. Insertar gráfico:

- Ir a la ficha "Insertar" → "Gráfico"
- Seleccionar "Diagrama de caja y bigotes"



4. Personalizar el gráfico:

- Título: "Comparación de Flexiones: Grupo A vs Grupo B"
- Eje Y: "Número de flexiones en 1 minuto"
- Eje X: "Grupos"
- Colores diferentes para cada grupo



Paso 6: análisis e interpretación del diagrama

Comparación visual:

Simetría:

- Grupo A: Casi simétrico (mediana \approx centro de la caja, bigotes de longitud similar)
- Grupo B: Ligeramente asimetría a la derecha (bigote derecho más largo que el izquierdo)

Dispersión:

- Grupo A: IQR = 5.5, Rango = 13 (35-22)
- Grupo B: IQR = 7, Rango = 22 (40-18)
- Conclusión: El Grupo B tiene mayor variabilidad en el rendimiento

Posición:

- Mediana del Grupo A (27.5) > Mediana del Grupo B (26)
- El Grupo A tiene mejor rendimiento central

Valores atípicos: Ningún grupo tiene valores atípicos claros

Paso 7: conclusión contextual**"El diagrama de caja revela que:**

1. El **Grupo A** tiene un rendimiento ligeramente superior (mediana más alta) y más consistente (menor dispersión).
2. El **Grupo B** muestra mayor variabilidad, con algunos estudiantes con rendimiento muy bajo (18 flexiones) y otros muy alto (40 flexiones).
3. La distribución del **Grupo B** es ligeramente asimétrica hacia la derecha, indicando que hay unos pocos estudiantes con rendimiento excepcionalmente bueno.
4. Ambos grupos tienen distribuciones razonablemente simétricas, lo que sugiere que los datos podrían aproximarse a una distribución normal."

Prohibida su reproducción





Construye el diagrama de caja para los datos de pulsaciones por minuto de 8 corredores después de un ejercicio intenso. Los valores son:

- 150, 155, 158, 160, 162, 165, 170, 185



PLANEA
Editorial

*Prohibida su
reproducción*

Integración y verificación de normalidad

La distribución normal es un modelo teórico en forma de campana simétrica (Gaussiana). Muchos fenómenos naturales y sociales se aproximan a ella. Podemos verificar la normalidad de forma visual (histograma, diagrama de caja simétrico) y numérica (relación entre media, mediana y moda; Regla Empírica).

Regla empírica (para distribuciones aproximadamente normales):

- ~68% de los datos está dentro de 1 desviación estándar de la media ($\bar{x} \pm s$).
- ~95% de los datos está dentro de 2 desviaciones estándar de la media ($\bar{x} \pm 2s$).
- ~99.7% de los datos está dentro de 3 desviaciones estándar de la media ($\bar{x} \pm 3s$).

Analiza el siguiente ejemplo de la aplicación de este concepto.

Un profesor quiere comparar el rendimiento en dos pruebas de diferente dificultad (Prueba A "Fácil" y Prueba B "Difícil") y ver si los resultados son normales.

Datos (sobre 100):

- **Prueba A:** Media=78, $s=5$, Histograma simétrico, Mediana=78.
- **Prueba B:** Media=65, $s=12$, Histograma con cola a la izquierda, Mediana=70.

Análisis comparativo:

1. Comparación de centros. El promedio fue mayor en la Prueba A (78 vs 65).
2. Comparación de dispersión. La Prueba B fue mucho más variable ($s=12$ vs $s=5$). Los resultados fueron más impredecibles.
3. Identificación de patrones/tendencias:
 - **Prueba A.** La coincidencia de Media \approx Mediana y la forma simétrica del histograma sugieren una distribución aproximadamente normal. Aplicando la regla empírica, esperaríamos que ~95% de estudiantes tenga entre $78 \pm (2 \cdot 5) =$ entre 68 y 88 puntos.
 - **Prueba B.** La Media (65) < Mediana (70) indica una asimetría negativa (sesgo a la izquierda). El histograma con cola a la izquierda lo confirma. Esto significa que hubo un grupo de notas muy bajas que "arrastraron" la media hacia abajo. No se distribuye normalmente.
4. Conclusión contextual:

La **Prueba A** discriminó de forma equilibrada. La **Prueba B** fue difícil para la mayoría (notas bajas), pero un grupo considerable logró rendir bien (notas alrededor de 70 o más), como muestra la mediana. El patrón no es normal, lo que podría indicar que la prueba tenía preguntas de dificultad muy desigual.



Práctica de aprendizaje



1. Tienes los pesos de perros de dos razas. Raza X: \bar{x} = 20 kg, s = 2 kg, mediana = 20 kg. Raza Y: \bar{x} = 25 kg, s = 6 kg, mediana = 23 kg. Describe las diferencias en centro, dispersión y forma

Handwriting lines for question 1

2. Para una variable con \bar{x} = 100 y s = 15, ¿entre qué valores esperarías encontrar al 68% central de los datos si es normal?

Handwriting lines for question 2

3. Un histograma muestra una barra muy alta a la izquierda y decrece rápidamente. El diagrama de caja tiene la mediana cerca de Q3. ¿Es normal? Describe la asimetría

Handwriting lines for question 3

4. Usando una hoja de cálculo, ingresa 30 datos simulados, por ejemplo, edades de familiares y genera su histograma y medidas. Describe lo que encuentres.

Handwriting lines for question 4

5. En un examen normal (\bar{x} = 70, s = 8), ¿aproximadamente qué porcentaje de estudiantes reprobaría (nota < 60)?

Handwriting lines for question 5



Estudio independiente

Responde las siguientes preguntas.

1. ¿Qué significa hacer un análisis exploratorio de datos no agrupados?

2. ¿Qué son las medidas de tendencia central y para qué sirven?

3. ¿Qué son las medidas de dispersión y qué información aportan?

4. ¿Qué representan los histogramas y los diagramas de caja y bigotes?

Prohibida su reproducción



Estudio independiente

5. ¿Cómo se puede verificar si los datos se distribuyen normalmente?

6. ¿Qué recursos tecnológicos pueden ayudar en este análisis?

Criterios	Nivel Avanzado (3 pts.)	Nivel Intermedio (2 pts.)	Nivel Básico (1 pt.)
Reconozco qué es el análisis exploratorio de datos	Analizo cómo se usan medidas y gráficos para interpretar datos no agrupados.	Explico que es buscar patrones en datos individuales.	Digo que es revisar datos.
Identifico medidas de tendencia central, dispersión y sus representaciones gráficas	Reconozco cómo estas medidas y gráficos permiten comparar, interpretar y verificar distribuciones.	Explico qué son la media, mediana, moda, rango y desviación.	Digo que muestran lo común y lo diferente.
Relaciono el análisis de datos con problemas reales y el uso de tecnología	Reflexiono sobre cómo aplicar el análisis estadístico y tecnológico para resolver problemas en distintos contextos.	Explico que ayudan a entender datos y tomar decisiones.	Digo que se usan gráficos.

Revisa tu desempeño:

9 puntos - Excelente.

De 6 a 8 puntos - Bien.

De 4 a 5 puntos - Suficiente.

3 puntos - Insuficiente.

Práctica transversal



El uso de las hojas de cálculo para el procesamiento y análisis de información dentro de la estadística es una herramienta poderosa, en la práctica transversal de esta unidad vas a reforzar tus conocimientos en el uso de la hoja de cálculo que desarrollaste en Cultura Digital, para el cálculo de medidas de tendencia central y dispersión.

El primer paso es observar a detalle el video del código QR, donde se explica el uso de la hoja de cálculo para realizar cálculos estadísticos.



Medidas de tendencia central y dispersión en un sólo paso | Estadística en Excel

Ahora que cuentas con las habilidades necesarias resuelve el siguiente planteamiento.

Se desea realizar el análisis de información de un grupo de la masa de 20 estudiantes de nivel medio superior, los resultados son:

51	55	55	53
61	62	57	62
65	55	59	61
65	54	57	54
64	54	51	54

Por medio de una hoja de cálculo realiza el cálculo de las medidas de tendencia central y dispersión de los datos anterior, así mismo, incluye un análisis de la información de acuerdo con los resultados obtenidos.

Envía por correo electrónico el documento a tu maestra(o) para su evaluación.





Taller de probabilidad y estadística

2



La Editorial Planea tiene como misión crear materiales didácticos de calidad, con los contenidos adecuados para impactar positivamente en la formación de los estudiantes, desarrollando sus conocimientos, habilidades y actitudes, que los transformen en jóvenes capaces de comprender su entorno e influir en él, aprender de manera autónoma a largo de su vida, ser consciente de sus destrezas para resolver problemas y aceptar retos que lo ayuden a alcanzar su metas, ser sensibles al arte y sus expresiones, asimismo activar la participación ciudadana que reafirme su conciencia cívica y ética, fomentando una actitud respetuosa a la interculturalidad, diversidad de creencias, valores e ideas, asumiendo un pensamiento crítico que ayude al desarrollo sustentable de su comunidad.

El libro de **Taller de Probabilidad y Estadística II**, está desarrollado bajo los Principios de la Nueva Escuela Mexicana, teniendo como eje rector el Nuevo Modelo Educativo de la Educación Media Superior y el programa de estudio por progresiones para el **Bachillerato General**, el cual propone los siguientes aprendizajes trayectoria de este **Recurso Sociocognitivo**:

- Valora la aplicación de procedimientos automáticos y algorítmicos, así como la interpretación de sus resultados para anticipar, encontrar y validar soluciones a problemas matemáticos, de áreas del conocimiento y de su vida personal.
- Adopta procesos de razonamiento matemático tanto intuitivos como formales tales como observar, intuir, conjeturar y argumentar, para relacionar información y obtener conclusiones de problemas (matemáticos, de las ciencias naturales, experimentales y tecnología, sociales, humanidades y de la vida cotidiana).
- Modela y propone soluciones a problemas tanto teóricos como de su entorno, empleando lenguaje y técnicas matemáticas.
- Explica el planteamiento de posibles soluciones a problemas y la descripción de situaciones en el contexto que les dio origen empleando lenguaje matemático y lo comunica a sus pares para analizar su pertinencia.

En la Editorial Planea tenemos un compromiso por desarrollar materiales que cumplan con las expectativas de las comunidades educativas.

Titulos relacionados

Pensamiento matemático 2
Andrés Herrera Lozano



Conciencia histórica México durante la expansión capitalista 2
Sofía Ivonne Noriega



Módulo 1 Administración Cultura emprendedora
Daniel Salazar Aguilar



771-159-1900

www.editorialplanea.com.mx