Universidad Zaragoza

1542

Trabajo Fin de Máster

Estadística en 4º de la ESO para Matemáticas Aplicadas

Statistics in the 4th year of ESO for Applied Mathematics

Autor

Asier Ruiz de Alegría Llop

Directora

Carmen Julve Tiestos

Facultad de Educación

2023

Índice

Definición del objeto matemático a enseñar

La estadística es un cuerpo matemático de la ciencia que pertenece a la recopilación, análisis, interpretación o explicación y presentación de datos. Algunos consideran que la estadística es una ciencia distinta en lugar de una rama de las matemáticas. Si bien muchas investigaciones científicas hacen uso de datos, la estadística se ocupa del uso de datos en el contexto de la incertidumbre y la toma de decisiones frente a la incertidumbre (Gordon, 1992, 14-25).

Al aplicar la estadística a un problema, es una práctica común comenzar con una población o proceso a estudiar. Las poblaciones pueden ser temas diversos como "todas las personas que viven en un país". Idealmente, los estadísticos compilan datos sobre toda la población. Esto puede ser organizado por institutos de estadística gubernamentales. Las estadísticas descriptivas se pueden utilizar para resumir los datos de la población. Los descriptores numéricos incluyen la media y la desviación estándar para datos continuos (como ingresos, la estatura de una población), mientras que la frecuencia y el porcentaje son más útiles para describir datos categóricos (como educación, color de ojos, partido político).

Cuando reunir datos de toda la población no es factible, se estudia un subconjunto elegido de la población llamado muestra . Una vez que se determina una muestra representativa de la población, se recopilan datos para los miembros de la muestra en un entorno de observación o experimental. Nuevamente, las estadísticas descriptivas se pueden usar para resumir los datos de la muestra. Sin embargo, extraer la muestra contiene un elemento de aleatoriedad; por lo tanto, los descriptores numéricos de la muestra también son propensos a la incertidumbre. Para sacar conclusiones significativas sobre toda la población, se necesitan estadísticas inferenciales que utilizan patrones en los datos de la muestra para sacar inferencias sobre la población representada mientras se tiene en cuenta la aleatoriedad. Estas inferencias pueden tomar la forma de responder preguntas de sí/no sobre los datos (pruebas de hipótesis), estimación de características numéricas de los datos, descripción de asociaciones de correlación dentro de los datos y modelado de relaciones dentro de los datos, por ejemplo, mediante análisis de regresión. La inferencia puede extenderse al pronóstico , la predicción y la estimación de valores no observados en la población que se está estudiando o asociados con ella.

Campo de problemas, técnicas y tecnologías asociadas al objeto matemático

La enseñanza de la estadística debe incluir campos de problemas, técnicas y tecnologías que estén relacionados con la modelización matemática. Estos componentes son esenciales para una adecuada enseñanza de la estadística en el aula.

Los campos de problemas a explorar en en 4º de la ESO para matemáticas orientadas a las enseñanzas aplicadas, como luego reflejaremos en la secuencia didáctica son: Muestreo, tablas de datos y frecuencias. Organización visual de los datos estadísticos. Medidas de centralización y de posición. Medidas de dispersión. Estadística bidimensional. A continuación los asociamos con sus técnicas y tecnologías.

CP 1. MUESTREO, TABLAS DE DATOS Y FRECUENCIAS

El **campo de problemas** del muestreo es una técnica utilizada en estadística para analizar una población a través del estudio de una muestra representativa de esa población. Al analizar una muestra en lugar de la población completa se pueden obtener resultados más rápidos y con menos esfuerzo. Además, al trabajar con una muestra se pueden hacer inferencias sobre la población completa.

Como **técnica**, una vez que se obtiene la muestra, se construyen tablas de datos y tablas de frecuencias para organizar y analizar los datos. Una tabla de datos es una representación gráfica de los datos, mientras que una tabla de frecuencias es una tabla que muestra la cantidad de veces que aparece cada valor en la muestra. Las tablas de frecuencias pueden ser absolutas o relativas, y pueden ser acumuladas o no acumuladas.

Como **tecnologías**, para poder llevar a cabo el muestreo es necesario definir la población y la muestra. La población es el conjunto completo de individuos o elementos que tienen algo en común, y que será objeto de estudio, mientras que la muestra es un subconjunto de la población elegida de manera aleatoria. La parte de la estadística que se encarga de seleccionar de forma adecuada la muestra se denomina, Teoría de muestras. La muestra debe ser representativa de la población, es decir, debe tener las mismas características que la población en su conjunto. La variable estadística es la característica que se mide u observa en la población o en la muestra. Una variable puede ser cuantitativa si toma valores numéricos o cualitativa si los valores que toma no lo son. Las variables cuantitativas que toman valores aislados se denominan variables discretas y las que pueden tomar cualquier valor de un intervalo de la recta real, variables continuas. Las frecuencias absolutas son el número de veces que aparece cada valor en la muestra, mientras que las frecuencias relativas son las frecuencias expresadas como porcentajes o proporciones. Las frecuencias acumuladas son el total de las frecuencias anteriores más la frecuencia actual.

Para llevar a cabo el muestreo y el análisis de datos, se pueden utilizar diversas tecnologías, como programas de análisis estadístico, como el conocido SPSS (Statistical Package for the Social Sciences). SPSS ofrece una amplia gama de herramientas y funciones estadísticas que facilitan la manipulación y el procesamiento de datos, lo cual resulta especialmente útil en investigaciones de gran envergadura o proyectos de análisis complejos. Aunque este programa es ampliamente utilizado en investigaciones académicas y profesionales, no se utilizará para la enseñanza de 4º de la ESO, ya que su nivel de complejidad y funcionalidades avanzadas podrían resultar innecesarias y abrumadoras para los estudiantes en esta etapa educativa. En su lugar, las hojas de cálculo, como Microsoft Excel o Google Sheets, ofrecen una opción más accesible, proporcionan funciones básicas de análisis y permiten organizar los datos de manera eficiente. Aunque su capacidad analítica es más limitada en comparación con programas especializados, las hojas de cálculo siguen siendo una herramienta ampliamente utilizada, especialmente en entornos más simples o cuando no se requieren técnicas estadísticas avanzadas.

Más allá de la elección del software utilizado, es fundamental considerar los principios éticos en el manejo de los datos. Esto implica **respetar la privacidad y la confidencialidad de los individuos o empresas involucradas y así se lo tenemos que comunicar a nuestros alumnos**. Por ejemplo, cuando se recopilan datos sensibles relacionados con la salud o la política, es crucial proteger la identidad de las personas y garantizar que no se revelen datos que puedan ser utilizados de manera perjudicial o discriminatoria. Esto implica garantizar el anonimato de los participantes, la confidencialidad de los datos recopilados y el uso responsable de los resultados obtenidos. En el caso de la Unión Europea, por ejemplo, el Reglamento General de Protección de Datos (RGPD) establece directrices estrictas sobre cómo las organizaciones deben recoger, procesar y almacenar datos personales. La ley exige garantizar la privacidad, el consentimiento informado y la seguridad en el manejo de los datos, así como permitir a los individuos ejercer sus derechos, como el acceso, la rectificación y la eliminación de sus datos personales.

CP 2. ORGANIZACIÓN VISUAL DE LOS DATOS ESTADÍSTICOS

El **campo de problemas** de la organización visual de los datos estadísticos se refiere a la presentación de los datos de una manera que sea fácil de entender y analizar.

Como **técnica**, esto incluye la elaboración de gráficos y diagramas que muestran las tendencias y patrones en los datos. Las técnicas utilizadas para este campo de problemas incluyen la elaboración de diferentes tipos de gráficos, como diagramas de barras, polígonos de frecuencias, diagramas de sectores, diagramas lineales e histogramas.

Como **tecnologías** diferenciaremos si se trata de una variable discreta o continua. En el caso de variables discretas utilizaremos el Diagrama de barras, que es un gráfico que se utiliza para comparar diferentes categorías o grupos.
Las barras se colocan verticalmente o horizontalmente y se utilizan para representar diferentes valores o cantidades. Polígono de frecuencias, que es un gráfico que se utiliza para mostrar la distribución de los datos en un conjunto de datos. Puede ser acumulado o no acumulado. Diagrama de sectores, que es un gráfico circular que se utiliza para mostrar la relación entre diferentes partes o categorías. Diagrama lineal, que es un gráfico que muestra la relación entre dos variables y se utiliza a menudo para mostrar tendencias o cambios a lo largo del tiempo. El histograma es la representación más apropiada si hablamos de variables

continuas. Se trata de diagrama de rectángulos en el que el área de cada rectángulo es proporcional a la frecuencia, de esta forma, la frecuencia de cada suceso viene representada por el área.

CP 3. MEDIDAS DE CENTRALIZACIÓN Y DE POSICIÓN

El **campo de problemas** de las medidas de centralización y de posición se utilizan para resumir y describir un conjunto de datos.

Las **técnicas** serían el cálculo de las medidas incluyendo la moda, la media, la mediana y los cuartiles.

Las **tecnologías** serían las definiciones de: Moda como el valor que más se repite en un conjunto de datos. Media, como el valor promedio de un conjunto de datos. Mediana, como el valor que ocupa el lugar central en un conjunto de datos ordenados. Cuartiles, como los valores que dividen un conjunto de datos ordenados en cuatro partes iguales. Así, el primer cuartil (Q1) es el valor de la variable que deja menores o iguales que él a la cuarta parte de los datos, es decir, un 25 % y por tanto el 75% serán mayores o iguales que él. La mediana es el segundo cuartil, que deja por debajo la mitad de los datos o un 50%. Finalmente, tercer cuartil (Q3) es el valor de la variable que deja menores o iguales que él a un 75% de los datos y mayores o iguales la cuarta parte..

CP 4. MEDIDAS DE DISPERSIÓN

El **campo de problemas** de las medidas de dispersión se utilizan para medir la variabilidad o dispersión de un conjunto de datos.

Como **técnicas** sería el cálculo de algunas medidas comunes de dispersión incluyen el recorrido o rango, el recorrido intercuartílico, la varianza, la desviación típica, el coeficiente de variación y la distribución, tanto la normal como al menos conocer la existencia otras (uniforme, binomial, poisson, t-student....)

Como **tecnologías** sería la definición del recorrido o rango como la diferencia entre el valor máximo y el valor mínimo de un conjunto de datos, del recorrido intercuartílico como la diferencia entre el tercer cuartil y el primer cuartil de un conjunto de datos ordenados, de la varianza como medida de la dispersión de un conjunto de datos en torno a la media, de la desviación típica como la raíz cuadrada de la varianza. Es una medida más fácil de interpretar que la varianza, ya que tiene las mismas unidades que los datos originales, el coeficiente de variación es la desviación típica dividida por la media, multiplicada por 100, que se utiliza para comparar la dispersión de diferentes conjuntos de datos que tienen diferentes unidades o escalas y la distribución de frecuencias, que es una representación que muestra la cantidad de veces que cada valor se presenta en un conjunto de datos.

Además, tenemos el concepto de asimetría estadística, la cual nos da información sobre la falta de simetría en la distribución de los datos. Una distribución perfectamente simétrica tendría una asimetría de cero. Si es positiva, la distribución tiene una cola larga a la derecha; si es negativa, la cola es larga a la izquierda. En la misma línea, la curtosis mide el apuntamiento de la distribución. Una curtosis positiva indica una distribución más apuntada (con colas más pesadas) que una distribución normal, mientras que una negativa sugiere una distribución más plana.

CP 5. ESTADÍSTICA BIDIMENSIONAL

El **campo de problemas** de las distribuciones bidimensionales se utilizan para analizar y entender la relación entre dos variables.

Algunas **técnicas** comunes utilizadas en este campo incluyen el calcular diferentes parámetros como la media la covarianza, el coeficiente de correlación y la recta de regresión lineal (y al menos conocer que existen otras no lineales) a partir de tablas de frecuencias y de contingencia, diagramas de dispersión y nubes de puntos. Las tablas de frecuencias y de contingencia son herramientas útiles para analizar la relación entre dos variables cualitativas.

Como **tecnología** se incluiría las definiciones de: tabla de frecuencias como muestra la frecuencia con la que aparece cada valor de una variable, mientras que una tabla de contingencia muestra la frecuencia con la que aparecen diferentes combinaciones de valores entre dos variables, un diagrama de dispersión como un gráfico que muestra la relación entre dos variables numéricas. Se utiliza para identificar patrones y tendencias en los datos, una nube de puntos como otro gráfico que muestra la relación entre dos variables numéricas. Cada punto en el gráfico representa una observación diferente en el conjunto de datos, la media como una medida de la tendencia central de los datos en una distribución bidimensional, la covarianza como una medida de la relación o ausencia de relación lineal entre dos variables, el coeficiente de correlación de Pearson como una medida de la fuerza y la dirección de la relación lineal entre dos variables y la recta de regresión lineal es una herramienta para ajustar una línea a un conjunto de puntos de datos y predecir valores futuros. También es importante que conozcan que si la relación entre dos variables es cuadrática (o polinómica de mayor órden) o logarítmica, o exponencial, etc., podríamos tener un coeficiente de correlación cercano a 0, pero esto puede significar que no hay relación(lineal) entre estas variables, pero puede haber otros tipos de correlación.

Estado de la enseñanza y aprendizaje de la estadística

La estadística es una materia que se enseña en la Educación Secundaria Obligatoria (ESO) en España. En concreto, forma parte del área de Conocimiento del Medio en educación primaria y de la asignatura de Matemáticas, que es obligatoria en todos los ciclos de la ESO.

En Aragón, con la entrada en vigor de la LOMLOE, se ha incluido una nueva optativa en 4º de la ESO, "Matemáticas para la toma de decisiones", cuyos saberes básicos se distribuyen en tres bloques: aritmética modular, teoría de grafos y la teoría de juegos. Precisamente, la presencia del azar en muchos de los juegos cotidianos para el alumnado hace que este contexto permite abordar contenidos o saberes propios del sentido estocástico vinculados a la inferencia y la incertidumbre en el juego.

El sentido estocástico se refiere a la habilidad para comprender y manejar situaciones aleatorias, inciertas o variables, lo que es especialmente útil en la inferencia estadística y la toma de decisiones. La aritmética modular, por ejemplo, puede utilizarse en criptografía y seguridad informática, donde la incertidumbre y el azar son factores clave. La teoría de grafos, por otro lado, ayuda a modelar y analizar relaciones y conexiones en sistemas complejos, lo que puede contribuir a tomar decisiones informadas en situaciones inciertas.

La teoría de juegos es un enfoque matemático para analizar situaciones de conflicto y cooperación entre individuos o grupos, que también implica incertidumbre y toma de decisiones. Esta teoría permite a los estudiantes entender cómo las partes involucradas pueden tomar decisiones óptimas en función de la información disponible, las acciones de los demás y la incertidumbre asociada.

En la ESO, la estadística se centra en el estudio de las medidas de tendencia central y de dispersión, las gráficas y los diagramas de frecuencias, los test de hipótesis y los intervalos de confianza. También se abordan otros temas como la regresión y el análisis de varianza.

Es importante mencionar que la enseñanza de la estadística en la ESO no se debe limitar a la adquisición de conocimientos teóricos, sino que se debe hacer hincapié en el desarrollo de habilidades prácticas como el análisis de datos y la toma de decisiones basadas en ellos. Además, se fomenta la comprensión de la importancia de la estadística en la vida cotidiana y en distintos ámbitos profesionales.

Autores como Ben-Zvi y Garfield (2004), comentan que algunos de estos desafíos incluyen la complejidad de algunas ideas y reglas estadísticas, la dificultad que tienen muchos estudiantes con las matemáticas subyacentes (como fracciones, decimales y fórmulas algebraicas), el contexto en muchos problemas estadísticos que puede confundir a los estudiantes y hacerles confiar en sus experiencias y sus intuitivos a menudo equivocados para producir una respuesta en lugar de seleccionar un procedimiento estadístico adecuado.

Batanero y Godino (2002) han identificado varios conflictos que dificultan a los estudiantes la comprensión de conceptos estadísticos básicos. **Estos conflictos incluyen la interpretación de tablas y gráficos, el conocimiento de las medidas de posición central, la comprensión de la variabilidad y las características de dispersión.**

Comprensión de tablas y gráficos estadísticos.

Los profesores a veces asumen que el proceso de creación de tablas y gráficos es sencillo y no dedican mucho tiempo a su enseñanza. Sin embargo, crear una tabla de frecuencias o un gráfico implica ya una primera reducción estadística, ya que se pierden los valores originales de cada dato individual al pasar a la distribución de frecuencias. Esto puede ser difícil de comprender para los estudiantes, ya que ellos entienden mejor las propiedades que se refieren a individuos individuales en lugar de al conjunto de datos. La habilidad para leer críticamente datos es importante en nuestra sociedad tecnológica, ya que encontramos tablas y gráficos en la prensa, el comercio y en diferentes materias escolares. Hay cuatro niveles de comprensión de gráficos: lectura literal, interpretación de datos, inferencias y valoración de datos. El objetivo de la educación estadística es llevar a cada estudiante al mayor nivel al que esté capacitado. Existen varios factores que afectan la comprensión de los gráficos y su dificultad y que deben tenerse en cuenta por los profesores, como el conocimiento previo del tema, el conocimiento previo de los conceptos matemáticos y el conocimiento previo del tipo de gráfico. Además, la claridad y precisión de la presentación de los datos también son importantes. Que los ordenen como quieran para que vean diferentes formatos (con palitos, por ejemplo) sólo posteriormente institucionalizar la representación en tablas.

Medidas de posición central

La media es uno de los conceptos estadísticos más importantes y tiene muchas aplicaciones prácticas en la vida diaria. Aunque puede parecer sencillo, calcular la media a veces puede ser complicado, como en el ejemplo de las 10 personas en un ascensor con diferentes pesos promedio. La media ponderada es a menudo necesaria en situaciones como calcular la puntuación promedio en un curso, la velocidad promedio o el índice de precios. También es importante tener en cuenta la frecuencia de cada valor al calcular la media a partir de una tabla de datos. También se pueden cometer errores al calcular la media, mediana y moda. Algunos de los errores más comunes incluyen tomar la mayor frecuencia absoluta en lugar del valor de la variable al calcular la moda, no ordenar los datos para calcular la mediana y calcular la moda en lugar de la mediana. Además, a veces se realiza el cálculo correctamente, pero no se entiende el proceso de cálculo. Muchas veces se utiliza la media para describir una población, por ejemplo, la nota media de una clase, pero no la describe completamente, hay que añadir características de dispersión. (Batanero, 2000, 41-58)

Características de dispersión

El estudio de una distribución de frecuencias no se puede limitar a sus promedios, ya que distribuciones con medias o medianas iguales pueden tener diferentes niveles de variabilidad. Es común ignorar la dispersión de los datos al comparar dos o más muestras o poblaciones. La desviación típica mide la intensidad con la que los datos se desvían respecto de la media, pero algunos libros de texto no enfatiza suficientemente esta propiedad y se centran más en la heterogeneidad entre las observaciones que en su desviación respecto de la posición central.

Pueden surgir dificultades en el cálculo de la desviación típica, como creer que no es necesario tener en cuenta los ceros o no ponderar los valores al calcular la desviación típica a partir de una tabla de frecuencias.

Análisis comparativo de libros de texto

Se han elegido, como se ve en la Figura 1, los siguientes libros de texto: **Anaya** Educación. Matemáticas 4º ESO orientadas a las ciencias aplicadas; **Marea Verde**. Matemáticas orientadas a las enseñanzas aplicadas: 4º de ESO; **Savia (SM).** Matemáticas orientadas a las ciencias aplicadas. 4 ESO. Se va a proceder al análisis cuantitativo y cualitativo de estos libros de texto.

Figura 1

Portadas de los libros de texto de las editoriales Anaya, Marea Verde y Savia SM

Anaya | Marea Verde | Savia (SM)

En las Tablas 1, 2 y 3 hemos realizado una recopilación de los tipos de actividades que aparencen y en que pagina

Tabla 1

Tipos de actividades que aparecen en el libro de Anaya

Anaya

Lección	Ejercicios	Ejemplos	Contextualizados	Páginas
1.1. Conceptos básicos	0	0	0	1
1.2. Tabla de frecuencias	2	2	2	2
1.3. Parámetros estadísticos: $\bar{x}$ y σ	2	1	0	2
1.4. Dependencia aleatoria y funcional	2	1	0	2
1.5. Correlación entre dos variables	2	2	1	2

1.6. Coeficiente de correlación lineal. Interpretación	2	0	0	2
1.7. Ejercicios y problemas	21	0	13	2
2.1. Distribuciones bidimensionales. Correlación	2	1	1	3
2.2. El valor de la correlación	2	1	1	2
2.3. La recta de regresión para hacer estimaciones	2	3	3	2
2.4. Ejercicios y problemas	11	0	4	1
Totales	**48**	**11**	**24**	**21**

Tabla 2

Tipos de actividades que aparecen en el libro de Marea Verde

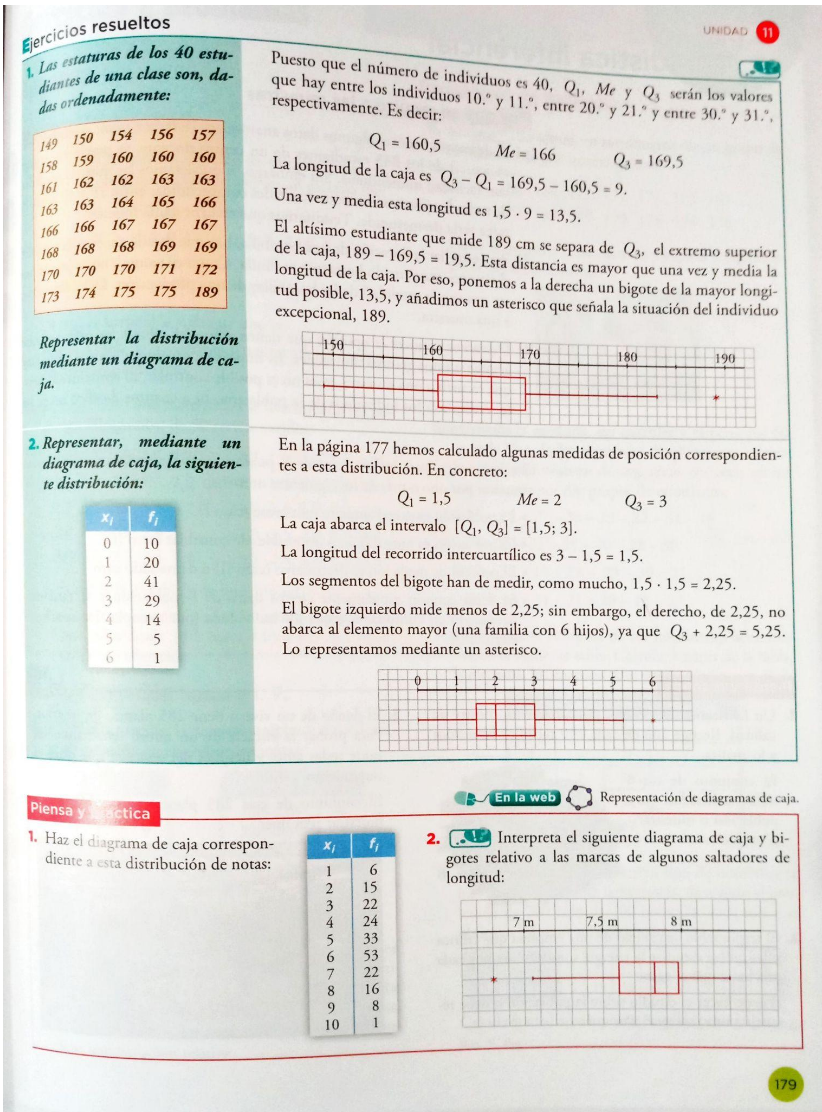
	Marea Verde			
Lección	Ejercicios	Ejemplos	Contextualizados	Páginas
1.1. Estadística. Muestras. Estudios Estadísticos	2	0	2	2
1.2. Variable Discreta. Tablas y Gráficos	5	1	4	3
1.3. Parámetros de Centralización y Dispersión	7	5	4	6
1.4. Diagrama de cajas	2	1	1	1
1.5. Variable Continua: Intervalos. Marcas de Clase. Histogramas	3	3	3	4
2.1. Datos Bidimensionales. Ideas Generales	0	0	0	1
2.2. Frecuencias Conjuntas	3	2	0	2
2.3. Diagrama de Dispersión y Recta de Regresión	1	1	0	2
2.4. Interpretación de la Recta de Regresión. Introducción a la Correlación	2	2	0	5
Totales	25	14	14	26

Actividades en el libro de texto de Marea Verde

En las Figuras 2, 3 y 4 hemos puesto varios ejemplos de actividades que aparecen en los libros de texto.

Figura 2

Ejemplos de actividades en Anaya



Ejercicios resueltos

UNIDAD 11

1. ***Las estaturas de los 40 estudiantes de una clase son, dadas ordenadamente:***

149	150	154	156	157
158	159	160	160	160
161	162	162	163	163
163	163	164	165	166
166	166	167	167	167
168	168	168	169	169
170	170	170	171	172
173	174	175	175	189

Representar la distribución mediante un diagrama de caja.

Puesto que el número de individuos es 40, Q_1, Me y Q_3 serán los valores que hay entre los individuos 10.º y 11.º, entre 20.º y 21.º y entre 30.º y 31.º, respectivamente. Es decir:

$$Q_1 = 160{,}5 \qquad Me = 166 \qquad Q_3 = 169{,}5$$

La longitud de la caja es $Q_3 - Q_1 = 169{,}5 - 160{,}5 = 9$.

Una vez y media esta longitud es $1{,}5 \cdot 9 = 13{,}5$.

El altísimo estudiante que mide 189 cm se separa de Q_3, el extremo superior de la caja, $189 - 169{,}5 = 19{,}5$. Esta distancia es mayor que una vez y media la longitud de la caja. Por eso, ponemos a la derecha un bigote de la mayor longitud posible, 13,5, y añadimos un asterisco que señala la situación del individuo excepcional, 189.

150 160 170 180 190

2. ***Representar, mediante un diagrama de caja, la siguiente distribución:***

x_i	f_i
0	10
1	20
2	41
3	29
4	14
5	5
6	1

En la página 177 hemos calculado algunas medidas de posición correspondientes a esta distribución. En concreto:

$$Q_1 = 1{,}5 \qquad Me = 2 \qquad Q_3 = 3$$

La caja abarca el intervalo $[Q_1, Q_3] = [1{,}5; 3]$.

La longitud del recorrido intercuartílico es $3 - 1{,}5 = 1{,}5$.

Los segmentos del bigote han de medir, como mucho, $1{,}5 \cdot 1{,}5 = 2{,}25$.

El bigote izquierdo mide menos de 2,25; sin embargo, el derecho, de 2,25, no abarca al elemento mayor (una familia con 6 hijos), ya que $Q_3 + 2{,}25 = 5{,}25$. Lo representamos mediante un asterisco.

0 1 2 3 4 5 6

En la web Representación de diagramas de caja.

Piensa y practica

1. Haz el diagrama de caja correspondiente a esta distribución de notas:

x_i	f_i
1	6
2	15
3	22
4	24
5	33
6	53
7	22
8	16
9	8
10	1

2. Interpreta el siguiente diagrama de caja y bigotes relativo a las marcas de algunos saltadores de longitud:

7 m 7,5 m 8 m

179

Figura 3

Ejemplos de actividades en Marea Verde

222 Estadística y probabilidad. 4ºA de ESO

1.2. Variable discreta. Tablas y gráficos

Tablas

Al hacer un estudio estadístico o realizar un experimento aleatorio la información obtenida se resume en una tabla o distribución de frecuencias.

Ejemplo:

- Preguntamos a 40 estudiantes de 4º si les gusta, o no, el fútbol. En la tabla del margen reflejamos los resultados.

Posibles resultados	Frecuencia absoluta
Les gusta	28
No les gusta	12
Total	40

Es una tabla de frecuencias absolutas.

Al dividir la frecuencia absoluta entre el número total tenemos la frecuencia relativa, así la frecuencia relativa de los que les gusta el fútbol es 28/40 = 0.7, y la de los que no les gusta el futbol es 12/40 = 3/10 = 0.3.

Posibles resultados	Frecuencias relativas	Porcentaje
Les gusta	0.7	70
No les gusta	0.3	30
Suma total	1	100

La **frecuencia absoluta** es el número de veces que se ha obtenido ese resultado.

La **frecuencia relativa** se obtiene dividiendo la frecuencia absoluta entre el número total de datos.

La suma de las frecuencias relativas es siempre igual a 1.

Multiplicando por 100 se obtienen los porcentajes.

Actividad resuelta

- *Se han obtenido los datos sobre el número de visitas que se han hecho de los Textos Marea Verde de Matemáticas en los meses indicados, y se han reflejado en una tabla. Haz una tabla de frecuencias absolutas, relativas y porcentajes, de frecuencias acumuladas absolutas y de frecuencias relativas acumuladas.*

Marea verde	Frecuencias absolutas	Frecuencias relativas	Porcentajes	Frecuencias acumuladas absolutas	Frecuencias acumuladas relativas
Septiembre	1 834	0.51	51	1 834	0.52
Octubre	956	0.26	26	2 790	0.77
Noviembre	432	0.12	12	3 222	0.89
Diciembre	389	0.11	11	3 611	1
TOTAL	3 611	1	100		

Observa que las **frecuencias acumuladas** se obtienen sumando la frecuencia anterior e indica, en este ejemplo, el número de visitas hasta ese momento.

Actividades propuestas

3. Copia en tu cuaderno y completa la siguiente tabla de frecuencias absolutas de los valores obtenidos al tirar un dado, con las frecuencias relativas y porcentajes, y con frecuencias acumuladas absolutas y frecuencias relativas acumuladas.

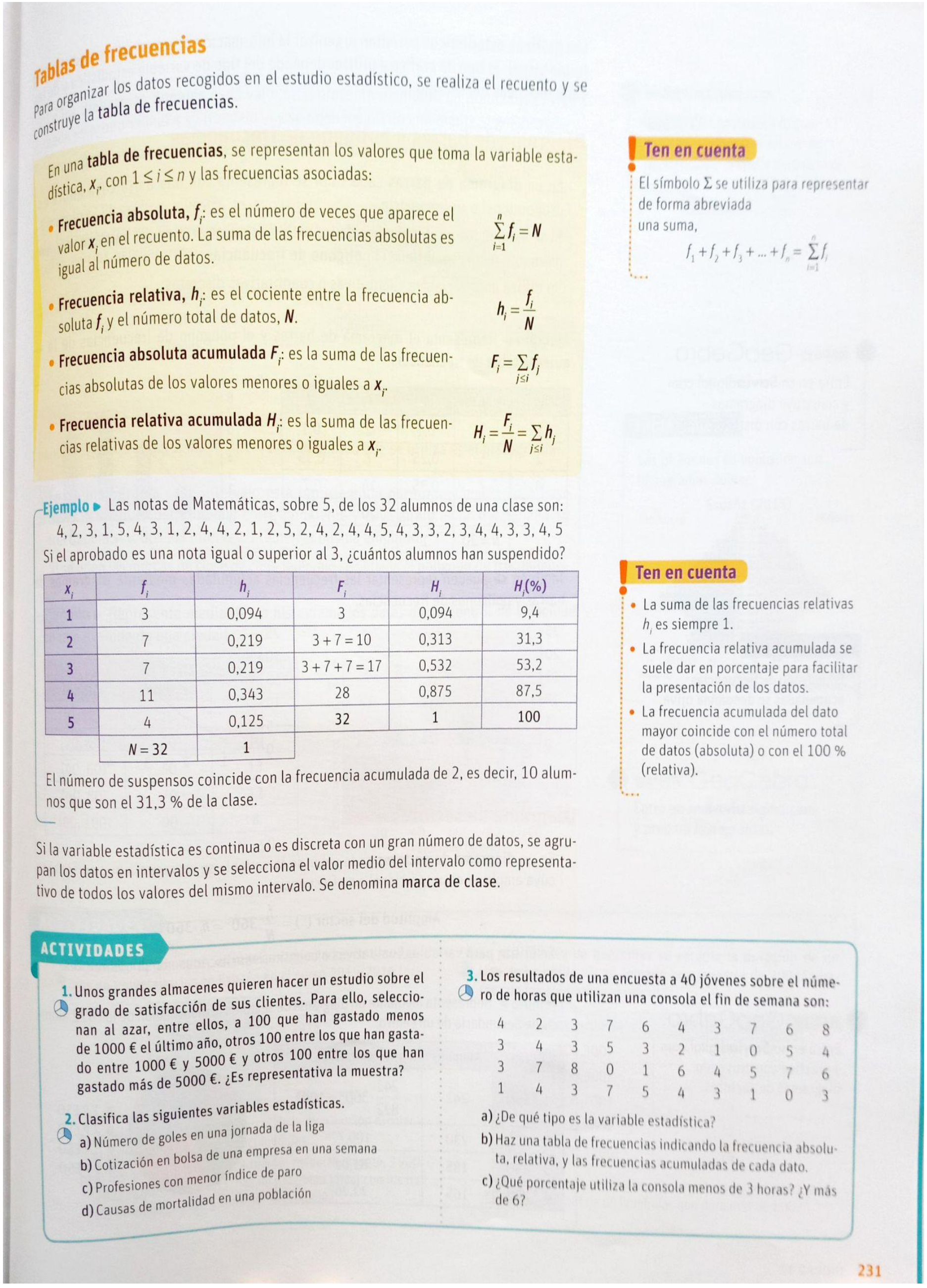
Resultados	Frecuencias absolutas
1	17
2	12
3	17
4	15
5	21
6	14

Autores: María Molero y Andrés García
Revisores: Raquel Caro y Nieves Zuasti
Ilustraciones: Banco de Imágenes de INTEF

Textos Marea Verde

Figura 4

Ejemplos de actividades en Savia(SM)



Tablas de frecuencias

Para organizar los datos recogidos en el estudio estadístico, se realiza el recuento y se construye la **tabla de frecuencias**.

En una **tabla de frecuencias**, se representan los valores que toma la variable estadística, x_i, con $1 \leq i \leq n$ y las frecuencias asociadas:

- **Frecuencia absoluta,** f_i: es el número de veces que aparece el valor x_i en el recuento. La suma de las frecuencias absolutas es igual al número de datos. $\sum_{i=1}^{n} f_i = N$
- **Frecuencia relativa,** h_i: es el cociente entre la frecuencia absoluta f_i y el número total de datos, N. $h_i = \frac{f_i}{N}$
- **Frecuencia absoluta acumulada** F_i: es la suma de las frecuencias absolutas de los valores menores o iguales a x_i. $F_i = \sum_{j \leq i} f_j$
- **Frecuencia relativa acumulada** H_i: es la suma de las frecuencias relativas de los valores menores o iguales a x_i. $H_i = \frac{F_i}{N} = \sum_{j \leq i} h_j$

Ten en cuenta

El símbolo Σ se utiliza para representar de forma abreviada una suma,

$$f_1 + f_2 + f_3 + ... + f_n = \sum_{i=1}^{n} f_i$$

Ejemplo Las notas de Matemáticas, sobre 5, de los 32 alumnos de una clase son:

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

Si el aprobado es una nota igual o superior al 3, ¿cuántos alumnos han suspendido?

x_i	f_i	h_i	F_i	H_i	$H_i(\%)$
1	3	0,094	3	0,094	9,4
2	7	0,219	3 + 7 = 10	0,313	31,3
3	7	0,219	3 + 7 + 7 = 17	0,532	53,2
4	11	0,343	28	0,875	87,5
5	4	0,125	32	1	100
	$N = 32$	1			

El número de suspensos coincide con la frecuencia acumulada de 2, es decir, 10 alumnos que son el 31,3 % de la clase.

Ten en cuenta

- La suma de las frecuencias relativas h_i es siempre 1.
- La frecuencia relativa acumulada se suele dar en porcentaje para facilitar la presentación de los datos.
- La frecuencia acumulada del dato mayor coincide con el número total de datos (absoluta) o con el 100 % (relativa).

Si la variable estadística es continua o es discreta con un gran número de datos, se agrupan los datos en intervalos y se selecciona el valor medio del intervalo como representativo de todos los valores del mismo intervalo. Se denomina **marca de clase**.

ACTIVIDADES

1. Unos grandes almacenes quieren hacer un estudio sobre el grado de satisfacción de sus clientes. Para ello, seleccionan al azar, entre ellos, a 100 que han gastado menos de 1000 € el último año, otros 100 entre los que han gastado entre 1000 € y 5000 € y otros 100 entre los que han gastado más de 5000 €. ¿Es representativa la muestra?

2. Clasifica las siguientes variables estadísticas.

a) Número de goles en una jornada de la liga

b) Cotización en bolsa de una empresa en una semana

c) Profesiones con menor índice de paro

d) Causas de mortalidad en una población

3. Los resultados de una encuesta a 40 jóvenes sobre el número de horas que utilizan una consola el fin de semana son:

4	2	3	7	6	4	3	7	6	8
3	4	3	5	3	2	1	0	5	4
3	7	8	0	1	6	4	5	7	6
1	4	3	7	5	4	3	1	0	3

a) ¿De qué tipo es la variable estadística?

b) Haz una tabla de frecuencias indicando la frecuencia absoluta, relativa, y las frecuencias acumuladas de cada dato.

c) ¿Qué porcentaje utiliza la consola menos de 3 horas? ¿Y más de 6?

231

Tabla 3
Tipos de actividades que aparecen en el libro de Savia(SM)

Savia (SM)

Lección	Ejercicios	Ejemplos	Contextualizados	Páginas
1.1. Conceptos elementales de estadísticas. Muestreo	3	3	0	2
1.2. Gráficos estadísticos	2	4	0	2
1.3. Medidas de centralización y posición	3	2	0	2
1.4. Medidas de dispersión	4	2	0	2
1.5. Interpretación conjunta de la media y la desviación típica	4	3	0	2
1.6. Organiza tus ideas	0	0	0	1
1.7. Actividades clave	0	2	0	1
1.8. Actividades	52	0	41	4
1.9. Ponte a prueba	4	1	1	2
2.1. Distribuciones bidimensionales	4	3	0	2
2.2. Covarianza y coeficiente de correlación lineal	4	3	0	2
2.3. Recta de regresión lineal	3	2	0	2
2.4. Organiza tus ideas	0	0	0	1
2.5. Actividades clave	0	1	0	1
2.6. Actividades	26	0	19	2
2.7. Ponte a prueba	4	2	3	2
Total	109	28	64	30

Al analizar los tres libros de texto, Anaya Educación, Marea Verde y Savia (SM), es importante destacar las diferencias en términos de cantidad de ejercicios y la importancia asignada a cada lección. Anaya proporciona un número sustancialmente mayor de ejercicios en comparación con Marea Verde y Savia (SM), lo que sugiere que este libro puede ofrecer más oportunidades para que los estudiantes practiquen y refuercen sus habilidades matemáticas. En cuanto a la importancia otorgada a cada lección, Anaya y Savia (SM) siguen una estructura similar en la distribución de páginas, aunque Anaya asigna más páginas en general.

En relación con la contextualización, es evidente que Anaya y Marea Verde le dan mayor importancia al contextualizar sus ejercicios, facilitando potencialmente la comprensión y el interés de los estudiantes al conectar las matemáticas con situaciones reales. Savia (SM), por otro lado, presenta una cantidad muy limitada de ejercicios contextualizados.

En términos de uso de tecnología, tanto Anaya como Marea Verde promueven el empleo del ordenador, además Savia (SM) promueve el uso de calculadora con ejemplos, por lo que aquí la editorial Savia está más completa, teniendo en cuenta que la calculadora si que suele estar permitida tanto en las clases como en el examen, al menos en los cursos superiores de la ESO, generalmente.

Además, es importante mencionar que **Savia (SM) incluye una cantidad mayor de fórmulas** matemáticas en comparación con los otros dos libros. Esto podría indicar un enfoque más teórico y un énfasis en el desarrollo de habilidades analíticas en matemáticas de ejercitación de técnicas con un mayor número de ejercicios.

En cuanto al enfoque histórico y las anécdotas, Anaya y Savia (SM) proporcionan un mayor contenido en este sentido que Marea Verde. Esto podría ser beneficioso para involucrar a los estudiantes en la materia y ayudarles a comprender cómo las matemáticas han evolucionado y se han aplicado a lo largo del tiempo.

Como reflexión, debemos tender a mostrar los objetos matemáticos a través de la resolución de problemas. Es importante la ejercitación de técnicas, pero lo es mucho más el análisis de los resultados obtenidos, más que las habilidades analíticas y la aplicación de fórmulas. Anaya es un buen libro intermedio entre la aproximación más académica de Savia y el enfoque meramente práctico de Marea Verde.

La estadística a lo largo de la educación obligatoria de un alumno

La estadística en 4º de la ESO no requiere conocimientos matemáticos avanzados para su correcto manejo, sino que solo se necesitan algunos conocimientos básicos de álgebra y análisis. Los conocimientos previos necesarios incluyen la capacidad para realizar operaciones con números naturales y decimales, fracciones, porcentajes y factor de proporcionalidad, comprensión del concepto de sumatorio, la habilidad para representar y entender funciones sencillas, la capacidad para distinguir intervalos abiertos y cerrados y la habilidad para resolver ecuaciones de primer grado.

A nivel estatal, según la LOMLOE, ley Orgánica 3/2020, de 29 de diciembre, por la que se modifica la Ley Orgánica 2/2006, de 3 de mayo, de Educación. Boletín Oficial del Estado, 340, de 30 de diciembre de 2020 y según su desarrollo curricular, y su concreción en la Orden ECD/1171/2022, de 2 de agosto, por la que se aprueban el currículo y las características de la evaluación de la Educación Secundaria Obligatoria y se autoriza su aplicación en los centros docentes de la Comunidad Autónoma de Aragón, La referencia al **sentido estocástico** es significativa porque muestra que el gobierno español se compromete a garantizar que los estudiantes adquieran las habilidades que necesitan para comprender y utilizar la probabilidad y la estadística. La probabilidad y las estadísticas son herramientas importantes para comprender el mundo que nos rodea, y también son esenciales para tomar decisiones informadas en una variedad de contextos. Al enseñar a los estudiantes sobre probabilidad y estadística, el gobierno español está ayudando a prepararlos para el éxito en el siglo XXI.

La orden ECD/1171/2022 en sus anexos, en la sección referente a la asignatura de Matemáticas, explica que el sentido estocástico es un elemento importante y distintivo de las matemáticas, especialmente en el ámbito de la estadística y la probabilidad. Este sentido se refiere a la capacidad para trabajar con la variabilidad de las situaciones frente al determinismo, lo que significa que es clave para crear una ciudadanía informada con suficientes conocimientos y competencias para que ante fenómenos aleatorios y **tratamiento e interpretación de datos e informaciones sean personas difícilmente manipulables y sean capaces de tomar decisiones y formarse opiniones de forma crítica y razonable**. En resumen, el documento destaca la importancia del sentido estocástico en el desarrollo de habilidades matemáticas valiosas para la vida diaria.

Los contenidos del bloque de estadística en **primaria** son: recogida y recuento de datos en situaciones de observación y registro e interpretación de datos en pictogramas en 1º; recogida, ordenación y clasificación de datos en función de un criterio y realización e interpretación de gráficos sencillos en 2º y 3º primaria; recogida, ordenación y clasificación de datos en función de más de un criterio y realización e interpretación de gráficos sencillos en 4º primaria; recogida y clasificación de datos cualitativos y cuantitativos, construcción de tablas de frecuencias absolutas e introducción intuitiva a conceptos como la media aritmética, rango, frecuencia y moda, y realización e interpretación de gráficos sencillos en 5º primaria; y recogida y clasificación de datos cualitativos y cuantitativos, construcción de tablas de frecuencias absolutas y relativas, introducción intuitiva a conceptos como la media aritmética, rango, frecuencia y moda; En 6º de primaria, los contenidos del bloque de estadística incluyen: recogida y clasificación de datos cualitativos y cuantitativos, construcción de tablas de frecuencias absolutas y relativas, introducción intuitiva a conceptos como la media aritmética, rango, frecuencia y moda, realización e interpretación de gráficos sencillos como diagramas de barras, poligonales y sectoriales, y análisis crítico de informaciones presentadas en gráficos estadísticos, aunque lamentablemente, según nos dice la experiencia, no es raro que la enseñanza de estos stemas se lleve a cabo de manera rápida y precipitada.

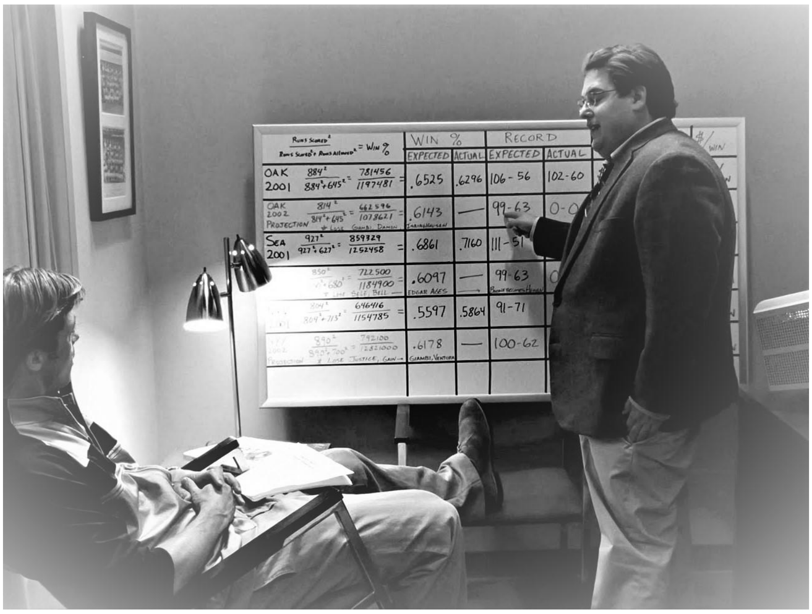
En la enseñanza **secundaria**, en 1º se trabajan conceptos básicos de estadística como la población y la muestra, las variables estadísticas (cualitativas y cuantitativas), las frecuencias absolutas y relativas, y la organización de datos en tablas. También se introducen gráficos estadísticos como diagramas de barras y de sectores y polígonos de frecuencias. Además, se estudian las medidas de tendencia central. En 2º se profundiza en el estudio de la población y la muestra, las variables estadísticas y las frecuencias absolutas y relativas, y se sigue trabajando con la organización de datos en tablas. También se estudian las medidas de dispersión y se profundiza en el uso de gráficos estadísticos como diagramas de barras y de sectores y polígonos de frecuencias. En 3º se introducen conceptos más avanzados de estadística como las fases y tareas de un estudio estadístico, la representatividad de una muestra y los parámetros de posición y de dispersión. También se trabaja con el cálculo, interpretación y propiedades de la media y la desviación típica y se utiliza el diagrama de caja y bigotes para analizar conjuntamente la media y la desviación típica.

Evaluación inicial

Para intentar mostrar la utilidad de la estadística y despertar su interés, como se ve en la Figura 5, se mostrará alguna escena de la película "Moneyball", protagonizada por Brad Pitt.

Figura 5

Escena de la película MoneyBall



Nota. Billy Beane(Brad Pitt) escuchando la explicación sobre la selección de fichajes de Peter Brand(Jonah Hill), personaje basado en el economista licenciado en Harvard Paul DePodesta

Es una introducción a la utilidad de la estadística para alumnos de 4º de la ESO porque presenta un caso de la historia reciente, atractivo -si son aficionados al deporte-, en el que el análisis estadístico se utiliza para tomar decisiones informadas y mejorar el rendimiento de un equipo deportivo mejorando sus fichajes y estrategia. La historia ilustra cómo los métodos estadísticos pueden desafiar la sabiduría convencional y proporcionar una ventaja competitiva. Mostrar una escena de "Moneyball" podría captar el interés de los alumnos y ayudarles a comprender la importancia y aplicabilidad de la estadística en la vida real.

Después de presentar la escena de "Moneyball", se realizará una corta evaluación inicial tipo test:

¿Qué es la entendemos por una estadística descriptiva?
a) Un método para tomar decisiones basadas en datos.
b) Un conjunto de técnicas para resumir y describir los datos.
c) Un enfoque para predecir el resultado de eventos futuros.
d) Un proceso para recopilar y analizar datos de una muestra.

¿Cuál de las siguientes es una medida que tiende a describir lo que hace la mayoría de la población?
a) Moda
b) Rango
c) Varianza
d) Desviación estándar

Si ordenas los datos el valor que queda en medio ¿Cómo se llama?
a) Moda
b) Media.
c) Mediana
d) Rango.

¿Cuál es la fórmula para calcular la media aritmética de un conjunto de datos?
a) Media = (Suma de todos los valores) / (Cantidad de valores)
b) Media = (Valor más alto + Valor más bajo) / 2
c) Media = (Suma de todos los valores) / (Cantidad de valores - 1)
d) Media = (Valor más alto - Valor más bajo) / 2

¿A qué te suena desviación?
a) La dispersión de los datos respecto a la media.
b) La diferencia entre el valor más alto y el más bajo.
c) La posición central de los datos.
d) La frecuencia con la que aparece cada valor.

"Al identificar el conocimiento previo y las habilidades de los estudiantes, los maestros pueden adaptar su enseñanza a las necesidades específicas de sus alumnos y garantizar que todos los estudiantes estén comprometidos y desafiados en el aula." (Chapin & Johnson, 2006, 25)

Razón de ser de la introducción escolar de la estadística

La estadística es fundamental en la vida diaria y en la organización social en general por ejemplo el IPC, el índice de precios al consumo que es un referente clave a la hora de establecer las subidas salariales las subidas de las pensiones con la consiguiente repercusión que tiene esto en el nivel de consumo y en el nivel de bienestar de los ciudadanos existen otras estadísticas más o menos bien conocidas: el PIB, producto interior bruto, las tasas de natalidad y mortalidad y la tasa de paro.

Estas estadísticas son producidas por instituciones públicas como el propio instituto nacional de estadística, INE, o las oficinas estadísticas de las comunidades autónomas todas ellas generan una serie de datos que van a ser la base para la futura planificación que es la clave de la gestión de los gobiernos los gobiernos locales de las comunidades autónomas

Se producen también estadísticas sobre todo relativas al ámbito del consumo de la opinión pública y de los seguros, las compañías aseguradoras y la propia seguridad social fijan las primas de riesgo y las cotizaciones mediante estadística.

Origen y evolución histórica de la estadística

Etimológicamente, estadística se deriva del latín statisticum collegium ("consejo de estado") y la palabra italiana statista ("estadista" o "político"). El *Statistik* alemán, introducido por primera vez por Gottfried Achenwall (1749), designaba originalmente el análisis de datos sobre el estado, significando la "ciencia del estado" (entonces llamada aritmética política). Adquirió el significado de la recopilación y clasificación de datos en general a principios del siglo XIX. Fue introducido al inglés en 1791 por Sir John Sinclair cuando publicó el primero de 21 volúmenes titulado "Cuenta estadística de Escocia" (Ball, 2006).

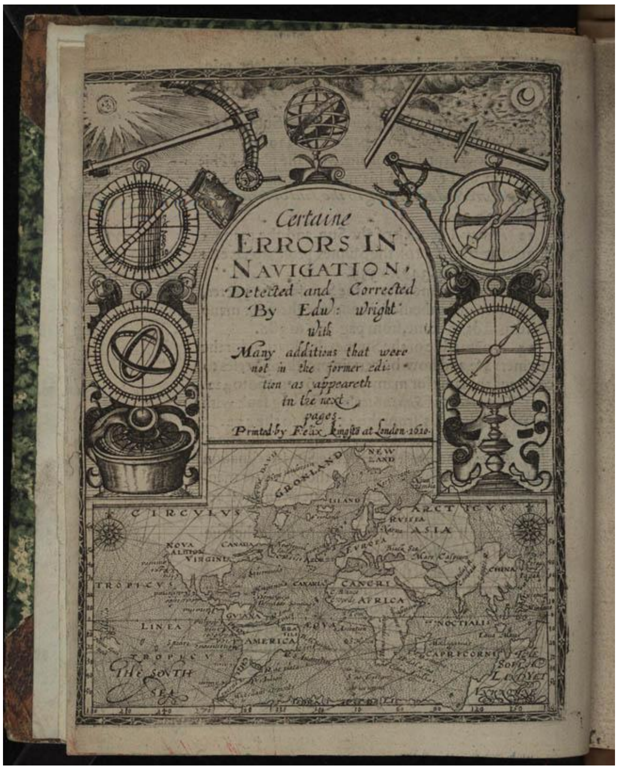
Muchos imperios, como el Imperio Romano y la dinastía Han, recopilaron datos sobre el tamaño de la población, la geografía y la riqueza. Los métodos estadísticos también se han utilizado desde al menos el **siglo V a.C**., como se describe en la Historia de la Guerra del Peloponeso de Tucídides (*History of the Peloponnesian War*, 1972, 204) para calcular la altura del muro de la ciudad de Platea.

El Juicio del Copón (Allen & Allen, 2012), una prueba de la pureza de las monedas que se lleva a cabo regularmente desde el **siglo XII** en la abadía de Westminster, se basa en métodos de muestreo estadístico.

Aunque es muy probable que el concepto de la mediana se haya utilizado antes en la práctica, el primer registro documentado de este concepto se encuentra en la obra de Edward Wright, "Certaine Errors in Navigation", publicada el **siglo XVI**, ver Figura 6. En una sección de su libro dedicada a determinar una localización con un compás, Wright introdujo la idea de la mediana. Él percibió que este valor, que separa la mitad superior de un conjunto de datos de la mitad inferior, era probablemente el más cercano a la verdad en un conjunto de observaciones. (Monmonier, 2004)

Figura 6

Portada de Certaine Errors in Navigation de Wright



El campo moderno de la estadística surgió a fines del siglo XIX y principios del XX en tres etapas (Walker, 1975).

La primera ola, el noruego Anders Nicolai Kiær introdujo el concepto de **muestreo estratificado** en 1895. Arthur Lyon Bowley introdujo nuevos métodos de muestreo de datos en 1906 cuando trabajaba en estadísticas sociales. Aunque las encuestas estadísticas de las condiciones sociales habían comenzado con "Life and Labor of the People in London" de Charles Booth (1889-1903) y "Poverty, A Study of Town Life" de Seebohm Rowntree (1901), la innovación clave de Bowley consistió en el uso de técnicas de muestreo aleatorio.

Francis Galton y Karl Pearson, ambos figuras prominentes en el desarrollo de la estadística que mantuvieron una relación de mentor y protegido, ver Figura 7.

Karl Pearson, que es mejor conocido por el desarrollo del coeficiente de correlación de Pearson (una medida de la correlación lineal entre dos variables), era en realidad un protegido de Francis Galton. Tras el retiro de Galton, Pearson continuó su trabajo, desarrollando y formalizando muchos de los conceptos y técnicas que Galton había esbozado de manera más general (Pearson, 2011).

Figura 7

Galton a la edad de 87 años, con K. Pearson

Francis Galton fue un pionero inglés en los campos de la estadística, la psicología y la biología. Hizo importantes contribuciones a la teoría de la regresión y la correlación, y es conocido como uno de los primeros defensores de la eugenesia, aunque esta última perspectiva ha sido ampliamente criticada y rechazada en tiempos más modernos. Aquí están algunos de los principales aportes de Galton:

Antropometría: Galton también fue un pionero en el campo de la antropometría, el estudio de las diferencias humanas y las medidas del cuerpo humano. Realizó una serie de estudios que buscaban correlacionar las medidas físicas con varias características mentales y de personalidad.

Eugenesia: Galton acuñó el término "eugenesia" para describir su teoría de que la inteligencia y otras características eran hereditarias y que la raza humana podría mejorarse a través de una cuidadosa reproducción selectiva. Sin embargo, esta teoría ha sido ampliamente desacreditada y es considerada altamente problemática desde una perspectiva ética. Pearson también ayudó a establecer en 1911 la Cátedra Galton de Eugenesia (más tarde renombrada como Cátedra Galton de Genética).

Cuestionario de huellas dactilares: Galton fue uno de los primeros en sugerir que las huellas dactilares podrían ser únicas para cada individuo y, por lo tanto, podrían ser utilizadas para la identificación personal. Su trabajo en este campo sentó las bases para el moderno uso de las huellas dactilares en la ciencia forense.

Silbato para perros de Galton (también conocido como **silbato ultrasónico**): Galton inventó este dispositivo, que produce un sonido de alta frecuencia que los humanos no pueden oír, pero que es audible para los perros y otros animales. Este silbato se usa comúnmente para el entrenamiento de perros.

Isobaras: Galton también fue uno de los primeros en utilizar isobaras en mapas meteorológicos. Las isobaras son líneas en un mapa meteorológico que unen puntos de igual presión. Estas son útiles para mostrar la estructura de sistemas de alta y baja presión, y ayudan a los meteorólogos a predecir cambios en el clima.

A lo largo de su vida, Galton demostró un interés y una curiosidad insaciables por el mundo natural y las ciencias, y estas son solo algunas de las formas en que su trabajo ha tenido un impacto duradero.

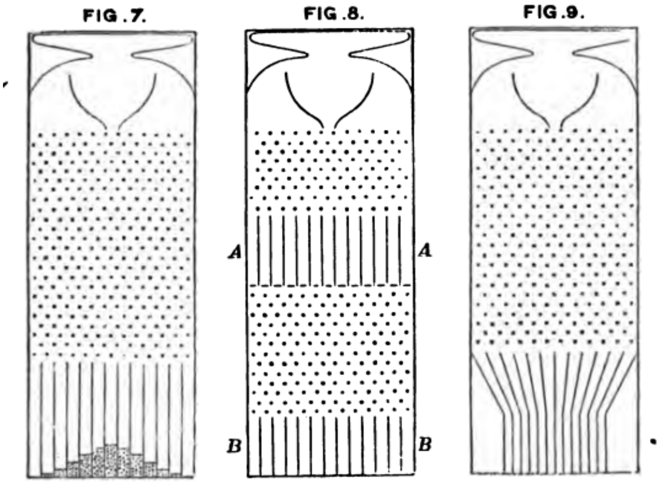
En cuanto a la **máquina de Galton,** ver Figura 8, es un dispositivo que se utiliza para demostrar el teorema central del límite en la estadística, que establece que la suma de un gran número de variables aleatorias independientes e idénticamente distribuidas tiende a distribuirse normalmente. La máquina consiste en una tabla vertical con varias filas de clavos y una serie de contenedores en la parte inferior. Cuando se sueltan las canicas en la parte superior de la tabla, rebotan de manera aleatoria en los clavos mientras caen. Aunque el camino de cada canica es aleatorio, cuando se dejan caer suficientes canicas, los contenedores al final tienden a llenarse de manera que forman una distribución normal (también conocida como curva de campana). Esto ilustra cómo los resultados aleatorios individuales pueden sumarse para formar un patrón predecible a nivel de población.

Figura 8

"Ilusitración mecánica de la causa de la curva de frecuencia" de Natural Inheritance, Galton, 1894

Frequency.—The Curve of Frequency, and that of Distribution, are convertible : therefore if the genesis of either of them can be made clear, that of the other becomes also intelligible. I shall now illustrate the origin of the Curve of Frequency, by means of an apparatus shown in Fig. 7, that mimics in a very pretty way the conditions

FIG. 7. FIG. 8. FIG. 9.



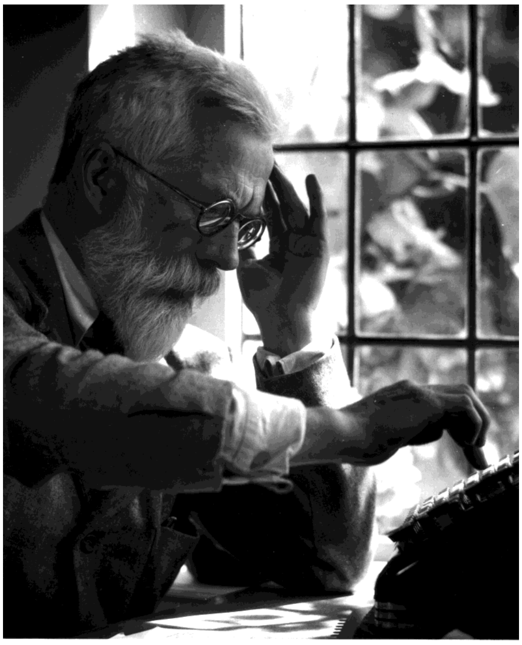
on which Deviation depends. It is a frame glazed in front, leaving a depth of about a quarter of an inch behind the glass. Strips are placed in the upper part to act as a funnel. Below the outlet of the funnel stand a

Existe un diseño físico de la exposición "Mathematica: A World of Numbers... and Beyond" de 1961 que se conserva hoy en día en el museo Henry Ford de Detroit.

La segunda ola de las décadas de 1910 y 1920 fue iniciada por William Sealy Gosset y alcanzó su culminación con las ideas de Ronald Fisher, ver Figura 9. Esto implicó el desarrollo de un mejor diseño de experimentos, modelos, pruebas de hipótesis y técnicas para usar con muestras de datos pequeños.

Figura 9

Gosset, William Sealy ('Student') 1876-1937, del Annals of Eugenics 9 (1939) y Fisher, Sir Ronald Aylmer 1890-1962, del R A Fisher, Collected Papers



William Sealy Gosset trabajó en la cervecería Guinness en Dublín, Irlanda, como maestro cervecero y estadístico. En aquel tiempo, Guinness tenía una política estricta que prohibía a sus empleados publicar cualquier tipo de investigación, debido a preocupaciones sobre la divulgación de secretos comerciales.

Sin embargo, el trabajo de Gosset en estadística era tan innovador y valioso para el campo de la estadística que se llegó a un acuerdo para permitirle publicar. La condición era que no podría usar su nombre real, y en su lugar, adoptó el seudónimo de "Student".

La "distribución t" se utilizaba en la cervecería Guinness para controlar la calidad de la cerveza. Cuando se mide la calidad, para no dañar la producción, se toman muestras pequeñas, por lo que es esencial un método estadístico que pueda manejar tales muestras y proporcionar estimaciones precisas. La distribución t de Student proporciona un marco para gestionar el mayor grado de incertidumbre que surge cuando se trabaja con muestras pequeñas y su uso en pruebas de hipótesis y en intervalos de confianza. Así, publicó su famosa distribución t-Student, que sigue siendo fundamental en estadística hoy en día, bajo este seudónimo. La "t" en la distribución t de Student no tiene un significado particular por sí misma. En sus escritos originales, utilizó la "t" para denotar la variable aleatoria que sigue la distribución.

Esencialmente, "t" es solo una letra que Gosset escogió para representar la estadística de prueba que creó, similar a cómo otras pruebas estadísticas tienen sus propias letras o símbolos asociados, como "z" para la distribución normal estándar.

La conexión entre Gosset y Fisher proviene de su correspondencia sobre el trabajo de Gosset con la distribución t. Fisher reconoció el valor del trabajo de Gosset y contribuyó a su difusión, así como a su extensión. En particular, Fisher es el responsable de la generalización de la distribución t de Student al caso de dos muestras, lo cual es fundamental en muchas pruebas estadísticas, incluyendo la prueba t de Student para muestras independientes. Así, aunque no trabajaron juntos en el mismo lugar, sus interacciones y la influencia mutua tuvieron un profundo impacto en el desarrollo de la estadística moderna.

Fisher dejó su puesto en el University College de Londres durante la Segunda Guerra Mundial y, después de un breve tiempo en Cambridge, aceptó la invitación para convertirse en profesor de la Universidad de Adelaide. En la Universidad de Adelaide, Australia, Fisher continuó su trabajo en genética y estadística de 1946 a 1957.

La tercera ola, que vio principalmente el refinamiento y la expansión de desarrollos anteriores, surgió del trabajo colaborativo entre Egon Pearson -hijo de Karl Pearson- y Jerzy Neyman, ver Figura 10, tuvieron una estrecha colaboración profesional que duró varias décadas.

Figura 10

Neyman, Jerzy 1894-1981 y Pearson, Egon 1895-1980 circa 1926 en la UCL.

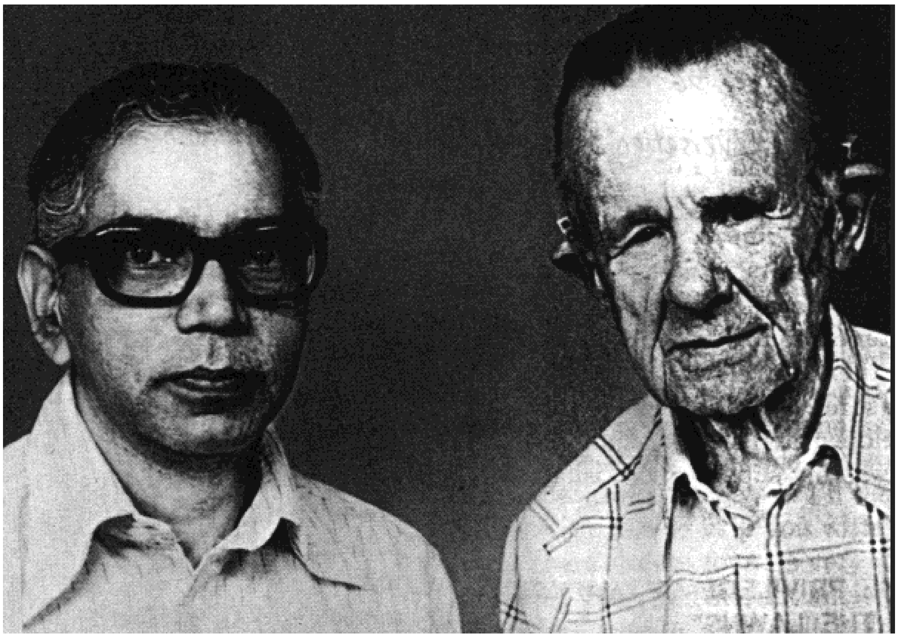
Se conocieron en la University College London (UCL), donde Neyman se unió a la facultad en la década de 1920. Juntos, desarrollaron lo que ahora se conoce como la teoría Neyman-Pearson de las pruebas de hipótesis.

Neyman más tarde se mudó a California, donde fundó el primer departamento de estadística de los Estados Unidos en la Universidad de California, Berkeley. Pearson, por su parte, siguió los pasos de su padre y se convirtió en una figura líder en estadística en el Reino Unido, continuando su carrera en UCL.

Para finalizar esta breve historia de la estadística, se van a citar estos matemáticos más contemporáneos:

Harald Cramér (1903-1985): profesor en la Universidad de Estocolmo, hizo avances en la teoría de la estimación, la teoría del muestreo y los métodos estadísticos. Es conocido por su trabajo en la desigualdad de Cramér-Rao, ver Figura 11, que establece un límite inferior para la varianza de cualquier estimador insesgado. También desarrolló la "distancia de Cramér-von Mises", que es una medida de discrepancia entre una distribución teórica y una distribución empírica. Sus contribuciones en la teoría de los procesos estocásticos y en el análisis asintótico de los estimadores también son notables.
C.R. Rao (1920-2021): Profesor en la Universidad de Calcuta, la Universidad de Cambridge y la Universidad Estatal de Pensilvania. Contribuciones en diseño de experimentos, análisis multivariado e inferencia estadística.

Figura 11
Rao, Cayampudi Radhakhrishna 1920-2021 con Cramer, Harald 1893-1985 de Conversation with C.R. Rao, Statistical Science, 2 (1987)



Bradley Efron (1938-presente), ver Figura 12: profesor emérito de Estadística y Bioestadística en la Universidad de Stanford. Efron es conocido por desarrollar el método bootstrap, que es una técnica de remuestreo ampliamente utilizada en el análisis de **Big Data**.

El objetivo principal del bootstrap es obtener una idea de la distribución muestral de un estadístico o parámetro de interés sin hacer suposiciones sobre la distribución subyacente de los datos. Esto es especialmente útil cuando la distribución exacta es desconocida o difícil de determinar. El proceso del método bootstrap es el siguiente:

A partir de una muestra original de datos, se generan múltiples muestras de datos, conocidas como muestras bootstrap. Estas muestras se obtienen tomando muestras aleatorias con reemplazo de la muestra original, lo que significa que se permite que los mismos datos aparezcan más de una vez en una muestra bootstrap y algunos datos pueden no aparecer en absoluto. Para cada muestra bootstrap, se calcula el estimador o estadístico de interés. Puede ser la media, la mediana, la desviación estándar u otro estadístico relevante para el problema en cuestión. Se repiten estos pasos un gran número de veces (por ejemplo, 1000 o más) para obtener una distribución aproximada de los estimadores o estadísticos de interés. Utilizando la distribución obtenida, se pueden calcular intervalos de confianza, realizar pruebas de hipótesis y realizar otras inferencias estadísticas.

Figura 12

Bradley Efron en su despacho de Berkeley en la década de los 90

Finalmente, enlazando la historia de la estadística con su relevancia actual, el análisis de datos masivos ha cobrado una relevancia vital en nuestra sociedad, y la estadística es la piedra angular que permite descifrar estos vastos conjuntos de datos. Empresas como Cabify, por poner un ejemplo, utilizan técnicas estadísticas para analizar y correlacionar datos, lo que les permite optimizar rutas y predecir demandas de forma eficiente. Estos procesos se basan en el sentido estocástico, ya que consideran la variabilidad y la incertidumbre inherente en los datos. Amazon también recurre a técnicas estadísticas para correlacionar las preferencias de los usuarios con su amplia biblioteca de contenido y productos, lo que permite personalizar las recomendaciones. Este uso de correlaciones y análisis estocásticos es un claro ejemplo de cómo la estadística puede mejorar el desempeño de las empresas en la sociedad.

Además, en campos como la medicina personalizada, la gestión del cambio climático, y la seguridad cibernética, el análisis estadístico, especialmente las técnicas de correlación y los modelos estocásticos, son fundamentales para interpretar y hacer predicciones. En resumen, el análisis estadístico es una herramienta esencial que está remodelando nuestra sociedad.

Metodología para su implementación

Adoptaremos una <u>metodología activa y orientada a la resolución de problemas</u> para impartir los temas de estadística a las sesiones. Los alumnos se organizan en grupos con el objetivo de fomentar el aprendizaje colaborativo. Esta estrategia permitirá que <u>los conceptos claves vayan aflorando</u> de manera natural durante el proceso de resolución de problemas, lo que fomentará una comprensión más profunda y aplicada de la estadística.

> El aprendizaje matemático se incrementa cuando los alumnos tienen oportunidades de confrontar y resolver problemas matemáticos interesantes y desafiantes, trabajar en colaboración con otros alumnos para explorar y elaborar conceptos y habilidades matemáticas y comunicar sus ideas y argumentos a otros. De hecho, la habilidad de trabajar y aprender en colaboración con otros es una habilidad que los alumnos necesitan desarrollar para su vida futura. (Silver, 2000, 21)

Esta cita respalda el uso de metodologías activas y colaborativas en la enseñanza de las matemáticas, ya que destaca cómo el aprendizaje matemático mejora cuando los alumnos trabajan juntos para resolver problemas y comunicar sus ideas.

A lo largo de estas sesiones, los alumnos trabajan en grupos y se enfrentan a problemas reales que requieren la aplicación de los conceptos estadísticos aprendidos. Además, al presentar sus resultados a sus compañeros y recibir retroalimentación, los alumnos pueden aprender de los errores y mejorar su comprensión de la estadística en un ambiente colaborativo y cercano.

La metodología activa se basa en el principio de que los alumnos aprenden de manera más efectiva cuando están involucrados activamente en el proceso de aprendizaje, en lugar de ser meros receptores pasivos de información. La metodología activa promueve el aprendizaje a través de la participación activa, la resolución de problemas, la colaboración y la reflexión crítica. Este enfoque puede ser particularmente beneficioso en la enseñanza de las matemáticas, ya que ayuda a los alumnos a desarrollar habilidades de pensamiento crítico, a comprender conceptos matemáticos de manera más profunda y a aplicarlos en situaciones del mundo real.

Diseño de la secuencia didáctica

La razón del diseño de la secuencia didáctica como se ha hecho es que se han investigado en diversos contextos, algunos de los cuales están cercanos a los intereses y preferencias de los estudiantes, como las preferencias musicales, de series de televisión o de medios de transporte. En general se han seguido las indicaciones de la asignatura del máster de **Diseño de actividades de aprendizaje de matemáticas** Arnal, A. y Muñoz, J.M. (2021). Se considera que problemas que no estén contextualizados o que puedan resultar más alejados de los intereses personales de los estudiantes resultan menos motivadores. Por otro lado, se han intentado evitar temas que aunque sean de su interés, pudiesen resultar demasiado polémicos para el desarrollo de las sesiones, o que representen algún tabú en el entorno multicultural habitual hoy en día.

> Es importante que los estudiantes estén expuestos a problemas y situaciones que sean relevantes para sus vidas y experiencia, y que les permitan ver cómo las matemáticas se aplican en su mundo cotidiano. La elección de situaciones y contextos que sean cercanos al alumno puede ayudar a motivar a los estudiantes y hacer que el aprendizaje de las matemáticas sea más significativo para ellos (OCDE, 2013, 34)

En relación a la complejidad de la didáctica propuesta en cada sesión, se podría justificar la elección de un campo de problemas como el transporte público, la música y las series de televisión en la medida en que estos son temas cercanos a los estudiantes, lo que les permitirá involucrarse en el aprendizaje de las matemáticas y comprender cómo las matemáticas se aplican en situaciones reales de su vida cotidiana. Por ejemplo, en la primera sesión se aborda el muestreo y la recolección de datos, lo que podría relacionarse con una encuesta realizada entre los estudiantes sobre sus preferencias musicales o sobre el uso que hacen del transporte público. En la segunda sesión se enseña a los estudiantes cómo representar visualmente los datos mediante gráficos, lo que podría relacionarse con la elaboración de gráficos sobre los medios de transporte preferidos por los estudiantes. En la tercera sesión se abordan medidas de centralización y de posición, lo que podría relacionarse con la identificación de tendencias en las preferencias musicales o en el uso del transporte público.

En la sesión de medidas de dispersión, se podría enseñar el uso de gráficas de cajas y bigotes para representar la variabilidad de los datos. Estas gráficas son útiles para comparar la variabilidad entre distintos grupos de datos y permiten identificar la presencia de valores atípicos. Se podría utilizar esta técnica para analizar la variabilidad en las preferencias musicales y en el uso del transporte público de los estudiantes, y para comparar esta variabilidad entre distintos grupos de estudiantes.

En las sesiones de estadística bidimensional se podría utilizar los mismos datos de géneros musicales en la sesión de estadística bidimensional, lo que permitiría a los estudiantes ver cómo se relacionan dos variables y cómo se puede analizar esta información estadísticamente.

Sesión 1. Evaluación inicial, Muestreo, tablas de datos y frecuencias

Tras la introducción con la secuencia de una película y la evaluación inicial que se han comentado en un capítulo anterior, se pasa a la parte de muestreo, tablas de datos y frecuencias mediante el siguiente ejercicio.

¿Cuál es la serie de televisión favorita de los alumnos de la clase?
Objetivo: Realizar un muestreo, organizar los datos en una tabla de frecuencias y calcular las frecuencias relativas y acumuladas.

Instrucciones:
Dividir a la clase en grupos de 3 alumnos.
Cada grupo tiene la tarea de identificar las series de televisión favoritas de 5 compañeros de clase, seleccionados al azar. Para asegurar una muestra representativa, sugerimos que cada miembro del grupo realice la consulta a uno o dos compañeros para garantizar una distribución adecuada de la muestra. Sin embargo, para evitar la posible tendencia de elegir principalmente a sus amigos, en lugar de seleccionar a los compañeros de clase directamente, que preparen papelitos numerados del 1 al 20 en una bola. Cada número representará a un estudiante en la clase. Los miembros del grupo deberán sacar uno o dos, según sea necesario para sumar un total de cinco.
Anotar los resultados en una tabla de datos, donde cada fila representa a un alumno y la columna contiene la serie de televisión favorita.
Organizar los datos en una tabla de frecuencias, donde las filas representan las diferentes series de televisión, y las columnas contienen la frecuencia absoluta, la frecuencia relativa y la frecuencia acumulada.
Calcular las frecuencias relativas y acumuladas de cada serie de televisión.
Comparar los resultados entre los grupos y discutir las diferencias y similitudes observadas.
Este ejercicio permite a los estudiantes familiarizarse con las técnicas de muestreo y la organización de datos en tablas de frecuencias mientras trabajan en un tema cercano a ellos, como sus series de televisión favoritas. Además, fomenta el trabajo en equipo y la comunicación entre los compañeros.
Después de explorar cómo los estudiantes organizan los datos con métodos que se les ocurran a ellos mismos, contando con palotes por ejemplo, es oportuno introducir el concepto de tablas de frecuencias para una representación más estructurada y precisa de los datos

Supongamos que un grupo de 3 alumnos encuestó a 5 compañeros de clase y obtuvo los siguientes resultados sobre sus series de televisión favoritas: Stranger Things, La Casa de Papel, The Mandalorian, Stranger Things y The Witcher.

Primero, organizamos los datos en una tabla de datos, ver Tabla 4:

Tabla 4
Relación alumno - serie favorita

Alumno	Serie favorita
1	Stranger Things
2	La Casa de Papel
3	The Mandalorian
4	Stranger Things
5	The Witcher

A continuación, ver Tabla 5, creamos una tabla de frecuencias con las columnas "Serie favorita", "Frecuencia absoluta (f)", "Frecuencia relativa (fr)" y "Frecuencia acumulada (F)":

Tabla 5
Relación serie favorita - frecuencias

Serie favorita	f	fr	F
Stranger Things	2	0.4	2
La Casa de Papel	1	0.2	3
The Mandalorian	1	0.2	4
The Witcher	1	0.2	5

La frecuencia relativa se calcula dividiendo la frecuencia absoluta por el total de respuestas (en este caso, 5). La frecuencia acumulada es la suma de las frecuencias absolutas hasta la fila actual. En este ejemplo, el grupo de alumnos encontró que la serie más popular entre los encuestados fue "Stranger Things", con una frecuencia absoluta de 2 y una frecuencia relativa del 40%.

Al resolver este tipo de ejercicios, los alumnos podrían enfrentar algunas **dificultades**, como:
Confusión entre tipos de frecuencias: Los alumnos pueden mezclar conceptos como frecuencia absoluta, relativa y acumulada.
Errores en el conteo de datos: Los estudiantes podrían cometer errores al contar las respuestas en la tabla de datos.
Interpretación de resultados: Pueden tener problemas para entender y describir el significado de los resultados obtenidos.
Organización de datos: Algunos alumnos pueden enfrentar dificultades para organizar y presentar los datos en tablas correctamente.
Comunicación en el grupo.

Aquí nos encontramos con que el tamaño de la muestra es muy pequeño, podemos organizarlo para ver a toda la población, por ejemplo que el grupo 1 pregunte al 2, el 2 al 3, y así sucesivamente hasta el último al primero, para unir todos los datos en una sola tabla.

Sesión 2. Organización visual de los datos estadísticos

Los estudiantes deben analizar los **tipos de transporte** que utilizan sus compañeros de clase **para llegar al instituto**.

Realizamos una encuesta en la clase, que hace el delegado preguntando en los 5 primeros minutos, preguntando a cada compañero qué tipo de transporte utiliza para llegar al instituto (a pie, en bicicleta, transporte público, coche, etc.).

Una vez que hayan recopilado los datos, los estudiantes deben organizarlos en una tabla de frecuencias, donde se muestre la frecuencia absoluta de cada tipo de transporte. A continuación, cada grupo debe elegir dos representaciones gráficas diferentes (por ejemplo, gráfico de barras y gráfico circular) para visualizar la información obtenida. Los alumnos deben crear los gráficos utilizando papel, lápiz y reglas. Por último, cada grupo presentará sus gráficos al resto de la clase y explicará sus observaciones sobre las preferencias de transporte de sus compañeros. El profesor puede guiar a los alumnos en un debate sobre las ventajas y desventajas de los diferentes tipos de gráficos y cómo la elección del gráfico afecta la interpretación de los datos. Por mostrar un posible desarrollo de la sesión, ver Tabla 6, supongamos que después de realizar la encuesta en una clase de 20 alumnos obtuvimos los siguientes resultados:

Tabla 6
Relación alumno - medio de transporte

Alumno	Medio de transporte
1	A pie
2	Bicicleta
3	Transporte público
4	Transporte público
5	Coche
6	A pie
7	A pie
8	Bicicleta
9	Coche
10	A pie
11	A pie
12	Transporte público
13	A pie
14	Coche
15	Transporte público
16	Bicicleta

17	A pie
18	Coche
19	A pie
20	Transporte público

Ahora, ver Tabla 7, vamos a organizar los datos en una tabla de frecuencias y a crear los gráficos de barras y circular, ver Figuras 13 y 14.

Tabla 7
Relación tipo de transporte - frecuencia absoluta

Tipo de transporte	Frecuencia absoluta
A pie	8
Bicicleta	3
Transporte público	5
Coche	4

Figura 13
Diagrama de barras

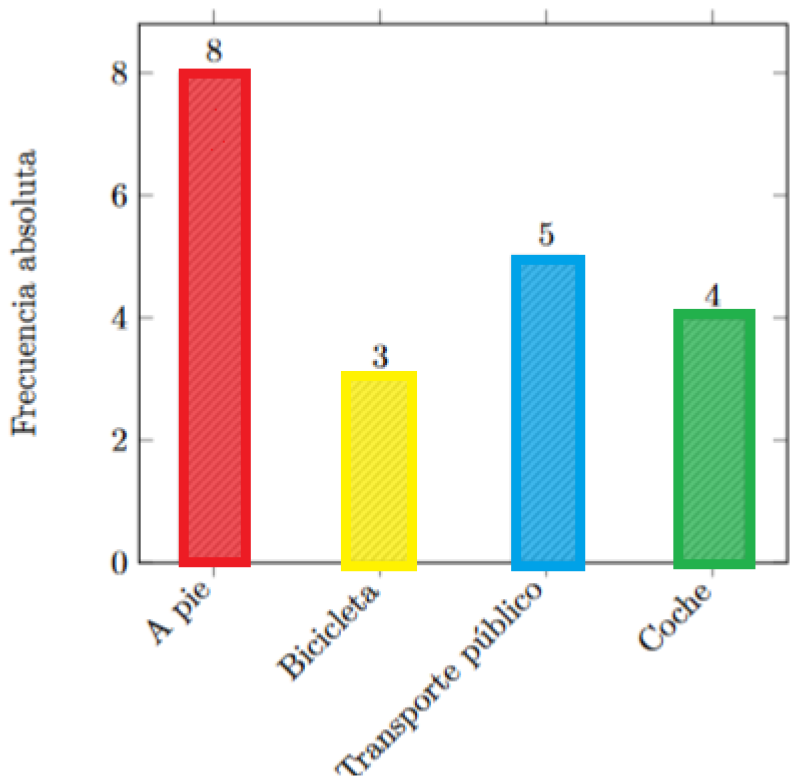
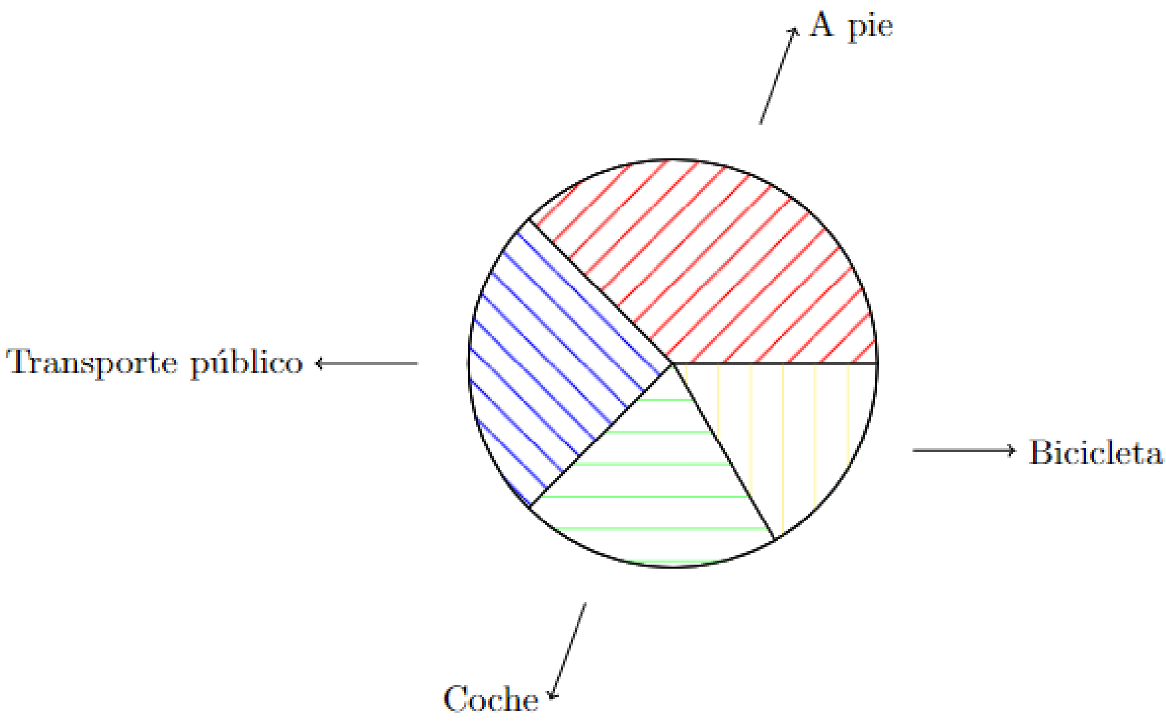


Figura 14
Gráfico circular



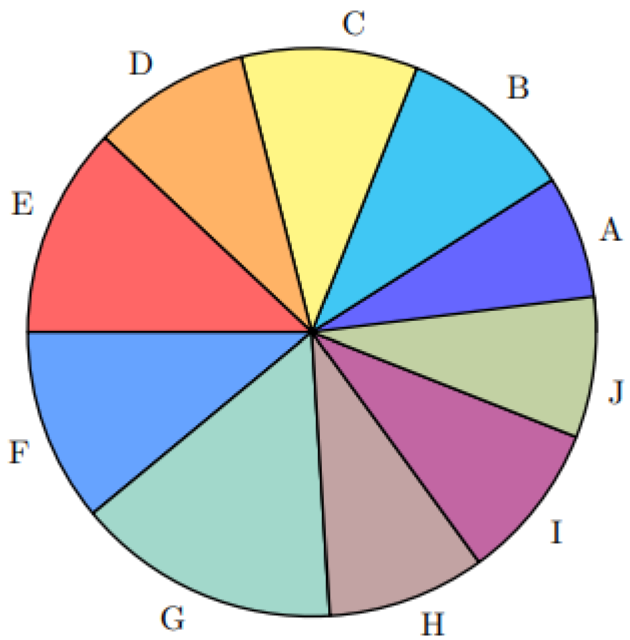
Una posible **discusión** en el aula sobre las ventajas y desventajas de los gráficos de barras y circulares podría incluir los siguientes puntos:

Ventajas del gráfico de barras:
Facilita la comparación de las frecuencias absolutas de cada categoría.
Es intuitivo y fácil de entender para la mayoría de las personas.
Puede representar datos en el eje horizontal o vertical, lo que permite adaptar la visualización según las necesidades.
Desventajas:
No muestra directamente las proporciones o porcentajes de cada categoría en relación con el total.
Puede ser difícil de interpretar si hay muchas categorías o si las diferencias en las frecuencias son pequeñas.

Ventajas del gráfico circular:
Muestra claramente las proporciones o porcentajes de cada categoría en relación con el total.
Es útil para visualizar rápidamente la composición de un conjunto de datos.
Puede ser más atractivo visualmente que un gráfico de barras.
Desventajas:

La comparación de las frecuencias absolutas de cada categoría puede ser más difícil que en un gráfico de barras.
Puede ser difícil de interpretar si hay muchas categorías o si las proporciones son muy similares.
La percepción del tamaño de las áreas en un gráfico circular puede ser complicada, especialmente si hay muchas categorías o sectores de tamaño similar, ver Figura 15.

Figura 15
Ejemplo de gráfico circular difícil de distinguir



La discusión en el aula podría llevar a los estudiantes a comprender en qué situaciones es más apropiado utilizar un gráfico de barras y en qué casos un gráfico circular es más adecuado. Además, es importante destacar que cada tipo de gráfico tiene sus propias ventajas y limitaciones, por lo que es fundamental seleccionar el gráfico adecuado según el contexto y el objetivo del análisis.

Sesión 3. Medidas de centralización y de posición

Se desea conocer las **preferencias musicales** de los alumnos de 4º de la ESO para adaptar su programa a los gustos de los estudiantes.

Se realiza una encuesta donde se pone la cantidad de tiempo que los estudiantes pasan escuchando diferentes géneros musicales a lo largo de una semana. Vamos a orientar nuestro enfoque para descubrir y entender conceptos como la media, la mediana y la moda. Por ejemplo, podríamos preguntarnos: ¿Cuál creen que es la música más en moda actualmente? ¿Y la más escuchada? Puede hacerla el delegado de clase, o apuntarlo alzando la mano en los primeros minutos.

La tabla siguiente muestra el tiempo en minutos que 20 estudiantes pasaron escuchando cada género musical que ellos mismos han puesto libremente en la encuesta. Una vez que se hayan calculado las medidas de centralización y posición, los estudiantes podrán comparar los géneros musicales en función de la cantidad de tiempo que pasan escuchándolos y discutir las preferencias musicales en el aula. También podrían analizar cómo se distribuyen los tiempos de escucha y qué géneros tienen una mayor variabilidad en el tiempo de escucha entre los estudiantes.

El **ejercicio** es calcular la media, la mediana y la moda del tiempo que los estudiantes pasan escuchando cada género musical.

Por mostrar un posible desarrollo de la sesión, supongamos que después de realizar la encuesta en una clase de 20 alumnos, obtuvimos los siguientes resultados, ver Tabla 8:

Tabla 8
Relación estudiante - minutos de música escuchados por género musical

Estudiante	Pop	Hip-Hop	Electrónica	Trap	Reggaeton
1	90	30	60	45	120
2	100	50	80	0	110
3	80	40	70	60	100
4	85	0	75	65	95
5	95	25	50	55	0
6	0	60	90	50	105
7	75	35	80	0	90
8	120	45	0	40	115
9	100	50	60	0	130
10	90	40	0	45	110
11	80	0	60	65	120
12	95	55	75	0	100
13	100	25	80	60	0
14	110	60	0	55	125
15	75	35	50	0	105
16	120	0	65	40	90
17	100	50	0	55	115
18	90	40	70	45	0
19	80	30	0	50	110
20	95	35	75	65	120

Para calcular la media de cada género musical, sumamos los valores de cada género y dividimos por el número total de estudiantes (20 en este caso). Las cuentas explícitas para cada género serían las siguientes:

Pop:
(90 + 100 + 0 + 85 + 95 + 0 + 75 + 120 + 100 + 90 + 80 + 95 + 100 + 110 + 75 + 120 + 100 + 90 + 80 + 95) / 20
Hip-Hop:
(30 + 50 + 0 + 0 + 25 + 60 + 35 + 45 + 50 + 40 + 0 + 55 + 25 + 60 + 35 + 0 + 50 + 40 + 30 + 35) / 20
Electrónica:
(60 + 80 + 0 + 75 + 50 + 90 + 80 + 0 + 60 + 0 + 60 + 75 + 80 + 0 + 50 + 65 + 0 + 70 + 0 + 75) / 20
Trap:
(45 + 0 + 0 + 65 + 55 + 50 + 0 + 40 + 0 + 45 + 65 + 0 + 60 + 55 + 0 + 40 + 55 + 45 + 50 + 80) / 20
Reggaeton:

(120 + 110 + 0 + 95 + 0 + 105 + 90 + 115 + 130 + 110 + 120 + 100 + 0 + 125 + 105 + 90 + 115 + 0 + 110 + 120) / 20

Realizando las operaciones y redondeando a dos decimales, obtenemos las **medias** para cada género en minutos.
Pop: 85
Hip-Hop: 33.25
Electrónica: 48.5
Trap: 37.5
Reggaeton: 88

Para calcular la mediana de cada género musical, primero necesitamos ordenar los valores de cada género en orden ascendente y luego encontrar el valor central, ver Tabla 9. Si hay un número par de valores, tomamos el promedio de los dos valores centrales.

Tabla 9
Minutos por género musical ordenados de menor a mayor

	1º	2º	3º	4º	5º	6º	7º	8º	9º	**10º**	**11º**	12º	13º	14º	15º	16º	17º	18º	19º	20º
Pop:	0	0	75	75	80	80	85	90	90	**90**	**95**	95	95	100	100	100	100	110	120	120
Hip-Hop:	0	0	0	25	25	30	30	35	35	**35**	**40**	40	45	50	50	50	55	60	60	60
Electró.:	0	0	0	0	50	50	60	60	60	**65**	**70**	75	75	80	80	80	80	90	95	120
Trap:	0	0	0	0	0	40	40	45	45	**50**	**50**	55	55	55	60	65	65	80	95	120
Reggaeton:	0	0	0	0	90	90	95	100	105	**105**	**110**	110	110	115	115	120	120	120	125	130

Debido a que hay 20 valores (un número par), tomamos el promedio de los dos valores centrales (10º y 11º) para calcular la **mediana**:
Pop: (90 + 95) / 2 = 92.5
Hip-Hop: (35 + 40) / 2 = 37.5
Electrónica: (65 + 70) / 2 = 67.5
Trap: (50 + 50) / 2 = 50
Reggaeton: (105 + 110) / 2 = 107.5

Para calcular la **moda**, encontramos el valor que ocurre con más frecuencia en cada género musical. Si hay varios valores con la misma frecuencia máxima, todos ellos serán modas.
Moda de Pop: 100 (cuatro veces)
Moda de Hip-Hop: 0, 50, 60 (tres veces cada uno)
Moda de Electrónica: 0, 80 (cuatro veces cada uno)
Moda de Trap: 0 (cinco veces)
Moda de Reggaeton: 0 (cuatro veces)

Tabla 10

Relación género, media, mediana y moda

Género	$\bar{x}$	Mediana	Moda
Pop	85.00	92.5	100
Hip-Hop	33.25	37.5	0, 50, 60
Electrónica	48.5	67.5	0, 80
Trap	37.5	50	0
Reggaeton	88	107.5	0

Alguna conclusión de la Tabla 10 que se podría sacar en la **discusión posterior** podría ser que en la tabla de resultados hay una gran diferencia entre la media y mediana del tiempo de escucha para algunos géneros musicales, como el Hip-Hop y la Electrónica, lo que indica que estos géneros tienen una distribución sesgada hacia valores bajos.

Por otro lado, el Reggaeton tiene una media mucho más alta que su mediana, lo que indica que algunos estudiantes pasan mucho más tiempo escuchando este género musical en comparación con el resto del grupo.

En una distribución sesgada hacia la izquierda, hay más valores bajos que altos, lo que hace que la media sea menor que la mediana, de hecho hay alumnos que no escuchan en absoluto algún género musical. En otras palabras, algunos valores extremadamente bajos en el conjunto de datos podrían estar influyendo en la media y disminuyendo su valor en comparación con la mediana.

Tener modas múltiples en un conjunto de datos puede indicar que la distribución no es unimodal, sino bimodal o multimodal. En una distribución bimodal o multimodal, hay más de un pico en la distribución de los datos, lo que sugiere que hay múltiples grupos o categorías dentro de los datos. La presencia de modas múltiples puede ser el resultado de diferentes subpoblaciones o factores que afectan los valores en el conjunto de datos. También podría ser una coincidencia en conjuntos de datos pequeños.

Es importante comunicar aquí a los alumnos que la media, la mediana y la moda son solo medidas resumidas de un conjunto de datos; para comprender más profundamente la distribución y las tendencias en un conjunto de datos, es útil utilizar gráficos y otras medidas estadísticas, como el rango intercuartil, la varianza y la desviación estándar, tema que se abordará en la siguiente sesión.

Sesión 4. Medidas de dispersión

En esta sesión, se introducirán las medidas de dispersión (rango, varianza y desviación típica) para describir cómo de dispersos están los datos alrededor de la media. Además, aprenderán a representar gráficamente los datos utilizando gráficos de cajas y bigotes.

Después de que los alumnos hayan completado sus cálculos y gráficos, los grupos presentarán sus resultados y compartirán sus observaciones sobre las diferencias en las distribuciones de los géneros musicales. Analizarán cómo las medidas de dispersión y los gráficos de cajas y bigotes pueden ayudar a entender la variabilidad y la tendencia central de los datos.

Es importante que los alumnos discutan sobre las diferencias que encuentren en los rangos, las varianzas y las desviaciones típicas de cada género musical, así como también identifiquen datos atípicos o patrones interesantes en los gráficos de cajas y bigotes.

Durante la discusión en clase, el profesor debe guiar a los alumnos para que entiendan cómo estas medidas de dispersión y gráficos pueden ser útiles en la toma de decisiones y en la interpretación de datos en contextos reales. También se puede enfatizar la importancia del trabajo en equipo.

Como este ejercicio conlleva la repetición de cuentas, se pueden dividir los grupos por filas por ejemplos, para que cada grupo solo tenga que hacer los cálculos de un género musical, para luego ponerlos en común y compararlos entre grupos si a varios les toca el mismo.

Ejercicio: En grupos de 3, y utilizando los datos de la tabla de géneros musicales que se halló en la sesión 3, calcular el rango, ver Tabla 11, la varianza y la desviación típica para cada género musical. A continuación, elaborar gráficos de cajas y bigotes para cada género musical para visualizar la distribución de los datos (para lo cual deberán de calcular los rangos intercuartílicos).

Tabla 11

Minutos por género musical ordenados de menor a mayor

	1º	2º	3º	4º	5º	6º	7º	8º	9º	10º	11º	12º	13º	14º	15º	16º	17º	18º	19º	**20º**
Pop:	**0**	0	75	75	80	80	85	90	90	90	95	95	95	100	100	100	100	110	120	**120**
Hip-Hop:	**0**	0	0	25	25	30	30	35	35	35	40	40	45	50	50	50	55	60	60	**60**
Electró.:	**0**	0	0	0	50	50	60	60	60	65	70	75	75	80	80	80	80	90	95	**120**
Trap:	**0**	0	0	0	0	40	40	45	45	50	50	55	55	55	60	65	65	80	95	**120**
Reggaeton:	**0**	0	0	0	90	90	95	100	105	105	110	110	110	115	115	120	120	120	125	**130**

Calculamos el **rango** para cada género musical:

Pop: 120 (máximo) - 0 (mínimo) = 120
Hip-Hop: 60 (máximo) - 0 (mínimo) = 60
Electrónica: 120 (máximo) - 0 (mínimo) = 120
Trap: 120 (máximo) - 0 (mínimo) = 120
Reggaeton: 130 (máximo) - 0 (mínimo) = 130
Como en todos los géneros hay algún alumno que escucha 0, coincide con cada valor máximo.

Para calcular la varianza de cada género musical necesitamos la media, ya calculada en la sesión anterior; Pop: 85, Hip-Hop: 33.25, Electrónica: 48.5, Trap: 37.5, Reggaeton: 88.

Se nos podría ocurrir calcular la **varianza** con la fórmula:

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

Por lo que en Pop, por ejemplo, realizaríamos estas cuentas:
$((90-85)^2 + (100-85)^2 + (0-85)^2 + (85-85)^2 + (95-85)^2 + (0-85)^2 + (75-85)^2 + (120-85)^2 + (100-85)^2 + (90-85)^2 + (80-85)^2 + (95-85)^2 + (100-85)^2 + (110-85)^2 + (75-85)^2 + (120-85)^2 + (100-85)^2 + (90-85)^2 + (80-85)^2 + (95-85)^2) / 20 = 952.5$

pero en seguida nos daremos cuenta en clase que implica menos cuentas la fórmula:

$$\sigma^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \mu^2$$

así evitamos hacer n-1 restas, quedando, para cada género
Pop: $(90^2 + 100^2 + 0^2 + 85^2 + 95^2 + 0^2 + 75^2 + 120^2 + 100^2 + 90^2 + 80^2 + 95^2 + 100^2 + 110^2 + 75^2 + 120^2 + 100^2 + 90^2 + 80^2 + 95^2)/20 - 85^2 = 952.5$
Hip-Hop: $(30^2 + 50^2 + 0^2 + 0^2 + 25^2 + 60^2 + 35^2 + 45^2 + 50^2 + 40^2 + 0^2 + 55^2 + 25^2 + 60^2 + 35^2 + 0^2 + 50^2 + 40^2 + 30^2 + 35^2)/20 - 33.25^2 = 378.19$
Electrónica: $(60^2 + 80^2 + 0^2 + 75^2 + 50^2 + 90^2 + 80^2 + 0^2 + 60^2 + 0^2 + 60^2 + 75^2 + 80^2 + 0^2 + 50^2 + 65^2 + 0^2 + 70^2 + 0^2 + 75^2)/20 - 48.5^2 = 1102.75$
Trap: $(45^2 + 0^2 + 0^2 + 65^2 + 55^2 + 50^2 + 0^2 + 40^2 + 0^2 + 45^2 + 65^2 + 0^2 + 60^2 + 55^2 + 0^2 + 40^2 + 55^2 + 45^2 + 50^2 + 80^2)/20 - 37.5^2 = 683.75$
Reggaeton: $(120^2 + 110^2 + 0^2 + 95^2 + 0^2 + 105^2 + 90^2 + 115^2 + 130^2 + 110^2 + 120^2 + 100^2 + 0^2 + 125^2 + 105^2 + 90^2 + 115^2 + 0^2 + 110^2 + 120^2)/20 - 88^2 = 2043.50$

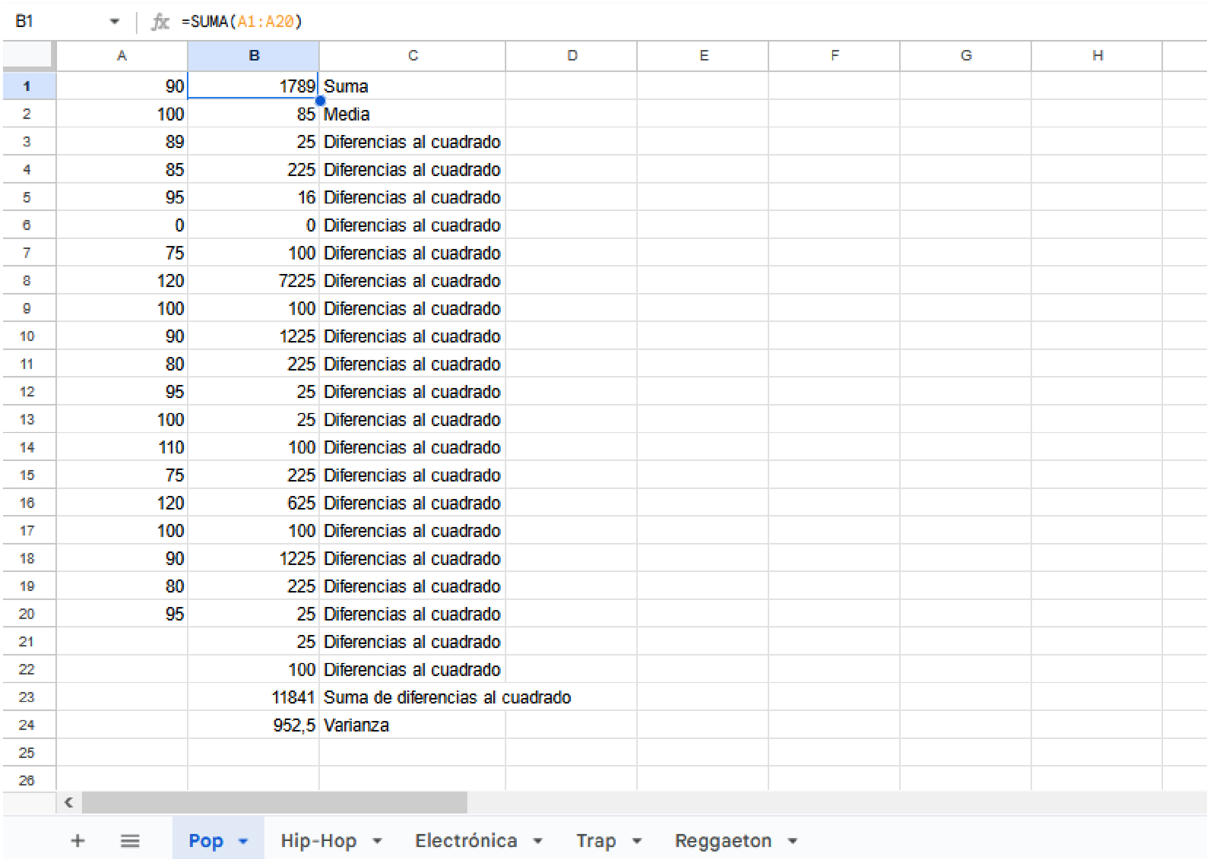
- UTILIZACIÓN DEL ORDENADOR -

Son cuentas repetitivas, que podrían desanimar, aún así recordemos que las habremos dividido entre los grupos, además, en el caso de los valores 0 no habría que hacer el cuadrado, los hemos puesto simplemente para asegurarnos de que no nos dejamos ningún valor. De todas formas aquí les vamos a ofrecer la opción de realizar las cuentas mediante **hojas de cálculo**, por ejemplo *Google Sheets*, ver Figura 16, lo que luego puede resultar en

un entregable evaluable. Si los datos están en las celdas A1 hasta A20, el procedimiento sería:
Calcular la suma total: En la celda B1, escribe =**SUM**(A1:A20) para calcular la suma total de los datos.
Calcular la media: En la celda B2, escribe =B1/20 para calcular la media (suma total dividida por el número de elementos, que en este caso son 20).
Calcular las diferencias al cuadrado: En la celda B3, escribe =(A1-B2)^2, esto calculará el cuadrado de la diferencia entre el dato y la media. Luego puedes arrastrar hacia abajo esta fórmula hasta la celda B22 para hacer el mismo cálculo para todos los datos.
Sumar todas las diferencias al cuadrado: En la celda B23, escribe =**SUM**(B3:B22) para sumar todas las diferencias al cuadrado.
Calcular la varianza: Finalmente, en la celda B24, escribe =B23/20 para obtener la varianza, que es la suma de las diferencias al cuadrado dividida por el número de elementos.
Este sería el procedimiento para calcular la varianza de los datos en las celdas A1 hasta A20 sin usar la función predefinida de Google Sheets, **VAR.P**, que está bien que la conozcan los alumnos pero que si la utilizas directamente no interiorizas el cálculo de la varianza.

Figura 16

Captura de pantalla de Google Sheets



B1 =SUMA(A1:A20)

	A	B	C	D	E	F	G	H
1	90	1789	Suma					
2	100	85	Media					
3	89	25	Diferencias al cuadrado					
4	85	225	Diferencias al cuadrado					
5	95	16	Diferencias al cuadrado					
6	0	0	Diferencias al cuadrado					
7	75	100	Diferencias al cuadrado					
8	120	7225	Diferencias al cuadrado					
9	100	100	Diferencias al cuadrado					
10	90	1225	Diferencias al cuadrado					
11	80	225	Diferencias al cuadrado					
12	95	25	Diferencias al cuadrado					
13	100	25	Diferencias al cuadrado					
14	110	100	Diferencias al cuadrado					
15	75	225	Diferencias al cuadrado					
16	120	625	Diferencias al cuadrado					
17	100	100	Diferencias al cuadrado					
18	90	1225	Diferencias al cuadrado					
19	80	225	Diferencias al cuadrado					
20	95	25	Diferencias al cuadrado					
21		25	Diferencias al cuadrado					
22		100	Diferencias al cuadrado					
23		11841	Suma de diferencias al cuadrado					
24		952,5	Varianza					
25								
26								

Pop | Hip-Hop | Electrónica | Trap | Reggaeton

Sacamos la raíz cuadrada a la varianza para saber la **desviación típica** para cada género musical,:
Pop: $\sqrt{952.5} \approx 30.86$
Hip-Hop: $\sqrt{378.19} \approx 19.44$
Electrónica: $\sqrt{1102.75} \approx 33.20$
Trap: $\sqrt{683.75} \approx 26.14$
Reggaeton: $\sqrt{2043.50} \approx 45.20$

En resumen, los resultados de la Tabla 12.

Tabla 12
Relación género - rango, varianza, desviación típica y media

Género	Rango	Varianza	σ	μ
Pop	120	952.50	30.86	85
Hip-Hop	60	378.19	19.44	33.25
Electrónica	120	1102.75	33.20	48.5
Trap	120	683.75	26.14	37.5
Reggaeton	130	2043.50	45.20	88

Aquí sería buen momento para **comentar** que la varianza es el promedio de las diferencias al cuadrado entre cada observación y la media. Sin embargo, al estar en unidades al cuadrado, puede resultar menos intuitiva que la desviación típica, que se expresa en las mismas unidades que los datos originales. Esto hace que sea más fácil de interpretar y comparar con las observaciones del conjunto de datos. En general, la desviación típica es más comúnmente utilizada debido a su facilidad de interpretación, mientras que la varianza se necesita cuando se trabaja con modelos estadísticos.

Para calcular los datos necesarios para elaborar **gráficos de cajas y bigotes** debemos de obtener el primer cuartil (Q1), la mediana (Q2) -ya calculada en la sesión anterior-, el tercer cuartil (Q3), el rango intercuartílico (IQR) y el bigote superior e inferior para cada género musical.

Datos ordenados de menor a mayor (marcamos en negrita los datos que marcan el límite de los cuartiles Q1 y Q3, ver Tabla 13)

Tabla 13

Minutos por género musical ordenados de menor a mayor

	1º	2º	3º	4º	**5º**	**6º**	7º	8º	9º	10º	11º	12º	13º	14º	**15º**	**16º**	17º	18º	19º	20º
Pop:	0	0	75	75	**80**	**80**	85	90	90	90	95	95	95	100	**100**	**100**	100	110	120	120
Hip-Hop:	0	0	0	25	**25**	**30**	30	35	35	35	40	40	45	50	**50**	**50**	55	60	60	60
Electró.:	0	0	0	0	**50**	**50**	60	60	60	65	70	75	75	80	**80**	**80**	80	90	95	120
Trap:	0	0	0	0	**0**	**40**	40	45	45	50	50	55	55	55	**60**	**65**	65	80	95	120
Reggaeton:	0	0	0	0	**90**	**90**	95	100	105	105	110	110	110	115	**115**	**120**	120	120	125	130

Recordemos que ya habíamos calculado en sesiones anteriores el **Q2 (Mediana)**: Pop:= 92.5, Hip-Hop: = 37.5, Electrónica = 67.5, Trap = 52.5, Reggaeton = 107.5

Q1 como media entre el 5º y el 6º dato: Pop: (80 + 80) / 2 = 80, Hip-Hop: (25 + 30) / 2 = 27.5, Electrónica: (50 + 50) / 2 = 50, Trap: (0 + 40) / 2 = 20, Reggaeton: (90 + 90) / 2 = 90

Q3 como media entre el 15º y 16º dato: Pop: (100 + 100) / 2 = 100, Hip-Hop: (50 + 50) / 2 = 50, Electrónica: (80 + 80) / 2 = 80, Trap: (60 + 65) / 2 = 62.5, Reggaeton: (115 + 120) / 2 = 117.5

IQR (Rango intercuartílico): Pop: 100 - 80 = 20, Hip-Hop: 50 - 40 = 10, Electrónica: 80 - 50 = 30, Trap: 62.5 - 20 = 42.5, Reggaeton: 117.5- 90 = 27.5

Calculamos hasta dónde tienen que llegar el límite superior e inferior de los bigotes:
Bigote Inferior Q1-IQR*1.5
Pop = 80 - 20*1.5 = 50 (0,0 **datos atípicos**), Hip-Hop = 40 - 10*1.5 = 25 (0,0,0 atípicos), Electrónica = 50 - 30*1.5 = 5 (0,0,0,0 atípicos), Trap = 40 - 25*1.5 = 2.5 (0,0,0,0 atípicos), Reggaeton: 90 - 27.5*1.5 = 48.75 (0,0,0,0 atípicos)
Bigote Superior Q3+IQR*1.5
Pop= 100+20*1.5 = 130 ~ ~ se queda **limitado** en 129 ya que es el valor máximo, Hip-Hop = 50 + 10*1.5 = 65 ~ 60 máx., Electrónica = 80 + 30*1.5 = 125 ~ 120 máx., Trap = 62.5 + 42.5*1.5 = 126.25 ~ 120 máx., Reggaeton = 117.5 + 27.5*1.5 = 158.75 ~ 130 máx.

Resumimos los datos en la Tabla 14 y mostramos el gráfico de cajas y bigotes en la Figura 17

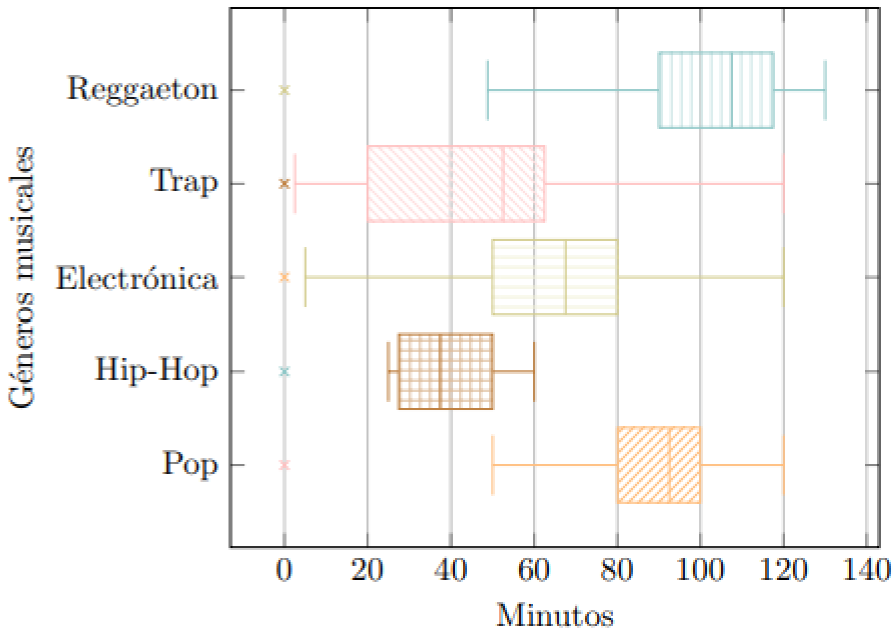
Tabla 14

Valores necesarios para elaborar el gráfico de cajas y bigotes

Valor	Pop	Hip-Hop	Electrónica	Trap	Reggaeton
Bigote Inferior	50	25	5	2.5	48.75
Q1	80	27.5	50	20	90
Q2	92.5	37.5	67.5	52.5	107.5
Q3	100	50	80	62.5	117.5
Bigote Superior	120	60	120	120	130
Datos Atípicos	0,0	0,0,0	0,0,0,0	0,0,0,0	0,0,0,0

Figura 17

Gráfico de cajas y bigotes



Una posible **discusión** que se podría tener en clase para interpretar estas gráficas sería: El género de música Pop tiene un rango intercuartílico (IQR) de 20 minutos, lo que indica una menor dispersión en comparación con los otros géneros, especialmente Trap y Reggaeton. La mayoría de los estudiantes escuchan entre 80 y 100 minutos de música Pop por semana.

El género Hip-Hop muestra una dispersión aún menor, con un IQR de 10 minutos, lo que indica que los hábitos de escucha entre los estudiantes son más consistentes en este género. La mayoría de los estudiantes escuchan entre 40 y 50 minutos de música Hip-Hop por semana. La música electrónica tiene un IQR de 30 minutos y un rango total de escucha de 5 a 120 minutos por semana, este rango más amplio sugiere una mayor diversidad en las preferencias de escucha entre los estudiantes en este género.

El género Trap muestra la mayor dispersión entre los géneros, con un IQR de 42.5 minutos y un rango total de escucha de 2.5 a 120 minutos por semana. Esto sugiere que las preferencias de escucha en este género varían significativamente entre los estudiantes.

El género Reggaeton también presenta una dispersión considerable, con un IQR de 27.5 minutos y un rango total de escucha de 0 a 130 minutos por semana. La mayoría de los estudiantes escuchan entre 90 y 117.5 minutos de música Reggaeton por semana.

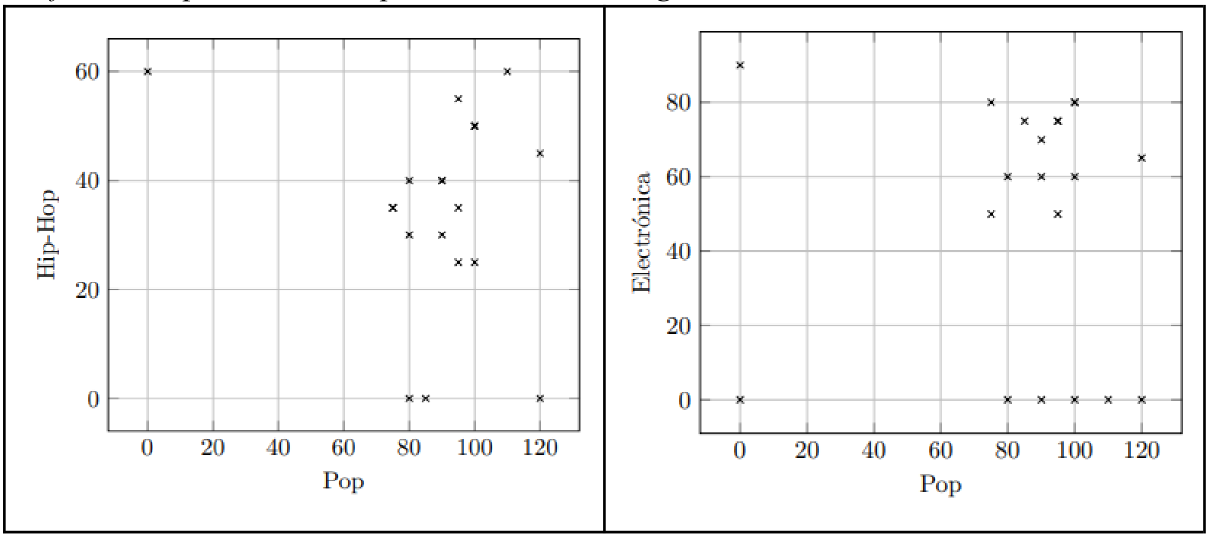
Todos los géneros tienen valores atípicos en el extremo inferior, lo que indica que hay estudiantes que escuchan una cantidad mucho menor de música en cada género que sus compañeros. Esto puede deberse a diferencias en las preferencias musicales o en la cantidad total de tiempo dedicado a escuchar música.

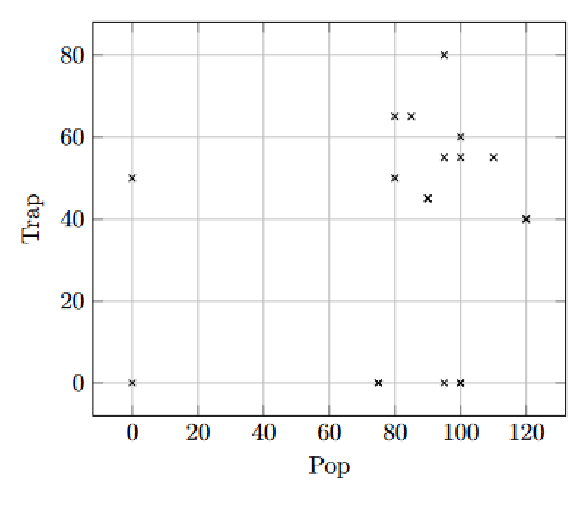
El género Pop es el más popular en términos de tiempo de escucha, seguido de cerca por Reggaeton. Los géneros Hip-Hop y Electrónica tienen rangos intercuartílicos más estrechos, lo que indica una menor dispersión en los hábitos de escucha entre los estudiantes. Por último, el género Trap presenta la mayor dispersión en los hábitos de escucha.

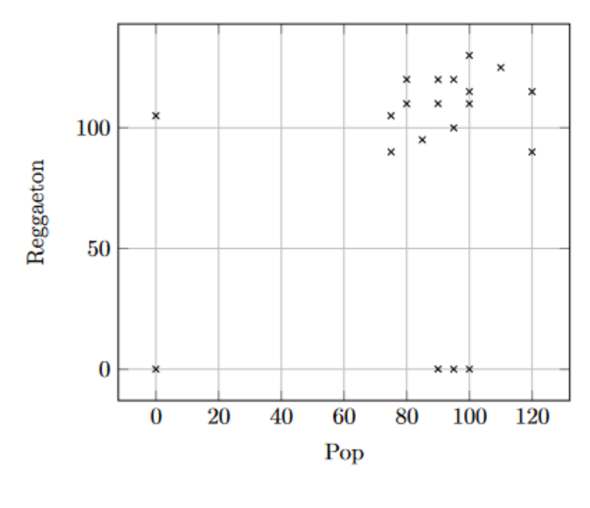
Sesión 5. Estadística bidimensional (parte 1)

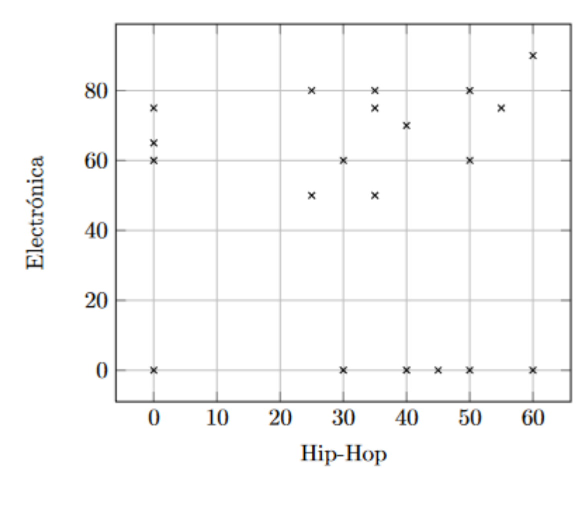
Seguimos con el ejercicio sobre los géneros musicales. Ahora queremos estudiar si existe relación entre el tiempo de escucha entre diferentes géneros. Utilizaremos la misma tabla de datos de la sesión 3. El **ejercicio** consiste en que divididos por grupos, seleccionamos los géneros musicales de dos en dos y representamos gráficamente, ver Figura 18, los datos en un diagrama de dispersión para cada par de géneros y se discute en grupo a raíz de la gráfica si hay alguna relación (fuerte o débil) aparente entre las dos variables. Después se pasa a calcular el coeficiente de correlación de Pearson para determinar si existe una relación lineal entre el tiempo dedicado a escuchar cada pareja de géneros.¿Creéis que existe alguna relación entre el tiempo dedicado a escuchar diferentes géneros musicales? ¿Por qué o por qué no? Comparte tus conclusiones y discute si la estadística bidimensional les ha proporcionado información útil. Recalcar aquí que los datos de las sesiones son de elaboración propia, y que en la sesión que se hiciese en clase, los datos serían diferentes y por tanto las conclusiones serían diferentes también.

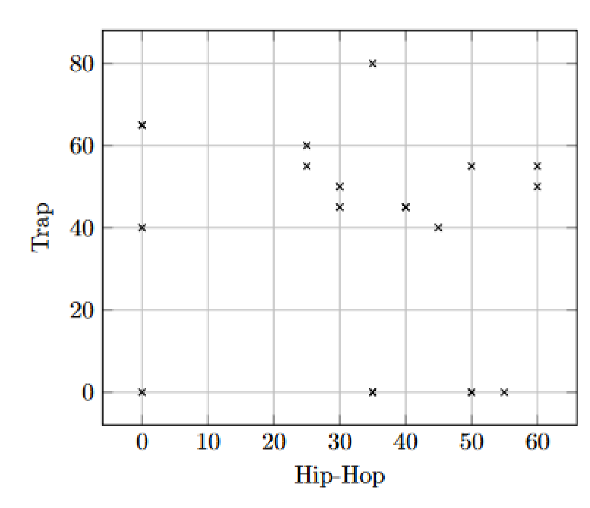
Figura 18
Gráfico de dispersión de tiempos de escucha entre géneros musicales

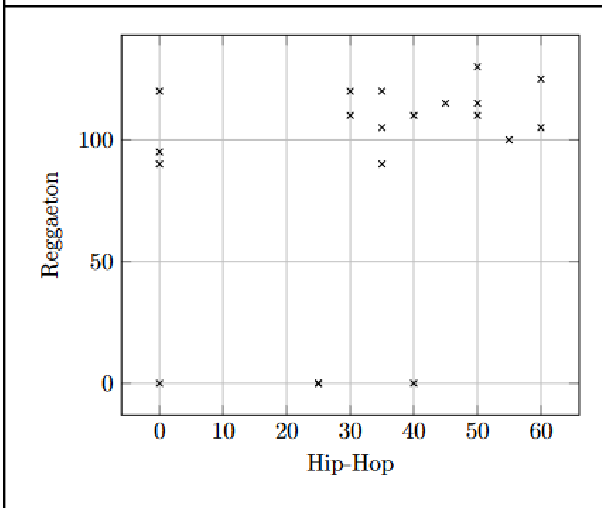


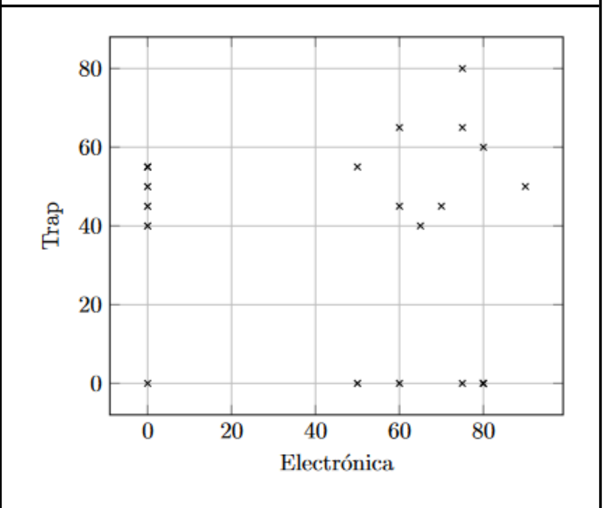

Trap
0
20
40
60
80
0 20 40 60 80 100 120
Pop

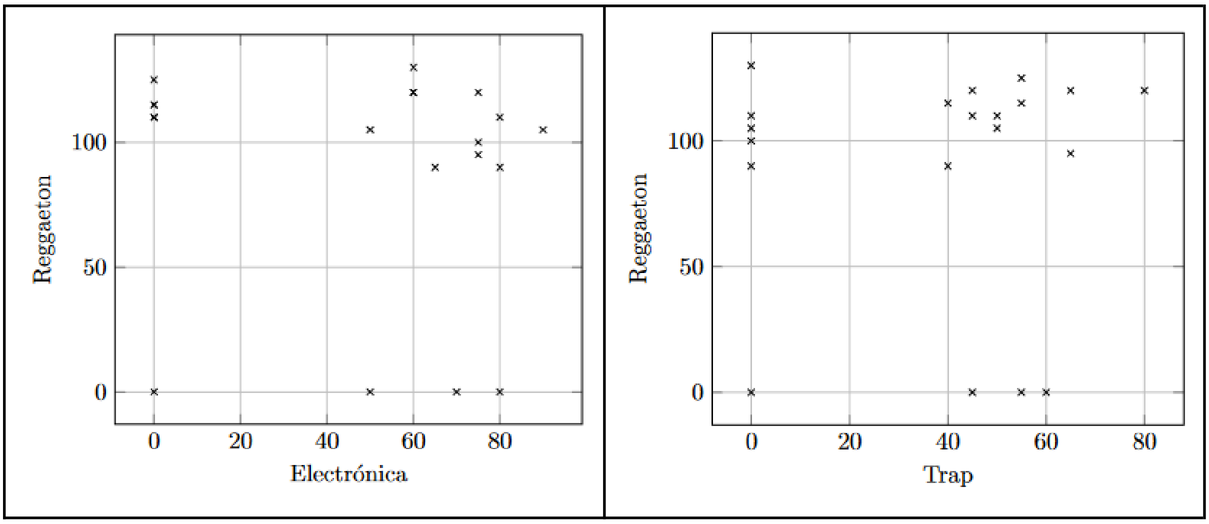

Reggaeton
0
50
100
0 20 40 60 80 100 120
Pop


Electrónica
0
20
40
60
80
0 10 20 30 40 50 60
Hip-Hop


Trap
0
20
40
60
80
0 10 20 30 40 50 60
Hip-Hop


Reggaeton
0
50
100
0 10 20 30 40 50 60
Hip-Hop


Trap
0
20
40
60
80
0 20 40 60 80
Electrónica



Para calcular el coeficiente de correlación de Pearson entre Pop y Hip-Hop necesitamos aplicar la fórmula

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Las medias de cada variable, 85 para Pop y 33.25 para Hip-Hop, ya se calcularon en la sesión 3.

Como puntualización, añadir que estas operaciones pueden presentar dificultades en la utilización de la calculadora, ya que se suele observar que la precedencia de operadores suele causar confusión en algunos alumnos.

Diferencia entre cada observación y su media correspondiente:
Pop - μ(Pop) = 5, 15, -85, 0, 10, -85, -10, 35, 15, 5, -5, 10, 15, 25, -10, 35, 15, 5, -5, 10
Hip-Hop - μ(Hip-Hop) = -3.25, 16.75, -33.25, -33.25, -8.25, 26.75, 1.75, 11.75, 16.75, 6.75, -33.25, 21.75, -8.25, 26.75, 1.75, -33.25, 16.75, 6.75, -3.25, 1.75

Sumatorio para la parte superior de la fracción:
∑((Pop - μ(Pop))(Hip-Hop - μ(Hip-Hop))) = -16.25 + 251.25 + 2828.75 + 0 -82.5 -2272.5 -17.5 + 411.25 + 251.25 + 33.75 + 166.25 + 217.5 -123.75 + 668.75 -17.5 + 1162.5 + 251.25 + 33.75 + 16.25 + 17.5 = 1450

Diferencia entre cada observación y su media correspondiente elevada al cuadrado
$\sum(\text{Pop} - \mu(\text{Pop}))^2$ = 25 + 225 + 7225 + 0 + 100 + 7225 + 100 + 1225 + 225 + 25 + 25 + 100 + 225 + 625 + 100 + 1225 + 225 + 25 + 25 + 100 = 19050

∑(Hip-Hop μ(Hip-Hop))^2 = 10.562 + 280.562 + 1105.562 + 1105.562 + 68.062 + 715.562 + 3.062 + 138.062 + 280.562 + 45.562 + 1105.562 + 473.062 + 68.062 + 715.562 + 3.062 + 1105.562 + 280.562 + 45.562 + 10.562 + 3.062 = 7563.75

Calculamos r
r = 1450 / √((19050)(7563.75)) = 0.1208
Aunque técnicamente es una correlación positiva, la relación entre sus variables es débil

- UTILIZACIÓN DEL ORDENADOR -

En esta sesión también vamos a ofrecer la opción de realizar las cuentas mediante **hojas de cálculo**, por ejemplo *Google Sheets*, lo que luego puede resultar en un entregable evaluable. Es importante tener en cuenta que *Google Sheets* ofrece funciones incorporadas para calcular el coeficiente de correlación de Pearson, como **CORREL**. Sin embargo, se desea hacer todo el proceso manualmente paso a paso:

Supongamos que tienes las observaciones de Pop en la columna A y Hip-Hop en la columna B. Para calcular las diferencias entre la media se colocarían en la columna C para Pop y la columna D para Hip-Hop.

Después de tener las diferencias, debes multiplicar las diferencias de Pop y Hip-Hop entre sí para cada observación y sumarlas todas. Eso sería el sumatorio para la parte superior de la fracción. En *Google Sheets*, puedes multiplicar las columnas C y D y sumarlas en una celda usando la función SUMPRODUCT. Por ejemplo, en la celda E1, podrías poner: =**SUMPRODUCT**(C1:C20, D1:D20) Esto te daría el resultado 1450.

Luego, necesitarás calcular la suma de los cuadrados de las diferencias para cada observación y su media correspondiente. Nuevamente, ya se te han dado los resultados de estos cálculos, pero normalmente podrías hacerlo elevando al cuadrado las diferencias que ya calculaste en los pasos 2 y 3 y luego sumando todos los resultados. Puedes hacer esto en Google Sheets usando la función SUMSQ. Por ejemplo, en la celda F1, podrías poner: =**SUMSQ**(C1:C20) Y en la celda G1, podrías poner:=**SUMSQ**(D1:D20) Esto te daría los resultados 19050 y 7563.75, respectivamente.

Finalmente, podrías calcular el coeficiente de correlación de Pearson, r, dividiendo el resultado del paso 3 por la raíz cuadrada del producto de los resultados del paso 4. En Google Sheets, puedes hacer esto en la celda H1, por ejemplo, con: =E1/**SQRT**(F1*G1) Esto te daría el resultado 0.1208.

Cada grupo de 3 alumnos puede hacer cada uno las posibles combinaciones entre géneros (no es necesario hacer todas las combinaciones, con una combinación por grupo bastaría,

pero hay de más porque algún grupo será más rápido que otro y puede hacer más de una), de resultados:
Pop vs Hip-Hop: 0.1208
Pop vs Electrónica: 0.1208
Pop vs Trap: 0.192
Pop vs Reggaeton: 0.249
Hip-Hop vs Electrónica: -0.054
Hip-Hop vs Trap: -0.1929
Hip-Hop vs Reggaeton: 0.3429
Electrónica vs Trap: -0.0633
Electrónica vs Reggaeton: -0.0994
Trap vs Reggaeton: 0.0031

Ante las preguntas del ejercicio, una posible respuesta que podrían dar los alumnos es que en general, la mayoría de los valores de r están cerca de 0, lo que indica que no hay una relación lineal fuerte entre los géneros musicales. La correlación más fuerte es entre Hip-Hop y Reggaeton (0.3429), pero sigue siendo una correlación débil.

Es importante tener en cuenta que el coeficiente de correlación de Pearson sólo mide las relaciones lineales entre las variables. Por lo tanto, aunque los valores de r pueden indicar que no hay una relación lineal fuerte entre los géneros musicales, podría haber otro tipo de relación entre ellos que no se captura mediante esta medida.

Aquí la estadística bidimensional proporciona información útil, pero no suficiente para concluir una relación clara entre el tiempo dedicado a escuchar diferentes géneros musicales sin complementar estos resultados con otro tipo de análisis o estudios que incluyan información adicional, como las preferencias personales de los individuos, su entorno sociocultural, entre otros factores que pueden influir en sus hábitos de escucha.

Si la correlación hubiese sido alta, eso indicaría una relación lineal fuerte entre las dos variables que se están comparando. En el caso de los géneros musicales, si, por ejemplo, hubiésemos encontrado una correlación de 0.8 entre Hip-Hop y Reggaeton, esto implicaría que hay una relación lineal fuerte entre estos dos géneros musicales.

En un contexto más práctico, si estamos viendo las preferencias musicales de una muestra de personas, una alta correlación podría sugerir que las personas que disfrutan del Hip-Hop tienden también a disfrutar del Reggaeton. Es importante recordar que la correlación no implica causalidad; una correlación fuerte entre dos variables no necesariamente significa que una causa la otra.

También es importante tener en cuenta que el valor de la correlación varía entre -1 y +1. Un valor de correlación cercano a +1 indica una fuerte correlación positiva (es decir, cuando una variable aumenta, la otra también tiende a aumentar), mientras que un valor cercano a -1 indica una fuerte correlación negativa (es decir, cuando una variable aumenta, la otra tiende a disminuir). Un valor cercano a 0, como ya hemos mencionado, indica una correlación débil o inexistente.

Sesión 6. Estadística bidimensional (parte 2)

En penúltima sesión de la secuencia didáctica nos enfocaremos en la regresión lineal y el uso de la línea de mejor ajuste para predecir valores futuros basados en las relaciones entre géneros musicales.

Ejercicio: Dado el conjunto de datos sobre el tiempo dedicado a escuchar diferentes géneros musicales, en grupos de tres, los estudiantes deben de escoger dos géneros musicales y calcular la ecuación de la línea de mejor ajuste. Utilizar la ecuación de la línea de mejor ajuste para predecir cuánto tiempo un estudiante escucharía un género musical en función del tiempo dedicado a escuchar otro género. Discutir si la predicción obtenida es razonable y si las correlaciones entre los géneros musicales se pueden utilizar para predecir el comportamiento de escucha de los estudiantes en el futuro. Presentar y discutir los resultados en clase.

Para escoger dos puntos del gráfico de dispersión y calcular la línea de mejor ajuste, primero seleccionemos dos puntos del conjunto de datos Hip-Hop y Reggaeton, ya que eran los de mayor coeficiente de relación. Supongamos que escogemos los siguientes dos puntos:
Punto A: (30, 120)
Punto B: (50, 110)

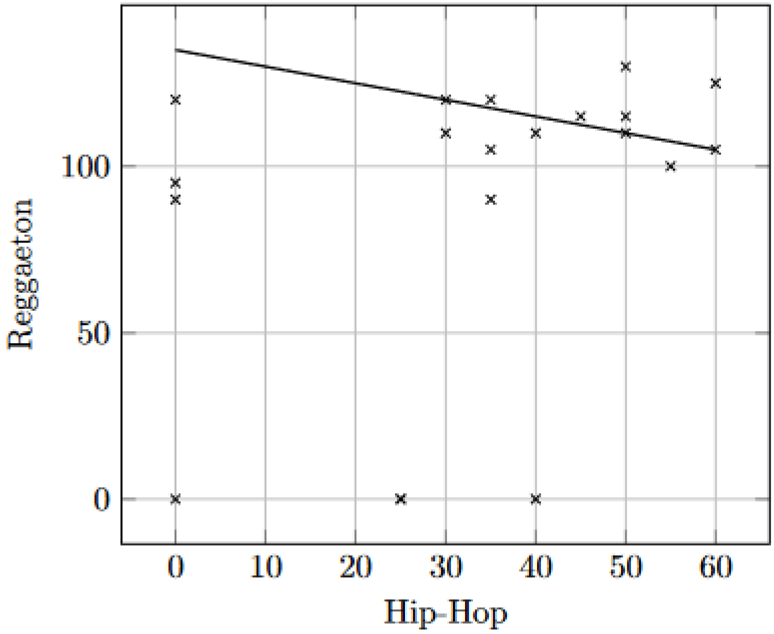
La ecuación de la recta que pasa por estos dos puntos se puede calcular de la siguiente manera:
Pendiente (m) = (y2 - y1) / (x2 - x1) = (110 - 120) / (50 - 30) = -10 / 20 = -0.5

Para encontrar la intersección en y (b), utilizamos uno de los puntos y la pendiente:
b = y - m * x
b = 120 - (-0.5) * 30 = 120 + 15 = 135

Entonces, la ecuación de la línea de mejor ajuste es:
y = -0.5x + 135
representada en la Figura 19.

Figura 19

Una posible recta de regresión



Aquí ha pasado algo interesante. En este caso, podemos distinguir los grupos de la siguiente manera:

Esquina superior derecha (Hip-Hop y Reggaeton altos): Estudiantes que escuchan mucho Hip-Hop y Reggaeton.

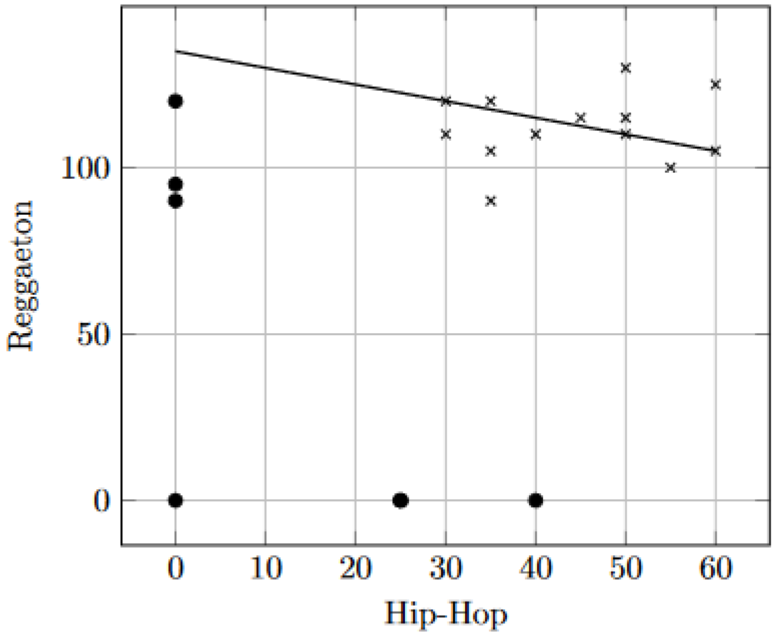
Esquina superior izquierda (Hip-Hop bajo, Reggaeton alto): Estudiantes que escuchan poco Hip-Hop pero mucho Reggaeton.

Esquina inferior (Hip-Hop y Reggaeton bajos): Estudiantes que escuchan poco Hip-Hop y Reggaeton.

Representado en la Figura 20.

Figura 20

Una posible recta de regresión, marcando grupos diferenciados



Lo que ha sucedido en este caso es que al seleccionar dos puntos de la esquina superior derecha para calcular la línea de mejor ajuste, hemos obviado la información de otros grupos de estudiantes con diferentes patrones de escucha. En este gráfico, podemos ver claramente la distribución de los tres grupos mencionados. La línea de mejor ajuste calculada anteriormente no representaba adecuadamente estos grupos, ya que se basaba en un subconjunto limitado de datos.

Este ejemplo ilustra la importancia de analizar cuidadosamente los datos y la distribución de los puntos en el gráfico de dispersión antes de realizar cualquier ajuste de línea. Se dejaría a los grupos de estudiantes que calculen y dividan en otros grupos otras relaciones de género musical que así lo considerasen.

Sesión 7. Repaso y resolución de dudas, evaluación final

Sesión de repaso: Para la sesión de repaso, el objetivo es consolidar los conocimientos adquiridos durante las sesiones anteriores y brindar a los estudiantes la oportunidad de aclarar dudas y practicar más ejercicios.

Revisión de conceptos clave: Comenzamos la sesión repasando brevemente los conceptos clave de cada sesión anterior haciendo preguntas a los estudiantes para asegurarnos de que comprendan los conceptos y animarlos a que hagan preguntas si necesitan más claridad.

Pasaremos a hacer el siguiente cuestionario, a mano alzada, entre grupos, para ayudar a refrescar conceptos:

¿Cuál de las siguientes representaciones gráficas es adecuada para mostrar las frecuencias de datos categóricas?
a) Histograma
b) Gráfico circular
c) Gráfico de barras
d) Tanto b como c

¿Qué medida de tendencia central se ve afectada más significativamente por valores atípicos en un conjunto de datos?
a) Media
b) Mediana
c) Moda
d) Ninguna de las anteriores

¿Qué medida de dispersión indica la diferencia entre el valor máximo y mínimo de un conjunto de datos?
a) Varianza
b) Desviación estándar
c) Rango
d) Coeficiente de variación

¿Cuál de las siguientes medidas de posición divide los datos en cuatro grupos iguales?
a) Cuartiles
b) Deciles
c) Percentiles
d) Ninguna de las anteriores

En una gráfica de cajas y bigotes, ¿qué representa la línea dentro de la caja?

a) Media
b) Mediana
c) Moda
d) Rango intercuartílico

Si el coeficiente de correlación de Pearson entre dos variables es cercano a 1, ¿qué tipo de relación tienen?
a) Relación positiva fuerte
b) Relación positiva débil
c) Relación negativa fuerte
d) Relación negativa débil

Resumen y reflexión: Para cerrar la sesión de repaso, haz un resumen de los conceptos clave y las habilidades que se han practicado. Pedimos a los estudiantes que reflexionen sobre lo que han aprendido y cómo pueden aplicar estos conocimientos en situaciones de la vida real. También puedes pedirles que compartan sus pensamientos sobre el proceso de aprendizaje y cómo pueden mejorar en el futuro.

Después pasaremos a la evaluación como prueba escrita que se detalla en una sección posterior de este documento.

Cronograma de la secuencia didáctica

A continuación, se presenta el cronograma para la planificación de la clase de matemáticas de 4º de la ESO del 17 al 28 de abril de 2023, distribuido en 7 sesiones de 50 minutos cada una, teniendo en cuenta que se dan 4 sesiones a la semana, como se ve en la Figura 21:

Figura 21
Abril de 2023

L	M	X	J	V	S	D
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30

Hay que tener en cuenta que el día 24 de abril no es lectivo (San Jorge, desplazado del día 23 por ser domingo).

Lunes 17 de abril: Evaluación inicial, Muestreo, tablas de datos y frecuencias
Tras la evaluación inicial, los estudiantes aprenderán cómo recolectar datos a través de diferentes técnicas de muestreo, cómo organizar los datos en tablas y cómo calcular las frecuencias.

Martes 18 de abril: Organización visual de los datos estadísticos
Durante esta sesión, se enseñará a los alumnos cómo representar visualmente los datos estadísticos mediante gráficos, como gráficos de barras, gráficos circulares, histogramas y polígonos de frecuencia.

Jueves 20 de abril: Medidas de centralización y de posición
En esta sesión, se introducirán las medidas de centralización (media, mediana y moda) y de posición (cuartiles, percentiles y deciles) para describir la tendencia central y la distribución de los datos.

Viernes 21 de abril: Medidas de dispersión
Los estudiantes aprenderán a calcular medidas de dispersión, como el rango, la varianza y la desviación estándar, para comprender la variabilidad de los datos.

Martes 25 de abril: Estadística bidimensional (Parte 1)
Se introducirá el concepto de estadística bidimensional, donde se analizan dos variables conjuntamente. En esta primera sesión, se abordarán la construcción de tablas de doble entrada y la representación gráfica de los datos.

Jueves 27 de abril: Estadística bidimensional (Parte 2)
En esta segunda sesión sobre estadística bidimensional, se discutirán las medidas de asociación y correlación entre dos variables, como la covarianza y el coeficiente de correlación de Pearson.

Viernes 28 de abril: Repaso y resolución de dudas, evaluación final.
En la última sesión, se llevará a cabo un repaso general de los temas tratados y se resolverán las dudas que puedan tener los estudiantes para asegurar una comprensión adecuada de los contenidos, tras lo que se hará una evaluación final.

Evaluación

Comenzaremos mostrando una cita que destaca la importancia de la evaluación como un proceso continuo y adaptativo que ayuda a ajustar la enseñanza según las necesidades de los estudiantes y también evaluar el programa de matemáticas en general.

> La evaluación en matemáticas debe ser una actividad continua que proporcione información sobre lo que los estudiantes saben, entienden y pueden hacer en matemáticas. Debe ser un proceso mediante el cual se monitoree el aprendizaje del estudiante y se ajuste la enseñanza para satisfacer las necesidades individuales de los estudiantes. La evaluación también debe ser utilizada para evaluar la efectividad del programa de matemáticas y para identificar las áreas que necesitan mejoras. (Bay-Williams et al., 2013, pp 82)

Para evaluar el examen escrito, se aplicará un enfoque de calificación basado en el modelo de tercios, según la propuesta de Gairín et al. (2012). Por errores en tareas generales auxiliares (como cálculos aritméticos o algebraicos), no se restará más de un tercio del puntaje total de la pregunta y se continuará con la corrección. Por errores en tareas auxiliares específicas (relacionadas con lo aprendido durante el curso pero que no constituyen el objetivo principal de la pregunta), no se restará más de dos tercios del valor de la pregunta y se continuará con la corrección. Por errores en las tareas principales de la pregunta, se aplicará una penalización completa y se dejará de corregir.

La evaluación incorporará elementos tanto formativos como sumativos, y se apoyará en evidencias recogidas a partir del cuaderno de clase y las hojas de cálculo generadas durante las sesiones. Este enfoque integral permitirá una evaluación más completa del progreso del estudiante

En este examen se repasan conceptos empleados en las 7 sesiones:

Ejercicio 1: Tablas de frecuencia y gráficos
Se realizó una encuesta a 20 alumnos de 4º de la ESO sobre la cantidad de tiempo que dedican a estudiar matemáticas durante la semana. Los resultados, en horas, son los siguientes:

3, 4, 5, 3, 6, 2, 3, 4, 6, 2, 5, 4, 3, 2, 7, 4, 6, 5, 3, 4

a) Construye una tabla de frecuencias con las columnas "Horas de estudio", "Frecuencia absoluta", "Frecuencia acumulada" y "Frecuencia relativa".

b) Crea un gráfico de barras a partir de la tabla de frecuencias.

Ejercicio 2: Medidas de centralización, posición y dispersión
Utiliza los datos del ejercicio 1 para responder a las siguientes preguntas:

a) Calcula la media, la mediana y la moda de las horas de estudio.

Media: $\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

Mediana: M_e

Moda: Mo

b) Calcula el rango, la varianza y la desviación típica de las horas de estudio.

Rango: $R = x_{max} - x_{min}$

Varianza: $s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}$

Desviación típica: $s = \sqrt{s^2}$

c) Haz un gráfico de bigotes a partir de los datos

Ejercicio 3: Estadística bidimensional y coeficiente de correlación
Se recopilaron datos de 10 alumnos sobre la cantidad de horas que pasan jugando videojuegos por semana (X) y sus calificaciones en matemáticas (Y):

X	2	4	6	4	3	5	1	6	7	3
Y	7	6	4	6	7	5	8	4	3	7

a) Crea una tabla de doble entrada con las variables X e Y y calcula el coeficiente de correlación de Pearson.

Coeficiente de correlación de Pearson: $r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}}$

b) Genera un diagrama de dispersión utilizando los datos proporcionados y describe la relación entre las dos variables.

Antes de ver la rúbrica, no podemos dejar de nombrar que en caso, esperemos que no, de que se produjera una falta de respeto grave -insultos, violencia, por ejemplo- o se detecte que un alumno está copiando o colaborando en el acto de copiar en un examen, se consultará el Reglamento de Régimen Interno para determinar la medida disciplinaria adecuada, como la suspensión o la expulsión del alumno. Dicho esto, que esperemos no sea necesario, se seguirán estos criterios de calificación según la Tabla 15:

Tabla 15

Criterios de calificación

Criterio	Puntuación máxima	Descripción
Examen	65%	Calificación obtenida en el examen, evaluando la comprensión y aplicación de los conceptos estudiados.
Realización de ejercicios	35%	Calidad y cantidad de ejercicios realizados a lo largo de la unidad

Esta rúbrica permite evaluar a los alumnos considerando diversos aspectos de su desempeño académico y no solo el resultado del examen. La realización de ejercicios en el cuaderno también son factores importantes que contribuyen al aprendizaje y al desarrollo del alumno.

Bibliografía

Alcaide Guindo, F. (2016). *Matemáticas orientadas a las enseñanzas aplicadas: 4 ESO*. SM.

Arnal, A. y Muñoz, J.M. (2021). *Diseño de actividades de aprendizaje de matemáticas*.[Apuntes académicos]. ADDUnizar.

Allen, M. R., & Allen, M. (2012). *Mints and Money in Medieval England*. Cambridge University Press.

Ball, P. (2006). *Critical Mass: How One Thing Leads to Another*. Farrar, Straus and Giroux.

Batanero, C. (2000). Significado y comprensión de las medidas de posición central. *Uno. Revista de Didáctica de las Matemáticas*, (25), 41-58.

Batanero, C., & Godino, J. D. (2002). *Estocástica y su didáctica para maestros* (Universidad de Granada. Departamento de Didáctica de la Matemática, Ed.). Universidad de Granada, Departamento de Didáctica de la Matemática.

Bay-Williams, J. M., Karp, K. S., & Van de Walle, J. A. (2013). *Elementary and Middle School Mathematics: Teaching Developmentally*. Pearson.

Ben-Zvi, D., & Garfield, J. (Eds.). (2004). *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*. Springer Netherlands.

Chapin, S. H., & Johnson, A. (2006). *Math Matters: Understanding the Math You Teach Grades K-8, 2nd Edition*.

Colera, J. (2016). *Matemáticas 4ºESO orientadas a las ciencias aplicadas*. Anaya Educación.

Gairin, J., Muñoz, J. y Oller, A. M. (2012). *Propuesta de un modelo para la calificación de exámenes de matemáticas*. Investigación en Educación Matemática XVI, (pp. 261-274).

Gordon, F. (1992). *Statistics for the Twenty-first Century* (F. Gordon & S. Gordon, Eds.). Mathematical Association of America.

Monmonier, M. (2004). *Rhumb lines and map wars : a social history of the Mercator projection*. University of Chicago Press.

Moya, P. (2016). *Matemáticas orientadas a las enseñanzas aplicadas 4º ESO*. Marea Verde.

Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, *97*(4), 557-625.

OECD. (2013). *PISA 2012 Results: What Students Know and Can Do (Volume I): Student Performance in Mathematics, Reading and Science*. OECD Publishing.

Orden ECD/1172/2022, de 2 de agosto, por la que se aprueban el currículo y las características de la evaluación de la Educación Secundaria Obligatoria y se autoriza su aplicación en los centros docentes de la Comunidad Autónoma de Aragón (Publicada en BOA el 11/08/2022)

Orden ECD/489/2016, de 26 de mayo, por la que se aprueba el currículo de la Educación Secundaria Obligatoria y se autoriza su aplicación en los centros docentes de la Comunidad Autónoma de Aragón. (Publicada en BOA el 2/6/2016)

Pearson, K. (2011). *The Life, Letters and Labours of Francis Galton: Volume 1, Birth 1822 to Marriage 1853*. Cambridge University Press.

Silver, E. A. (2000). *Principles and Standards for School Mathematics: An Overview*. National Council of Teachers of Mathematics.

Thucydides. (1972). *History of the Peloponnesian War*.

Walker, H. M. (1975). *Studies in the History of Statistical Method*. Arno Press.

Wallter, W. (1938). *The Founder of Statistics* (Vol. 5(4)). Review of the International Statistical Institute.