Tema 19

Estadística descriptiva





Intención didáctica: Desarrollar la capacidad de analizar e interpretar conjuntos de datos a través de la aplicación de una variedad de procedimientos y estrategias.

Eje: Análisis de datos

Aprendizaje esperado: Recolecta, registra y lee datos en tablas y gráficas de barras. Toma decisiones con base en el uso y la interpretación de la moda, la media aritmética y el rango en un conjunto de datos.

Orientaciones matemáticas generales

La estadística descriptiva no es un fin en sí, sino un medio para entender un contexto más amplio y generar nuevas preguntas de indagación sobre ese contexto. Hay muchos posibles escenarios, pero el más común es cuando los datos a analizar representan una **muestra** que viene de una población más grande. En este caso lo que nos interesa es lo que los datos nos pueden decir de la **población** de dónde provienen.

Conscientes de esta relación que los datos tienen con el contexto de dónde vienen, es importante fomentar en los estudiantes una mirada que va más allá de la mera descripción de los datos, y que se enfoca en identificar características y tendencias más generales. Aunque en quinto de primaria todavía no se habla de población y muestra de manera formal, es conveniente introducir la idea a los estudiantes de que los datos son representativos de algo más grande o general y promover reflexiones al respecto.

La estadística y la probabilidad estudian fenómenos **aleatorios**. La esencia de la estadística es la variabilidad y los datos que se analizan son valores particulares de unas **variables** más generales, cuyo comportamiento es lo que verdaderamente nos interesa. El concepto de una variable es muy abstracto y complejo para un estudiante de quinto de primaria, pero se puede hablar de **características** de las personas u objetos de interés como primer paso en el desarrollo del concepto de variable.

Lo que se busca es generalizar de los datos particulares al concepto general. Un ejemplo podría ser comparar las alturas de un grupo de personas adultas nacidas en el siglo XVIII con las alturas de un grupo de personas adultas nacidas en el siglo XXI. Después de analizar los datos de los dos grupos de personas específicas se puede dirigir la discusión hacia el concepto de altura en general con preguntas como ¿Creen que las personas de hoy son más altas que las personas del siglo XVIII? Aquí estamos refiriendo a la altura de las personas en general, no nada más de las personas en el conjunto de datos.

Una de las características más importantes de la estadística descriptiva es la de resumir un conjunto de datos en uno o dos números. Generalmente estos se tratan de **medidas de tendencia central** (moda, mediana y media aritmética) que describen **dónde** se ubican los datos y **medidas de dispersión** que describen que tanta **variabilidad** existe entre los datos. En quinto de primaria únicamente se ven las medidas de tendencia central, pero es muy importante empezar a hablar de la variabilidad en los datos, aunque sea de manera informal, porque es una característica crítica en la estadística. Un promedio no nos dice mucho sin tener una idea de la variabilidad en los datos. El problema inicial de la sesión 1 se trata exactamente de eso.



Orientaciones didácticas generales

La estadística descriptiva tiene dos aspectos fundamentales. El primero es la parte procedimental, que consta de construir tablas, crear gráficas y calcular medidas con la finalidad de **organizar, comunicar** y **resumir** los valores que componen un conjunto de datos. El segundo es la interpretación de los datos, que incluye identificar las **características** y **tendencias** generales de estos, **comparar** dos o más conjuntos de datos, plantear **conjeturas** con base a estos y generar nuevas **preguntas**, entre otros.

Se puede argumentar que el segundo aspecto no es estrictamente descriptivo, siendo más bien inferencia, sin embargo, no conviene separar estos dos aspectos, ya que al enfocarse exclusivamente en el primero genera en los estudiantes una apreciación equivocada de lo que se trata la estadística, reduciéndolo a un conjunto de técnicas mecánicas que no tienen un propósito claro. Por lo tanto, se recomienda siempre generar discusiones y reflexiones alrededor de los datos que se analizan, para fomentar, en los estudiantes, la curiosidad de profundizar sobre el contexto de los datos y no quedar en lo superficial.

La parte procedimental no es completamente automática. Siempre hay que tomar decisiones y hacer elecciones. ¿Conviene más una tabla o una gráfica? ¿En qué orden organizo los valores? ¿Qué tipo de gráfica es más apropiada? ¿Cuál medida de tendencia central es más adecuada? En quinto de primaria únicamente se trabajan las gráficas de barras, pero no sobra preguntar a los estudiantes si se les ocurre otra manera de comunicar los datos. Cuando los datos son cuantitativos, es lógico organizar los valores por su orden numérico, sin embargo, hay que decidir qué hacer con la información que no aparece en los datos. Si los datos son cualitativos, no se puede calcular la media aritmética, pero la mediana sí, siempre y cuando estos tengan un orden. En fin, hay que promover una reflexión constante respecto a lo que se comunica.

La parte interpretativa es más difícil. No hay reglas fijas o procedimientos estándares. Se requiere aplicar el pensamiento crítico y el juicio, basados en la experiencia, pero nunca es demasiado temprano para empezar a desarrollar estas habilidades. En las orientaciones de cada sesión se mencionan aspectos a discutir con los estudiantes, pero en general se puede fijar en los siguientes puntos:

- La dispersión (variación) que muestra una gráfica
- Valores extremos, que pueden ser mucho más pequeños o grandes que los demás datos
- Diferencias entre dos o más grupos de datos en donde se mide la misma variable en dos o más diferentes poblaciones
- Diferencias entre la moda, mediana y media aritmética de un conjunto de datos
- La identificación de subgrupos dentro de un conjunto de datos

Sobre todo, la idea es que surjan nuevas preguntas y conjeturas más que de tratar de llegar a conclusiones contundentes.

Estrategias: Escribe tu nombre y apellido

Es muy común que los estudiantes olviden escribir su nombre y apellido cuando presentan un examen. El estrés del momento y su preocupación por contestar bien las preguntas los distraen, o bien lo dejan para el final y luego se les acaba el tiempo.

Hay que formar hábitos con ellos para evitar que esto pase. Exija que siempre pongan su nombre completo y la fecha en todos los trabajos que realicen. No acepte ni califique un trabajo que no tenga nombre y fecha.

Cuando se están preparando para presentar un examen, insista que la primera cosa que hagan al empezar el examen sea escribir su nombre en la hoja de respuestas, antes de leer las preguntas o cualquier otra cosa.

Suena obvio y muy sencillo, pero suele pasar con frecuencia y es algo completamente evitable.



Sesión 1: Orientaciones matemáticas

La primera sesión se enfoca en los diferentes tipos de datos y en las gráficas de barras. La distinción más importante en cuanto a tipos de datos es entre **cualitativos**, que básicamente son categorías sin valor numérico, y **cuantitativos**, que sí tienen valores numéricos.

Dentro de cualitativos se puede distinguir entre los que tienen un **orden** natural y los que no tienen orden. Por ejemplo, en el ejercicio 3 los estudiantes pueden contestar "Pobre", "Regular", "Bueno" o "Excelente". Estos valores claramente no tienen un valor numérico (aunque en ocasiones se pueden sustituir por 1, 2, 3 y 4 para calcular promedios), sin embargo, sí tienen un orden. En contraste, en el inciso b del ejercicio 1, los valores son los colores de los coches, los cuales no tienen un orden.

Los datos cuantitativos pueden ser **discretos** o **continuos**. Los datos discretos casi siempre se obtienen de **conteos** (el número de hermanos, el número de goles, la edad en años, etc.), mientras que los datos continuos generalmente corresponden a variables físicas como **tiempo**, **distancia**, **temperatura**, **peso**, o una combinación de ellas. Los datos continuos presentan un nivel de dificultad mayor por la necesidad de agruparlos en rangos (por ejemplo, de 1 cm a 2 cm) por lo que no hay ejercicios en este tema que involucren datos continuos. Un ejemplo interesante que es conocido por los estudiantes es el de las calificaciones. Estos son datos cuantitativos pero discretos, ya que hay un número finito de posibles valores, sin embargo, no surgen de un conteo.

Las gráficas de barras son comunes y muy útiles para comunicar de manera visual la distribución de un conjunto de datos. Estrictamente no se deben usar para datos continuos, que requieren un trato distinto. Las barras deben ser separadas, del mismo color y del mismo ancho. Los valores de la variable de interés deben colocarse en el eje horizontal y las frecuencias (números de veces que cada valor se repite) se representan en el eje vertical.

Un error muy común es confundir las frecuencias con los valores de la variable de interés, en particular cuando las frecuencias caen dentro del mismo rango que los valores, como en el problema inicial. En este caso, por ejemplo, el equipo D anota 2 goles cuatro veces, y 4 goles una vez. El número de goles (2 y 4) es la variable de interés y la frecuencia (4 y 1) simplemente nos dice cuántas veces se repitió cada valor de esa variable. En las orientaciones didácticas hay sugerencias de cómo ayudarles a los estudiantes a distinguir entre los valores y las frecuencias.

Sesión 1: Orientaciones didácticas

El problema inicial está diseñado para ayudar a los estudiantes a leer e interpretar las gráficas de barras, y para introducir el concepto de dispersión en un conjunto de datos. La intención de la primera pregunta es verificar que pueden leer las gráficas correctamente. Pregunte a los estudiantes cuál equipo anotó más goles y anote las respuestas sin comentarlas. A menos que todos puedan leer correctamente las gráficas, pida a un estudiante que se haya equivocado que explique su procedimiento para calcular el total de goles por equipo e invite a los demás a comentar sobre esta explicación, para llegar a una explicación de cómo leer las gráficas.

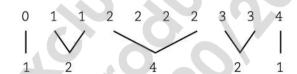
Es muy común que los estudiantes confundan los valores de la variable de interés (número de goles) con las frecuencias. Una manera de ayudarles a tener claridad sobre esto es pedir que para cada equipo escriban todos los valores representados en la gráfica en una lista así:

Equipo A: 0 0 0 0 0 4 4 4 4 4

• Equipo B: 2 2 2 2 2 2 2 2 2 2

• Equipo C: 0 0 1 1 2 2 3 3 4 4

• Equipo D: 0 1 1 2 2 2 2 3 3 4



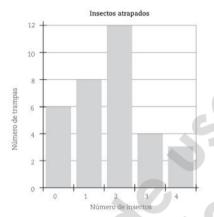
Para hacer todavía más claro de dónde vienen las frecuencias, se puede agrupar los datos que tienen el mismo valor, como en el ejemplo de arriba para el equipo D. Es importante enfatizar que la gráfica no implica un orden específico de los datos. Es decir, que los goles del equipo A podrían haber sido anotados en el orden: {0, 0, 0, 0, 0, 4, 4, 4, 4, 4, 4, 4, 4, 4, 0, 0, 0, 0, 0, 0, 0, 4, 0, 4, 0, 4, 0, 4, 0, 4, 0, 4} entre muchos otros, todos correspondientes a la misma gráfica.

Ahora se puede abrir una discusión sobre la interpretación de las gráficas. Se podría argumentar que C y D son las más parecidas, porque las dos tienen todos los valores presentes de 0 a 4. A y B son las menos parecidas porque B tiene únicamente el valor 2, repetido 10 veces, mientras A tiene los dos extremos (0 y 4 goles). Si se habla de consistencia, el equipo B es el más consistente porque siempre anota 2 goles. Sin embargo, viendo las gráficas podría pensarse que el más consistente es el equipo C, porque la gráfica es uniforme. Valdría la pena tomar unos minutos para reflexionar sobre esto, a manera de introducir el concepto de variabilidad. El equipo con menos variabilidad es B, luego D, luego C y por último A.

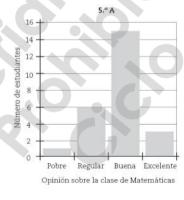
En el inciso f del ejercicio 2 se requiere que los estudiantes calculen el número total de insectos atrapados. Esto se puede hacer escribiendo todos los valores y sumándolos $(0+0+0+0+0+0+1+1+\dots)$. Guíelos a ver que el cálculo se simplifica utilizando la multiplicación $(0\times 6+1\times 8+\dots)$. En el ejercicio 3, a primera vista, parece que en los dos grupos el mismo número de estudiantes calificaron la clase como "Buena". Sin embargo, en el grupo A nada más hay 25 estudiantes mientras en el grupo B hay 30, entonces la proporción es mayor en el grupo A. Por otro lado, la proporción que la calificaron "Excelente" es mucho mayor en el grupo B. En este tipo de datos, muchas veces juntamos ambos valores (Buena y Excelente), dado que estos representan evaluaciones positivas. En ese caso, serían 18 de 25 en el A (aproximadamente $\frac{3}{4}$) y 25 de 30 en el B $(\frac{5}{6})$ con la conclusión de que el grupo B la considera mejor.

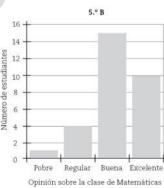
Sesión 1: Solucionario

- Para cada uno de los siguientes ejemplos, identifica si los datos son cualitativos sin orden, cualitativos con orden o cuantitativos. Escribe, por lo menos, tres posibles valores para cada uno de los ejemplos.
 - A) El precio de un kilogramo de papa en diferentes tiendas. Cuantitativos.
 - B) El color de los coches que pasan por un lugar durante determinado tiempo. Cualitativos sin orden.
 - La calificación de los estudiantes de un grupo en un examen de Matemáticas.
 Cuantitativos.
 - **D)** La respuesta de un grupo de clientes de un restaurante a quienes se les pregunta su nivel de satisfacción con la comida. **Cualitativos con orden.**
- 2. La siguiente gráfica muestra el número de insectos atrapados en una serie de trampas que se colocaron en el desierto. Usa la gráfica para contestar las siguientes preguntas.



- A) ¿Cuántas trampas quedaron vacías? 6
- B) ¿Cuántas trampas tuvieron un insecto? 8
- c) ¿Cuál fue el mayor número de insectos atrapados en una trampa? 4
- D) ¿Cuál fue el número más frecuente de insectos atrapados? 2
- E) ¿Cuántas trampas había en total? 33
- F) ¿Cuántos insectos se atraparon en total? 56
- 3. Las siguientes gráficas muestran lo que piensan los estudiantes de los grupos de 5.º A y 5.º B sobre su clase de matemáticas.





- A) ¿Cuál grupo tiene más estudiantes? ¿Cómo sabes? 5.º B. Sumando las frecuencias.
- B) ¿Cuál es la respuesta más común de los estudiantes? Buena
- C) ¿A cuál grupo le gusta más su clase de Matemáticas? Justifica tu respuesta.
 5.º B, porque 25 de 30 calificaron la clase como "Buena" o "Excelente", y en 5.º A fueron 18 de 25.