LA MEDIANA EN LA EDUCACIÓN SECUNDARIA OBLIGATORIA: ¿UN CONCEPTO SENCILLO?

Belén Cobo y Carmen Batanero

UNO 23, 85-96, 2000.

Resumen

Se presenta un análisis conceptual y didáctico de la mediana, incluyendo sus definiciones métodos de cálculo y propiedades. El interés del análisis de justifica por la inclusión en los currículo de la ESO del análisis exploratorio de datos donde la mediana y estadísticos de orden cobran un papel significativo.

Introducción

En la actualidad nos encontramos en un proceso de reforma de la enseñanza, en el que observamos cómo se pretende dar más importancia a la formación estocástica de los alumnos. Este fenómeno no ocurre sólo en España; también en otros países se han dedicado grandes esfuerzos a diseñar materiales didácticos para la enseñanza de la estadística en los niveles no universitarios.

En particular los nuevos diseños curriculares recomiendan la introducción del análisis exploratorio de datos en la enseñanza secundaria, porque esta nueva filosofía de aplicación de la estadística es muy apropiada para conseguir algunos de los nuevos objetivos para la enseñanza de las matemáticas, como la resolución de problemas abiertos, el trabajo interdisciplinar y con datos reales, el uso de sistemas de representación múltiple y el trabajo con ordenadores o calculadoras gráficas.

Por otro lado, en el análisis exploratorio de datos se da más peso a los estadísticos de orden, que consideran la posición relativa de ciertos elementos dentro del conjunto de datos. Entre ellos se encuentran la mediana, cuartiles, percentiles y rangos de percentiles. También se introducen ideas nuevas como la de "valor atípico" y representaciones gráficas basadas en los estadísticos de orden, como el gráfico de la caja.

En este contexto hemos iniciado una investigación relacionada con los estadísticos de orden, dirigida a evaluar la posible dificultad que puede tener el implantar el análisis exploratorio de datos en la enseñanza secundaria (Cobo, 1998). Aunque el estudio de la mediana y percentiles ya se incluía en los currículos anteriores, su uso era más limitado que el que se propone ahora. Por otro lado, aunque existen bastante investigaciones sobre la comprensión que los alumnos tienen de la media aritmética (que se describen, por ejemplo, en Batanero, 2000; Batanero y cols., 1994, Tormo, 1995, Batanero, Godino y Estepa, 1998), las posibles dificultades de comprensión de la mediana y otros estadísticos de orden no han sido investigadas.

Nuestra hipótesis es que estos temas podrían resultar excesivamente difíciles para los alumnos de secundaria, por estar relacionados con el razonamiento proporcional y la ideas de orden y distribución que con frecuencia causan dificultad a los alumnos. Por otro lado, cuando se analiza el tema con detalle, observamos múltiples definiciones, propiedades e incluso algoritmos de cálculo para un mismo concepto.

En este trabajo presentamos un análisis detallado desde el punto de vista matemático y didáctico de uno de los estadísticos de orden citados: la mediana, llevado a cabo a partir de los textos de Calot (1974) y Batanero y cols. (1988) y apoyándonos también en análisis similares realizados en relación a la media aritmética (Tormo, 1995). Este estudio nos parece necesario para identificar dificultades específicas que los estudiantes de Enseñanza Secundaria pueden encontrar en su comprensión conceptual y en su uso en la resolución de problemas de análisis exploratorio de datos.

La mediana ¿una definición sencilla?

Cuando preguntamos a los alumnos que es la mediana nos suelen contestar que "el punto medio" o el "centro de la distribución". Los alumnos parecen comprender que la mediana es el centro de "algo", pero con frecuencia la identifican como el centro del recorrido de la variable o el valor que ocupa la posición central, incluso aunque el conjunto de datos no esté ordenado. Sin embargo, no es sencillo darles una definición clara y concisa, que no les lleve a confusión.

En primer lugar los alumnos deben comprende que, como cualquier resumen estadístico, la mediana se refiere a todo el conjunto de datos, y no a ninguno de los individuos particulares. Comprender esto requiere un cambio de perspectiva (pasar a la perspectiva estadística), que consiste en atender a las características de los agregados y no a las de los individuos. Decir que un colectivo tiene una cierta tendencia o referirse a uno de sus resúmenes estadísticos implica que el colectivo es una colección de individuos idénticos que varían respecto a la propiedad de interés. La comprensión de dicho estadístico implicará también la de la variabilidad de los datos respecto a su valor.

Existen diferentes definiciones equivalentes de la mediana, usadas en los textos de secundaria, que reproducimos a continuación:

D1. *Si suponemos ordenados de menor a mayor todos los valores de una variable estadística, se llama mediana al valor de la variable tal que existen tantos datos con valores de la variable superiores o iguales como inferiores o iguales a él.* Lo esencial para hallar la mediana no es el valor numérico de la variable, sino la posición relativa u ordenación de los datos. Es, por tanto, un estadístico aplicable a datos ordinales, como por ejemplo, el orden de llegada a una meta y no sólo a medidas numéricas como peso o tiempo. La definición es, asimismo, ambigua puesto que se basa en una desigualdad no estricta (mitad de los datos menor o igual que la mediana). Por ello, los estudiantes podrían tener dificultades en saber cuál de las dos mitades (inferior o superior) engloban los casos centrales cuando aquellos tienen el mismo valor de la variable.

D2. *La mediana es el valor de la variable estadística que divide en dos efectivos iguales a los individuos de la población supuestos ordenados por valor creciente del carácter.* Esta definición incorpora el término de efectivo y es un puente hacia la tercera definición basada en la noción de frecuencia relativa acumulada y curva de distribución (empírica).

D3. *La mediana es el valor de la variable estadística tal que la ordenada del diagrama acumulativo de frecuencias absolutas es igual a n/2, siendo n el número de datos*. Se hace referencia aquí al diagrama acumulativo de frecuencias absolutas que se debe conocer previamente. Esta es la definición más abstracta, al presentarla como raíz de una ecuación. Esta raíz es única cuando la función de distribución es estrictamente creciente (caso de variables continuas) pero puede no ser única si el diagrama presenta escalones (lo que ocurre para las variables discretas).

D4. *La mediana es el valor de la variable estadística tal que la ordenada de la representación gráfica de las frecuencias relativas acumuladas es igual a 1/2*. Esta definición se deduce directamente de D3 y tiene la ventaja de no depender del número de datos. Utiliza la idea de frecuencia relativa y su representación gráfica. Tanto en el caso D3 como en el D4 hay que tener en cuenta que la representación de las frecuencias acumuladas es diferente según se agrupen o no los datos en intervalos de clase.

Como se deduce de estas definiciones, la mediana es un valor central. Sin embargo cada una de las definiciones presenta dificultades y ambigüedades. Veremos que estas persisten al tratar de calcular la mediana.

¿Cómo calcular la mediana?

La principal dificultad que nos encontramos es que algoritmo de cálculo de la mediana no es único, sino que depende del tipo de datos, de la forma de presentación de los mismos, e incluso del número de ellos. Los alumnos deben aprender diversos algoritmos y cuando aplicarlos. El valor obtenido no siempre es único, lo que provoca problemas de comprensión a los estudiantes. Describimos los diferentes casos que podemos encontrar.

Para el cálculo de la mediana, distinguiremos entre datos no agrupados y agrupados en clases. Como indica Schuyten (1991) hay un gran número de pasos, no siempre bien comprendidos, desde la definición de la mediana hasta su método de cálculo.

Datos no agrupados en clases

Si el *número de datos es impar*, la mediana es el valor de la variable, individuo u observación que ocupa el centro de la tabla, supuestos éstos ordenados por valores crecientes o decrecientes de la variable. Para aplicar este método, que se deduce directamente de la definición D1, el alumno ha de ser capaz de ordenar los datos y hallar el valor central de los mismos.

Por ejemplo, la mediana de los valores 1, 7, 3, 5, 5 es igual a 5.

Si el *número de valores es par*, la mediana es la media aritmética de los dos valores que se encuentren en el centro de la tabla. Al método anterior se añade ahora la dificultad de que existen dos valores que cumplen la definición de mediana. Esta dificultad se soslaya tomando como mediana el valor medio de los dos valores, práctica que no queda directamente justificada a partir de ninguna definición. Como contrapartida se obtiene como mediana un valor que no coincide con ninguno de los datos, a menos que los dos datos centrales tengan el mismo valor.

Así la mediana de 1, 7, 3, 5, 5, 4 es igual a 4.5.

Si *los valores se presentan en una tabla de frecuencias*, es útil calcular las frecuencias acumuladas para hallar la mediana, lo que podría servir para introducir este nuevo concepto. El cálculo de la mediana se puede hacer en este caso gráficamente a partir del diagrama acumulativo de frecuencias, que el alumno debe ser capaz de representar correctamente. Esto entraña una serie de dificultades, que analizaremos separadamente para los datos agrupados y no agrupados en intervalos.

Datos no agrupados en intervalos

Las frecuencias no acumuladas, tanto absolutas como relativas se definen sólo en un conjunto de valores discretos, que es el conjunto de valores de la variable. Las frecuencias acumuladas, por el contrario, se definen en un conjunto continuo de valores que es el conjunto de números reales. Por ello, mientras que la representación gráfica de las frecuencias no acumuladas es una gráfica discreta, ahora nos encontramos con la gráfica de una función de variable real.

La función obtenida presenta discontinuidades de salto para cada valor de la variable. Es, en consecuencia, un tipo de gráfica que los alumnos pueden no haber trabajado en el tema de las funciones.

Para la construcción de la gráfica, se consideran constantes los valores de esta función entre un valor y otro de la variable estadística. Este es un punto bastante difícil de comprender por parte de los estudiantes.

Una vez realizada la gráfica, procedemos al cálculo de la mediana. Para ello basta tener en cuenta que la frecuencia acumulada que corresponde a la mediana ha de ser igual a $n/2$, o bien, que la frecuencia relativa acumulada es igual a 1/2. Es decir, hemos de partir de las definiciones D3 o D4, que son más complejas.

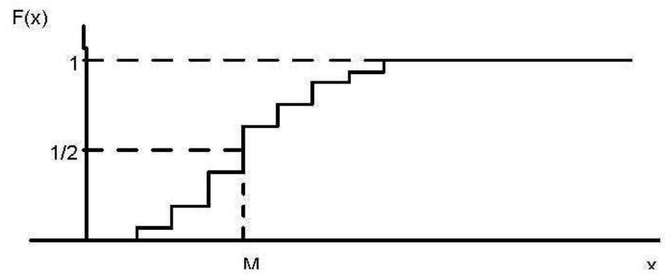
En general, la ecuación $F(M) = 1/2$ no tiene solución, puesto que la función $F(x)$ varía por saltos y es posible que nos encontremos en uno de los dos casos siguientes:

1. Si ningún valor posible x_i corresponde en la gráfica a una frecuencia relativa acumulada $F(x_i)=1/2$ (situación que se presenta en la Figura 1) se considera como valor de la mediana el valor x_i tal que:

$$F(x_i - 0) < 1/2 < F(x_i + 0),$$

es decir tal que: $n_1 + n_2 + ... + n_{i-1} < n/2 < n_1 + ... + n_i$.

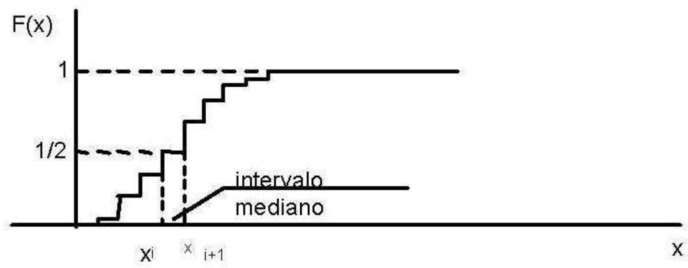
Figura 1. Cálculo de la mediana en variables no agrupadas (caso 1)



El valor 1/2 corta a la gráfica precisamente en el salto que tiene la curva de distribución para uno de los valores de la variable. Ello es debido a que, en este caso todos los valores de la variable comprendidos entre el lugar n_{i-1} y n_i son iguales a x_i y uno de ellos ocupa exactamente el lugar $n/2$. Por tanto este valor x_i cumple la definición de mediana y es el único valor que la cumple.

Si uno de los valores x_i corresponde a $F(x_i) = 1/2$ -lo que ocurre solamente si el total n de la población es par- la mediana está indeterminada entre los valores x_i y x_{i+1}, ya que cualquiera de los valores de x incluidos en el intervalo (x_i, x_{i+1}) cumple la definición de mediana (Figura 2).

Figura 2. Cálculo de la mediana en variables no agrupadas (caso 2)



El intervalo (x_i, x_{i+1}) se denomina mediano y suele tomarse como mediana la media aritmética de estos dos valores, lo que tampoco queda justificado en ninguna de las definiciones.

En la tabla estadística, la mediana se determina a partir de la columna que da las frecuencias (o las frecuencias absolutas) acumuladas, repitiendo el proceso que hemos descrito y finalizando, por tanto, en uno de los casos anteriores. El proceso sería más complejo al no contar con el apoyo de la representación gráfica.

Datos agrupados en intervalos

La agrupación de los datos implica el trabajo con valores aproximados, en lugar de con los datos originales. Por ello, los valores obtenidos para los diferentes estadísticos, como la mediana, son sólo valores aproximados. Además, esta aproximación cambia con la amplitud elegida de los intervalos de clase, lo que también es fuente de dificultades para el alumno acostumbrado a que un procedimiento matemático conduzca siempre a una solución única.

Es preciso también interpretar el significado de las frecuencias, que ahora se refieren a intervalos de valores y no a valores aislados. Si los datos están agrupados en clases, se calculan las frecuencias acumuladas de las clases, comenzando el proceso obteniendo la clase mediana. Una vez calculadas estas frecuencias, se representa el polígono acumulativo de frecuencias y, mediante éste, se determina, gráfica o analíticamente, el valor de la variable cuya frecuencia acumulada es *n/2*. Dicho valor es la mediana.

Se debe construir correctamente esta representación gráfica y esto causa en ocasiones problemas al no comprender con claridad por qué se usa una representación gráfica diferente que en el caso de las variables no agrupadas. Otro problema frecuente es recordar que la frecuencia acumulada en un intervalo se refiere al extremo superior de dicho intervalo. La gráfica se construye uniendo mediante segmentos lineales los valores de las frecuencias acumuladas correspondientes al extremo superior de cada uno de los intervalos de clase. En consecuencia, se obtiene la gráfica de una función continua, lineal, a trozos (Figura 3).

La ecuación F(M)=1/2 tiene siempre en este caso una raíz única que, en general, se sitúa entre dos extremos de clase, o bien coincide con uno de ellos. La determinación de su valor no es, sin embargo, sencilla, puesto que se trata de resolver una ecuación que no está dada mediante una expresión algebraica. El alumno debe recurrir a un procedimiento gráfico o numérico. La clase número i es la clase mediana si:

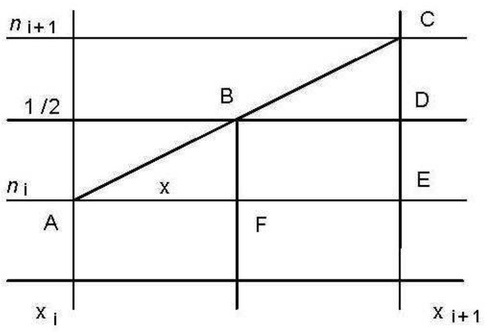
$$F(e_{i-1}) < 1/2 < F(e_i),$$

es decir si:

$$n_1 + ... + n_{i-1} < 1/2 < n_1 + ... + n_i .$$

Se determina la clase mediana a partir de las frecuencias absolutas acumuladas o de las frecuencias acumuladas y por interpolación lineal se obtiene la mediana. Aparece aquí un nuevo concepto que el alumno no suele conocer y que se soslaya haciendo uso del razonamiento proporcional, que es siempre conflictivo para un porcentaje de alumnos, siguiendo la siguiente regla de cálculo basada en el Teorema de Thales.

Figura 3. Cálculo de la mediana con datos agrupados



$$Me=x_i+x$$

$$\frac{x}{AE}=\frac{BF}{CE} \quad \frac{x}{x_{i+1}-x_i}=\frac{\frac{1}{2}-n_i}{n_{i+1}-n_i}$$

$$Me=x_i+x$$

$$\frac{x}{AE}=\frac{BF}{CE} \quad \frac{x}{x_{i+1}-x_i}=\frac{\frac{1}{2}-n_i}{n_{i+1}-n_i}$$

Puesto que $x_{i+1} - x_i$ es la amplitud del intervalo, CE la frecuencia en el intervalo mediano y BF la diferencia entre 1/2 y la frecuencia relativa acumulada en el intervalo mediano, obtenemos la cantidad que hay que sumar al extremo inferior del intervalo mediano para calcular la mediana.

Este método de cálculo, aparentemente sencillo, se basa en la comprensión del significado de polígono de frecuencias acumuladas y en el correcto razonamiento proporcional del alumno, así como en su comprensión del concepto de semejanza de triángulos.

Datos presentados en un diagrama de tallo y hojas

Si los datos se encuentran representados en un diagrama de tallo y hojas se efectúa un recuento desde el tallo menor (arriba), anotando el número de hojas de cada tallo y acumulándolo a los anteriores, hasta que se supere el valor de *n/2*, siendo n el número total de datos; en ese momento se comienza el mismo recuento empezando por el tallo mayor (abajo) hasta llegar al tallo en que nos detuvimos antes. La mediana se encontrará en el tallo cuyo recuento supera el valor de *n/2*, y sólo habrá que buscar el dato central de los valores de la distribución que se encuentra en este tallo. Si el número de datos el impar, la mediana será el valor que ocupa exactamente la posición central; mientras que si el número de datos es par, la mediana será la media aritmética de los dos valores que se encuentren exactamente en el centro de los datos.

Este método tiene la ventaja de que se visualiza con bastante la claridad el significado de mediana como medida de posición de un conjunto de datos. Está basado en la búsqueda, entre todos los datos ordenados, de aquel que ocupa la posición central, como mostramos en el ejemplo siguiente:

n/2=12'5

		recuento	
1	223455789	9	
2	566**7**8	14	Me =27
3	556779	11	
4	12446	5	

Propiedades características de la mediana

Además de la definición y algoritmos de cálculo, es importante el estudio de las propiedades de la mediana, que puede ser contemplada desde diversos puntos de vista: como el resultado de un cálculo (el valor obtenido en el cálculo de la mediana), como operador que a una distribución asigna un número y como un resumen estadístico o parámetro que caracteriza una distribución. Para cada uno de estos puntos de vista podemos analizaremos las propiedades más relevantes de la mediana.

Propiedades numéricas: Son las que se deducen cuando consideramos la mediana como un número, el valor obtenido al calcular la mediana.

N1. *La mediana está comprendida entre el valor mínimo y el valor máximo de los datos*. Nunca podemos obtener un valor fuera del rango de la variable. Aunque esta propiedad parece trivial, sin embargo, Tormo (1993) ha indicado que los alumnos de primaria tienen a veces dificultades en reconocer esta misma propiedad para el caso de la media.

N2. *La mediana puede no coincidir con ninguno de los valores de los datos*, cuando ocurre el caso de indeterminación. Esto hace que incluso pueda obtenerse un número que no tenga un sentido real, como obtener 1.5 como número mediano de niños por familia.

N3. *La mediana no contempla todos los valores de los datos*. Es una operación extraña ya que los valores concretos no influyen en el cálculo, sino tan solo el valor tomado como mediana.

N4. *La mediana es invariante si se disminuye una observación inferior a ella o si se aumenta una superior*. Es consecuencia de la propiedad anterior y ocurre que podemos variar el valor de algunos datos sin que cambie la mediana, lo que no ocurre con la media.

Propiedades algebraicas: Son las que se deducen cuando consideramos la mediana como una operación, la operación de calcular la mediana.

A1. *No es una operación interna en el conjunto numérico empleado*, ya que pueden obtenerse valores en otro sistema numérico diferente al dado.

A2. *Conserva los cambios de origen y de escala.* Ello hace que ésta se exprese en la misma unidad de medida que los datos.

A3. *No tiene elemento neutro ni elemento simétrico*. Cualquier dato que añadamos al conjunto hace variar (potencialmente, al menos) el valor de la mediana.

A4. *No tiene la propiedad asociativa*: Por ejemplo: Me (1, 2, 3) = 2 ; Me ((1,2), 3) = 2'25

A5. *Es conmutativa.* No se ve afectada por el orden de aparición de los datos.

***Propiedades estadísticas*:** Son las que aparecen cuando consideramos la mediana como una medida de posición central

E1. *Es una medida de tendencia central, aunque puede no coincidir con el centro del recorrido*.

E2. *La mediana es un representante o valor típico de un colectivo* y proporciona información global sobre la muestra.

E3. *Es un estadístico resistente*: con pequeñas fluctuaciones de la muestra no cambia su valor. **E4**. *En una distribución simétrica la mediana coincide con la media y la moda* (en distribuciones unimodales).

E5. *Si la distribución es asimétrica a la derecha el orden en que aparecen es moda-mediana-media, y si es asimétrica a la izquierda el orden es media-mediana-moda* (para distribuciones unimodales).

E6. *Si la distribución es asimétrica es preferible la mediana a la media como medida de tendencia central.*

E7. *Existe mediana en distribuciones en las que los datos son ordinales*.

E8. *Es preferible la mediana en distribuciones con datos agrupados en intervalos en los que al menos uno es abierto.*

Conclusiones

El análisis realizado revela la complejidad semiótica (Godino y Batanero, 1998) del tema, que también aparece si cambiamos la mediana por los percentiles o sus rangos. Hemos visto que hay distintas definiciones y métodos de cálculo, numerosas propiedades tanto numéricas, como algebraicas y estadísticas, así como relaciones con otros conceptos estadísticos, que hacen que su estudio sea más complejo de lo que pueda parecer a simple vista.

Pensamos por ello que es necesario analizar con detenimiento la ubicación de este tema dentro del currículo, atendiendo a la dificultad que presenta para los alumnos en el tramo de edad en el que está previsto. Por nuestra parte hemos iniciado una investigación orientada a evaluar la dificultad real de comprensión del tema por parte de estos alumnos. Asimismo, pensamos que se necesita mucha labor de diseño curricular para secuenciar convenientemente este contenido, incluyendo las componentes conceptuales, computacionales e interpretativas analizadas en este trabajo.

Referencias

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

Batanero, C., Estepa, A y Godino, J. (1988*). Curso de estadística basado en el uso de ordenadores*. Jaén: Los autores.

Batanero, C., Godino, J. D., y Estepa, A. (1998). Building the meaning of statistical association through data analysis activities. In *PME CONFERENCE* (Vol. 1, pp. 1-221).

Batanero, C., Godino, J. Green, D. Holmes, P. y Vallecillos, A. (1994). Errors and difficulties in understanding statistical concepts. *International Journal of Mathematics Education in Science and Technology*, 25(4), 527-547.

Calot, G. (1974). *Curso de Estadística Descriptiva*. Madrid: Paraninfo.

Cobo, B. (1998). *Estadísticos de orden en la enseñanza secundaria.* Memoria de Tercer Ciclo. Universidad de Granada.

Godino, J. D., y Batanero, C. (1998). Funciones semióticas en la enseñanza y aprendizaje de las matemáticas. *IX Seminário de Investigaçao em Educaçao Matemática*, 25-45.

Tormo, C. (1995). Dificultades del alumnado respecto a la media aritmética. Uno, 5, 29-36. Grao. Barcelona.

Schuyten, G. (1991). Statistical thinking in Psychology and Education. En Vere-Jones (eds.) *Proceedings of the International Conference on Teaching Statistics* (Voorburg, TheNetherlands: International Statistical Institute), 486-490.