

Comprensión de la mediana por estudiantes universitarios

University students' understanding of median

Ana Esther Madrid @, Silvia M. Valenzuela-Ruiz @, Carmen Batanero @, José A. Garzón-Guerrero @

Universidad de Granada (España)

Resumen ∞ La mediana es una medida de posición central muy utilizada en el análisis exploratorio de datos y la inferencia no paramétrica, por lo que su enseñanza se incluye en los cursos universitarios de estadística. Con la finalidad de identificar los conflictos semióticos que se producen en el aprendizaje del tema, se ha llevado a cabo un estudio de evaluación de la comprensión de la mediana en 148 estudiantes de Ciencias de la Actividad Física y del Deporte. Se presentan los resultados obtenidos tras analizar las respuestas abiertas a un cuestionario de cuatro tareas relacionadas con la definición, cálculo y propiedades de la mediana, identificando sus conflictos semióticos conceptuales, procedimentales y notacionales, algunos de los cuáles no han sido descritos en la investigación previa.

Palabras clave ∞ Mediana; Comprensión; Estudiantes universitarios; Conflictos semióticos

Abstract ∞ The median is a central tendency statistics widely used in exploratory data analysis and non-parametric inference, which is why its teaching is included in university statistics courses. To identify semiotic conflicts in the topic, the results of a study to evaluate the understanding of the median in 148 students of Physical Activity and Sport Sciences are presented. The open-ended responses to a questionnaire of four tasks related to the definition, calculation and properties of the median are analysed, identifying the students' conceptual, procedural, and notational semiotic conflicts, some of which have not been described in previous research.

Keywords ∞ Median; Understanding; University students; Semiotic conflicts

Madrid, A. E., Valenzuela-Ruiz, S. M., Batanero, C. & Garzón-Guerrero, J. A. (2022). Comprensión de la mediana por estudiantes universitarios. *AIEM - Avances de investigación en educación matemática*, 22, 1-21. <https://doi.org/10.35763/aiem22.3902>

1. INTRODUCCIÓN

La mediana es un estadístico de orden, que indica el valor de la variable que ocupa la posición central dentro de un conjunto de datos una vez ordenados y, como tal, tiene un papel relevante en el análisis exploratorio de datos (Cox, 2017). Esta medida de posición central es preferible a la media en presencia de valores atípicos y distribuciones asimétricas, y tiene un amplio uso en inferencia no paramétrica, que se utiliza en muestras pequeñas, datos ordinales o no normales (Gibbons y Chakraborti, 2020). Aunque aparentemente simple, su aprendizaje no es siempre sencillo (Mayén, Díaz y Batanero, 2009).

La investigación relacionada con su comprensión es escasa y, generalmente, se centra en estudiantes de educación secundaria, siendo necesarios estudios que evalúen la comprensión de los estudiantes universitarios una vez estudiado el tema. Para completar esta carencia, el objetivo de este trabajo es evaluar la comprensión de los estudiantes que cursan el Grado de Ciencias de la Actividad Física y del Deporte de la Universidad de Granada sobre la mediana, sus propiedades y método de cálculo.

A continuación, se exponen los fundamentos y antecedentes del trabajo, se describe el método de análisis y los resultados, finalizando con algunas conclusiones.

2. MARCO TEÓRICO Y ANTECEDENTES

La investigación se basa en el enfoque ontosemiótico (EOS) (Godino et al., 2007, 2019) y en los antecedentes descritos a continuación.

2.1. Marco teórico

Godino et al. (2021) analizan diversas teorías de significado en educación matemática comparándolas con la propuesta en el EOS, donde el significado de un objeto matemático (en este trabajo, la mediana) es el sistema de prácticas que realiza una persona (significado personal), o una institución (significado institucional), para resolver las situaciones-problemas relacionadas con el objeto. Se entiende por institución a un conjunto de personas interesadas por resolver una misma clase de situaciones-problemas, aplicando herramientas compartidas. En este marco teórico se asume la siguiente clasificación de objetos primarios que intervienen en la actividad matemática:

- Campos de problemas: Situaciones cuya solución requiere prácticas matemáticas de donde surge el objeto y pueden ser intra matemáticas o extra matemáticas. Un ejemplo en el caso de la mediana sería comparar dos conjuntos de datos ordinales.
- Procedimientos que se emplean para resolver las situaciones. Así, es característico en el cálculo de la mediana ordenar la serie de datos para determinar el valor central.

- Lenguaje: Dado que los objetos matemáticos son inmateriales se necesitan representarlos para operar con ellos; en el trabajo con la mediana usamos el término “mediana” o las notaciones Me , P_{50} , D_5 y Q_2 , así como gráficos y tablas.
- Conceptos (mediana y otros relacionados con ella), como percentil o frecuencia acumulada.
- Proposiciones: relaciones y propiedades entre conceptos, por ejemplo, que la mediana es menos sensible a los valores atípicos que la media.
- Argumentos requeridos para justificar las soluciones de los problemas o comunicar la solución.

Los objetos descritos se relacionan formando correspondencias que Godino et al. (2007) denominan funciones semióticas y tienen tres componentes: la expresión, su contenido y el criterio o regla de correspondencia. Los problemas de interpretación que, en ocasiones, surgen de dichas funciones semióticas llevan en el EOS a definir el *conflicto semiótico* (Godino et al., 2007), como discordancia entre los significados atribuidos a una expresión por dos sujetos (personas o instituciones). Si la disparidad se produce entre significados institucionales, se habla de conflictos semióticos de tipo epistémico, y si se produce entre el significado personal de un sujeto y el de la institución, serían conflictos semióticos cognitivos. Además, los conflictos se pueden clasificar en *conceptuales* (confusión entre conceptos o propiedades), *notacionales* (uso o interpretación del lenguaje matemático) y *procedimentales* (si afectan a los procedimientos).

2.2. Antecedentes

Los principales antecedentes son los trabajos de Mayén, Batanero y Díaz (2009) y Mayén, Díaz y Batanero (2009). Basándose en el EOS, Mayén, Díaz y Batanero (2009) analizan los conflictos semióticos asociados al cálculo de la mediana en un conjunto de datos no agrupados en 518 estudiantes mexicanos de 4.º curso de Educación Secundaria Obligatoria, ESO (15-16 años) y 2.º curso de Bachillerato (17-18 años), mientras que Mayén, Batanero y Díaz (2009) estudian los conflictos semióticos en 643 estudiantes mexicanos, que incluye a la muestra anterior, al comparar dos conjuntos de datos ordinales. En esta sección, clasificamos los resultados de estas y otras investigaciones en función del tipo de objeto matemático cuya comprensión se evalúa.

Comprensión de campos de problemas. Para trabajar con la mediana hay que saber en qué situaciones debe utilizarse. No obstante, sólo el 30,7 % de la muestra estudiada por Mayén, Díaz y Batanero (2009) la elige como el mejor representante de un conjunto de datos con valores atípicos. También son pocos los estudiantes (12,4 %), que en Mayén, Batanero y Díaz (2009) utilizan la mediana para comparar dos conjuntos de datos ordinales, ya que el 33,7 % utiliza la media, el 12,1 % utiliza la moda, el 27,1 % da un valor no relacionado con las medidas de tendencia central y el 14,6 % no responde. Resultados similares son reportados por Boaventura y Fernandes (2004) en estudiantes de Bachillerato; Groth y Bergner (2006) en futuros

profesores de educación primaria; y Gea et al. (2016) con futuros profesores de educación secundaria.

Respecto al *lenguaje*, los estudiantes confunden los términos media, mediana y moda o sus representaciones simbólicas (Cobo, 2003; Mayén, Díaz y Batanero, 2009), mientras que Gea et al. (2017) describen dificultades en futuros profesores en la interpretación de gráficos relacionados con la mediana.

Hay una variedad de interpretaciones incorrectas de la *definición de mediana*; por ejemplo, Barr (1980) indica que muchos estudiantes (17 a 21 años) piensan que la mediana es el centro del conjunto de datos sin ordenar. Mayén, Díaz y Batanero (2009) encontraron estudiantes que conciben la mediana como centro geométrico de la distribución. Batanero et al. (2020) advierten que algunos estudiantes piensan que la mediana deja por debajo el 50 % del rango de variación de la variable y no del número de datos en la muestra.

Comprensión de propiedades. Las propiedades algebraicas de la mediana no son sencillas para algunos estudiantes; por ejemplo, se le asignan propiedades inexistentes como la existencia de elemento neutro o simétrico (Cobo, 2003; Mayén, Díaz y Batanero, 2009). Además, hay dificultades en la comprensión del efecto de los valores atípicos sobre el valor de la mediana, así como en su relación con la media y moda en distribuciones asimétricas (Freitas et al., 2018).

Procedimientos de cálculo. Algunos estudiantes no usan los valores de las frecuencias para calcular la mediana (Barr, 1980) o no resuelven el caso de indeterminación (Cobo, 2003; Mayén, Díaz y Batanero, 2009; Zawojewski y Shaughnessy, 2000). Por otro lado, Carvalho (2001), Cobo (2003) y Mayén, Díaz y Batanero (2009) encuentran estudiantes que calculan la mediana usando las frecuencias en lugar de los datos. Otros (Cobo, 2003) que tienen dificultades para calcular la mediana a partir de un gráfico.

La finalidad de este trabajo es analizar si los conflictos semióticos relacionados con la mediana descritos en trabajos previos se producen también en estudiantes universitarios. Para ello, se analizan las respuestas abiertas a un cuestionario que incluye algunos ítems de estos estudios y otros nuevos.

3. METODOLOGÍA

La muestra estuvo compuesta por 148 alumnos que cursaban la asignatura de Estadística (2.º curso) del Grado en Ciencias de la Actividad Física y del Deporte de la Universidad de Granada. Aunque su formación previa en estadística es muy diversa, ya que acceden al grado por diferentes estudios, los cuestionarios fueron aplicados después de haberse explicado el tema en la asignatura citada. Más concretamente, se estudió su definición, su cálculo a partir de datos aislados, de la tabla de frecuencias o representación gráfica y sus propiedades. El cuestionario fue aplicado en una sesión de clase, dos semanas después de haber concluido el tema, para tener tiempo de estudiarlo y afianzarlo, informando a los participantes de la finalidad del mismo. Los estudiantes tuvieron el tiempo necesario para completar el cuestionario que, generalmente, no excedió de 60 minutos.

El cuestionario se presenta en el Anexo y consta de cuatro ítems abiertos. El primero ha sido adaptado de Mayén, Díaz y Batanero (2009). Analiza la comprensión de la definición de mediana y su cálculo para un número reducido de datos no agrupados. En la primera pregunta, basta ordenar los datos y tomar el valor central; en la segunda, hay dos valores centrales y se debe tomar la media de estos. En la tercera, se espera que respondan que la media aritmética no es buen representante, pues el valor 43 está muy alejado del resto, por lo que sería más adecuado considerar la mediana.

El segundo ítem evalúa el cálculo de la mediana a partir de un gráfico. Dado que la distribución es aproximadamente simétrica, el valor de la mediana coincidirá con la moda, la cual vale 7, que también se puede calcular mediante la tabla de frecuencias acumuladas. El tercer ítem evalúa el cálculo de la mediana a partir de una tabla de datos. Para ello, se deben obtener las frecuencias acumuladas de cada valor de la variable y calcular la posición central de los datos. La mediana es el valor de la variable situado en esta posición; puesto que hay dos valores centrales, se toma como la mediana su media aritmética, que es 2. El cuarto ítem ha sido adaptado del utilizado por Mayén, Batanero y Díaz (2009) y pide comparar dos conjuntos de datos de una variable ordinal; como no es posible utilizar las medias, habría que calcular las medianas de los dos conjuntos de datos para compararlos. La respuesta esperada es que el segundo profesional es mejor, al tener un mayor valor de la mediana.

En la Tabla 1 se especifican los contenidos evaluados en el cuestionario, que abarcan los diferentes objetos primarios considerados en el EOS (campos de problemas, conceptos, procedimientos y proposiciones). La corrección del lenguaje y argumentos utilizados se evalúa en todos los ítems.

Tabla 1. Contenidos evaluados por el cuestionario

Tipos de objetos primarios	Objetos considerados	Ítem 1	Ítem 2	Ítem 3	Ítem 4
Campos de problemas	Elección del mejor representante	x			x
	Comparación de datos ordinales				x
Conceptos	Definición de la mediana	x	x	x	x
	Cálculo en datos aislados	x			x
Procedimientos	Indeterminación en el cálculo	x	x		
	Cálculo a partir de una tabla			x	
	Cálculo a partir de un gráfico		x		
	Cálculo en variables ordinales				x
	Lectura de gráficos		x		
Proposiciones	Efecto de un valor atípico sobre la mediana	x			
	Representatividad de la mediana en datos ordinales				x

4. RESULTADOS Y DISCUSIÓN

Las respuestas a cada apartado de los diferentes ítems fueron clasificadas mediante un proceso cíclico e inductivo propio del análisis de contenido (Neuendorf, 2017). En primer lugar, se diferenciaron las respuestas correctas y dentro de las incorrectas se comenzó por analizar los conflictos semióticos identificados en los antecedentes. En dicho proceso se encontraron nuevos conflictos semióticos no descritos en la investigación previa. Para asegurar la fiabilidad de la clasificación, esta fue revisada por dos de los autores y se discutieron los casos discordantes hasta llegar a un acuerdo.

A continuación, se presenta el análisis y resultado de cada ítem, clasificando los conflictos semióticos en conceptuales, procedimentales o notacionales, según impliquen confusión entre conceptos y/o propiedades, procedimientos o lenguaje matemático. Los conflictos de cada tipo se van numerando como C1, C2, P1, P2, N1, N2, etc., repitiéndose el mismo código cada vez que se describe el mismo conflicto. Puesto que algunos estudiantes presentan más de un conflicto en su respuesta, la suma de los porcentajes (que se han calculado respecto al número de estudiantes) en las siguientes tablas es, en ocasiones, mayor que 100 %.

4.1. Respuestas al Ítem 1

En el cálculo de la mediana con datos sin agrupar (preguntas a y b) se siguió el siguiente criterio de clasificación de las respuestas.

Cálculo correcto de la mediana: El estudiante ordena los datos y elige el valor que corresponde a la posición central del conjunto de datos, resolviendo además el caso de indeterminación en la pregunta b).

4.1.1. Conflictos semióticos conceptuales

C1. Suponer que la mediana es el valor central del conjunto de datos sin ordenar. Implica un conflicto semiótico en la definición de la mediana, también encontrado en Barr (1980), Cobo (2003) y Mayén, Díaz y Batanero (2009), consistente en no tener en cuenta que se precisa ordenar los datos para hallar la mediana, como valor central.

E1: 15, 25, 17, 19, 16, 26, 18, 19, 24; Me=16

C2. Confundir la mediana y el lugar que ocupa cuando se ordena la serie de datos, es decir, tomar el número de orden del dato central, en vez del valor de la variable que corresponde a dicho lugar. Este conflicto no lo hemos encontrado en los antecedentes.

$$E2: Me = \frac{N+1}{2} = \frac{9+1}{2} = 5$$

4.1.2. Conflictos semióticos procedimentales

P1. No resolver el caso de indeterminación en el apartado b) donde hay un número par de datos y, por tanto, se debiera tomar como mediana la media aritmética de los

mismos. En estos casos, los estudiantes dan dos valores o un intervalo de valores como mediana. Este conflicto fue identificado por Mayén, Díaz y Batanero (2009).

E3: Me= entre 19 y 24 días.

Otros errores de cálculo, así, en el siguiente ejemplo, una vez determinado el valor que correspondería a la mediana lo divide entre dos.

$$E4: Me = \frac{n+1}{2} = \frac{9+1}{2} = 5; Me = \frac{x_{(5)}}{2} = \frac{19}{2} = 9.5$$

4.1.3. Conflictos semióticos notacionales

N1. Confundir los términos *media* y *mediana* y, en consecuencia, calcular la media. Lo interpretamos como conflicto semiótico notacional, pues pensamos que la confusión se produce únicamente a nivel de terminología y no a nivel de concepto. Este conflicto fue también encontrado por Cobo (2003).

$$E5: Me = \frac{N15+25+17+19+16+20+18+19+24}{9} = 19.88$$

En la Tabla 2 se presentan los resultados obtenidos en las dos primeras preguntas del Ítem 1, en las que la mayor parte de los estudiantes da la respuesta correcta, aunque alrededor de la cuarta parte tiene algún conflicto o error de cálculo, que aparecen en un porcentaje importante. Los resultados son mejores que los de Cobo (2003) con estudiantes españoles de ESO (38 % de respuestas correctas) y de Mayén, Díaz y Batanero (2009) con estudiantes mexicanos de Bachillerato (68 % de respuestas correctas) en un problema similar. Las respuestas son algo mejores en el segundo apartado, al contrario de lo que ocurría en Cobo (2003).

Tabla 2. Distribución de respuestas en los apartados a y b del Ítem 1

Respuesta	Apartado a		Apartado b	
	Frecuencia	%	Frecuencia	%
Cálculo correcto	108	73,0	113	76,4
C1. Mediana como valor central del conjunto de datos no ordenado	9	6,1	9	6,4
C2. Confundir la mediana y el lugar que ocupa	7	4,7	3	2,1
P1. No resolver la indeterminación			3	2,1
Otros errores de cálculo	18	12,2	20	13,5
N1. Confundir los términos media y mediana	6	4,1		
No responde			2	1,4

En el apartado c se pide elegir un valor central representativo cuando hay valores atípicos, clasificándose las respuestas con el siguiente criterio:

Respuesta y argumento correctos. Se indica que la media aritmética no sería un buen representante, debido a la presencia de un valor atípico, que no afecta a la mediana.

4.1.4. Conflictos conceptuales

C3. Creer que la variabilidad de los datos afecta el valor de la media, lo cual es incorrecto, pues dos conjuntos de datos de igual media pueden tener diferente variabilidad. Este conflicto no lo hemos encontrado en los antecedentes.

E6: No, porque hay una diferencia muy grande entre el que ha realizado menos deporte y el que ha practicado más deporte.

C4. No percibir el efecto del valor atípico sobre la media. Este conflicto semiótico fue descrito por Mayén Díaz y Batanero (2009).

E7: Si, porque a pesar de que hay un dato que podríamos considerar atípico, la media aritmética no se vería influida.

C5. Suponer que la media siempre es un buen representante de la variable, en cualquier circunstancia, conflicto conceptual identificado por Mayén, Díaz y Batanero (2009);

E8: Si, ya que así calcularíamos el número medio de práctica deportiva de estos chicos.

Tabla 3. Distribución de respuestas en el apartado c del Ítem 1

Respuesta	N.º de estudiantes	%
Respuesta y argumento correcto	59	39,9
C3. La variabilidad en los datos afecta a la media	15	10,1
C4. No se percibe el efecto del valor atípico sobre la media	33	22,3
C5. La media siempre es un buen representante	9	6,1
Otros argumentos incorrectos	17	11,5
No responde	15	10,1

El porcentaje de respuestas correctas a esta pregunta es muy inferior al de las dos anteriores (Tabla 3), siendo ligeramente mayor que el encontrado por Mayén, Díaz y Batanero (2009) con estudiantes de bachillerato mexicano (36,9 % de respuestas correctas). Se identifica un nuevo conflicto semiótico C3 consistente en suponer que el valor de la media depende de la variabilidad de los datos y esta sería mejor representante de un conjunto de datos, sólo en el caso de que haya poca variabilidad.

4.2. Respuestas al Ítem 2

Las respuestas a este ítem, en que se pide calcular la mediana a partir de un diagrama de barras, se analizaron de la siguiente manera:

Respuesta correcta a partir de las frecuencias acumuladas: Cuando el estudiante lee correctamente el gráfico, lo traduce a una tabla de frecuencias, calcula las frecuencias acumuladas y toma como mediana el valor al que corresponde la frecuencia acumulada $N/2=24$, obteniendo el valor correcto 7.

Respuesta correcta, aplicando la igualdad de mediana y moda, al ser la distribución simétrica. Esta respuesta apunta a un alto grado de comprensión, al citar una

de las propiedades de la mediana, que permite eludir el cálculo y obtenerla aplicando la propiedad.

E9: $Me=7$, ya que el gráfico es prácticamente simétrico, por lo que tanto la media, moda y mediana, rondarían el punto central, que en este caso son 7 horas.

4.2.1. Conflictos semióticos conceptuales

C2. Confundir la mediana y el lugar que ocupa cuando se ordena la serie de datos, dando como mediana el valor $N/2$, es decir, muestra el conflicto C2 ya señalado.

$$E10: Me = \frac{47+1}{2} = 24$$

4.2.2. Conflictos semióticos procedimentales

P1. No resolver el caso de indeterminación, proporcionando un intervalo de valores como mediana. El estudiante obtiene dos valores que cumplen la definición y no resuelve la indeterminación. Fue identificado en Mayén, Díaz y Batanero (2009).

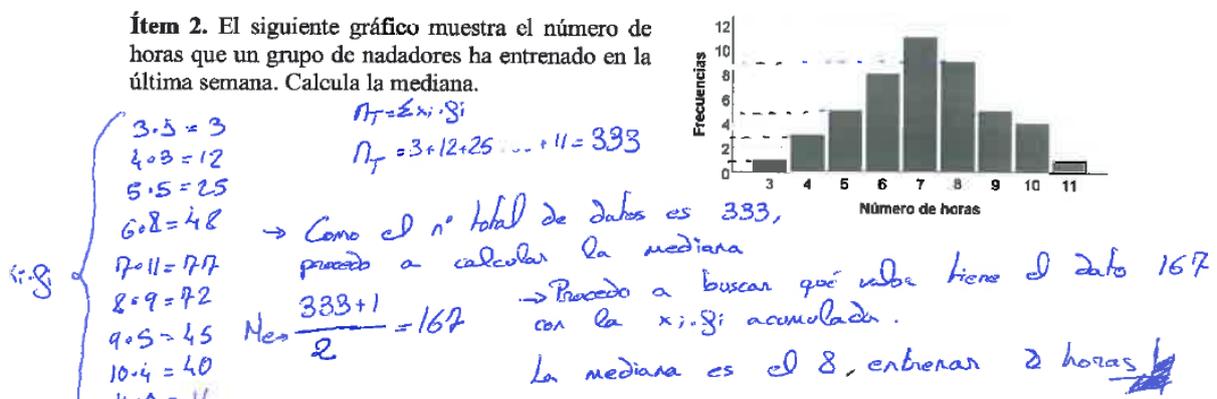
$$E11: Me = [7,8].$$

P2. No considerar las frecuencias en el cálculo de la mediana, calculando la mediana como si todos los valores tuviesen frecuencia unidad. Debido a ello, toman como tamaño de la muestra el número de valores diferentes que toma la variable. Este conflicto fue descrito por Barros (2003).

$$E12: Me = \frac{9+1}{2} = 5 \rightarrow Me = X(5) = 7$$

P3. Calcular el tamaño de la muestra de manera incorrecta. En estas respuestas se confunde el tamaño muestral con las horas empleadas por el total de nadadores, es decir, estos estudiantes calculan el tamaño muestral como la suma del producto de los valores por sus frecuencias. Es un conflicto procedimental que no hemos encontrado descrito en los antecedentes (Figura 1).

Figura 1. Cálculo incorrecto del tamaño de la muestra (E10)



Errores de cálculo, se han incluido en esta categoría aquellas respuestas que son incorrectas debido exclusivamente a errores de cálculo, es decir, aquellas en las que los alumnos usan las frecuencias acumuladas y entienden que la mediana es el dato central. pero cometen algún error al calcularla.

Otras respuestas incorrectas, son aquellas que presentan conflictos distintos de los mencionados anteriormente y que no los consideramos suficientemente representativos por ser muy poco frecuentes. El estudiante del ejemplo E13 ha ordenado los datos en función de sus frecuencias en lugar de en función de los valores de la variable para calcular la mediana.

$$E13: 3,11, 4, 10, 5, 9, 6, 8, 7; Me = \frac{10+5}{2} = 7.5$$

Los resultados de este ítem se presentan en la Tabla 4. El 49,3 % de los estudiantes ha respondido correctamente, aunque algunos no justifican el cálculo, mientras el resto muestra errores de cálculo, no responde o algún tipo de conflicto semiótico. Entre ellos se destacan no tener en cuenta las frecuencias y confundir la mediana con el lugar que esta ocupa al ordenar la 5) = serie de datos. Cabe destacar que en 8 respuestas se han presentado estos dos conflictos a la vez y se han contabilizado en ambas categorías. En comparación con el Ítem 1, el cálculo de la mediana a partir de un gráfico ha sido mucho más difícil que a partir de datos aislados.

Tabla 4. Distribución de respuestas en el Ítem 2

Respuesta	N.º de estudiantes	%
Respuesta correcta a partir de las frecuencias	36	24,3
Respuesta correcta aplicando simetría	9	6,1
Respuesta correcta sin justificar	28	18,9
C2. Confundir la mediana y el lugar que ocupa	14	9,5
P1. No resolver la indeterminación	2	1,4
P2. No considerar las frecuencias	28	18,9
P3. Calcular el tamaño muestral como suma del producto de los valores por sus frecuencias	4	2,7
Errores de cálculo	11	7,4
Otras respuestas incorrectas	14	9,5
No responde	10	6,8

4.3. Respuestas al Ítem 3

En este ítem, en que se pide calcular la mediana a partir de una tabla de frecuencias, se clasificaron las respuestas en las siguientes categorías:

Respuesta correcta. Se determinan los dos valores centrales que aparecen en el conjunto de datos, utilizando las frecuencias acumuladas y los valores proporcionados en la tabla de datos y se calcula la mediana como media de los dos valores centrales, resolviendo la indeterminación.

E14: La mediana se encuentra entre el dato 20 y 21, por lo que $Me=2$ lesionados.

4.3.1. Conflictos conceptuales

C2. *Confundir la mediana y el lugar que ocupa cuando se ordenan los datos.* Este conflicto ya se ha analizado en ítems anteriores.

$$E15: Me = \frac{N+1}{2} = \frac{41}{2} = 20,5$$

C6. *Confundir frecuencia y valor de la variable*, por lo que se calcula el tamaño muestral N como suma de los valores y no como suma de las frecuencias. Fue identificado por Carvalho (2001) y Mayén, Batanero y Díaz (2009).

$$E16: N=28; \frac{N+1}{2} = \frac{29}{2} = 14,5: Me \text{ se encuentra en } 5 \text{ lesionados}$$

4.3.2. Conflictos procedimentales

P1. *No resolver el caso de indeterminación*. Al estar los datos agrupados en una tabla de frecuencias, algunos estudiantes no han determinado los dos datos centrales o no han sabido resolver la indeterminación. Se trata del conflicto procedimental P1 ya comentado anteriormente. Además, en el ejemplo siguiente, se confunde valor de la variable y frecuencia, que es el conflicto semiótico conceptual (C6), analizado en el párrafo anterior.

$$E17: Me = \frac{N+1}{2} = 20,5; Me = \frac{X(20)+X(21)}{2} = \frac{10+10}{2} = 10$$

P2. *No considerar las frecuencias en el cálculo de la mediana*. Se calcula la media de las dos modalidades centrales, es decir, no incluyen las frecuencias de cada valor en el cálculo de la mediana. Este conflicto fue identificado por Barros (2003).

$$E18: Me = \frac{N+1}{2} = \frac{9}{2} = 4,5 \text{ } M(3,4); \frac{3+4}{2} = 3,5$$

P3. *Calcular el tamaño muestral de manera incorrecta*, como suma del producto de los valores por sus frecuencias. Conflicto procedimental P3, detectado y explicado en el ítem anterior.

$$E19: Me = \frac{N+1}{2} = \frac{79+1}{1} = 40; \text{ como } 52 \text{ es el valor más cercano que lo supera, la mediana es } 3.$$

Otras respuestas erróneas: Por ejemplo, el estudiante calcula la media aritmética y, en lugar de dividir por el tamaño muestral, divide por el número de modalidades.

$$E20: 7+11+20+21+4+10+6+7=86/8=10,75$$

Tabla 5. Distribución de respuestas del Ítem 3

Respuesta	N.º de estudiantes	%
Respuesta correcta	48	32,4
C2. Confundir la mediana con el número de orden	8	5,4
C6. Confundir frecuencia con variable	6	4,1
P1. No resolver la indeterminación	21	14,2
P2. No considerar las frecuencias	43	29,1
P3. Cálculo del tamaño muestral como suma de producto de valores por frecuencias	8	5,4
Otros errores	14	9,5
No responde	14	9,5

Fue difícil para los estudiantes calcular la mediana a partir de una tabla de frecuencias, pues sólo la tercera parte de las respuestas fueron correctas (Tabla 5), por lo cual la dificultad de cálculo aumenta aún más que en el caso del gráfico. Entre los principales conflictos semióticos destacan el no tener en cuenta las frecuencias para calcular la mediana P2 (29,1 %) y no resolver la indeterminación (14,2 %).

4.4. Respuestas al Ítem 4

En este ítem, adaptado de Mayén, Batanero y Díaz (2009), se analiza el uso de la mediana para comparar dos conjuntos de datos de una variable ordinal.

4.4.1. Apartado a: Comparación de datos ordinales

Respuesta correcta comparando las medianas. Se elige la mediana como medida de tendencia central para comparar los dos conjuntos de datos, considerando mejor profesional aquel que tiene mayor mediana. También hemos considerado correcto un estudiante que compara las medianas, aunque no lo justifica y otro que compara las modas y justifica correctamente.

4.4.2. Conflictos conceptuales

C7. Pensar que se puede calcular la media en variables ordinales. Se asignan valores numéricos a cada categoría para calcular la media y se señala como mejor profesional al que tiene mayor media (Figura 2). Esta estrategia es incorrecta, puesto que la media no está definida para variables ordinales. Este conflicto fue descrito por Cobo (2003) y también se encontró en el trabajo de Mayén, Batanero y Díaz (2009).

Figura 2. Comparación de medias en el problema 4 (E21)

$$P = -1 \quad I = 0 \quad M = 1 \quad R = 2 \quad P_1 = -1 + 0 + 0 + 1 + 1 + 2 + 2 - 1 - 1 + 0 + 0 + 0 + 1 + 2 + 2 - 1 + 0 + 0 + 4 \cdot 2 = 14 \rightarrow \frac{14}{23} = 0'60$$

$$P_2 = 5 \cdot -1 + 2 \cdot 0 + 6 \cdot 1 + 4 \cdot 2 = 9 \rightarrow \frac{9}{17} = 0'52$$

El 1, porque si se le asigna una puntuación en base a los resultados la del P₁ será más positivo

C8. Suponer que las muestras no son comparables por tener distinto tamaño, o elegir la muestra de mayor tamaño, como se muestra en los ejemplos siguientes. Este conflicto no lo hemos encontrado en los antecedentes.

E22: No se puede saber porque es una variable cualitativa ordinal y además no tenemos el mismo número de muestras.

E23: El promedio más adecuado sería el profesional 1 ya que al haber más datos es más representativo.

4.4.3. Conflictos procedimentales

P4. No utilizar las medidas de tendencia central para comparar dos conjuntos de datos, comparando en lugar de ello las frecuencias de algunas categorías. Este conflicto fue también encontrado en Batanero et al. (1996) en su trabajo sobre la asociación estadística, quienes indicaron que los estudiantes mostraban una concepción local

de la asociación, al suponer que esta se puede determinar únicamente a partir de datos aislados. También aparece en Cobo (2003) y Mayén, Díaz y Batanero (2009). Lo dividimos en dos, según compare frecuencias relativas o absolutas.

P4.1. Compara las frecuencias relativas de una o varias categorías. Se consideran todas o algunas categorías individualmente, comparando las frecuencias relativas de cada profesional. Fue también hallada en Mayén, Batanero y Díaz (2009) y se observa en la respuesta de E24. Otros estudiantes comparan las frecuencias relativas de varias categorías agrupándolas, como ocurre en E25.

E24: El profesional 2, ya que, según las respuestas, 6 de 17 han contestado mejor, y del profesional 1 solo 3 personas de 23, y los dos profesionales han obtenido la respuesta peor por igual.

E25: El profesional 2, porque ha mejorado o recuperado a 10 clientes de 17 posibles, mientras que el profesional 1 ha mejorado a 11 pero de 23 posibles.

P4.2. Compara frecuencias absolutas de una o varias categorías. Estas respuestas son similares a la anterior, a la vez que utilizan las frecuencias absolutas sin tener en cuenta que las muestras no tienen el mismo tamaño, por lo que estarían en el primer nivel de comprensión del concepto de distribución según Watson y Moritz (1999). En ocasiones, algunos estudiantes, como E27, agrupan las frecuencias absolutas de varias categorías para compararlas.

E26: El profesional 1 ya que hay mayor número de recuperados.

E27: He interpretado que los resultados buenos son los “Mejor” y “Recuperado”, por lo que, tras realizar una tabla contando los diferentes resultados, he llegado a la conclusión de que el profesional 1 ha dado mejores resultados (12).

4.4.4. Errores de cálculo de tipo diverso

Como se muestra en la Tabla 6, apenas hay respuestas correctas en este apartado (6,8 %, porcentaje muy similar al encontrado en Mayén, Díaz y Batanero, 2009). La mayoría de los estudiantes de la muestra se limita a comparar frecuencias (absolutas, 22,3 % y relativas 43,3 %), mostrando en el primer caso un nivel muy bajo de comprensión de la distribución (Watson y Moritz, 1999), y en ambos casos la concepción local de la asociación (Batanero et al. 1996), unidos al correspondiente conflicto procedimental.

Tabla 6. Distribución de respuestas en el apartado a del Ítem 4

Respuesta	N.º de estudiantes	%
Respuesta correcta	10	6,8
C7. Pensar que se puede calcular la media en v. ordinales	6	4,1
C8. Suponer que las muestras no son comparables o elegir la muestra de mayor tamaño	13	8,8
P4.1. Comparar frecuencias relativas de una o varias categorías	64	43,3
P4.2. Comparar frecuencias absolutas de una o varias categorías	33	22,3
Errores de cálculo	9	6,1
No responde	13	9,0

4.4.5. Apartado b: Promedio preferible para comparar una variable ordinal

Respuesta correcta: La mediana. El estudiante indica que es preferible la mediana como mejor promedio para representar los datos, por no estar la media definida para variables ordinales, lo cual es la respuesta correcta.

4.4.6. Conflictos conceptuales

C7. Pensar que se puede calcular la media en variables ordinales, no discriminando las variables en que se puede o no calcular este estadístico. Fue identificado por Carvalho (2001) y Mayén, Batanero y Díaz (2009).

E28: El promedio más apropiado es la media, ya que tiene en cuenta todos los valores, con el mismo peso otorgado, y siendo estos determinantes en el ámbito de la salud.

C9. Creer que no se pueden ordenar los datos ordinales y por tanto no se puede calcular la mediana. En este caso, el estudiante que elige la moda ya supone que no se podría calcular ni la media ni la mediana. Este conflicto no ha sido descrito anteriormente.

E29: La moda, porque no pueden ser ordenados, es una variable ordinal.

4.4.7. Conflictos notacionales

Se muestran en las respuestas en las que los estudiantes no eligen uno de los promedios para representar el conjunto de datos. Los alumnos muestran no reconocer el término promedio y proporcionan otras respuestas, mostrando conflictos notacionales, que se han clasificado de la siguiente forma:

N2. Confundir los términos promedio y porcentaje. No se ha encontrado en los antecedentes.

E30: Hallar el porcentaje, ya que los dos profesionales se ocupan de un número de casos distintos.

N3. Toma como promedio un valor de la variable. No se ha encontrado en los antecedentes.

E31: El promedio “mejor”, ya que hay más o menos los mismos casos de igual y peor juntos que de recuperados.

N4. Confundir promedio con una representación gráfica. No se ha encontrado en los antecedentes. En el ejemplo siguiente, se propone un diagrama de sectores como promedio, aunque no se usa el nombre adecuado del gráfico:

E32: Una circunferencia de sectores, ya que son datos cualitativos.

Otras respuestas incorrectas. En el ejemplo, el estudiante, en lugar de indicar un promedio, ha indicado que sería más adecuado el del primer profesional, por tener éste una muestra de mayor tamaño.

E33: El promedio más adecuado sería el profesional 1 ya que al haber más datos es más representativo.

La Tabla 7 muestra la frecuencia de las respuestas en este apartado, en el que un elevado porcentaje de alumnos (39,9 %) no responde, mostrando la dificultad que para ellos ha tenido elegir un representante apropiado del conjunto de datos ordinales. También es alto el porcentaje de alumnos que presentan un conflicto notacional (25 %), mientras que solo el 6,8 % consideran correctamente que el promedio adecuado es la mediana.

Tabla 7. Distribución de respuestas en el apartado b del ítem 5

Respuesta	N.º de estudiantes	%
Respuesta correcta: Mediana	10	6,8
C7. Pensar que se puede calcular la media con datos ordinales	15	10,1
C9. Creer que no se pueden ordenar datos ordinales	11	7,4
N2. Confundir promedio con porcentaje	8	5,4
N3. Dar como promedio un valor de la variable	2	1,4
N4. Confundir promedio con representación gráfica	27	18,2
Otras respuestas incorrectas	15	10,1
No responde	59	39,9

5. SÍNTESIS DE RESULTADOS

Para finalizar el análisis, la Tabla 8 recoge los porcentajes de respuestas correctas en todos los ítems y apartados.

Tabla 8. Porcentaje de respuestas correctas en cada apartado de los diferentes ítems

Ítem	Contenido	Correcta
1a	Cálculo de la mediana con un número impar de valores	73,0
1b	Cálculo de la mediana con un número par de valores	76,4
1c	Valor representativo con datos atípicos	39,9
2	Cálculo de la mediana desde un diagrama de barras	49,3
3	Cálculo de la mediana desde tabla	32,4
4a	Comparación datos ordinales	6,8
4b	Promedio en variable ordinal	6,8

Los ítems que presentaron mayor dificultad (porcentaje de respuestas correctas menor que 25 %) fueron el Ítem 4a (comparación de variables ordinales) y el Ítem 4b (resumen estadístico preferible para comparar una variable ordinal). Tuviron dificultad media (entre 25 % y 50 % de respuestas correctas) el Ítem 3 (cálculo de la mediana a partir de una tabla de frecuencias), Ítem 1c (elección de un valor central representativo cuando hay valores atípicos), e Ítem 2 (cálculo de la mediana a partir de un diagrama de barras). Fueron sencillos con más de 70 % de respuestas correctas el Ítem 1a e Ítem 1b (cálculo de la mediana con un número impar o par de datos).

Como resumen del análisis de conflictos semióticos realizados, reproducimos a continuación la lista de todos los identificados a lo largo del capítulo. Como se ha descrito, en el marco teórico, se consideran conflictos semióticos las disparidades entre el significado dado a un objeto matemático por un estudiante, respecto al asumido correcto en la institución matemática. La amplia lista de conflictos muestra que el tema sigue siendo difícil para los estudiantes de la muestra.

5.1. Conflictos semióticos conceptuales

- C1. Suponer que la mediana es el valor central del conjunto de datos sin ordenar (6,1 % de estudiantes en el Ítem 1a y 6,4 % en el Ítem 1b). Este conflicto fue identificado por Barr (1980), Cobo (2003) y Mayén, Díaz y Batanero (2009).
- C2. Confundir la mediana y el lugar que ocupa cuando se ordena la serie de datos (las frecuencias de aparición en los diferentes ítems son 4,7 % en el Ítem 1a, 2,1 % en el Ítem 1b, 9,5 % en el Ítem 2, 5,4 % en el Ítem 3, Ítem 4). No encontrado en los antecedentes.
- C3. Creer que la variabilidad de los datos afecta el valor de la media (Aparece en el 10,1 % de las respuestas del Ítem 1c). No descrito en trabajos anteriores.
- C4. No percibir el efecto del valor atípico sobre la media (Lo presentan el 22,3 % de estudiantes en el Ítem 1c). Fue descrito por Mayén, Díaz y Batanero (2009).
- C5. Suponer que la media siempre es un buen representante de la variable (Se identifica en el 6,1 % de estudiantes en el Ítem 1c), detectado en Mayén, Díaz y Batanero (2009).
- C6. Confundir frecuencia y valor de la variable, por lo que calculan el tamaño muestral N como suma de los valores. Fue identificado por Carvalho (2001) y Mayén, Batanero y Díaz (2009) (4,1 % en el ítem 3).
- C7. Pensar que se puede calcular la media en variables ordinales, no discriminando las variables en que se puede o no calcular este estadístico (4,1 % de estudiantes en el Ítem 4a y 10,1 % en el Ítem 4b). Descrito por Cobo (2003).
- C8. Suponer que las muestras no son comparables por tener distinto tamaño, o se elige la muestra de mayor tamaño para comparar dos muestras (8,8 % de estudiantes en el ítem 4a y 7,4 % en el ítem 4b). No identificado previamente.
- C9. Creer que no se pueden ordenar los datos ordinales y por tanto no se puede calcular la mediana (7,4 % en el Ítem 4b). Identificado en este estudio.

5.2. Conflictos semióticos notacionales

- N1. Confundir los términos media y mediana (Aparece en el 4,1 % de estudiantes en el Ítem 1a). Fue también descrito por Cobo (2003).
- N2. Confundir los términos promedio y porcentaje (5,4 % en el Ítem 4b). No descrito anteriormente.
- N3. Tomar como promedio un valor de la variable (1,4 % en el Ítem 4b). No descrito en la investigación previa.

N4. Confundir promedio con una representación gráfica (18,2 % Ítem 4b). No descrito en la investigación previa.

5.3. Conflictos semióticos procedimentales

P1. No resolver el caso de indeterminación, dando dos valores o un intervalo de valores como mediana (Se observa con porcentaje 2,1 % en el Ítem 1b, 1,4 % en el Ítem 2, y en el 14,2 % en el Ítem 3). Descrito por Mayén, Batanero y Díaz (2009), siendo mucho menos frecuente en nuestros estudiantes, entre los que se dio preferentemente al determinar la media a partir de un gráfico.

P2. No considerar las frecuencias en el cálculo de la mediana (Se mostró en el 18,9 % de estudiantes en el Ítem 2, Ítem 4, Ítem 6a). Este conflicto es descrito por Barros (2003).

P3. Calcular el tamaño muestral de manera incorrecta como la suma del producto de las modalidades por sus frecuencias (2,7 % en el Ítem 2, 5,4 % en el Ítem 3). No se ha descrito anteriormente.

P4. No utilizar las medidas de tendencia central para comparar conjuntos de datos, comparando, en lugar de ello, frecuencias de algunas categorías (Ítem 4a). Descrito por Batanero et al. (1996), Cobo (2003) y Mayén, Díaz y Batanero (2009). Aparece en el ítem 4 y se ha dividido en P4.1, cuando se usan frecuencias relativas en la comparación, que aparece en el 43,3 % de los estudiantes, y al P.2, que se da en el 22,3 % de los mismos que utilizan únicamente frecuencias absolutas.

La mayor parte de estos conflictos semióticos ya han sido identificados en trabajos previos, pero también se han encontrado otros nuevos que constituyen una aportación de este trabajo, en concreto, los conflictos C2, C3, C8, C9, N2, N3, N4 y P3. Aunque la mayor parte de los conflictos tienen baja incidencia, algunos son más frecuentes, como C3, C4, C8, C9, N4, P1, P2, y P4. Es importante que sean considerados por los profesores en su enseñanza, para promover una comprensión lo más completa posible del tema.

6. CONCLUSIONES

Los resultados obtenidos en los apartados a y b del Ítem 1 muestran el aprendizaje del cálculo de la mediana en un conjunto pequeño de datos aislados con un número par e impar de valores, con más del 75 % de respuestas correctas, doblando los resultados de Cobo (2003) con estudiantes de ESO y superando los resultados a los obtenidos en una tarea similar por Mayén, Díaz y Batanero (2009). Aunque también se obtienen respuestas correctas al elegir entre media y mediana para representar un conjunto de datos con un valor atípico (Ítem 1c) en mayor frecuencia que dichos estudios, todavía la mitad de la muestra participante en el presente trabajo toma la media como mejor representante. Las respuestas incorrectas confirman los conflictos semióticos identificados por Mayén, Díaz y Batanero (2009) y dejan entrever nuevos conflictos no descritos en la investigación previa, que se han analizado en la sección anterior.

En el Ítem 2, que pide calcular la mediana en un diagrama de barras, una proporción importante de estudiantes confundió la mediana con el centro geométrico del gráfico; en otros casos, al tratar de aplicar directamente la definición de mediana para calcularla, ordenaron los valores de la variable en función de su frecuencia, en lugar de utilizar el orden de sus valores.

Menor proporción de respuestas correctas hubo al trabajar con una tabla de frecuencias (Ítem 3), donde casi la tercera parte de la muestra no utilizó las frecuencias para el cálculo de la mediana y otros estudiantes no resolvieron el caso de indeterminación. Finalmente, la mayor dificultad se dio al comparar datos de una variable ordinal (ítem 4), en que son minoría las respuestas correctas y aparecen una variedad de conflictos semióticos descritos en los apartados anteriores.

Estos resultados, aunque mejores que los descritos con estudiantes de secundaria en la investigación previa, indican la necesidad de poner atención al concepto de mediana en la enseñanza universitaria, pues los resultados obtenidos indican que la comprensión tampoco es completa en los estudiantes de la muestra. La lista de conflictos, que amplía los descritos en Mayén, Batanero y Díaz (2009) y Mayén, Díaz y Batanero (2009), apunta a la complejidad del concepto mediana, no siempre tenida suficientemente en cuenta en la enseñanza del tema. Esperamos que el análisis realizado en este trabajo permita a los profesores comprender esta complejidad y estar atentos a los posibles conflictos de los estudiantes, para ayudarles a alcanzar una comprensión suficiente del tema.

AGRADECIMIENTOS

Proyecto PID2019-105601GB-I00 / AEI / 10.13039/501100011033 y Grupo FQM126 (Junta de Andalucía).

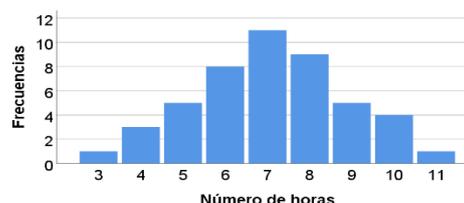
ANEXO. CUESTIONARIO

Ítem 1. El número de días que en los últimos meses 9 chicos practicaron deporte fueron:

15, 25, 17, 19, 16, 26, 18, 19, 24

- ¿Cuál es la mediana de este conjunto de datos?
- ¿Cuál es la mediana si añadimos un chico que practicó deporte 43 días?
- Para estos datos y sin realizar nuevos cálculos ¿sería la media aritmética un buen representante de los 10 datos? Razona la respuesta y si lo crees conveniente propón otro representante.

Ítem 2. El siguiente gráfico muestra el número de horas que un grupo de nadadores ha entrenado en la última semana. Calcula la mediana.



Ítem 3. La siguiente tabla muestra el número de usuarios lesionados en una muestra de 40 gimnasios por un uso inadecuado de los aparatos. ¿Cuál es la mediana del número de lesionados?

N.º lesionados	0	1	2	3	4	5	6	7
N.º gimnasios	7	11	10	7	1	2	1	1

Ítem 4. En una clínica de fisioterapia se está recogiendo información sobre el estado de los clientes una vez recibido su tratamiento. En la clínica los tratamientos son impartidos por dos profesionales, por lo que se divide a los clientes en dos grupos, según el profesional por el que fueron atendidos. Las respuestas dadas por los clientes encuestados han sido codificadas según los códigos P (peor), I (igual), M (mejor) y R (recuperado) y son mostradas en la siguiente tabla:

P1	P	I	I	M	M	R	R	P	P	P	I	I	I	M	R	R	P	I	I	R	R	R	R
P2	R	R	P	P	I	M	I	M	P	P	R	M	M	R	P	M	M						

- ¿Qué profesional ha proporcionado mejores resultados a sus pacientes? ¿Por qué?
- ¿Cuál sería el promedio más apropiado para representar estos datos? ¿Por qué?

REFERENCIAS

- Batanero, C., Estepa, A., Godino, J. D. & Green, D. R. (1996). Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education* 27(2), 151-169. <https://doi.org/10.2307/74959>
- Batanero, C., Valenzuela-Ruiz, S. M. & Begué, N. (2020). Estadísticos de orden y razonamiento proporcional. *UNIÓN*, 16(60), 233-244.
- Barr, G. V. (1980). Some student ideas on the median and the mode. *Teaching Statistics*, 2(2), 38-41. <https://doi.org/10.1111/j.1467-9639.1980.tb00381.x>
- Barros, P. (2003). *Os futuros professores do 2.º ciclo e a estocástica – Dificuldades sentidas e o ensino do tema*. Associação de Professores de Matemática.
- Boaventura, M. G. & Fernandes, J. (2004). Dificuldades de alunos do 12.º ano nas medidas de tendência central: O contributo dos manuais escolares. En J. A. Fernandes (Ed.), *Actas do I Encontro de Probabilidades e Estatística na Escola* (pp. 103-126). Universidad de Miño, Portugal.
- Carvalho, C. (2001). *Interações entre pares: Contributos para a promoção do desenvolvimento lógico e do desempenho estatístico no 7º ano de escolaridade*. Tesis doctoral sin publicar. Universidad de Lisboa.
- Cobo, B. (2003). *Significado de las medidas de posición central para los estudiantes de secundaria*. Tesis doctoral sin publicar. Universidad de Granada.
- Cox, V. (2017). Exploratory data analysis. En V. Cox (Ed.), *Translating statistics to make decisions* (pp. 47-74). Apress. https://doi.org/10.1007/978-1-4842-2256-0_3
- Freitas, A., Figueiredo, T. S., Silva, N. & Miranda, M. C. (2018). Dificuldades na aprendizagem da mediana e quartis por alunos do 8.º ano de escolaridade: estudo comparativo Fórmula versus gráfico. *Indagatio Didactica*, 10(2), 109-132. <https://doi.org/10.34624/id.v10i2.11313>
- Gea, M. M., Arteaga, P. & Cañadas, G. (2017). Interpretación de gráficos estadísticos por futuros profesores de Educación Secundaria. *Avances de Investigación en Educación Matemática*, 12, 19-37. <https://doi.org/10.35763/aiem.v1i12.189>

- Gea, M. M., Batanero, C., Fernandes, J. A. & Arteaga, P. (2016). Interpretación de resúmenes estadísticos por futuros profesores de educación secundaria. *REDIMAT*, 5(2), 135-157. <https://doi.org/10.4471/redimat.2016.1902>
- Gibbons, J. D. & Chakraborti, S. (2020). *Nonparametric statistical inference*. CRC Press.
- Godino, J. D., Batanero, C. & Font, V. (2007). The onto-semiotic approach to research in mathematics education. *ZDM. The International Journal on Mathematics Education*, 39(1-2), 127-135.
- Godino, J. D., Batanero, C. & Font, V. (2019). The onto-semiotic approach: Implications for the prescriptive character of didactics. *For the Learning of Mathematics*, 39(1), 38-43. <https://doi.org/10.1007/s11858-006-0004-1>
- Godino, J. D., Burgos, M. & Gea, M. M. (2021). Analysing theories of meaning in mathematics education from the onto-semiotic approach. *International Journal of Mathematical Education in Science and Technology*, 51, 1-28. <https://doi.org/10.1080/0020739X.2021.1896042>
- Groth, R. E. & Bergner, J. A. (2006). Preservice elementary teachers' conceptual and procedural knowledge of mean, median, and mode. *Mathematical Thinking and Learning*, 8, 37-63. https://doi.org/10.1207/s15327833mtl0801_3
- Mayén, S., Batanero, C. & Díaz, C. (2009). Conflictos semióticos de estudiantes mexicanos en un problema de comparación de datos ordinales. *Revista Latinoamericana de Investigación en Matemática Educativa*, 12(2), 151-178.
- Mayén, S., Díaz, C. & Batanero, C. (2009). Conflictos semióticos de estudiantes con el concepto de mediana. *Statistics Education Research Journal*, 8(2), 74-93. <https://doi.org/10.52041/serj.v8i2.396>
- Neuendorf, K. A. (2017). *The content analysis guidebook*. Sage. <https://doi.org/10.4135/9781071802878>
- Watson, J. M. & Moritz, J. B. (1999). The development of concepts of average. *Focus on Learning Problems in Mathematics*, 21(4), 15-39.
- Zawojewski, J. S. & Shaughnessy, J. M. (2000). Take time for action: Mean and median: Are they really so easy? *Mathematics Teaching in the Middle School*, 5(7), 436-440. <https://doi.org/10.5951/MTMS.5.7.0436>

∞

Ana Esther Madrid

Universidad de Granada (España)
anaesther@ugr.es

Silvia M. Valenzuela-Ruiz

Universidad de Granada (España)
svalenzuela@ugr.es

Carmen Batanero

Universidad de Granada (España)
batanero@ugr.es

José A. Garzón-Guerrero

Universidad de Granada (España)
jgarzon@ugr.es

Recibido: 3 de agosto de 2021

Aceptado: 17 de enero de 2022

University students' understanding of median

Ana Esther Madrid @, Silvia M. Valenzuela-Ruiz @, Carmen Batanero @,
José A. Garzón-Guerrero @

Universidad de Granada (España)

The median is a central tendency statistics which is widely used in exploratory data analysis because its value is scarcely affected by the presence of outliers in the data. It is also employed in many non-parametric methods, which are preferable in small samples, asymmetric distributions and other situations where the assumptions of parametric methods are not valid. Due to these reasons, the median is studied both in high school and university levels. Although this concept is apparently simple, previous research with secondary and high school students and prospective primary school teachers have described a variety of difficulties in understanding its definition and properties and in its computation. The aim of this study was to complete this research at university level, where previous investigation is scarce. To achieve this goal we analyse the responses to a questionnaire with four items of 148 students of Physical Activity and Sport Sciences at the University of Granada, Spain, after having studied this topic. The tasks proposed in the questionnaire include the computation of the median with an odd and even number of isolated data, the computation of the median with data given in a table and data presented in a graph, the comparison of two ordinal data sets and two questions concerning which statistics (mean or median) is preferable as a central tendency measure in a given situation.

The research is supported on the ontosemiotic approach to mathematics cognition and instruction (OSA) where the meaning of a mathematical object is considered as the system of practices carried out when solving problems related to the object. This theoretical framework distinguishes between personal (relative to a person) and institutional (for a given institution) meaning of a mathematical object. More specifically we consider the idea of semiotic conflict as any disparity between the meaning given to a mathematical object (problem, language, concept, property, procedure) by two subjects, in our research the teaching institution and the student.

The analysis of the open-ended responses to the questionnaire served to identify a wide list of semiotic conflicts related to the students' interpretation of concepts (conceptual conflicts), procedures (procedural conflicts) and language (notational conflicts) linked to the median. Part of these conflicts have not been described in previous research.