DIFICULTADES DE ESTUDIANTES MEXICANOS EN LA COMPARACIÓN DE DATOS ORDINALES

Silvia Mayén, Instituto Politécnico Nacional, México

Carmen Batanero, Universidad de Granada, España

Carmen Díaz, Universidad de Huelva, España

RESUMEN

En este trabajo analizamos las respuestas de 643 estudiantes mexicanos a un problema de comparación de datos ordinales. Analizamos las respuestas abiertas, clasificándolas según la medida de tendencia central empleada y las dificultades detectadas. Mediante el test Chi- cuadrado estudiamos la dependencia entre respuesta y grupo de estudiante. Observamos mejores resultados en los alumnos de Secundaria, quienes utilizan más la mediana y la moda.

ABSTRACT

In this paper we analyse the responses given by 643 Mexican students to a problem involving the comparison of ordinal data. We analyse the open responses, taking into account the central measure used in the comparison and the students' difficulties. We observe better results in secondary school students who use median and mode more frequently, although they also tend to omit the response more frequently.

Mayén, S., Batanero, C., Díaz, C. (2009). Dificultades de estudiantes mexicanos en la comparación de datos ordinales. En M.J. González, M.T. González & J. Murillo (Eds.), *Investigación en Educación Matemática XIII* (pp. 301-310). Santander: SEIEM.

INTRODUCCIÓN

Diversos autores como Pollatsek, Lima y Well (1981), Barr (1989), Cai (1995), Watson y Moritz (2000), Cobo y Batanero (2000) o Cobo (2003) describen errores y dificultades en relación a las medidas de posición central, en particular sobre la media aritmética. Las tareas propuestas se refieren a datos medidos en escala de intervalo, donde diferencias numéricas iguales corresponden a las mismas diferencias de cantidad en la magnitud subyacente.

Actualmente se da un gran peso a la mediana y otros estadísticos que se usan en representaciones como el gráfico de caja. Aunque estos estadísticos son también apropiados para datos medidos en escala de intervalo, toman su sentido especialmente con datos ordinales, que representan orden en una cierta magnitud, pero diferencias iguales de orden pueden no corresponder a las mismas diferencias de cantidad en la magnitud.

El propósito de este trabajo es analizar las dificultades de los estudiantes mexicanos al comparar dos conjuntos de datos ordinales. Con ello, continuamos las investigaciones previas de Cobo (2003) en su trabajo con estudiantes españoles.

MARCO TEÓRICO Y ANTECEDENTES

Godino, Batanero y Font (2007) consideran diversos objetos (campos de problemas, definiciones, proposiciones, lenguaje, procedimientos y argumentos) en la actividad matemática e indican que, en el trabajo matemático, los símbolos (significantes) remiten a entidades conceptuales (significados). Toman de Eco (1995) la noción de función semiótica como correspondencia que pone en juego tres componentes:

- Un plano de expresión (objeto inicial, considerado como el signo);
- Un plano de contenido (objeto final, significado del signo, esto es, lo representado);
- Un criterio o regla de correspondencia, esto es un código interpretativo que relaciona expresión y contenido

Los autores destacan que cualquier tipo de objeto, entre los que hemos descrito puede aparecer en la función semiótica tanto en el plano de la expresión como en el del contenido. Denominan *conflicto semiótico* a las interpretaciones de expresiones matemáticas por parte de los estudiantes que no concuerdan con las pretendidas por el profesor o investigador. Estos conflictos no se deben a falta de conocimientos, sino a no haber relacionado adecuadamente los dos términos de una función semiótica. El objetivo de este trabajo es determinar los conflictos semióticos de estudiantes de la muestra en relación con la mediana, en una situación de comparación de datos ordinales.

Comprensión de la mediana

Las investigaciones previas indican que los estudiantes interpretan la mediana como el centro de "algo", pero no comprenden a qué se refiere este "algo" (Barr, 1980). Encuentran difícil aceptar que se pueden emplear dos algoritmos diferentes de cálculo para la mediana o que puedan obtenerse valores distintos en el cálculo con datos agrupados al variar la amplitud de los intervalos de clase; tampoco comprenden cómo pasar de la definición de la mediana a su cálculo (Schuyten, 1991).

Los alumnos tienen dificultad al calcular la mediana si parten de las representaciones gráficas ya que no están acostumbrados a las funciones discontinuas a saltos. Al interpolar para hallar el valor de la mediana, surgen errores por fallo de razonamiento proporcional. No tienen tampoco suficiente dominio del manejo de las desigualdades que aparecen asociadas a la definición y cálculo de la mediana (Estepa, 2004). Otros errores en el cálculo de la mediana (Carvalho 1998; 2001) son: no ordenar los datos para calcular la mediana, confundir frecuencia con el valor de la variable o calcular la moda en vez de la mediana.

Comparación de distribuciones

Konold, Pollatsek, Well y Gagnon (1997) observaron que el uso de las medidas de tendencia central para comparar distribuciones no es intuitivo. Los estudiantes se centran en las frecuencias absolutas y no en las relativas, o comparan las frecuencias de los valores de la variable que coinciden en ambos grupos. Watson y Moritz (2000) indican que, en un primer nivel, los estudiantes son capaces de comparar conjuntos de igual tamaño, mientras en el segundo nivel se comparan conjuntos de datos de diferente tamaño usando un razonamiento proporcional.

Mientras que las investigaciones anteriores se han centrado en datos medidos en escala de intervalo, Cobo (2003) utiliza un ítem que contiene datos ordinales como parte de su cuestionario sobre comprensión de las medidas de tendencia central, pero no realiza un análisis completo de las respuestas. En nuestro trabajo nos centraremos únicamente en este ítem para completar la investigación de Cobo (2003) y comprobaremos si sus resultados se confirman en el contexto mexicano.

MÉTODO

La muestra estuvo compuesta por 643 estudiantes mexicanos: 162 de Secundaria (dos centros) que habían estudiado por primera vez el tema, en el curso en que se aplicó el cuestionario y 481 de Bachillerato (seis centros) que lo estudiaron otra vez el año en que se aplicó el cuestionario. En ambos grupos la mediana se definió como: "el valor que divide a un conjunto ordenado de datos, de tal manera que resulten igual número de datos arriba y debajo de él". Se realizaron problemas de cálculo e interpretación de la mediana con datos cuantitativos agrupados y no agrupados y también unos pocos ejemplos de cálculo en datos ordinales (ver Figura 1).

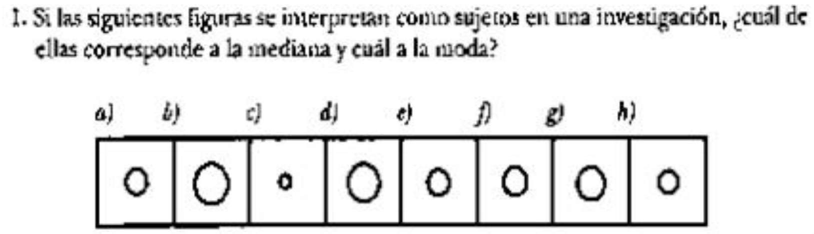

Figura 1. Ejemplo de problema con datos ordinales trabajado por los estudiantes

A continuación se reproduce el problema propuesto, elegido por ser el que registró mayor número de respuestas incorrectas en un estudio comparado previo (Mayén, Cobo, Batanero y Balderas, 2007), tanto en alumnos mexicanos como españoles, lo que sugiere una dificultad no específica del contexto educativo. Antes de resolverlo, se les explicó a los estudiantes el significado de las diferentes calificaciones.

Problema. Un profesor califica a sus alumnos del siguiente modo: I: Insuficiente, A: Aprobado, N: Notable, S: Sobresaliente. En la siguiente tabla tenemos las notas que ha puesto a dos grupos de alumnos:

Grupo 1: I A A N N S S I I I A A A N S S I A A S S S S

Grupo 2: S S I I A N A N I I S N A S I N N

A. ¿Qué grupo ha obtenido mejores notas?

B. ¿Cuál sería la medida de tendencia central más apropiada para representar estos datos?

La solución óptima se analiza en la Tabla 1, dividida en unidades de análisis. En la columna izquierda representamos los pasos en la solución y en la derecha las funciones semióticas que establece el alumno. Como observamos en la tabla, la actividad pedida requiere la comprensión de muchos conceptos, propiedades, procedimientos y representaciones, unido a una argumentación análisis-síntesis. Esta complejidad hace esperar conflictos en los diferentes pasos del proceso. Hacemos notar que, además de esta respuesta correcta, también se han considerado correctas las respuestas en que el alumno resuelve el problema comparando adecuadamente las modas, que están definidas en un conjunto de datos ordinales y parcialmente correctas si transforma los datos a numéricos y compara correctamente las medias.

Pasos	Funciones semióticas
Un profesor califica a sus alumnos del siguiente modo: I=Insuficiente, A=Aprobado, N=Notable, S=Sobresaliente. En la siguiente tabla tenemos las notas que ha puesto a dos grupos de alumnos:	– El alumno ha de relacionar el enunciado (lenguaje) con una situación real. – Se debe asociar el signo = con una propiedad matemática (igualdad). – Se establece un código que se usará en el futuro, A es igual que Aprobado, etc. (correspondencia entre dos elementos lingüísticos y entre cada uno de ellos y la calificación del alumno).
Grupo 1: I A A N N S S I I I A A A N S S I A A S S S S Grupo 2: S S I I A N A N I I S N A S I N N	– Se establece una relación entre cada símbolo (lenguaje) y el valor correspondiente de una variable ordinal (concepto). Este valor sería el correspondiente a un alumno imaginario en el problema (unidad estadística). – Se pone en correspondencia el conjunto de datos dados en cada grupo con la distribución de datos en el grupo.
a) ¿Qué grupo ha obtenido mejores notas?	– Convenio implícito sobre el sentido de "mejores", puesto que ninguno de los dos grupos tiene todas sus notas mejores que el otro. La pregunta establece una correspondencia implícita con un campo de problemas "comparar dos grupos", y su solución que viene dada por la comparación de las medianas.
b) ¿Cuál sería la medida de tendencia central más apropiada para representar esos datos? Para resolver el problema, y puesto que los datos son ordinales, lo más adecuado sería la mediana.	– Se supone un conocimiento implícito en el alumno sobre el tipo de variable en los datos. – Se requiere que el alumno reconozca que en una variable ordinal la mediana es la medida de tendencia central más adecuada en este conjunto de datos.
Para calcularla en cada grupo, el alumno tendría que ordenar previamente los valores, como se muestra a continuación: Grupo 1: I I I I I A A A A A A A N N N S S S S S S S S Grupo 2: I I I I I A A A N N N N N S S S S	– El alumno ha de establecer la relación entre el concepto mediana y su definición (centro de un conjunto ordenado). – La ordenación de los símbolos remite a una ordenación de los valores de la variable.
El primer grupo tiene 23 elementos, y el alumno central, que ocupa la posición 12 tiene un aprobado. En el segundo grupo el valor de la mediana es notable, puesto que el elemento central es el 9, al haber 17 elementos.	– El alumno ha de hallar el centro en cada grupo, puesto que el número de elementos es diferente y no ocupan el mismo lugar. Se ha de poner en correspondencia el concepto "centro" con un procedimiento "hallar el centro". – La mediana es el valor de la variable del elemento que ocupa el lugar central. Se relaciona el concepto de mediana con el procedimiento de cálculo. – El alumno también tiene que llevar a cabo el procedimiento operando con los símbolos previamente ordenados y encontrar el que ocupa el lugar central; ha de relacionar el signo correspondiente del elemento que ocupa este lugar central con su significado (aprobado, etc.). – Finalmente se comparan las dos medianas (valores), hallando la mayor de ellas. El grupo correspondiente tiene "mejores" notas. Se establece una relación entre "tener mayor mediana" y ser el mejor grupo.

Tabla 1. Análisis de la solución correcta al problema

RESULTADOS

Recogidas las respuestas de los estudiantes, se compararon las respuestas similares. Con un proceso similar al mostrado en la Tabla 1, se realizó un análisis semiótico de las respuestas típicas en cada categoría para inferir los objetos matemáticos que el estudiante usa en su resolución. Mediante un proceso inductivo de revisión y comparación se clasificaron las respuestas, considerando en primer lugar la medida de tendencia central (media, mediana, moda o ninguno) que se usa, y en segundo lugar la existencia de conflictos semióticos semejantes. A continuación se describen las categorías.

C1.Respuestas basadas en la media aritmética. El estudiante realiza una transformación de los datos ordinales para convertirlos en cuantitativos y compara las medias para resolver el problema. *No discrimina bien los diferentes tipos de variable estadística* pues la transformación conserva el orden en la variable, pero no la diferencia entre dos valores consecutivos. Otro conflicto es no tener en cuenta que *la media no tiene sentido en los datos ordinales*. Las variantes son:

- *C1.1. Calcula correctamente la media.* Transforma los datos en cuantitativos, y toma la media y no la mediana para resolver el problema. Por ello, la solución no coincide con la esperada, aunque podemos considerarla parcialmente correcta.
- *C1.2. Confunde valor central con valor de la variable*. En la segunda parte del problema el alumno da como medida de posición central más adecuada un valor de la variable, por ejemplo N.
- *C1.3. Calcula la media de frecuencias relativas*. Usa el valor de la frecuencia relativa y no el de la variable en el cálculo de la media, confundiendo estos dos conceptos.
- *C1.4. Halla el número esperado de alumnos por categoría si no hubiese diferencias en los grupos.* Halla para cada categoría de la variable una frecuencia media, dividiendo el número total de alumnos por categoría en ambos grupos entre dos. Es decir, define un "grupo medio", que sería el número esperado de alumnos por categoría si se dividen el número de insuficientes, aprobados, etc. en dos partes iguales. Hay un conflicto respecto a la definición de media que es interpretado como una "distribución media".
- *C1.5. Halla la media del número de alumnos por categoría dentro de cada grupo*, dividiendo el total de alumnos en el grupo entre el número de categorías. Es decir, el estudiante halla el número esperado de alumnos por categoría en una distribución uniforme dentro de cada uno de los grupos. La conclusión obtenida es incorrecta, pues la frecuencia esperada es siempre mayor en el primer grupo al ser mayor el tamaño de la muestra, independientemente de los valores de la variable.
- *C1.6. Indica que habría que calcular la media en cada grupo pero no la calcula.*

C2. Respuestas basadas en la mediana. El alumno identifica correctamente que se trata de un problema resoluble por medio de la mediana. Sin hacer transformaciones de los datos, la calcula o trata de calcularla directamente usando los datos ordinales, por lo que el estudiante diferencia datos ordinales y datos medidos en escala de razón. Encontramos las siguientes variantes:

- *C2.1. Cálculo correcto de las medianas.* Se trataría de la solución correcta al ítem reproducida en la Tabla 1.
- *C2.2. Calcula las medianas en los grupos pero no finaliza el problema.* Un estudiante calcula las medianas pero no da respuesta sobre cuál es el mejor grupo.
- *C2.3. Proporciona una respuesta correcta sin calcular la mediana*. Muestra conocimientos de la definición de la mediana y su propiedad, de ser un estadístico que tiene en cuenta el orden de los datos, no las calcula.

C3. El alumno utiliza las modas para hacer la comparación, calculándola correctamente. Aunque no llega a la solución óptima, es una estrategia correcta, pues la moda está definida para variables ordinales.

C4. Otras respuestas. El estudiante no usa las medidas de posición central para resolver el problema, mostrando una deficiente adquisición de la idea de distribución. Hemos encontrado las siguientes variantes:

- *C4.1. Compara porcentajes o frecuencias relativas de categorías*. Estepa (2004) denomina *concepción local de la asociación,* la conducta consistente en comparar valores aislados en dos muestras. Los estudiantes que siguen esta estrategia pueden llegar a la respuesta correcta, dependiendo de qué valor comparen. Una variante es *comparar porcentajes*, agrupando algunas categorías.
- *C4.2. Comparar frecuencias absolutas* de un valor en cada grupo para resolver el problema. Además de la concepción local de la asociación (Estepa, 2004), estos estudiantes estarían en el primer nivel de comprensión del concepto de distribución según Watson y Moritz (2000).
- *C4.3. Dar como respuesta un valor de la variable* (por ejemplo “notable”) sin hacer referencia a medidas de posición central. Estos estudiantes confunden medida de posición central con el valor de la variable-

C5. Dar como mejor uno de los grupos sin justificar la respuesta.

Categorías de argumentos empleados	Frecuencia	%
C1.1. Compara correctamente las medias	98	15,3
C1.2. Confunde media con valor de la variable	5	,8
C1.3. Calcula la media de frecuencias relativas	16	2,5
C1.4. Halla la frecuencia esperada por categoría	12	1,9
C1.5. Halla el número esperado por categoría en una distribución uniforme	21	3,3
C1.6. Indica que hay calcular la media, pero no la calcula	65	10,1

C2.1. Compara correctamente las medianas	34	5,3
C2.2. Calcula las medianas pero no finaliza el problema	1	,2
C2.3. Respuesta correcta sin justificarla	45	7,0
C3. Compara correctamente las modas	78	12,1
C4.1. Compara porcentajes o frecuencias relativas	39	6
C4.2. Compara frecuencias absolutas	14	2,2
C4.3. Dar un valor de la variable	111	17,3
C5. No justifica la respuesta o no contesta	104	16,2
Total	643	100,0

Tabla 2. Frecuencias y porcentajes de respuestas

En la Tabla 2, mostramos la variedad de respuestas, la mayoría relacionadas con la media. El uso de la mediana en datos ordinales no es intuitivo para los estudiantes, lo que coincide con los resultados de Cobo (2003). Lo más frecuente es dar un valor cualquiera de la variable sin utilizar medidas de posición central. La segunda respuesta más frecuente es utilizar la media correctamente y la tercera tratar de utilizarla, pero no ser capaz de un cálculo correcto.

En la Tabla 3 se resumen los datos sumando las respuestas que hacen referencia a una medida de posición central específica y comparando por grupo. A pesar de tratarse de variables ordinales, la tercera parte de estudiantes trata de utilizar la media en la respuesta. Observamos mejores resultados en los alumnos de Secundaria pues usan más la mediana y la moda, aunque también dan más respuestas en blanco. Las diferencias fueron estadísticamente significativas en el test Chi cuadrado de independencia entre grupo y respuesta (Chi = 48,04; 4 g.l., p=0,0001).

Respuesta	Secundaria N=162	Bachillerato N=356	Total
Media	26,5	39,0	35,1
Mediana	18,5	12,4	14,3
Moda	19,8	3,9	8,9
Otro	16,7	30,1	25,9
No responde	18,5	14,6	15,8

Tabla 3. Porcentajes de respuestas en el ítem clasificadas por grupo

CONCLUSIONES

El estudio indica que la comparación de datos ordinales, incluso en un contexto familiar para el estudiante, como es el de calificaciones, no es intuitiva. Incluso es menos intuitivo para los estudiantes de Bachillerato que para los de Secundaria, de modo que la enseñanza no parece ayudar a desarrollar esta intuición en nuestros estudiantes. Dado el interés señalado de los datos ordinales en la vida diaria y el análisis exploratorio de datos, sería necesario utilizar problemas similares al que presentamos en este trabajo en la enseñanza Secundaria y el Bachillerato.

Nuestro análisis confirma la existencia de los siguientes conflictos descritos por Cobo (2003) en relación con la comprensión de las medidas de posición central: a) no usar medidas de tendencia central para comparar dos conjuntos de datos; b) suponer definida la media en datos ordinales; y c) no discriminar datos ordinales y numéricos. Hemos encontrado además, los siguientes conflictos no descritos en las investigaciones previas:

- *Relacionados con definiciones de objetos matemáticos:* Confundir las medidas de posición central con valor de la variable, frecuencias absolutas con porcentajes; y valor de la variable con frecuencia.
- *Relacionados con propiedades de las medidas de posición central:* No se es consciente que el cambio de escala cambia los promedios.

De estos conflictos se deducen errores al aplicar un procedimiento*:* Calcular la media de las frecuencias (al no diferenciar frecuencia y valor); establecer una correspondencia que no conserva la escala de medida (al no diferenciar escalas); establecer correspondencias diferentes en grupos que se quiere comparar (no tener en cuenta que el cambio de escala cambia el promedio); usar una correspondencia que transforma un conjunto variable en otro constante (no diferenciar entre variable y constante).

Esta lista indica puntos a mejorar en la enseñanza que abarcan no sólo la necesidad de trabajar con los estudiantes con datos ordinales, sino aspectos conceptuales y procedimentales relacionados con la media y mediana y con las ideas aún más elementales de variable estadística y distribución. La semejanza de algunos resultados con los obtenidos por Cobo (2003) con alumnos españoles de menor edad sugiere que los conflictos descritos no son específicos de ninguno de los dos sistemas educativos, sino son compartidos por estudiantes mexicanos y españoles y se mantienen con la edad. Esperamos que el análisis mostrado en este trabajo sea útil a los profesores e incida en la mejora de la enseñanza del tema.

Agradecimientos: Este trabajo forma parte del proyecto SEJ2007-60110 (MEC - Feder) y Grupo PAI FQM-126.

BIBLIOGRAFÍA

Barr, G. V. (1980). Some student's ideas on the median and the mode. *Teaching Statistics,* (2), 38-41.

Cai, J. (1995). Beyond the computational algorithm. Students' unsderstanding of the arithmetic average concept. En L. Meira (Ed.). *Proceedings of the 19th PME Conference* (v.3, pp. 144-151). Universidade Federal de Pernambuco, Recife, Brazil.

Carvalho, C. (1998). Tarefas estadísticas e estratégias de resposta. Comunicación presentada en el *VI Encuentro en Educación Matemática de la Sociedad Portuguesa de Ciencias de la Educación.* Castelo de Vide, Portugal.

Carvalho, C. (2001). *Interaçao entre pares. Contributos para a promoçao do desenvolvimiento lógico e do desempenho estatístico no 7º ano de escolaridade.* Tesis Doctoral. Universidad de Lisboa.

Cobo, B., Batanero, C. (2000). La mediana en la educación secundaria obligatoria: ¿un concepto sencillo? *UNO* 23, 85-96.

Cobo, B. (2003). *Significado de las medidas de posición central para los estudiantes de secundaria.* Departamento de Didáctica de la matemática. Universidad de Granada.

Eco, U. (1995). *Tratado de semiótica general*. Barcelona: Lumen, 1976.

Estepa, A. (2004). Investigación en educación estadística. La asociación estadística. En R. Luengo (Ed.). *Líneas de investigación en Educación Matemática,* (pp. 227-255). Badajoz: Servicio de Publicaciones de la Federación Española de Sociedades de Profesores de Matemáticas. Universidad de Extremadura.

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.

Konold, C., Pollatsek, A., Well, A., Gagnon, A. (1997). Students analyzing data: Research of critical barriers. En J. B. Garfield y G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics.* Voorburg, The Netherlands: International Statistical Institute.

Mayén, S., Cobo, B., Batanero, C., Balderas, P. (2007). Comprensión de las medidas de posición central en estudiantes mexicanos de bachillerato. *UNION, 9,* 187-201

Pollatsek, A., Lima, S., Well, A. D. (1981). Concept or computation: Students' understanding of the mean. *Educational Studies in Mathematics,* 12, 191-204.

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

Watson, J. M., Moritz, J. B. (2000). The longitudinal development of understanding of average. *Mathematical Thinking and Learning.* 2(1 y 2), 11-50.