

LAS MEDIDAS DE DISPERSIÓN EN EL ESTUDIO DE LA INFERENCIA ESTADÍSTICA

Batanero, C.^a, López-Martín, M.M.^b, González-Ruiz, I^c, Díaz-Levicoy, D.^d

^aUniversidad de Granada (España), ^b Universidad de Cantabria (España)
batanero@ugr.es, mariadelmarlopez@ugr.es, ignacio.gonzalezruiz@unican.es,
dddiaz01@hotmail.com

Resumen

Dentro del conjunto de medidas utilizadas en el análisis de datos encontramos las medidas de dispersión. Aunque estas medidas son esenciales para obtener conclusiones sobre algunas de las herramientas empleadas tanto en la estadística descriptiva como en la inferencia estadística, las investigaciones sobre ellas en educación estadística son escasas. En el currículo español vigente, dentro de los diversos niveles educativos, su estudio se introduce, inicialmente, vinculándolo al concepto de distribución de un conjunto de datos para posteriormente generalizarla, de forma progresiva, al estudio de las distribuciones de probabilidad, estimación de parámetros y el contraste o prueba de hipótesis. La finalidad de este trabajo es analizar cómo se introduce y desarrolla el concepto de dispersión en el estudio de la inferencia estadística en el currículo de segundo curso de bachillerato de la modalidad de Humanidades y Ciencias Sociales.

Palabras clave: *Inferencia estadística, medidas de dispersión, currículo de bachillerato.*

INTRODUCCIÓN

Las medidas de variabilidad, o comúnmente conocidas como medidas de dispersión, juegan, junto a las medidas centrales, posición y forma, un papel importante en el análisis de datos. La presencia de dispersión en valores, tanto poblacionales como muestrales, es un hecho real presente en cualquier investigación. Estepa y del Pino (2013) justifican la importancia que tiene la dispersión citando a Cobb y Moore (1997), quienes indican que los datos no son sólo números, sino que son números con un contexto.

El estudio de dichas medidas alcanza una gran relevancia en la formación estadística. Por ejemplo, Snee (1990) establece que el estudio de la dispersión es un concepto fundamental en estadística. Otros autores, como Wild y Pfannkuch (1999) indican que este tipo de medidas adquieren gran relevancia como uno de los componentes básicos en el pensamiento estadístico. En este sentido, Moore (1990, p. 135) lista cinco elementos fundamentales en el pensamiento estadístico, de los cuales tres están asociados a la variabilidad aleatoria: (1) percepción de su ubicuidad en el mundo que nos rodea; (2) competencia para su explicación, identificando los factores de las que depende; (3) habilidad de cuantificarla que implica comprender y saber aplicar el concepto de dispersión.

Aunque, tal y como han puesto de manifiesto Sánchez, Borim y Coutinho (2011), la dispersión es un tema fundamental dentro de la estadística, las investigaciones en Educación Estadística sobre estas son escasas. La finalidad de este trabajo es analizar cómo se trata la dispersión dentro de la inferencia estadística con el fin de orientar a los profesores en su labor docente.

Fundamentos teóricos y metodología

El análisis realizado en este trabajo se fundamenta sobre las ideas del enfoque onto-semiótico (Godino, Batanero y Font, 2007), que asume que los objetos matemáticos surgen de las prácticas (acciones u operaciones) como respuesta a situaciones problemáticas extra o intra matemáticas. En

este enfoque, el término “objeto matemático” tiene un significado muy amplio y puede verse como el conjunto de prácticas asociadas a dicho objeto y puede ser asimismo institucional y personal.

De acuerdo con nuestro referente teórico, las prácticas matemáticas se caracterizan mediante los objetos que intervienen en ellas, que pueden ser de diferente naturaleza: situación-problema, lenguaje, conceptos, proposiciones, procedimientos y argumentos. Todos estos objetos están relacionados, entre sí, formando configuraciones, que serán epistémicas si son propias de una institución matemática o de enseñanza y cognitivas si son específicas del alumno. El aprendizaje del alumno depende sustancialmente del significado institucional presentado en la enseñanza, por lo que es importante analizar el significado institucional de referencia fijado en las directrices curriculares para poder prever las posibles dificultades del aprendizaje del estudiante.

Nuestro trabajo es un estudio de tipo teórico. Para llevarlo a cabo se analizó el contenido de las directrices que fijan el currículo (MECD, 2015), centrándonos en el segundo curso de bachillerato de la modalidad de Humanidades y Ciencias Sociales, donde aparece el tema. Con un proceso inductivo, típico de la metodología cualitativa, se analizaron los objetos matemáticos implícitos en la descripción de contenidos, criterios de evaluación y estándares de aprendizaje descritos en dicho documento. Posteriormente se analizaron recursos interactivos en Internet que podrían aplicarse en el estudio del tema.

INFERENCIA ESTADÍSTICA

El estudio de las medidas de dispersión dentro del ámbito de la inferencia estadística aparece, de una forma más detallada, en el segundo curso de bachillerato de la modalidad de Humanidades y Ciencias Sociales (ver Tabla 1); coincidiendo con el momento en que el alumno comienza a familiarizarse con contenidos propios de la inferencia estadística.

Tabla 1. Contenidos relacionados con inferencia estadística (MECD, 2015)

2º Bachillerato, Ciencias Sociales	Población y muestra. Métodos de selección de una muestra. Tamaño y representatividad de una muestra. Estadística paramétrica. Parámetros de una población y estadísticos obtenidos a partir de una muestra. Estimación puntual. Media y desviación típica de la media muestral y de la proporción muestral. Distribución de la media muestral en una población normal. Distribución de la media muestral y de la proporción muestral en el caso de muestras grandes. Estimación por intervalos de confianza. Relación entre confianza, error y tamaño muestral. Intervalo de confianza para la media poblacional de una distribución normal con desviación típica conocida. Intervalo de confianza para la media poblacional de una distribución de modelo desconocido y para la proporción en el caso de muestras grandes.
---	---

En este contexto, los principales problemas de interés radican en: (1) estimar los parámetros de las distribuciones de probabilidad en una población (particularmente sobre la distribución de la media muestral en una población normal, la distribución de la media muestral y de la proporción muestral en el caso de muestras grandes) partiendo de la información que se extrae de una muestra aleatoria obtenida de la misma población; (2) contrastar una hipótesis sobre los parámetros de la distribución de la población mediante intervalos de confianza o test de hipótesis –en concreto, calcular un intervalo de confianza para la media poblacional de una distribución normal con desviación típica conocida, un intervalo de confianza para la media poblacional de una distribución de modelo desconocido y para la proporción en el caso de muestras grandes–. Este tipo de problemas van asociados a la introducción de nuevos conceptos relacionados con la idea de dispersión, como pueden ser, distribución muestral, error de estimación e intervalo de confianza.

Dado que el razonamiento inferencial parte de los datos para detectar tendencias o fuentes de variabilidad con el fin de extrapolar los datos y realizar inferencias de muestras a poblaciones, el

alumno deberá tener claro los instrumentos necesarios para conseguir tal objetivo. Sin embargo, la adaptación a los nuevos conceptos y razonamientos, relacionados con las medidas de dispersión, conllevan ciertas dificultades de entendimiento (Vallecillos y Batanero, 1997; Castro Sotos, Vanhoof, Van den Noortgate y Onghena, 2007; Canal y Behar, 2010).

En general, la principal dificultad que encuentran los estudiantes cuando se enfrentan a actividades de inferencia estadística es saber qué variable está siendo estudiada y, como consecuencia, determinar la distribución con la que se distribuye dicha variable. Las distribuciones más empleadas en esta tarea son: (1) la distribución de los datos de la muestra seleccionada con el fin de realizar inferencia sobre la población. En esta situación las medidas que describen dicha distribución son la media aritmética (o media) \bar{x} y la desviación típica de la muestra S_x ; (2) la distribución de probabilidad de la población. En general, éste suele seguir una distribución gaussiana o normal de parámetros μ y σ los cuales, en la mayoría de los casos pretendemos estimar o contrastar; (3) la distribución de probabilidad de la variable aleatoria media muestral. De todas las muestras, del mismo tamaño, que se extraen de una población se verifica que la media muestral es una variable aleatoria con una distribución de probabilidad. Por ejemplo, si consideramos una población que se distribuye según una distribución $N(\mu, \sigma)$ y consideramos una muestra aleatoria simple de tamaño n , se tiene que la variable media muestral, denotada habitualmente como \bar{X} , se distribuye según una distribución normal con media μ y desviación típica $\frac{\sigma}{\sqrt{n}}$, es decir, $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

Por otro lado, dentro de este contexto, la estimación de los parámetros desconocidos de una población, con una fiabilidad o un error prefijado (por estimación puntual o por intervalos), genera ciertas dificultades de entendimiento. Ejemplo de ello son:

- La relación que existe entre el aumento o disminución del tamaño de una muestra y la precisión (o error de estimación) que se tiene a la hora de obtener una estimación aproximada del parámetro en estudio, mediante su correspondiente intervalo de confianza. Dado que en un intervalo de confianza, el error en la estimación está inversamente relacionado con el tamaño muestral, se tiene que a medida que aumenta el tamaño de la muestra ganamos precisión en la estimación del parámetro. Este hecho provoca que la dispersión existente entre el verdadero valor del parámetro y su correspondiente estimador sea menor o mayor en función del tamaño muestral.
- La relación existente entre la varianza (o desviación típica) de la variable de la población y la varianza (o desviación típica) de la variable media muestral en función al tamaño muestral considerado. En este caso se tiene que a medida que aumenta el tamaño de la muestra seleccionada la dispersión de la distribución de la variable aleatoria, media muestral, es inferior que la descrita por la distribución de la población (suponiendo que la población se distribuye según una distribución gaussiana).

Es necesario señalar que, aunque son muchos los estudiantes que aprenden a realizar los cálculos necesarios para obtener estimaciones de los parámetros puntualmente, por intervalos o por test de hipótesis, no siempre consiguen comprender todos los procesos ni los conceptos que llevan a cabo dichos cálculos. Este hecho, con origen en las enseñanzas de bachillerato, perdura en el ámbito universitario, lo cual nos advierte de la importancia que supone una introducción óptima de los conceptos inferenciales. Algunas de las investigaciones que dan fe de estas observaciones son las llevadas a cabo por Alvarado (2007), Olivo (2008) y Vallecillos (1994), que describen dificultades de comprensión del teorema central del límite, el intervalo de confianza y el contraste de hipótesis, respectivamente.

IMPLICACIONES DIDÁCTICAS DE LA DISPERSIÓN EN INFERENCIA ESTADÍSTICA

La dificultad que encierra la percepción que pueden tener los estudiantes con respecto a la dispersión asociada a los conceptos de la inferencia estadística, puede ser resuelta hoy en día a través de un ambiente computacional, resultando más atractivo si se trabaja en contextos de datos reales (Sánchez, Borim y Coutinho,2011). De esta forma, mediante un razonamiento informal el estudiante desarrolla los conceptos necesarios para comprender, por ejemplo, el efecto directo que tiene un aumento o disminución del tamaño muestral en la precisión de la estimación de los parámetros poblacionales mediante intervalos de confianza, de un modo más sencillo e intuitivo.

Como recursos de apoyo para los procesos de enseñanza-aprendizaje destacamos las posibilidades de los simuladores. Estas herramientas digitales sirven de apoyo en los procesos de transferencia de conocimiento permitiendo asimilar de una forma más correcta los conceptos estadísticos. Ejemplo de ello lo encontramos en www.rossmance.com/applets/samplemeans/samplemeans.html, donde el alumno puede simular la distribución de la media muestral y ver como varía la dispersión de la distribución en función del tamaño seleccionado (véase Figura 1). En diferentes ventanas el simulador representa la distribución poblacional (izquierda-superior), la distribución de datos en la última muestra simulada (derecha-superior), la distribución muestral empírica de todas las medias que se van obteniendo en sucesivas muestras (izquierda – inferior) y la misma distribución muestral estandarizada, que se va aproximando a la distribución normal $n(0,1)$ cuando se repite el proceso de muestreo.

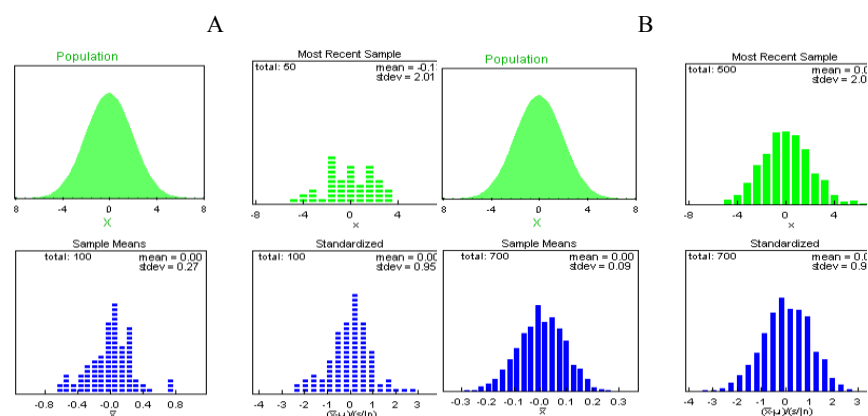


Figura 14. Simulación A: $n=50$ y B: $n=500$

CONCLUSIONES

A lo largo de los niveles educativos, el currículo recoge desde la Educación Primaria el primer contacto con la idea de dispersión desde una perspectiva descriptiva, en el que el principal interés de estudio radica únicamente en analizar un conjunto de datos sin llegar a generalizar los resultados del análisis. Una vez que el estudiante cursa el segundo año de bachillerato de la modalidad de Humanidades y Ciencias Sociales debe haber comprendido los diversos significados de la dispersión con el fin de construir un significado global de la misma y para que no aplique una serie de fórmulas que quizás no comprenda. En esta etapa los estudiantes han de manejar a la vez variables estadísticas y variables aleatorias, junto con sus distribuciones y las distribuciones muestrales. Por lo que el uso de herramientas informáticas y simuladores puede servir como hilo conductor entre los contenidos que los estudiantes deben aprender y los significados de éstos.

En resumen, es importante que el profesor sea consciente del papel destacado de la dispersión en estadística. Puesto que, como indicó Moore (1990), la estadística es la ciencia de los datos y estos están caracterizados por la variabilidad, la habilidad para percibir, medir y explicar la dispersión de los datos y de los modelos que utilizamos para describirlos es la clave del razonamiento estadístico.

Agradecimientos: Proyecto EDU2013-41141-P (MEC) y grupo FQM126 (Junta de Andalucía).

Referencias

- Alvarado, H. (2007). *Significados del Teorema Central del Limite en la Enseñanza de la Estadística en Ingeniería. Tesis doctoral. Universidad de Granada.*
- Canal, G. y Behar, R. (2010). *The confidence intervals: a difficult matter, even for experts. En C. Reading (Ed.), Proceedings of the Eighth International Conference on Teaching Statistics. Voorburg, The Netherlands: International Statistical Institute.*
- Castro Sotos, A. L., Vanhoof, S., Van den Noortgate, W. y Onghena, P. (2007). *Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. Educational Research Review 2, 98-113.*
- Cobb, G. W. y Moore, D. S. (1997). *Mathematics, statistics, and teaching. The American Mathematical Monthly, 104, 801-823.*
- Estepa, A., y del Pino, J. (2013). *Elementos de interés en la investigación didáctica y enseñanza de la dispersión estadística. Números, 83, 43-63.*
- Godino, J. D., Batanero, C. y Font, V. (2007). *The onto-semiotic approach to research in mathematics education. ZDM. International Journal on Mathematics Education, 39(1-2), 127-135.*
- MECD (2015). *Real Decreto 1105/2014 de currículo básico de la Educación Secundaria Obligatoria y del Bachillerato. Madrid: Autor.*
- Moore, D. S. (1990). *Uncertainty. En L. A. Steen (Ed.), On the shoulders of giants: New approaches to numeracy (pp. 95-137). Washington, D. C.: National Academy Press.*
- Olivo, E. (2008). *Significado de los intervalos de confianza para los estudiantes de ingeniería en México. Tesis doctoral. Universidad de Granada.*
- Sánchez, E., Borim, C. y Coutinho, C. (2011). *Teachers' understanding of variation. En C. Catanero, G. Burrill, & C. Reading (Eds.), Teaching statistics in school mathematics-challenges for teaching and teacher education (pp. 211-221). New York: Springer.*
- Snee, R. D. (1990). *Statistical thinking and its contribution to total quality. American Statistician, 26, 116-121.*
- Vallecillos, A. (1994). *Estudio teórico-experimental de errores y concepciones sobre el contraste estadístico de hipótesis en estudiantes universitarios. Tesis doctoral. Universidad de Granada.*
- Vallecillos, A. y Batanero, C. (1997). *Conceptos activados en el contraste de hipótesis estadísticas y su comprensión por estudiantes universitarios. Recherches en Didactique des Mathématiques, 17, 29-48.*
- Wild, C. J. y Pfannkuch, M. (1999). *Statistical thinking in empirical enquiry. International Statistical Review, 67(3), 223-263.*