

ARTÍCULOS DE INVESTIGACIÓN

DOI: 10.24844/EM3001.02

Niveles de razonamiento sobre variación estadística de estudiantes de nivel medio superior al resolver problemas en un contexto de riesgo

Reasoning Levels about Statistical Variation of High School Students Solving Problems in Risk Context

José Antonio Orta Amaro¹
Ernesto Sánchez Sánchez²

Resumen. En esta investigación se formuló la siguiente pregunta: ¿Qué aspectos del razonamiento de los estudiantes sobre la variabilidad emergen mediante su actividad en la resolución de problemas ubicados en un contexto de riesgo? Para responderla se diseñó un cuestionario basado en dos problemas de comparación de conjuntos de datos en contexto de riesgo, uno sobre un juego hipotético y otro sobre tratamientos médicos. Los problemas fueron respondidos por 59 alumnos de bachillerato de 16 a 17 años de edad que aún no habían tomado el curso de Estadística y probabilidad. Con base en la taxonomía SOLO y en las respuestas de los participantes, se propuso una jerarquía para definir los principales niveles de razonamiento presentes en las respuestas. Se muestra que el desarrollo del razonamiento en sus niveles más bajos solo considera un valor de cada conjunto, y en sus niveles más altos articula varios datos de cada conjunto e integra consideraciones sobre la variabilidad, en particular, en forma de riesgo. En el análisis de las respuestas se revela la importancia de que los estudiantes consideren la media aritmética.

Fecha de recepción: 6 de marzo de 2017. **Fecha de aceptación:** 30 de octubre de 2017

¹ Escuela Nacional para Maestras de Jardines de Niños jaortaa@gmail.com

² Departamento de Matemática Educativa. Cinvestav, Instituto Politécnico Nacional esanchez@cinvestav.mx

Palabras clave: *Variación estadística; contexto de riesgo; razonamiento; nivel medio superior.*

Abstract. In this research, the following question was raised: What aspects of students' reasoning on the variability emerge through their activity in the resolution of problems in a risk context? To answer it, a questionnaire was designed based on two problems about comparing groups in risk context, one about a hypothetical game and another on medical treatments. The problems were answered by 59 high school students (16 to 17 years old) who had not yet taken a Statistics and probability course. Based on the solo taxonomy and in the participant' answers, a hierarchy was proposed to define the main reasoning levels present in the answers. It is shown that the development of reasoning in its lowest levels only considers one value of each group, and at their highest levels articulates several data of each group and integrates considerations about the variability, interpreted like risk. The analysis of the answers shows the importance that students consider the arithmetic mean.

Keywords: *Statistical variation; risk context; reasoning; high school students.*

INTRODUCCIÓN

La variabilidad es una realidad observable que está presente en todas partes y afecta todos los aspectos de la vida (Wild y Pfannkuch, 1999). Un objetivo central de la estadística es describir, medir y explicar la variabilidad no determinística en sus diversas manifestaciones, por ejemplo en los datos, las muestras o en las distribuciones (Shaughnessy, 2007). Los términos *variabilidad* y *variación* suelen utilizarse como sinónimos, no obstante, conviene utilizar la palabra *variación* para referirse a la descripción o medida de la entidad observable llamada *variabilidad* (Reading y Shaughnessy, 2004).

Aunque la variabilidad es central para los estadísticos, no se suele comunicar su importancia a los estudiantes, quienes aprenden las medidas de variación como fórmulas para encontrar números, sin que se construya su significado; por ejemplo, para ellos no es clara la relación de las medidas de variación con los demás conceptos básicos de la estadística, a saber: gráficas, medidas de tendencia central, datos bivariados, probabilidad, etcétera. Además, no siempre es fácil

entender cómo las expresiones matemáticas de las medidas de variación, sobre todo de la desviación estándar, permiten medir la dispersión de un conjunto de datos. Por su dificultad y la sobrevaloración que se otorga a las medidas de tendencia central, el estudio sobre la variabilidad ha estado durante mucho tiempo prácticamente ausente en la enseñanza de la probabilidad y la estadística (Shaughnessy, 1997).

En consecuencia, la investigación didáctica sobre la variabilidad es relativamente reciente. En los años 90 algunos autores como Green (1991) y Shaughnessy (1997) hicieron un llamado, cada uno por su parte, sobre la necesidad de investigar cómo razonan los estudiantes en problemas donde la variabilidad estadística está presente, y cómo pueden mejorar su razonamiento. Desde entonces se han buscado contextos en los cuales plantear tareas que propicien, durante el proceso de su solución, el desarrollo de ideas relevantes sobre la variabilidad. Se ha encontrado que situaciones de muestreo repetido, de juegos de azar, de datos acerca del clima y de mediciones repetidas de magnitudes físicas permiten formular problemas para motivar la reflexión de los alumnos sobre la variabilidad.

Un contexto que aún no se ha explorado suficientemente es el de riesgo; es decir, de situaciones de incertidumbre en las que hay resultados que pueden causar pérdidas o daños. No obstante, las pocas experiencias que se han realizado utilizando problemas en este contexto sugieren que es apropiado para comprometer a los estudiantes en su solución (Kahneman y Tversky, 2000). ¿Qué aspectos del razonamiento de los alumnos sobre la variabilidad emergen mediante su actividad en la resolución de problemas ubicados en un contexto simple de riesgo? El objetivo del este estudio es explorar los niveles de razonamiento que muestran los participantes en su respuestas a tareas en las que intervienen la variabilidad y el riesgo.

ANTECEDENTES

El tema de la variabilidad estadística en la investigación didáctica de los niveles básicos es relativamente reciente. Durante la primera década de este siglo, una serie de estudios fueron configurando un programa de investigación acerca del razonamiento de los alumnos de niveles básicos y preuniversitarios sobre la variación estadística. Entre los rasgos de este programa cabe destacar el tipo de tareas que se utilizan para explorar el razonamiento de los estudiantes y los contextos en los que se inscriben.

Varios estudios han explorado el razonamiento sobre la variabilidad en problemas de muestreo repetido con estudiantes tanto de primaria como de secundaria: Shaughnessy *et al.* (1999), Torok y Watson (2000), Kelly y Watson (2002) han analizado el razonamiento con el *Problema de los Dulces*; se trata de preguntas sobre la variabilidad del número de dulces rojos en una muestra de tamaño 10 de una urna que contiene dulces de diferentes colores, con 50% de dulces rojos. En sus respuestas a tales problemas los jóvenes muestran ideas en diferentes niveles sobre centro y dispersión. Watson, Kelly, Callingham y Shaughnessy (2003) utilizaron datos de 746 estudiantes de grados 3, 5, 7 y 9 para sugerir un modelo de comprensión de la variabilidad consistente en 4 niveles. Usaron tareas en diferentes contextos para explorar el razonamiento de los estudiantes, por ejemplo, predecir las frecuencias de ocurrencia de cada cara en 60 lanzamientos de un dado; predecir la frecuencia en la que se detienen 10 (o 50) giros de una ruleta, y otras en contextos escolares. El modelo resultante describe niveles de comprensión de la variabilidad: 1. Prerrequisitos 2. Reconocimiento parcial, 3. Aplicaciones, 4. Aspectos críticos.

El clima es otro contexto para propiciar la reflexión sobre la variabilidad. Torok y Watson (2000) y Watson y Kelly (2005) utilizaron un problema basado en los datos de varios años de los promedios de temperatura máxima diaria. Se pedía a los alumnos que explicaran el significado de “temperatura máxima diaria” y que elaboraran gráficas. Reading (2004) usó un problema en el que estudiantes de 8º a 12º grados analizaron datos de lluvia y de temperatura para decidir en qué semana del año convenía llevar a cabo un evento social, con el objetivo de minimizar la probabilidad de mal clima. Watson y Kelly (2004) implementaron una tarea en contexto de azar basada en una ruleta con solo dos posibles resultados y probabilidad $\frac{1}{2}$ para alumnos de grados 3º, 5º, 7º y 9º (8-15 años). Petrosino, Lehrer, y Schauble (2003) propusieron actividades de registro de medidas repetidas de magnitudes físicas (altura de un asta bandera, longitud de un lápiz y la altura máxima que podía alcanzar un cohete a escala) con estudiantes de 4º grado (9 y 10 años). Los niños elaboraron distribuciones de errores de medida, juzgando la precisión del proceso de medida mediante el análisis de la variabilidad de la distribución. Ben-Zvi (2004) observó a alumnos de 7º grado frente a una tarea de comparación de grupos y con apoyo de tecnología (Excel); un grupo consistía en las longitudes de apellidos hebreos y el otro la longitud de apellidos norteamericanos.

En los estudios mencionados los sujetos participantes, en su mayoría, son de nivel primaria y secundaria; además, los problemas no se refieren a cálculos

de ninguna medida de dispersión, sino al análisis de los datos para comprender un fenómeno o tomar una decisión. Un rasgo notable de los estudios mencionados es que, a pesar de que las tareas utilizadas se relacionan con la incertidumbre y el azar, no emergen de los estudiantes razonamientos de probabilidad; en los estudios de los autores mencionados, estos se enfocaron más en el análisis de datos que en el azar. Es probable que no buscaran que emergieran ideas de probabilidad, por considerar que sus sujetos no tenían aún la madurez suficiente para hacerlo. En problemas de toma de decisiones en contextos de riesgo es evidente el papel de la probabilidad; de hecho, las situaciones paradigmáticas de riesgo las incluyen explícitamente.

En efecto, en uno de los estudios de Kahneman y Tversky (1984), los investigadores proponen a los participantes elegir entre las dos opciones siguientes:

- A. Ganar \$1000 con probabilidad de 85% o perder todo con probabilidad de 15%
- B. Ganar \$800 con certeza, es decir, con probabilidad 1.

En la teoría clásica de la toma de decisiones se resuelven estas tareas mediante el principio de *maximización de la utilidad* (Chernoff y Moses, 1959), consistente en comparar la utilidad de cada opción y elegir aquella que ofrezca mejores utilidades. Si la función utilidad se identifica con la ganancia esperada (como se hace en muchas situaciones), entonces la opción A debería ser elegida porque ofrece una ganancia de \$850, mientras que la B de solo \$800. Sin embargo, no siempre es conveniente identificar la función utilidad con la ganancia, pues la gente muchas veces racionalmente prefiere B, renunciado a obtener la máxima ganancia (esperada), pero evitando el *riesgo* que implica optar por A; esta respuesta contraria a la norma es un rasgo psicológico que se conoce como *aversión al riesgo*. En la presente investigación se buscaron problemas para relacionar de manera más evidente la variabilidad en los datos con la incertidumbre, y encontramos en las situaciones de riesgo de los estudios de Kahneman y Tversky una pista para formular problemas que permitieran hacer dicha conexión.

MARCO CONCEPTUAL

El marco conceptual expone brevemente los principales conceptos que permitan entender el sentido de la investigación. En primer lugar, se precisan algunos

conceptos relacionados con el *contenido estadístico* en juego, en especial la variación, el riesgo y la probabilidad. En segundo lugar, se expone la idea de razonamiento como el aspecto de la actividad cognitiva del alumno que interesa estudiar. Finalmente, se mencionan algunos aspectos del modelo SOLO utilizado para organizar los datos y construir una jerarquía de razonamiento.

CONTENIDO ESTADÍSTICO EN JUEGO

En esta investigación se pone atención al razonamiento que exhiben los estudiantes frente a tareas que implican aspectos del concepto de variación o dispersión de un conjunto de datos. En particular, el estudio se relaciona con dos medidas de variación, el *rango* y la *desviación media*. Dado un conjunto de datos numéricos, el rango es simplemente el máximo menos el mínimo de los datos, mientras que la desviación media es la *media aritmética* de las desviaciones absolutas de cada valor respecto a la *media aritmética* de los datos. Hemos subrayado el concepto de media aritmética para hacer notar que se utiliza dos veces en la definición del concepto de desviación media (lo mismo ocurre en la definición de desviación estándar). Esto pone de manifiesto que el estudio de las medidas de variación, como la desviación media y la desviación estándar, presuponen el concepto de media aritmética, la cual se define como la suma de todos los datos, dividida entre el número de estos. Conviene notar que una de las medidas de dispersión más utilizadas en estadística es la desviación estándar, no obstante, su construcción es mucho más compleja y, como veremos, está más alejada de la posibilidad de que los estudiantes de los niveles preuniversitarios la entiendan como una medida de variación; en todo caso, entender e interpretar el rango y la desviación media como medidas de variación puede ser un antecedente importante para la comprensión de la desviación estándar.

RIESGO

Según Wild y Pfannkuch (1999) en estadística deben considerarse tres mensajes sobre la variación: 1) Es omnipresente, 2) Puede tener serias consecuencias prácticas, y 3) La estadística es un medio para entender un mundo acosado por la variación. Desarrollar el contenido del segundo mensaje puede ser útil para

motivar a que los alumnos se comprometan en el estudio de la variación. En este sentido, en algunas situaciones de *riesgo* se pueden apreciar de manera clara consecuencias de la variación. De acuerdo con Fischhoff y Kadvaný (2011) el riesgo está presente cuando se tienen potenciales resultados no deseados que pueden traer como consecuencia pérdidas o daños. Definir el riesgo significa especificar los resultados valiosos y los no deseados en un orden que refleje el valor que se les atribuye. El análisis del riesgo ofrece información para la toma de decisiones; cuando este análisis resulta en un conjunto ordenado de posibles resultados, la regla para tomar decisiones es simple: elegir la opción cuyo resultado produzca la mayor cantidad del valor (dinero, descanso, acres de tierras húmedas, etcétera) y menor cantidad de daño (pérdidas, dolor, reducción de tiempo de vida) (Fischhoff y Kadvaný, 2011: 65). En nuestra investigación se proponen situaciones de riesgo con el objetivo de que en el proceso de solución emerjan algunas consideraciones del estudiante sobre la variación presente en los datos.

PROBABILIDAD

En psicología un paradigma de investigación sobre la toma de decisiones en situaciones de riesgo consiste en analizar las respuestas de los sujetos a enunciados concisos que solo se apoyan en las nociones de *probabilidad* y *ganancia*, como el ejemplo que se mencionó en la introducción. Esto es un indicador de que, en general, en el análisis de las situaciones de riesgo se involucra implícita o explícitamente una noción de probabilidad. Por las características de los problemas que se utilizaron para explorar el razonamiento de los participantes, conviene recordar que hay diferentes concepciones de probabilidad (Batanero, Henry, Parzys, 2005). Aunque estos autores distinguen cuatro acercamientos a la probabilidad, una demarcación radical y suficiente para el caso presente, de acuerdo con Carranza y Fuentealba (2010), y que ha llevado a hablar de la *dualidad* de la probabilidad a lo largo de su historia (Hacking, 1975), es entre su significado frecuencial y su significado bayesiano. El primero se refiere al fenómeno de “estabilización de la frecuencia de aparición de un fenómeno [evento]”, y el segundo, a la noción de grado de “certeza de la veracidad de una proposición” (Carranza y Fuentealba, 2010: 57). Estos autores afirman que en una situación dada no se elige libremente uno u otro significado, sino que dependen del contexto en el que se requieren las consideraciones de

probabilidad. Es importante, entonces, mencionar las condiciones que permiten formular un juicio de probabilidad con base en cada significado, esto con el fin de considerarlo en el análisis de los problemas que enfrentan los alumnos, y de las respuestas que proponen.

Las situaciones en las que es aplicable el enfoque frecuencial de probabilidad son experiencias aleatorias que pueden ser repetidas bajo condiciones similares; además, se debe satisfacer la condición de regularidad de las frecuencias, es decir, que para cada evento sus frecuencias relativas converjan a un valor fijo cuando el número de repeticiones de la experiencia se repite indefinidamente. Para aplicar el enfoque bayesiano “es suficiente que dispongamos de una proposición y que desconozcamos su veracidad” (Carranza y Fuentealba, 2010: 59), pues esto permite asignar una medida que refleje la creencia del sujeto acerca de la veracidad de la proposición. Veremos que, en el caso presente, los estudiantes eligen entre tres opciones con base en una estimación subjetiva acerca de las opciones que perciben.

RAZONAMIENTO

El razonamiento surge de diversas maneras, en especial cuando: a) Se busca y apoya la solución de un problema, b) Se analizan las posibilidades para tomar una decisión, o c) Se justifica la verdad de una proposición. Aunque el razonamiento es un proceso mental, las pautas principales que lo constituyen pueden expresarse oralmente o por escrito; esta característica permite estudiarlo, pues las expresiones pueden ser registradas, analizadas, discutidas y compartidas. Cuando los alumnos explican o justifican por escrito (u oralmente) la respuesta a una tarea, o la toma de una decisión, exhiben ciertos rasgos característicos de su razonamiento. En la presente investigación se les pide que tomen una decisión y ofrezcan los argumentos que la motivan. Una buena decisión en cada una de las tareas planteadas en este estudio depende de la apreciación y tratamiento de la variabilidad en los datos, pero ésta es compleja y depende de diversos componentes. Se analizan entonces sus razonamientos sobre la variabilidad en función de las componentes pertinentes que incorporan en sus justificaciones. Un método que puede aplicarse al análisis de los componentes que constituyen un razonamiento sobre la solución de una tarea ha sido desarrollado por Biggs y Collis (1982 y 1991), dando lugar al *Modelo SOLO*, que se introduce en la siguiente sección.

Las concepciones, sesgos y creencias sobre conceptos presentes en las situaciones o problemas que resuelven los estudiantes tienen una fuerte influencia en su razonamiento. En particular, con relación al riesgo han sido identificadas actitudes de los sujetos que influyen en sus respuestas y no necesariamente coinciden con el razonamiento normativo. Kahneman y Tversky (2000) señalan que el ejemplo paradigmático de la decisión con riesgo es la participación en un juego cuyos resultados generan posibles ganancias o pérdidas económicas de acuerdo con ciertas probabilidades específicas. En estas situaciones han encontrado la *aversión al riesgo* y la *propensión al riesgo*. La primera consiste en evitar, en la medida de lo posible, situaciones desfavorables, aunque esto presuponga renunciar a una ganancia que podría compensar el riesgo. Su contraparte es la *propensión al riesgo*, consistente en preferir una situación en función de resultados favorables, a pesar de que implique mayor riesgo.

MODELO SOLO

El modelo SOLO (Structure of Observed Learning Outcome), elaborado por Biggs y Collis (1982, 1991), se ubica dentro de las teorías neo-piagetianas. Parte de supuestos constructivistas, es decir, sostiene que el sujeto construye su manera de entender el mundo mediante su involucramiento en actividades diversas; además, supone que el desarrollo cognitivo muestra un patrón general común en diferentes campos o dominios, y dicho desarrollo se produce a través de varias etapas o estadios organizados de manera jerárquica. En particular, en contextos escolares y académicos las actividades suelen estar mediadas por el discurso, es decir, referirse a situaciones evocadas. En el marco de Biggs y Collis (1982) se definen las etapas de desarrollo de manera diferente a Piaget, pues aunque se corresponden, las caracterizan en términos de *modos de representación* y no en términos de esquemas lógicos. Los autores destacan que, dado un contenido específico y especificable, la calidad del aprendizaje que puede desplegar un sujeto respecto a ese contenido depende tanto de factores externos –por ejemplo, las características del proceso de enseñanza por el que ha pasado– como de factores intrínsecos –su motivación, habilidades y conocimiento previo–, y que no necesariamente se corresponden con los rasgos que caracterizan el modo de representación al que se supone pertenece. Por esta razón, para la enseñanza es útil encontrar formas de evaluar la calidad del aprendizaje que logran los sujetos en relación con contenidos

definidos, y esto se logra enfocando la evaluación en sus respuestas a preguntas específicas.

La Taxonomía SOLO propone las bases para discriminar entre las diferentes respuestas de los estudiantes, en función de la calidad del aprendizaje que los sujetos muestran acerca de un determinado contenido. Se proponen 5 niveles en los cuales se pueden clasificar las respuestas: *Prestructural*, *Uniestructural*, *Multiestructural*, *Relacional* y *Abstracto Extendido*. Los niveles centrales son definidos en función de si las respuestas contienen los aspectos relevantes a la solución de la tarea. Una solución es una estructura formada por diferentes componentes adecuadamente relacionadas; si una respuesta solo incluye una componente, entonces es Uniestructural; si incluye dos o más componentes pero no las relaciona de manera adecuada es Multiestructural, y si incluye varias componentes de manera integrada es Relacional. Cuando una pregunta no contiene ninguna componente pertinente a la tarea es Prestructural. El nivel Abstracto Extendido se define cuando una respuesta Relacional se generaliza hacia otras tareas; generalmente, para detectar respuestas en este nivel es necesario ampliar las tareas. Por otro lado, conviene indicar que en este estudio las respuestas que se analizan consisten en las justificaciones que ofrecen los alumnos a una elección entre dos situaciones; en consecuencia, se analiza su razonamiento, por lo cual lo que se construye es una Jerarquía SOLO de razonamiento.

MÉTODO

Esta investigación se enfoca en el razonamiento de los estudiantes frente a enunciados que se refieren a situaciones de riesgo. En general, en la investigación sobre los procesos de enseñanza-aprendizaje escolares se estudian los comportamientos, razonamiento, creencias o dificultades de los sujetos frente a diferentes tipos de tareas formuladas en el discurso, combinando lenguaje matemático y lenguaje coloquial; muchas de las tareas se enriquecen ubicándolas en contextos extra-matemáticos. Hay dos propósitos o funciones del uso de los contextos extra-matemáticos en situaciones de enseñanza, a saber: 1) El contexto extra-matemático ofrece una oportunidad para ampliar el sentido de los conceptos matemáticos, 2) Los conceptos matemáticos ofrecen una oportunidad para mejorar la comprensión de situaciones extra-matemáticas (Niss, Blum, Galbraith, 2007). En el primer caso, las situaciones

extra-matemáticas son un referente que ayuda a darle sentido a las ideas matemáticas; en el segundo, se muestra la potencia de la matemática para entender el mundo real. En la presente investigación, la intención es la primera, utilizar un contexto de riesgo para darle sentido a la consideración de la variabilidad en los datos.

Como se quiere explorar el razonamiento de los estudiantes sobre la variabilidad en contexto de riesgo, es importante: 1) Determinar el nivel de los alumnos que se van a examinar, y 2) Utilizar tareas que propicien que externen sus ideas sobre la variabilidad.

PARTICIPANTES

Los participantes fueron 59 estudiantes de nivel medio superior (16 a 17 años de edad) de una escuela pública de la Ciudad de México, quienes aún no habían tomado el curso de Estadística y probabilidad, en ese grado. No obstante, en el nivel educativo anterior (secundaria) debieron haber abordado los temas de medidas de tendencia central y dispersión (SEP, 2011), por lo cual se puede suponer que saben calcular la media aritmética, la mediana y la moda.

INSTRUMENTOS

Para explorar los razonamientos de los participantes, los autores diseñaron dos problemas en situaciones de riesgo que implican comparaciones de conjuntos de datos. Para hacer las comparaciones conviene calcular las medias aritméticas de los conjuntos y compararlas. Los datos del Problema 1 se han elegido de manera que tengan la misma media aritmética y diferente dispersión. En consecuencia, los conjuntos a comparar en el Problema 1 son equivalentes en términos del beneficio obtenido, pero son diferentes en términos del riesgo que implican. En el Problema 1 (ver Figura 1) se presenta una situación en la que se ofrecen dos muestras de resultados de las ganancias de dos juegos; a partir de los datos, los estudiantes deben decidir en qué juego les conviene jugar.

En una feria, se invita a los asistentes a participar en uno de dos juegos. Juan puede participar en un juego, pero no en ambos. Para saber por cuál decidirse observa, anota y ordena los resultados de dos muestras de 10 personas que han participado en cada juego. Las pérdidas (-) o premios (+) en efectivo que han obtenido las 20 personas se muestran en las siguientes listas:

Juego 1:
 15 -21 -4 50 -2 11 13 -25 16 -4

Juego 2:
 120 -120 60 -24 -21 133 -81 96 -132 18

a) Si tienes la posibilidad de participar en un solo juego ¿Cuál juego elegirías?
 ¿Por qué?

Figura 1. Problema 1.

Debe notarse que en ambos juegos hay diferentes valores, los negativos representan pérdidas y los positivos ganancias; también que no se da información sobre el mecanismo del juego ni de la distribución de probabilidad y de las ganancias correspondientes.

En consecuencia, no se dan las condiciones necesarias para hacer estimaciones de probabilidad con base en un enfoque frecuencial, por ejemplo, no se puede determinar el espacio de todos los posibles resultados (ganancias) de un juego y sus probabilidades. No obstante, en la medida en que es una situación de incertidumbre, se requiere considerar probabilidades desde un enfoque bayesiano. En efecto, los sujetos pueden considerar los eventos “El juego A es mejor que el juego B”, “El juego A es equivalente al juego B” y “El juego A no es mejor que el juego B” (donde “Es mejor” quiere decir “se gana más”) y buscar argumentos en la información dada para favorecer la probabilidad de alguno de tales eventos. En el análisis de las respuestas de los estudiantes podremos determinar cuáles son los argumentos que ellos consideran para tomar su decisión, lo cual revelará sus estimaciones cualitativas acerca de la probabilidad de los eventos mencionados.

A continuación hacemos un análisis de una solución al problema que aporta argumentos para apoyar una decisión utilizando herramientas estadísticas. Como en muchos problemas estadísticos, no se puede afirmar que es una solución determinista correcta, sino solo que es una solución racional que busca

incorporar toda la información. En primer lugar, se puede aplicar el principio de razón insuficiente, formulado por Laplace. Como no tenemos más que la información de 10 resultados del juego y no hay ninguna razón para inclinarse a pensar que alguno es más probable que el otro, asignamos a cada valor presente de la variable “ganancia” la misma probabilidad (véanse Tabla 1 y Tabla 2):

Tabla 1. Distribución de la ganancia del Juego 1.

Ganancia o pérdida	-25	-21	-4	-2	11	13	15	16	50	Total
Probabilidad	1/10	1/10	2/10	1/10	1/10	1/10	1/10	1/10	1/10	1

Tabla 2. Distribución de la ganancia del Juego 2.

Ganancia o pérdida	-132	-120	-81	-24	-21	18	60	96	120	133	Total
Probabilidad	1/10	1/10	1/10	1/10	1/10	1/10	1/10	1/10	1/10	1/10	1

Dos valores destacados de las distribuciones son el centro y la dispersión, el primero puede representarse con la media y el segundo con el rango o la desviación media (no consideramos la desviación estándar por ser más compleja e innecesaria en este caso). Otros dos valores importantes son el mínimo y el máximo, pues informan sobre el rango de las ganancias en cada juego. En la Tabla 3 se enlistan los 5 valores mencionados de cada distribución:

Tabla 3. Valores importantes de la distribución.

	Juego 1	Juego 2
Valor máximo	50	133
Valor mínimo	-25	-132
Media aritmética	4.9	4.9
Rango	75	265
Desviación media	16.1	80.5

Los valores de la Tabla 3 y sus interpretaciones se utilizan para apoyar algunos de los enunciados que permiten la toma de una decisión. Por ejemplo, un argumento es el siguiente: como la media aritmética de las ganancias de cada juego es 4.9, y asumiendo una hipotética interpretación frecuencial del juego, si se jugara un gran número de juegos, a la larga la ganancia que se obtendría con cada uno sería la misma. Pero, como se ha dicho, no hay indicios suficientes en el problema para asumir el enfoque frecuencial, de modo que conviene pensar que se va a jugar solo una vez. En estas circunstancias se impone llevar a cabo un razonamiento bayesiano. ¿Qué otros argumentos se pueden derivar de los datos para identificar la mejor decisión? El argumento basado solamente en la comparación de los máximos (Elegir el Juego 2 porque se gana hasta 132, mientras que en el juego 1 solo 50) es sesgado, pues únicamente tiene en cuenta los extremos máximos e ignora los demás datos. Un sesgo similar es el simétrico, que solo observa los mínimos (las pérdidas), llevando a tomar la decisión de elegir el Juego 1 “porque se pierde menos”. En cambio, la consideración simultánea de máximo y mínimo lleva a ponderar el riesgo. En este caso, la elección de uno u otro juego va a depender de las preferencias de los sujetos acerca del riesgo; los adversos al riesgo preferirán el Juego 1, mientras quienes sean propensos al riesgo preferirán el Juego 2.

En la Figura 2 se presenta el Problema 2; los datos se refieren a tiempo de vida en años después de seguir un tratamiento u otro; se agregan además las gráficas de las distribuciones.

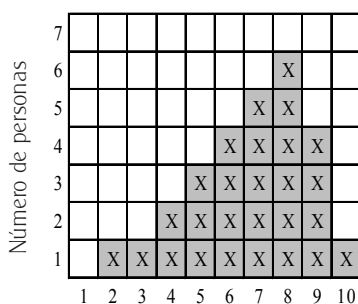
Considera que debes aconsejar a una persona que padece una enfermedad grave, incurable y mortal, pero que es tratable con medicamentos que pueden extender la vida por varios años más. Es posible elegir entre tres tratamientos. Las personas tienen diferentes reacciones a las medicinas, para algunas tienen el resultado previsto, mientras que para otras pueden ser más benéficas o más perjudiciales. En las siguientes listas se muestran los años que han vivido varios pacientes que se han tratado con una de las opciones mencionadas; cada dato de las listas corresponde al tiempo que ha sobrevivido un paciente con el respectivo tratamiento. Después se muestran las gráficas correspondientes a los tratamientos.

Tratamiento 1

Tiempo vivido en años por persona

7	6	5	9	8	4	6
	8	9	5	8	7	5
7	3	7	6	4	8	7
9	8	2	8	10	9	6

Gráfica del tratamiento 1



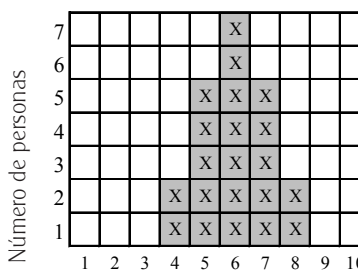
Número de años vividos por persona

Tratamiento 2

Tiempo vivido en años por persona

6	5	6	7	6	5	6
4	7	5	5	7	6	7
8	6	4	6	8	5	7

Gráfica del tratamiento 2



Número de años vividos por persona

¿Qué tratamiento aconsejarías? Explica tu respuesta

Figura 2. Problema 2.

Bajo supuestos similares a los planteados en el problema anterior, se pueden determinar las distribuciones de probabilidad de cada tratamiento (véanse Tablas 4 y 5):

Tabla 4. Distribución del tiempo de vida después de seguir el Tratamiento 1.

Tiempo de vida	2	3	4	5	6	7	8	9	10	Total
Probabilidad	1/27	1/27	2/27	3/27	4/27	5/27	6/27	4/27	1/27	1

Tabla 5. Distribución del tiempo de vida después de seguir el Tratamiento 2.

Tiempo de vida	4	5	6	7	8	Total
Probabilidad	2/21	5/21	7/21	5/21	2/21	1

En la Tabla 6 se han organizado los 5 valores de la distribución que permiten formular argumentos para la toma de una decisión.

Tabla 6. Valores importantes de la distribución.

	Tratamiento 1	Tratamiento 2
Valor máximo	10	8
Valor mínimo	2	4
Media aritmética	6.7	6.0
Rango	8	4
Desviación media	1.61	0.86

Comparando las medias aritméticas se puede argumentar que el Tratamiento 1 ofrece mayor probabilidad de vivir un poco más, por lo que, solo con este argumento, se puede preferir ese tratamiento. Sin embargo, tiene el inconveniente de que hay un caso en que el tiempo de vida es solamente de 2 años y otro en el que es de 3 años. Por otro lado, en el Tratamiento 1 hay un caso que vivió 9 años y otro que vivió 10; mientras en el Tratamiento 2 el que mayor tiempo ha vivido registra 8 años de vida. El Tratamiento 1 puede ser más prometedor, pero

se corre más riesgo; en cambio, el Tratamiento 2 es conservador, con menos riesgo. Nuevamente, la decisión dependerá de las actitudes ante el riesgo de quien toma la decisión. Si es adverso al riesgo optará por el Tratamiento 2, si es propenso al riesgo optará por el Tratamiento 1.

RESULTADOS

El análisis consistió en identificar aquellas componentes pertinentes en las respuestas, en un ir y venir entre un análisis a priori de las soluciones esperadas, así como una comparación constante de las respuestas obtenidas para identificar las componentes que permitieran la construcción de una jerarquía SOLO. En este proceso se construyó la jerarquía SOLO de la Tabla 7.

Tabla 7. Jerarquía de razonamiento SOLO para los problemas.

Preestructural
Las respuestas clasificadas en este nivel tienen justificaciones circulares (por ejemplo, “porque se gana más” o “porque se vive más”), o en el argumento no se utiliza ninguna de las ideas estadísticas pertinentes al problema.
Uniestructural
Eligen un dato de cada conjunto y los comparan, por ejemplo: solo los máximos, solo los mínimos o solo las modas.
Multiestructural
Consideran los máximos y lo mínimos de cada conjunto y los comparan; esto se manifiesta con expresiones como “se gana más, pero se pierde menos”. Comparan las sumas de los datos de cada conjunto y notan que son iguales; esto los lleva a decir que es indiferente elegir cualquiera de los dos conjuntos.
Relacional
Consideran la media y la dispersión. Notan que las ganancias globales son iguales en ambos conjuntos, entonces consideran el riesgo. La elección depende de sus preferencias sobre éste: Los adversos al riesgo eligen el conjunto con menos variabilidad y los propensos al riesgo el de mayor variabilidad.

Lo que caracteriza el nivel de razonamiento reflejado en las respuestas de las tareas planteadas es la cantidad de información que el estudiante es capaz de extraer y utilizar de cada conjunto de datos para elaborar argumentos que apoyen una decisión: en el nivel más bajo de razonamiento no utilizan la información; en el más alto, logran hacer una síntesis adecuada de la información de cada conjunto. En la jerarquía de la Tabla 7 se describen las características de los niveles de manera precisa y, enseguida, se ofrecen ejemplos que muestran el tipo de respuestas que se clasifican en cada nivel.

PRESTRUCTURAL

En este nivel se agruparon las respuestas con argumentos circulares o no pertinentes. Los primeros consisten en elegir un conjunto de datos y justificar la elección diciendo que se “gana más” o “se vive más”. Este tipo de respuesta muestra que aunque notan algo en los conjuntos que los lleva a tomar una decisión, no pueden expresarlo con claridad; lo que se espera de un argumento es que den razones, que aclaren por qué se gana más o se vive más. Las respuestas que no contienen elementos pertinentes consistieron en buscar y comparar proporciones irrelevantes, o buscar patrones de la ocurrencia de pérdidas y ganancias, por ejemplo observar que hay 5 valores negativos y 5 positivos.

Un estudiante cuya respuesta al Problema 1 se encuentra en el nivel Preestructural, elige el Juego 1 y argumenta: *Porque es en donde se presentan mayores oportunidades de ganar más premios*. Una respuesta al Problema 2 que se clasifica en este nivel elige el Tratamiento 1 y comenta: *mis años de vida se extienden*.

Conviene destacar que los participantes entendieron bien el sentido de la pregunta, a diferencia de otros estudios en que las respuestas de nivel Preestructural se producen –en gran medida– porque no se entiende la pregunta, o el contexto es tan familiar que la respuesta se basa en experiencias personales sin relación con el problema, aquí ellos saben que la respuesta consiste en elegir el juego o el tratamiento más favorable. En las respuestas clasificadas en este nivel toman una decisión; no obstante, su elección tiene argumentos circulares o inapropiados.

UNIESTRUCTURAL

Un estudiante elige el Juego 2, y argumenta:

Si llegas a ganar, las ganancias son mucho más grandes que las del Juego 1 y el índice de ganadores es con un 50% de posibilidades, por lo cual pienso que valdría la pena el Juego 2.

Se puede deducir que compara los máximos de cada grupo y elige el Juego 2 porque en éste la ganancia es mayor (133) que en el 1 (50); menciona además un hecho irrelevante, pues en ambos juegos la mitad corresponde a pérdidas y la mitad a ganancias.

Con relación al problema de los tratamientos médicos, un alumno elige el Tratamiento 1, y argumenta:

La enfermedad es mortal, el punto cumbre del pensamiento es vivir más tiempo y en el [Tratamiento] 1 el máximo de tiempo que se puede conseguir es de 10 años.

Esta respuesta solo considera los valores máximos de cada conjunto, 10 y 8 años (aunque no se menciona explícitamente el 8) y considera que con el Tratamiento 1 se vive más por este hecho. Ignora que con el Tratamiento 1 también hay casos que solamente han vivido 2 años.

MULTIESTRUCTURAL

En el siguiente ejemplo, el participante elige el Juego 1 y argumenta:

Porque en caso de perder no perdería tanto, lo máximo que se han llevado en este juego es 50 y lo más que perdieron fue 25, por lo que podría decirse que hay más que ganar, sin perder tanto; en cambio, en el Juego 2 se gana 133 pero se pierde 132, es más arriesgado y, en caso de perder, no convendría.

Evalúa que el Juego 2 es más arriesgado que el Juego 1 mediante la consideración de los máximos y los mínimos, y su decisión refleja aversión al riesgo, pues evita el juego donde se puede perder más, sin importar el monto de lo que se podría ganar. Esta respuesta sería relacional si además del

máximo y el mínimo se utilizaran, en la comparación, los centros de cada distribución.

En el problema de los tratamientos, el estudiante del siguiente ejemplo eligió el Tratamiento 1, argumentando:

Ya que me arriesgaría porque podría vivir 2 años pero también hasta 10 años, en cambio, con el otro tratamiento, sí podré tener un poco más de seguridad de vivir 4 años, mínimo, pero solo lo máximo son 8 años; así que por eso elegí el Tratamiento 1.

Al igual que el caso anterior, también considera los máximos y los mínimos del tiempo de vida después de cada tratamiento y nota que el 1 es el más arriesgado, pero también que ofrece posibilidades de vivir más. Al elegir el Tratamiento 1 muestra propensión al riesgo, como él mismo lo expresa cuando dice "... me arriesgaría".

RELACIONAL

En el problema de los juegos un ejemplo de este nivel es la respuesta de un estudiante que eligió el Juego 2, argumentando:

Sumamos las ganancias (+105) y las pérdidas (-56) del Juego 1; y del Juego 2, 427 contra -378. Podemos ver que en ambos hay una diferencia de 49; yo prefiero el 2 ya que aunque en el 1 ganas poco, también pierdes poco; solo vas a jugar una vez, así que vale la pena ya que podrías ganar más.

En esta respuesta se considera indirectamente la media aritmética al percibir que la suma de los valores (ganancias – pérdidas) es 49 en ambos juegos; el participante también nota el riesgo al decir que "en el 1 ganas poco, también pierdes poco". Es interesante notar que el estudiante hace explícito que "solo vas a jugar una vez", con lo cual muestra que implícitamente su razonamiento es bayesiano. Como se decide por el Juego 2 y con la aclaración de que solo se va a jugar una vez y vale la pena [arriesgar], muestra que es propenso al riesgo.

Con respecto al Problema 2, hemos elegido el siguiente ejemplo en el que el alumno prefiere el Tratamiento 1, con el siguiente argumento:

Tiene mayor promedio de vida el Tratamiento 1 respecto al 2; aunque alguna gente vive menos en el primero, también puede vivir más (arriesgando todo).

Promedio de años de vida en Tratamiento 1: 6.7 años

Promedio de años de vida Tratamiento 2: 6 años.

Calcula la media aritmética de cada conjunto y detecta el riesgo. Muestra que es propenso al riesgo con el comentario “arriesgando todo”.

Conviene aclarar que la jerarquía es un esquema simplificado de los diferentes niveles de razonamiento de los estudiantes, pero las respuestas reales ofrecen variantes en las que se combinan razonamientos del esquema con otras dos características; una de ellas son las expresiones que reflejan que se percibe el riesgo, “es un juego muy arriesgado” o “con este tratamiento se corre más riesgo”; evaluamos que, en general, estas consideraciones mejoran la calidad de la respuesta. Es decir, una respuesta de un nivel puede ser clasificada en el nivel inmediato superior gracias a tales consideraciones. La otra característica de varias respuestas va en el sentido inverso y se presenta cuando hay expresiones incompletas o errores en los cálculos, si esto ocurre se evalúa inferior la calidad de la respuesta. Por ejemplo, una respuesta que considera dos o más datos y que, de acuerdo a la jerarquía, se clasificaría en Multiestructural, es clasificada en el nivel Uniestructural debido a las expresiones incompletas o errores en los cálculos. Teniendo en cuenta tales consideraciones, cada respuesta fue clasificada en alguno de los niveles SOLO, y las frecuencias de respuesta en cada nivel se presentan en las Tablas 8 y 9.

Tabla 8. Distribución de frecuencias de respuestas por nivel SOLO al Problema 1.

Nivel	Preestructural	Uniestructural	Multiestructural	Relacional	Total
No. respuestas	20	16	20	3	59
Porcentaje	34%	27%	34%	5%	100%

Tabla 9. Distribución de frecuencias de respuestas por nivel SOLO al Problema 2.

Nivel	Preestructural	Uniestructural	Multiestructural	Relacional	Total
No. respuestas	22	8	26	3	59
Porcentaje	37%	14%	44%	5%	100%

Se puede observar que en ambas preguntas poco más de la tercera parte de las respuestas cae en el nivel Preestructural. Esto significa que los estudiantes no saben argumentar su elección con base en la información que proporcionan los datos. Algunos probablemente ven en estos rasgos que les hacen creer que una opción les conviene sobre la otra, pero no lo expresan de manera adecuada, sino solo con respuestas circulares o expresiones vagas. Las frecuencias en el nivel Uniestructural indican que solamente tienen en cuenta un dato de cada conjunto, y la disminución de la frecuencia en el caso del Problema 2 muestra que hay menos probabilidad, en la situación de los tratamientos, de conformarse con un solo dato; esto mismo explica que aumente la frecuencia de respuestas en el nivel Multiestructural del Problema 2. Se puede conjeturar que los alumnos son más sensibles al riesgo en el caso de tiempo de vida que en el de ganancias monetarias. La tercera parte en el Problema 1 y poco menos de la mitad en el Problema 2 se clasifican en Multiestructural, es decir, no pocos estudiantes se esfuerzan por explorar en los datos. Únicamente 3 respuestas de cada pregunta contienen rasgos de los dos factores clave de los datos, a saber, la media y la dispersión, y llegan a estas consideraciones no mediante una aplicación mecánica de fórmulas preconcebidas, sino a través de un análisis de la situación. En este sentido es que se conjetura que los problemas ayudan al desarrollo del razonamiento acerca de los datos y su variabilidad.

CONCLUSIONES

Una de las principales dificultades para resolver problemas que involucran la comparación de conjuntos de datos es concebir a estos últimos como agregados (Bakker y Gravemeijer, 2004), es decir, no como elementos individuales y aislados, sino cada conjunto como un todo con propiedades que los distinguen. Al hacerlo, la media, el rango y la desviación media se asumen como parámetros que caracterizan y representan al conjunto. Al emerger algunas consideraciones de riesgo en los argumentos que los estudiantes llevan a cabo, implícitamente realizan estimaciones de probabilidad de tipo bayesiano; esto estimula a que –en lugar de aplicar fórmulas que muchas veces no entienden– busquen argumentos en los datos que apoyen sus decisiones, y con ellos cobren sentido los diferentes estadísticos que pueden entrar en juego. En la jerarquía que se ha construido se observa un progreso que va de elegir elementos aislados del conjunto para hacer la comparación (máximo, mínimo), a la consideración

simultánea de más de un dato (máximo y mínimo), e incluso de todos los datos (su suma). Gracias al contexto de riesgo los análisis que los estudiantes hicieron con los datos tenían un significado relacionado con la búsqueda del mayor beneficio potencial contenido entre los conjuntos de datos. De esta manera, ellos progresan al considerar los conjuntos como agregados y asociarles la media y la variación, dándoles un significado en términos de su contenido y no solo un significado operacional.

La interpretación de la variación como indicadora de riesgo contribuye a valorarla y entender lo que Wild y Pfannkuch (1999: 235) llaman el segundo mensaje que conlleva: “[la variación] puede tener serias consecuencias prácticas”. En efecto, en muchas de sus respuestas los alumnos indican que perciben las diferencias en términos del riesgo presente en los juegos o en los tratamientos. Consideramos que esto muestra el potencial que tiene este tipo de problemas para ofrecer la oportunidad a los estudiantes de valorar la importancia de percibir y analizar la dispersión de los datos.

En el estudio se reveló el papel fundamental que juega la media aritmética, y reafirma que “es imposible considerar la variabilidad sin también considerar el centro ya que ambas ideas son necesarias para darle significado al análisis de datos” (Garfield y Ben-Zvi, 2008: 203). El objetivo inicial del presente trabajo se enfocaba fundamentalmente en la variación de los datos, sobre todo del rango y la desviación media; no obstante, en el análisis de las respuestas se notó que los participantes no utilizaron la media aritmética. La estrategia de sumar los datos de cada conjunto y compararlos fue la más cercana a la consideración de la media, pero omitía el análisis de la variación. Casi ningún estudiante usó en sus respuestas la media aritmética como representante de todo el grupo de datos y, por tanto, que se puede utilizar para hacer comparaciones entre grupos; esto sugiere que los problemas que hemos diseñado pueden complementarse con otros similares en los que la media aritmética no sea igual en ambos conjuntos; de esta manera, antes de entrar en el análisis de la variación se compararían las medias aritméticas.

REFERENCIAS

- Batanero, C., Henry, M. y Parzysz, B. (2005). The Nature of Chance and Probability. In: G. A. Jones (ed.). *Exploring Probability in School: Challenges for Teaching and Learning* (pp. 15-37). New York: Springer.

- Bakker, A. y Gravemeijer, K. P. E. (2004). Learning to Reason about Distribution. In: D. Ben-Zvi & J. Garfield (eds.). *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking* (pp. 147-168). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Ben-Zvi, D. (2004). Reasoning about Variability in Comparing Distributions. *Statistics Education Research Journal*, 3(2), 42-63.
- Biggs, J. B. y Collis, K. (1982). *Evaluating the Quality of Learning: The solo Taxonomy*. New York, Academic Press.
- Biggs, J. y Collis, K. (1991). Multimodal Learning and the Quality of Intelligent Behavior. En: H. Rowe (ed.). *Intelligence, Reconceptualization and Measurement* (pp.57-76). New Jersey: Laurence Erlbaum Assoc.
- Carranza, P. y Fuentealba, J. (2010). Dualidad de la probabilidad y estadística. *UNION*, 24, 57-68.
- Chernoff, H. y Moses, L. (1959). *Elementary Decision Theory*. Dover: New York.
- Fischhoff, B. y Kadvany, J. (2011). *Risk, a Very Short Introduction*. Oxford: United States, New York.
- Garfield, J. y Ben-Zvi, D. (2008). *Developing Students' Statistical Reasoning: Connecting Research and Teaching Practice*. New York: Springer.
- Green, D. R. (1991). A Longitudinal Study of Pupil's Probability Concepts. *Technical Report ME90/91*. Loughborough: University of Technology.
- Hacking, I. (1975). *The Emergence of Probability*. Cambridge: Cambridge University Press.
- Kahneman, D. y Tversky, A. (1984). Choices, Values, and Frames. *American Psychologist*, 39(4), 341-350.
- Kahneman, D. y Tversky, A. (2000). *Choices, Values, and Frames*. Cambridge: Russell Sage Foundation.
- Kelly, B. A. y Watson, J. M. (2002). Variation in a Chance Sampling Setting: The Lollies Task. En: B. Barton, K. C. Irwin, M. Pfannkuch & M. O. J. Thomas (eds.). *Mathematics Education in the South Pacific (Proceedings of the 26th Annual Conference of the Mathematics Education Research Group of Australasia, Auckland, NZ, vol. 2 pp. 366-373)*. Sydney: merga.
- Niss, M., Blum, W. y Galbraith, P. (2007). Introduction. In: W. Blum, P. L. Galbraith, H-W. Henn & M. Niss (eds.). *Modelling and Applications in Mathematics Education. The 14th ICMI Study*. New York: Springer.
- Petrosino, A., Lehrer, R. y Schauble, L. (2003). Structuring Error and Experimental Variation as Distribution in the Fourth Grade. *Mathematical Thinking and Learning*, 5(2&3), 131-156.
- Reading, C. (2004). Student Description of Variation While Working with Weather Data, *Statistics Education Research Journal*, 3(2), 84-105.

- Reading, C. y Shaughnessy, M. (2004). Reasoning about Variation. In: D. Ben-Zvi and J. Garfield (eds.). *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 201-226). The Netherlands: Kluwer Academic Publishers.
- SEP. (2011). *Programas de Estudio 2011 Guía para el maestro*. Educación Básica Secundaria. Matemáticas. México. Secretaría de Educación Pública:
- Shaughnessy, J. M. (1997). Missed Opportunities in Research on the Teaching and Learning of Data and Chance. In: F. Bidulph & K. Carr (eds.). *Proceedings of the Twentieth Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 6-22). Rotorua, N. Z.: University of Waikata.
- Shaughnessy, J. M., Watson, J., Moritz, J. y Reading, C. (1999). School Mathematics Students' Acknowledgement of Statistical Variation. En: C. Maher (Chair), *There's More to Life than Centers. Pre-session Research Symposium, 77th. Annual National Council of Teachers of Mathematics Conference*, San Francisco, CA.
- Shaughnessy, J. M. (2007). Research on Statistics Learning and Reasoning. In: F. K. Lester Jr (ed.). *Second Handbook of Research on Mathematics Teaching and Learning* (pp. 957-1009). Reston: The National Council of Teachers of Mathematics.
- Torok, R. y Watson, J. (2000). Development of the Concept of Statistical Variation: An Exploratory Study. *Mathematics Education Research Journal*, 12(2), 147-169.
- Watson, J., Kelly, B. A., Callingham, R. A. y Shaughnessy, J. M. (2003). The Measurement of School Students' Understanding of Statistical Variation. *International Journal of Mathematical Education in Science and Technology*, 34(1), 1-29.
- Watson, J. M. y Kelly, B. A. (2004). Statistical Variation in a Chance Setting: A Two-year Study. *Educational Studies in Mathematics*, 57, 121-144.
- Watson, J. M. y Kelly, B. A. (2005). The Winds are Variable: Students' Intuitions about the Weather. *School Science in Mathematics*, 105(5), 252-259.
- Wild, C. y Pfannkuch, M. (1999). Statistical Thinking in Empirical Enquiry. *International Statistical Review*, 67(3), 223-265.

