ISSN 0103-636X

Vinculación de la Variable Aleatoria y Estadística en la Realización de Inferencias Informales por parte de Futuros Profesores*

Coordination between Random Variable and Statistical Variable in the Execution of Informal Inferences by Pre-serviceTeachers

Blanca Ruiz**
Carmen Batanero***
Pedro Arteaga****

Resumen

Las inferencias sobre una variable aleatoria se realizan a partir del análisis de la distribución de la variable estadística asociada, cuyos estadísticos proporcionan información sobre los correspondientes parámetros de la distribución de la variable aleatoria. En un doble proceso de modelación, la variable estadística modela los datos observados y la variable aleatoria el comportamiento de la variable estadística en la población de donde se extrajo la muestra. En este artículo analizamos el trabajo de 101 futuros profesores de educación primaria con estos dos modelos en una tarea abierta. La mayoría de participantes construyen la distribución de la variable estadística, calculando correctamente las medidas de tendencia central y dispersión. Observamos problemas, sin embargo, en la realización de inferencias informales sobre la variable aleatoria por una parte importante de los

* Agradecimientos: Este trabajo forma parte de la beca FPU AP2007-03222 (MEC-Feder), proyecto EDU2010-14947 (MICIN) y Grupo PAI FQM126. Junta de Andalucía.
** Maestría en Ciencias/Cicata (IPN), México. Profesora del Departamento de Matemáticas, Instituto Tecnológico y de Estudios Superiores de Monterrey. Dirección del correo postal: Campus Monterrey. Garza Sada 2501, 64849 Monterrey, México. E-mail: bruiz@itesm.mx
*** Doctorado en Matemáticas/Universidad de Granada, España. Catedrática de Universidad del Departamento de Didáctica de la Matemática. Facultad de Educación. Dirección del correo postal: Campus de Cartuja. 18071, Granada, España, E-mail: batanero@ugr.es
**** Maestría en Educación Matemática/Universidad de Granada, España. Becario predoctoral del Departamento de Didáctica de la Matemática. Facultad de Educación. Dirección del correo postal: Campus de Cartuja. 18071, Granada, España. E-mail: parteaga@ugr.es

futuros profesores, quienes en pocos casos completan el ciclo de modelación interpretando los resultados matemáticos en el contexto del problema.

Palabras-clave: Variable Aleatoria. Variable Estadística. Distribuciones. Proyectos. Formación de Profesores.

Abstract

Inference on random variables is carried out by analyzing the distribution of the associated statistical variable, whose statistics summaries provide information on the random variable distribution. In a double modeling process, the statistical variable models the observed data and the random variable the statistics variable behavior in the population from which the sample was taken. In this paper, we analyze the work with these two models in an open-ended task by 101 pre-service primary school teachers. Most participants build the statistics variable distribution and correctly compute the central position and spread statistics. However, we observed problems in making informal inferences on the random variable by a significant proportion of pre-service teachers, few of whom completed the modeling cycle and interpreted the mathematical results in the problem context.

Keywords: Random Variable. Statistical Variable. Distributions. Projects. Teacher Education.

1 Introducción

Una de las tareas fundamentales en el análisis de datos es la realización de inferencias, lo que implica el uso coordinado de las variables estadística (distribución de datos) y aleatoria (distribución de probabilidad). Mientras que la variable estadística es un primer modelo matemático que representa los datos obtenidos en una muestra, la variable aleatoria supone un segundo nivel de modelización, al imaginar que la toma de datos se extiende al total de la población de donde se tomó la muestra. La importancia de la variable aleatoria fue resaltada por Heitele (1975), quien la incluyó en su lista de diez ideas fundamentales en la enseñanza de la estadística. Miller (1998) señala su relevancia en posteriores nociones estadísticas, como las distribuciones de probabilidad, el modelo de regresión o la obtención de estimadores. La comprensión de la relación entre la distribución de frecuencias y la distribución de probabilidad permite la realización de inferencias que, finalmente, han de interpretarse en el contexto donde se tomaron los datos. Esta comprensión y coordinación del carácter dual de la distribución comprende las ideas de aleatoriedad, independencia, tendencia, valor esperado y variabilidad (SHAUGHNESSY, 2007). Todos estos conceptos se incluyen en las nuevas orientaciones curriculares para la Educación Primaria en España (ESPAÑA, 2006) y, por tanto, será necesario asegurar la comprensión adecuada de los mismos por parte de los futuros profesores que han de enseñar este tema.

Varios autores sugieren la necesidad de desarrollar en los estudiantes capacidad de realizar inferencias informales, para preparar un estudio posterior más formalizado del tema en la Universidad. Dicho razonamiento sobre inferencia informal incluye la discriminación entre posición central y variabilidad en las distribuciones de datos (RUBIN; HAMMERMAN; KONOLD, 2006). Rossman (2008) sugiere usar la simulación para conectar la aleatoriedad que los estudiantes perciben en el proceso de recogida de datos con la inferencia. Él propone los siguientes pasos: a) Comenzar con una hipótesis sobre los datos; b) usar la simulación para concluir que los datos observados son poco plausibles si la hipótesis es cierta; y c) rechazar la hipótesis inicial basándose en los resultados. Otros elementos de inferencia informal incluyen razonar sobre las características de una población con base en los datos y razonar acerca de posibles diferencias entre dos poblaciones a partir de las diferencias observadas en dos muestras de datos de la misma (ZIEFFLER et al., 2008).

En este trabajo evaluamos la vinculación y comprensión de estos modelos por parte de una muestra de estudiantes que cursan la carrera de profesor de Educación Primaria en España. La evaluación se realiza a partir de sus producciones escritas en una tarea abierta de análisis de datos en la que deben realizar inferencias sobre las diferencias entre dos variables aleatorias a partir de la comparación de las correspondientes variables estadísticas. En primera instancia, presentamos algunos antecedentes de nuestro trabajo.

2 Investigaciones previas

El primer antecedente de investigación sobre la variable aleatoria la tenemos en el trabajo de Piaget e Inhelder (1951), quienes plantean a niños de diferentes edades experimentos para conocer su concepción de las distribuciones normales y uniforme. Concluyen que, en la adolescencia, la comprensión progresiva de la aleatoriedad conduce a la percepción de la distribución de una serie de resultados en un experimento aleatorio (primeras intuiciones sobre la distribución de una variable aleatoria).

La mayoría de trabajos posteriores se centran en la comprensión que los estudiantes muestran de la idea de distribución de una variable estadística, utilizando tareas de comparación de dos distribuciones de datos. Por ejemplo, Konold et al. (1997) estudian a un grupo de estudiantes universitarios cuando comparan dos distribuciones de datos, encontrando que algunos, ni siquiera usan intuitivamente la media, además, se centran en las frecuencias absolutas y no en las relativas, incluso cuando las muestras eran de tamaños muy diferentes. Algunos autores sugieren que la mayor parte de los estudiantes no ven el dato como un valor de una variable aleatoria, y, en consecuencia, no perciben los resúmenes estadísticos (como media o rango) como propiedades de la distribución (BAKKER; GRAVEMEIJER, 2004).

Por su parte, Shaughnessy y Ciancetta (2002) analizan el trabajo de los estudiantes cuando les piden que escriban una distribución con un valor dado de la media o mediana. Observan que estos se limitan a repetir valores muy similares o equidistantes de la medida de posición central. Otro requisito para comprender la distribución es la idea de variabilidad, que está siempre presente en los datos y tiene múltiples significados (READING; SHAUGHNESSY, 2004), entre otros los siguientes: variabilidad de resultados posibles en un experimento aleatorio; variabilidad en los datos recogidos; variabilidad en una variable aleatoria; variabilidad en las muestras o la distribución muestral.

Batanero, Estepa y Godino (1997) averiguan la forma en que los estudiantes comparan dos distribuciones. Como estrategias correctas encuentran que los estudiantes comparan las medias o bien, reducen el conjunto de datos a una sola variable, restando los valores correspondientes, en el caso de muestras relacionadas, y luego comparan la media con cero. También encuentran estrategias incorrectas como comparar valores aislados en las dos distribuciones.

En una serie de trabajos, Watson y colaboradores (WATSON; MORITZ, 1998; WATSON, 2001) clasifican las estrategias que usan los estudiantes cuando comparan distribuciones. Utilizan el modelo jerárquico SOLO para definir niveles de comprensión. En el primer nivel de la jerarquía, los estudiantes son capaces de comparar conjuntos de igual tamaño, mientras que en el segundo se comparan conjuntos de datos de diferente tamaño. Las estrategias de los estudiantes incluyen razonamiento proporcional y comparación de las gráficas y de las medidas de posición central de los dos grupos.

Respecto a profesores en formación, Canada (2008) confronta sus razonamientos con los de estudiantes de secundaria cuando comparan distribuciones de datos con la misma media y diferente dispersión. Aunque el grupo de futuros profesores tuvo mejor rendimiento que el de los estudiantes, el 35% de los mismos consideraron que dos conjuntos de datos con la misma media eran iguales, aunque la dispersión fuese muy diferente.

La idea de distribución de una variable aleatoria también ha sido investigada, aunque no con tanta intensidad. Truran (1994) observa la comprensión de los estudiantes de la magnitud de las desviaciones de las frecuencias relativas obtenidas en una serie de experimentos con respecto a su probabilidad teórica. Antes de realizar los experimentos, pregunta a los estudiantes los resultados que esperan en muestras pequeñas de una población; al obtener valores extremos, los estudiantes se sorprendieron ya que casi todos esperaron una distribución de frecuencias prácticamente igual que la de la probabilidad en la población. La conclusión del autor es que los estudiantes no comprenden la relación entre frecuencia y probabilidad ni la variabilidad de una variable aleatoria.

En los estudios que hemos descrito se estudia la distribución de una o varias variables estadísticas o aleatorias independientemente. Kazak y Confrey (2007) indican que la idea de distribución se puede ver de dos maneras: a) una visión estadística como agregado de un conjunto de datos (distribución de la variable estadística); y b) una visión probabilística como conjunto de posibles resultados de un experimento aleatorio en el que aparece una variable involucrada (distribución de la variable aleatoria). Utilizando diversos tipos de tarea como análisis de distribuciones espaciales de objetos reales (como flores o rebaños de animales), replicaciones de los experimentos de Piaget y otros en que aparece la distribución Binomial, concluyen que la comprensión intuitiva de la idea de distribución estadística puede evolucionar a un modelo más formal de distribución de una variable aleatoria, sobre todo si se realizan tareas de simulación.

En lo que sigue, tratamos de complementar las anteriores investigaciones centrándonos en futuros profesores y utilizando una tarea abierta. Analizamos también su competencia para completar el ciclo de modelación, aspecto no tenido en cuenta en las investigaciones citadas. Analizamos, en primer lugar, la tarea propuesta. Después, presentamos nuestros resultados sobre formación de la distribución de la variable estadística, cálculo e interpretación de las medidas de tendencia central y dispersión de la misma y realización de inferencias sobre la variable aleatoria.

3 Material y método

Los datos se tomaron a partir de los informes escritos en una tarea abierta diseñada por Godino et al. (2008) como actividad de clase y extra-clase para futuros profesores. El objetivo de la actividad es que los mismos estudiantes (futuros profesores) analicen sus producciones para que decidan si la clase tienen buenas intuiciones sobre la aleatoriedad. La actividad gira alrededor de la comparación del número de caras obtenidas en dos grupos de secuencias de 20 lanzamientos de una moneda equilibrada: en el primer grupo de ellas, *secuencias simuladas*, los participantes inventan los resultados de la secuencia sin lanzar realmente la moneda y el segundo grupo, *secuencias reales,* se obtiene al anotar los resultados lanzando 20 veces la moneda. Cada participante realiza el experimento de inventar una secuencia simulada y de lanzar la moneda para generar una secuencia real. Al finalizar los experimentos se elabora el listado de datos del grupo completo (ver 0) y el formador de profesores pide a los participantes que comparen las distribuciones del número de caras en las secuencias real y simulada. También pidió que analizasen los datos individualmente, y produjeran un informe escrito en el que justificasen si la clase en su conjunto tenía o no buenas intuiciones sobre la aleatoriedad, con base en el análisis de los datos. Los estudiantes tuvieron libertad de elegir los objetos, gráficos o resúmenes, estadísticos que considerasen convenientes o bien de usar ordenadores.

Tabla 1 - Ejemplo de la tabla obtenida por los *m* estudiantes participantes.

Estudiante	1	2	3	4	…	m
Número de caras en Secuencia Simulada	10	12	11	10	…	11
Número de caras en Secuencia Real	11	11	11	8	…	8

En esta actividad aparecen dos variables aleatorias y sus correspondientes variables estadísticas:

- La variable aleatoria η_r: es el número de caras en 20 lanzamientos de la moneda equilibrada. Es una variable aleatoria Binomial con parámetros *n=20 p=q=1/2*. Su media, sería igual a *np=10* y su varianza igual a *npq=5*. Se manifiesta en la distribución teórica que deben seguir las secuencias de los lanzamientos de las monedas.
- La variable aleatoria η_s: es el número de caras en la secuencia de longitud 20 inventada por los estudiantes. La usamos como modelo matemático para reflejar las *intuiciones colectivas* de los estudiantes sobre los experimentos aleatorios. No sería una variable Binomial, pues la investigación didáctica (entre ellos, SERRANO, 1996) muestra que los ensayos inventados por las personas en este tipo de experimentos no son independientes, aunque la probabilidad de cara sigue siendo $p = 1/2$.

Puesto que cada estudiante realiza el experimento y cuenta el número de caras obtenidas, tanto en la secuencia real como en la simulada, tenemos una serie de *m* ensayos de cada una de las variables aleatorias anteriores, siendo *m* el número de estudiantes, y por lo tanto obtenemos muestras de dos variables estadísticas:

- La variable estadística Y_r son los resultados de una muestra de *m* valores de la variable aleatoria η_r. El estudiante usará esta variable estadística para sacar algunas conclusiones sobre la variable aleatoria $\eta_{r,}$ puesto que sus conocimientos matemáticos no le permiten trabajar directamente con η_r.
- La variable estadística Y_s o resultados de una muestra de *m* valores de la variable aleatoria η_s. El estudiante ha de usar esta variable estadística para sacar algunas conclusiones sobre la variable aleatoria $\eta_{s,}$ puesto que, a priori, no se conoce la distribución de η_s para este grupo de estudiantes.

El estudiante ha de comparar las características de las distribuciones de las variables Y_r y Y_s para realizar una inferencia sobre las semejanzas o diferencias entre η_r y η_s. Se espera que los futuros profesores construyan un gráfico de las distribuciones o calculen y comparen algunos de sus estadísticos (por ejemplo, los mostrados en la 0). También, de este análisis, podrán deducir que el grupo de estudiantes tiene buenas intuiciones sobre las medidas de tendencia central, que tendrán que ser aproximadamente iguales en las dos distribuciones de datos, pero no respecto a la dispersión, que esperamos que sea mayor en la secuencia real.

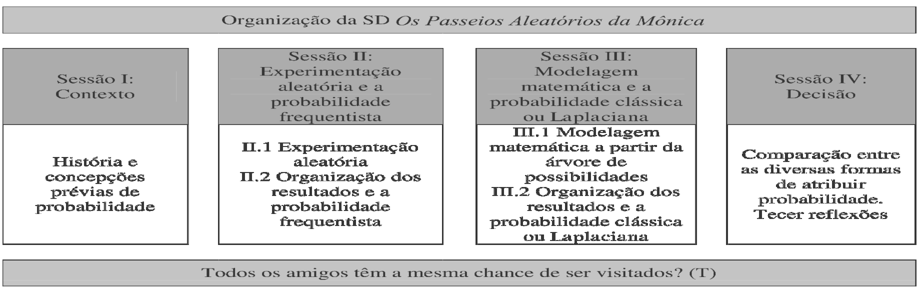


Figura 1 - Distribución y estadísticos del número de caras en las secuencias reales y simuladas

4 Resultados y discusión

Una vez que se tuvo el trabajo de los estudiantes (informes por escrito), se llevó a cabo un análisis cualitativo para deducir categorías en las respuestas. A continuación presentamos los principales resultados sobre la representación de las distribuciones de datos, y cálculo de estadísticos de la misma, así como sobre las inferencias que hacen sobre la variable aleatoria y la forma en que completan el ciclo de modelación.

4.1 Representación de la Distribución de la Variable Estadística

Para poder obtener conclusiones sobre las intuiciones del grupo, es necesario formar y representar las distribuciones de las variables estadísticas Y_r y Y_s. De este modo se inicia un primer ciclo de modelación, se pasa de la realidad (experimentos realizados por cada alumno) al modelo matemático de la variable estadística y al trabajo con el mismo. Ello implica el uso implícito de una configuración de objetos matemáticos (GODINO; BATANERO; FONT, 2007) que varían, dependiendo de la representación producida de la distribución de datos. De acuerdo con Arteaga (2008), dichas representaciones pueden clasificarse en cuatro niveles de complejidad:

N1. Sólo se representa los resultados de su experimento individual. En este nivel están aquellos estudiantes que se limitan al análisis de sus propios datos. No habiendo comprendido el propósito del proyecto o no siendo capaces de realizar un análisis global de los datos, tratan de resolver la pregunta para su caso particular (si él mismo tiene una buena intuición). Consideran que la variable estadística está vinculada con el experimento *lanzamiento de una moneda*, y por lo tanto, los valores que tomaría serán: *Cara* o *Cruz*, sin tomar en cuenta los resultados de sus compañeros. Los 20 lanzamientos son analizados como repeticiones del experimento aleatorio.

N2. Representan todos los datos sin llegar a la distribución de frecuencias. Son los estudiantes que no agrupan los valores similares del número de caras obtenidos en las secuencias reales o simuladas del total de alumnos en la clase para formar la distribución de frecuencias. En lugar de ello, en el gráfico representan el valor (o valores) del número de caras en las secuencias de 20 ensayos obtenidos para cada estudiante en el orden en que los datos fueron obtenidos, que es un orden arbitrario. Aunque llegan al modelo de variable estadística, la asocian con el número del estudiante (una etiqueta) que obtuvo ese dato no con la frecuencia con la que se presentó. Es decir, no calculan las frecuencias asociadas a cada valor.

N3. Alumnos que forman la distribución separada de cada variable. Son quienes forman tablas de frecuencias para las variables Y_r y Y_s y a partir de ellas producen los distintos gráficos o representan directamente un gráfico de cada uno de los valores diferentes de la variable con sus frecuencias. Por tanto pasan del conjunto de datos al modelo de variable estadística y de esta a su distribución de frecuencias. El representar por separado las dos distribuciones dificulta la comparación, sobre todo en caso de no usar la misma escala de representación en ambos gráficos.

N4. Representación conjunta de las dos distribuciones. El alumno llega a formar las distribuciones de las variables y sus distribuciones y las representa conjuntamente en el mismo gráfico, lo cual facilitará su comparación.

El resultado del análisis del Nivel de complejidad en las producciones de los estudiantes, se presenta en la 0. En ella se puede observar que de un total de 101 participantes, 88 (87,1%) representan sus datos a través de gráficos, lo que indica la necesidad sentida de los futuros profesores de producir un gráfico y llegar, mediante un proceso de transnumeración a un conocimiento no disponible en los datos brutos (WILD; PFANNKUCH, 1999). La mayoría de las representaciones de los datos (45,54%) son de distribuciones separadas de las dos variables (nivel 3), el 24,75% de los estudiantes representan conjuntamente las dos distribuciones. Son pocos los que analizan sólo sus propios datos (nivel 1) y sólo 14,85% estudian los valores obtenidos por cada futuro profesor, caso a caso, sin llegar a formar la distribución. En consecuencia, el modelo de variable estadística e incluso el de distribución es usado al menos a un nivel intuitivo por los participantes para resolver la tarea propuesta.

Tabla 2 - Nivel de complejidad en la representación de los datos

Nivel de complejidad	Frecuencia	Porcentaje (%)
N1. Representa sólo sus datos	2	1.98
N2. Representa dato a dato	15	14.85
N3. Representa dos distribuciones separadas	46	45.54
N4. Representación conjunta de dos distribuciones	25	24.75
No representa los datos	13	12.87
Total	101	100

4.2 Cálculo de resúmenes de la variable estadística

Heitele (1975) señala tres conceptos fundamentales en relación a la variable aleatoria: su distribución, esperanza y variabilidad. En consecuencia, estos tres componentes serán también fundamentales en la variable estadística. A continuación analizamos las medidas de posición central y dispersión calculadas por los participantes.

4.2.1 Medidas de posición central

De un total de 101 estudiantes, 91 calcularon alguna medida de tendencia central. Esto indica una primera aproximación a la idea de distribución en los estudiantes, incluso en aquellos que no representaron gráficamente la distribución, pues pasaron del dato aislado a un resumen estadístico del conjunto de datos. En la 0 vemos que los estadísticos calculados con preferencia son la media (90) y la moda (80); la mediana fue calculada por 72 participantes. En general el cálculo de los medidas de posición central es correcto, lo que indica un buen dominio de los algoritmos de cálculo, aunque aparecen algunos errores: no ordenar los datos en la mediana (17), falta de ponderación o error de cálculo en media (11) y no tener en cuenta el caso de bimodalidad para la moda (3). Estos errores coinciden con los ya señalados en estudiantes de secundaria en investigaciones previas como la de Cobo (2003).

4.2.2 Dispersión

El uso de medidas de dispersión es menor que el de medidas de tendencia central; la mayoría de los que lo usan calculan el rango (60), 54 de ellos correctamente. Algunos estudiantes toman incorrectamente los mismos valores del máximo mínimo y rango en la secuencia real y simulada. Unos pocos calculan además la desviación típica (25), varianza (15) y coeficiente de variación (1). Es evidente que las medidas de dispersión son menos intuitivas para los estudiantes, por el menor uso que se hace de las mismas, como también se confirma en la investigación de Silva y Coutinho (2008).

Tabla 3 - Estadísticos calculados por los participantes (m=101)

	Correcto	Incorrecto	Total
Media	79	11	90
Mediana	55	17	72
Moda	77	3	80
Rango	54	6	60
Otra medida de dispersión	41	–	41

4.3 Interpretación de la distribución de la variable estadística

El proceso de modelación requiere no sólo la construcción y el trabajo con el modelo matemático, sino también la interpretación de sus resultados en relación a la situación modelada (HENRY, 1997). En la 0 clasificamos la submuestra de futuros profesores que llega a construir la distribución (n=88), según el tipo interpretación que hacen de la misma.

Tabla 4 - Clasificación de estudiantes, según interpretación de la distribución

Tipo de interpretación	Frecuencia	Porcentaje
Correcta	29	33
Parcialmente correcta	37	42
Incorrecta o no interpreta	22	25
Total	88	100

Una tercera parte de ellos hace una interpretación correcta y completa de la distribución de la variable estadística, detectando las tendencias y variabilidad de los datos. Un ejemplo se muestra a continuación:

> *Centrándome en el gráfico observamos que los valores de las caras en la secuencia simulada se concentran en torno a 10 y 11, mientras que en la real, los valores están más disgregados, aunque también vemos que se concentran en un núcleo 8-15, fluctuando al alza y a la baja. De esta forma en la secuencia real comprobamos que 10 caras se obtiene el mayor número de veces, tanto en la simulada (20) como en la real (14).... la horquilla de la secuencia real es mucho más amplia que la de la simulada* (Alumno FL).

Un 42% de los estudiantes hacen una interpretación parcialmente correcta de la distribución, analizando tan sólo la tendencia sin tener en cuenta la variabilidad o bien al contrario, comparando sólo los rangos de variación, sin tener en cuenta las tendencias. El siguiente ejemplo lo consideramos parcialmente correcto porque no compara la posición central.

> *En cuanto al número de caras reales que ha obtenido el grupo es mucho más variable, son más homogéneos que en la simulación. Para poder realizar esta afirmación hemos usado el rango* (Alumno JI).

El resto de los futuros profesores (25%) hacen una interpretación incorrecta, o se limitan a presentar la distribución sin comentar nada sobre ella. En el siguiente ejemplo, la interpretación errónea se debe a errores conceptuales, respecto a la dependencia funcional lineal:

> *Todos los gráficos presentan una dependencia funcional lineal de – a + o viceversa, según las circunstancias* (Alumno PF).

En el siguiente ejemplo, el futuro profesor hace un análisis de los valores de las frecuencias creyendo que analiza la variable de interés y saca sus conclusiones sobre esa variable. Representa los valores de la frecuencia en el eje X. Este error ya fue descrito en Ruiz (2006). Este estudiante describe:

> *En la secuencia real las frecuencias son más consecutivas, no se repite la misma cara tantas veces y se repiten menos veces y distintos números, por ello las barras están más juntas* (Alumno CB).

Esto es, en las secuencias reales el número de caras toma más valores diferentes (puesto que hay más variabilidad), las frecuencias de cada valor son menores y también hay más frecuencias diferentes (*se repiten menos veces y distintos números*). Al representar los valores que aparecen en secuencia real, pero no en la simulada, (como el 13, que tiene frecuencia nula en la secuencia simulada), él concluye que el valor mínimo para su variable de interés en la secuencia simulada es cero. A su vez, en esa misma secuencia el valor de 10 caras lo obtuvieron 14 estudiantes (valor máximo de su variable) así que los valores de las frecuencias en este gráfico varían más que en la secuencia real. El alumno, por lo tanto, expresa que en un gráfico *las frecuencias son más consecutivas*. La interpretación tiene varios errores, además de los señalados, pues confunde *consecutivo* con *distante* o *separado*. Por otro lado, compara las frecuencias aisladas (*se repite menos veces y distintos números*) en lugar de prestar atención al valor del número de caras. Es claro que su variable de interés es la frecuencia.

4.4 Realización de Inferencias sobre la Variable Aleatoria y Segundo Ciclo de Modelización

Una vez que observaron las diferencias entre las dos variables estadísticas Y_r y Y_s, se espera que los participantes realicen un segundo proceso de modelación, generalizando las diferencias observadas en las dos variables estadísticas a diferencias en las variables aleatorias asociadas, realizando una inferencia informal sobre las semejanzas o diferencias entre η_r y $\eta_{s.}$ Finalmente esperamos que interpreten en la realidad los resultados de este segundo modelo, obteniendo conclusiones sobre las intuiciones del grupo respecto a la aleatoriedad.

Tabla 5 - Conclusiones obtenidas por los participantes sobre las intuiciones

Conclusión	Frecuencia
Correcta	4
Parcialmente correcta	24
Incorrecta o no concluye	73
Total	101

Los resultados sobre las conclusiones obtenidas se presentan en la 0, donde observamos que la realización de dicha inferencia para obtención de la conclusión es la tarea más difícil para los participantes, ya que sólo una tercera parte obtiene una conclusión, al menos parcial. Solo cuatro futuros profesores completan el proceso de inferencia y el ciclo de modelización llegando a la conclusión de que, por un lado, el grupo tiene buena intuición respecto al promedio de número de caras y por otro las intuiciones sobre la variabilidad de la variable aleatoria número de caras es pobre en los estudiantes. A continuación reproducimos una de estas conclusiones.

> *En cuanto al número de caras, las intuiciones del aula fueron aproximadas a la realidad, pero no del todo, ya que la desviación típica nos indica que los datos se distancian.* (Alumna CG).

Anteriormente, la alumna había escrito: *La media entre la simulada y la real se asemejan,.... la mediana y la moda dan los mismos datos*.

Veinticuatro participantes llegan a una conclusión parcial, debido a que sólo comparan las medidas de posición central sin tener en cuenta la dispersión, o al contrario.

> *La intuición de mis compañeros observando la tabla del nº de caras es buena, ya que los valores más repetidos en la secuencia simulada coinciden con los valores de la secuencia real: 10 y 11 son las más repetidas. La media de*

> *las dos secuencias es 10, por lo tanto creo que la intuición es buena* (Alumno TG).

También hemos considerado conclusión parcial cuando aparece imprecisión de lenguaje, como en el siguiente ejemplo, donde el alumno visualiza tanto la diferencia en la dispersión como al similitud en los valores centrales, pero no llega a concluir claramente sobre la intuición de los estudiantes.

> *Respecto a las gráficas construidas sobre el número de caras he de comentar que se parece observar algunos cambios en la secuencia simulada donde encontramos cuatro posibilidades y en la real hay diez posibilidades. Con esto podríamos decir que los alumnos no han sido intuitivos, pero no creo que esto sea así, ya que si nos detenemos atentamente en las gráficas podemos ver que los valores de 9, 10, 11 y 12 en ambas han sido dados por un mayor número de alumnos que en los demás casos* (Alumno HC).

El siguiente alumno reconoce la diferencia de dispersión en las distribuciones, pero interpreta el problema como la búsqueda de diferencia de intuiciones entre los estudiantes. Es decir, concluye que las intuiciones son similares en distintos estudiantes, pero no llega a relacionar estas intuiciones con las características del fenómeno aleatorio. Tampoco hace observaciones sobre las medidas de posición central.

> *Realmente las intuiciones de los alumnos han sido más o menos muy parecidas. No hay muchas irregularidades. Pero cuando las comparamos con sus secuencias reales, se puede ver a simple vista en el gráfico que existen grandes irregularidades. En la secuencia real el número máximo de caras es 16 y el mínimo 7, sin embargo en la simulada el máximo es 13 y el mínimo 8; su recorrido es más pequeño que el real* (Alumno MM).

El resto no llega a una conclusión, o bien hace una conclusión incorrecta. Los estudiantes no siempre conectan los resultados del trabajo matemático con la situación problemática, es decir, no ven las implicaciones de lo obtenido en el análisis estadístico sobre las intuiciones de los estudiantes.

> *Comparando los datos me he dado cuenta que son muchos los resultados entre los compañeros los que coinciden, pero aún así, sigo pensando que es mera casualidad, porque en la simulada hemos puesto lo que hemos querido (Alumno EL).*

Otros estudiantes, aún cuando conectan el trabajo con el modelo matemático (distribución) con el problema real, fallan en la obtención de conclusiones debido a que no han comprendido la pregunta planteada y suponen que una buena intuición ha de corresponder a obtener los mismos resultados en las secuencias real y aleatoria. En la siguiente respuesta, el estudiante muestra una concepción correcta del azar (no se puede prever) y otra incorrecta al tratar de evaluar el número de coincidencias o la diferencia de valores obtenidos en cada estudiante en los experimentos, en lugar de comparar directamente las distribuciones de las variables. Este estudiante compara caso a caso no utilizando la distribución al hacer la comparación.

> *Partiendo del estudio de los gráficos, se podría decir que en valores absolutos, la previsión del grupo no ha sido demasiado desafortunada. Un juego de azar es imposible de prever con total exactitud, pero las aproximaciones sumadas a los aciertos son mayores a las previsiones muy alejadas del resultado real* (Alumno LG).

En otros falta la capacidad de análisis para detectar las diferencias, lo que les lleva a concluir que los resultados en las dos secuencias son similares. En el siguiente ejemplo, además de darse este caso, no se llega a relacionar el resultado con la intuición de los estudiantes.

> *Como conclusión, podemos ver que los resultados son prácticamente los mismos tanto en la real como en la simulada, de lo que podemos deducir que lo real y lo simulado es muy parecido ya que lo que te inventes puede ser prácticamente igual a lo real (Alumno AC).*

5 Discusión y Conclusiones

La mayoría de los futuros profesores muestra una comprensión de la variable estadística y llega a formar y representar su distribución en una proporción aceptable. Un alto porcentaje ha calculado algún estadístico, principalmente medidas de posición central (media, mediana y moda) y en menor medida de dispersión. El cálculo de estadísticos es, en general, correcto, con pocas excepciones, aunque también aparecen algunos de los errores descritos en la investigación de Cobo (2003) y Mayén (2006), pero con mucha menor frecuencia que en aquellos estudios. Por el contrario, son pocos los futuros profesores que son capaces de generalizar los resultados obtenidos con la variable estadística para realizar una inferencia informal sobre la variable aleatoria. El razonamiento sobre la variabilidad, que es otro de los componentes esenciales del razonamiento estadístico, es difícil para la mayoría.

Muchas tareas cotidianas del profesor en la clase de matemáticas, tales como "indagar lo que los estudiantes conocen, elegir y manejar representaciones de las ideas matemáticas, seleccionar y modificar los libros de texto, decidir entre modos posibles de acción" (BALL; LUBIENSKI; MEWBORN, 2001, p. 453) dependen de su razonamiento y conocimientos matemáticos. En consecuencia, sería necesario atender a estos problemas en la formación de los profesores de educación primaria, pues una mejora de la educación de los niños exige una buena formación del profesor.

Para resolver el problema planteado, además de coordinar las ideas de variable estadística y aleatoria, los estudiantes han de interpretar los resultados del trabajo matemático realizado con ellas en el contexto del problema (traducir estos resultados a lo que indican respecto de las intuiciones de los estudiantes). Es precisamente este último paso (puesta en relación del resultado con la pregunta planteada) el que ha causado más dificultad, por la falta de familiaridad de los futuros profesores con actividades de modelación. Hay, aquí, otro punto en que es necesario incrementar la competencia de los profesores, si queremos que la educación estadística de los alumnos llegue a ser una realidad.

Referencias

ARTEAGA, P. **Análisis de gráficos estadísticos elaborados en un proyecto de análisis de datos**. 2008. 100 f. Disertación (Maestría en Didáctica de las Matemáticas) – Departamento de Didáctica de la Matemática, Universidad de Granada, Granada. 2008.

BALL, D.; LUBIENSKI, S.; MEWBORN, D. Research on teaching mathematics: the unsolved problem of teachers' mathematical knowledge. In: RICHARDSON, V. (Org.). **Handbook of research on teaching**. Washington, DC: American Educational Research Association, 2001. p. 433-456.

BAKKER, A.; GRAVEMEIJER, K. Learning to reason about distribution. In: GARFIELD, J.; BEN ZVI, D. (Orgs.). **The Challenge of Developing Statistical Literacy, Reasoning and Thinking**. Dordrecht: Kluwer, 2004. p. 147-168.

BATANERO, C.; ESTEPA, A.; GODINO, J. Evolution of Students' Understanding of Statistical Association in a Computer-Based Teaching Environment. In: GARFIELD, J.; BURRILL, G. (Orgs.). **Research on the role of technology in teaching and learning statistics.** Voorburg: International Statistical Institute, 1997. p. 183-198.

CANADA, Daniel. Conceptions of distribution held by middle school students and preservice teachers. In: JOINT ICMI/IASE STUDY: TEACHING STATISTICS IN SCHOOL MATHEMATICS. CHALLENGES FOR TEACHING AND TEACHER EDUCATION. ICMI STUDY 18 AND 2008 IASE ROUND TABLE CONFERENCE, 18th y 4th, 2008, Monterrey. **Proceedings…** Monterrey: ICMI e IASE, 2008. CD-Rom.

COBO, B. **Significado de las medidas de posición central para los estudiantes de secundaria**. 2003. 303 f. Tesis (Doctorado en Didáctica de las Matemáticas) – Departamento de Didáctica de la Matemática, Universidad de Granada, Granada, 2003.

ESPAÑA. Ministerio de Educación del Gobierno de España-MEC. Real Decreto 1513/ 2006, de 7 de diciembre, **por el que se establecen las enseñanzas mínimas correspondientes a la Educación primaria**. Madrid: MEC, 2006. Disponible en: <http://www.boe.es/boe/dias/2006/12/08/pdfs/A43053-43102.pdf>. Acceso en: 24 feb. 2010.

GODINO, J.; BATANERO, C.; FONT, V. The onto-semiotic approach to research in mathematics education. **ZDM. The International Journal on Mathematics Education,** Berlín, v. 39, n.1-2, p. 127-135, mar. 2007.

GODINO, J.; BATANERO, C.; ROA, R.; WILHELMI, M. Assessing and developing pedagogical content and statistical knowledge of primary school teachers through project work. In: JOINT ICMI/IASE STUDY: TEACHING STATISTICS IN SCHOOL MATHEMATICS. CHALLENGES FOR TEACHING AND TEACHER EDUCATION. ICMI STUDY 18 AND 2008 IASE ROUND TABLE CONFERENCE, 18th y 4th, 2008, Monterrey. **Proceedings…** Monterrey: ICMI e IASE, 2008. CD-Rom.

HEITELE, D. An epistemological view on fundamental stochastic ideas. **Educational Studies in Mathematics**, Netherlands, v. 6, n. 2, p. 187-205, jul. 1975.

HENRY, M. Notion de modéle et modélization en l'enseignement. In: COMISSION INTER-IREM STATISTIQUE ET PROBABILITÉ. **Enseigner les probabilités au lycée**. Francia: Inter-IREM, 1997. p. 77-84.

KAZAK, S.; CONFREY, J. Elementary school students' intuitive conceptions of random distribution. **International Electronic Journal of Mathematics Education,** Turquía, v. 2, n. 3, p. 227-244, oct. 2007. Disponible en: <http://www.iejme.com>. Acceso en: 24 feb. 2010.

KONOLD, C.; POLLATSEK, A.; WELL, A.; GAGNON, A. Students analyzing data: Research of critical barriers. In: GARFIELD, J.; BURRILL, G. (Orgs.). **Research on the role of technology in teaching and learning statistics.** Voorburg, The Netherlands: International Statistical Institute, 1997. p. 151-167.

MAYEN, S. **Comprensión de medidas de posición central en estudiantes mexicanos de Bachillerato**. Sept. 2006. 100 f. Trabajo de Investigación Tutelada (Certificado) – Departamento de Didáctica de las matemáticas, Universidad de Granada, Granada, 2006.

MILLER, T. The random variable concept in introductory statistics. In: INTERNATIONAL CONFERENCE ON TEACHING STATISTICS-ICOTS, 5th, 1998, Singapore. **Proceedings…** Singapore: International Association for Statistical Education, 1998, p. 1221-1222. Disponible en: <http://www.stat.auckland.ac.nz/~iase/publications/2/Topic9h.pdf>. Acceso en: 24 feb. 2010.

PIAGET, J.; INHELDER, B. **La genèse de l'idée de hasard chez l'enfant**. Paris: Presses Universities de France, 1951. 232 p.

READING, C.; SHAUGHNESSY, M. Reasoning about variation. In: GARFIELD, J.; BEN ZVI, D. (Orgs.). **The Challenge of Developing Statistical Literacy, Reasoning and Thinking**. Dordrecht: Kluwer, 2004. p. 201-226.

ROSSMAN, A. Reasoning about Informal Statistical Inference: One Statistician's View. **Statistics Education Research Journal**, Netherlands, v. 7, n. 2, p. 40-58, nov. 2008.

RUBIN, A.; HAMMERMAN, J.; KONOLD, C. Exploring informal inference with interactive visualization software. In: INTERNATIONAL CONFERENCE ON TEACHING STATISTICS-ICOTS, 7th, 2006, Salvador, Brasil. **Proceedings…** Salvador: International Association for Statistical Education., 2006. Disponible en: <http://www.stat.auckland.ac.nz/~iase/publications/17/2D3_RUBI.pdf>. Acceso en: 24 feb. 2010.

RUIZ, B. **Un acercamiento cognitivo y epistemológico a la didáctica del concepto de variable aleatoria**. Diciembre, 2006. 209 f. Disertación (Maestría en Ciencias en Matemática Educativa) – Centro de Investigación en Ciencia Aplicada y tecnología Avanzada, Instituto Politécnico Nacional, México, 2006.

SERRANO, L. **Significados institucionales y personales de conceptos matemáticos ligados a la aproximación frecuencial de la enseñanza de la probabilidad**. 1996. 223 f. Tesis (Doctorado en Didáctica de las Matemáticas) – Departamento de Didáctica de la Matemática, Universidad de Granada, Granada, 1996.

SHAUGHNESSY, M. Research on statistics learning and reasoning. In: LESTER, F. (Org.). **Second Handbook of Research on Mathematics Teaching and Learning.** Greenwich, CT: Information Age Publishing, Inc., and NCTM, 2007. p. 957-1009.

SHAUGHNESSY, M.; CIANCETTA, M. Students' understanding of variability in a probability environment. In: INTERNATIONAL CONFERENCE ON TEACHING STATISTICS-ICOTS, 6th, 2002, Cape Town, South Africa. **Proceedings…** Cape Town: International Association for Statistical Education, 2002. Disponible en: <http://www.stat.auckland.ac.nz/~iase/publications/1/6a6_shau.pdf>. Acceso en: 24 feb. 2010.

SILVA, C. B.; COUTINHO, C. Q. S. Reasoning about variation of a univariate distribution: a study with secondary mathematics teachers. In: JOINT ICMI/IASE STUDY: TEACHING STATISTICS IN SCHOOL MATHEMATICS. CHALLENGES FOR TEACHING AND TEACHER EDUCATION. ICMI STUDY 18 AND 2008 IASE ROUND TABLE CONFERENCE, 18th y 4th, 2008, Monterrey. **Proceedings…** Monterrey: ICMI e IASE, 2008. CD-Rom.

TRURAN, J. Children's intuitive understanding of variance. In: INTERNATIONAL CONFERENCE ON TEACHING STATISTICS-ICOTS, 4th, 1994, Marrakech, Morocco. **Proceedings…** Marrakech: International Association for Statistical Education, 1994.

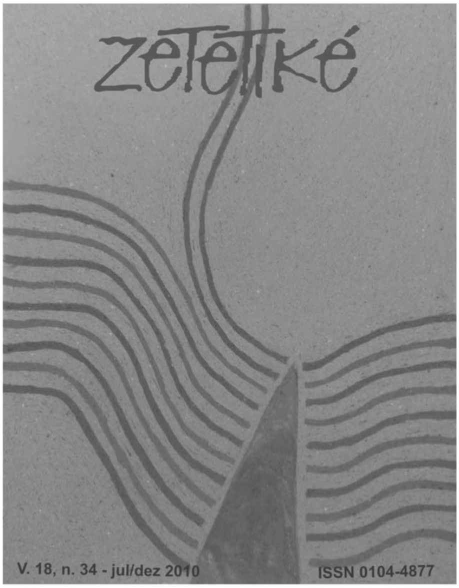
WATSON, J. Longitudinal development of inferential reasoning by school students. **Educational Studies in Mathematics**, Netherlands, v. 47, n. 3, p. 337-372, sept. 2001.

WATSON, J; MORITZ, J. The beginning of statistical inference: Comparing two data sets. **Educational Studies in Mathematics**, Netherlands, v. 37, n. 2, p. 145-168, nov. 1998.

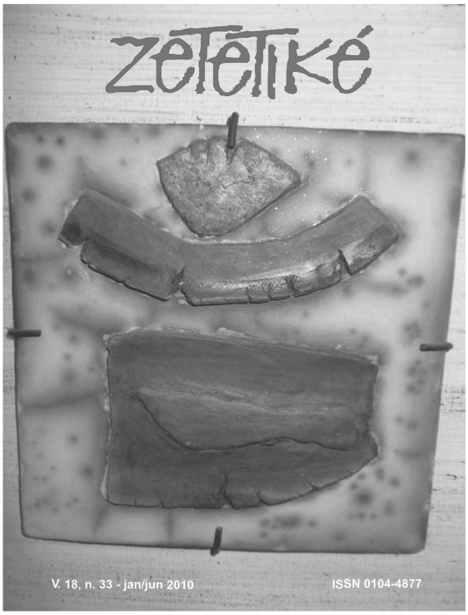
WILD, C.; PFANNKUCH, M. Statistical thinking in empirical enquiry (con discusión). **International Statistical Review**, Netherlands, v. 67, n. 3, p. 223-265, dic. 1999.

ZIEFFLER, A.; GARFIELD, J.; DELMAS, R.; READING, C. A framework to support research on informal inferential reasoning. **Statistics Education Research Journal**, Netherlands, v. 7, n. 2, p. 40-58, nov. 2008.

Submetido em Março de 2010.
Aprovado em Julho de 2010.


zetetiké
V. 18, n. 34 - jul/dez 2010
ISSN 0104-4877


zetetiké
Número Temático 2010
Linguagem e práticas culturais:
perspectivas para a Educação Matemática
ISSN 2176-1744


zetetiké
V. 18, n. 33 - jan/jun 2010
ISSN 0104-4877