

## DIDÁCTICA DE LA ESTADÍSTICA: DISTRIBUCIONES DE MUESTREO

Tomás Sánchez y Armando Albert

INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE MONTERREY

tsanchez@itesm.mx, albert@itesm.mx

**Resumen.** *La necesidad social de un aprendizaje escolar de probabilidad y estadística más adecuado a los usos y prácticas de las distintas áreas profesionales ha llevado a mirar más de cerca la naturaleza del problema. Una de los conceptos fundamentales de la probabilidad y estadística es el concepto de distribución de muestreo. Existen varios estudios sobre los errores y dificultades frecuentes de los estudiantes al respecto, así como significados asociados a ellas. Sin embargo, hacen falta aún estudios encausados a conocer más de cerca los procesos cognitivos que se dan en su aprendizaje que sirvan de fundamento para las nuevas propuestas didácticas. Este estudio se inicia en este camino.*

**Palabras Clave:** distribuciones de muestreo, didáctica, estadística.

### Introducción

La Probabilidad y Estadística ocupan actualmente en la actividad científica y de ingeniería un lugar cada vez más importante. Es por eso que el aprendizaje escolar de estas ciencias constituye una gran preocupación de estadísticos, investigadores y educadores. Una de las ideas fundamentales de la Probabilidad y Estadística son las distribuciones de muestreo. Ellas son, además, la base de la estadística inferencial. Sin embargo, no son pocas las investigaciones que manifiestan las graves dificultades que se presentan para su aprendizaje en la escuela (Vallecillos, 1996; Chance, delMas y Garfield, 2004; Inzunza, 2006).

En un curso de probabilidad y estadística, las distribuciones de muestreo se estudian después de haber abordado las distribuciones para una variable aleatoria discreta, como la binomial, poisson, hipergeométrica, binomial negativa, entre otras y las

distribuciones para variable continua como exponencial, gamma, normal, entre otros y suelen ser el antecedente inmediato y esencial al estudio de la estadística inferencial Vallecillos (1995), pues cualquier procedimiento inferencial, nos dice ella, implica conocer la distribución muestral de algún estadístico.

Esta investigación aborda distribución muestral del estadístico  $\bar{X}$  y sus propiedades. Muchos son los elementos que intervienen para formar este concepto; como variable aleatoria. Por ejemplo, el de la distribución normal como modelo teórico al que converge la distribución de los promedios de las muestras seleccionadas bajo las condiciones establecidas por el Teorema Central del Límite, y características esenciales como su valor esperado y la variabilidad de los promedios. De la teoría se sabe que la media de la distribución de los promedios de las muestras coincide con la media de la población original sin importar el tamaño de la muestra, no así la variabilidad que se ve afectada por el número de elementos en las muestras seleccionadas y con ello la tendencia de la distribución. El tamaño de muestras que componen una distribución muestral, finalmente, será de suma importancia en la estimación y pruebas de hipótesis; temas centrales de la inferencia estadística.

Distintas investigaciones dan cuenta de las dificultades para los estudiantes de aprender distribuciones de muestreo debido, no sólo a su complejidad en sí mismas, sino, como reportan Chance, delMas y Garfield (2004), porque los estudiantes suelen tener un entendimiento incompleto de conceptos relacionados con la distribución y la variabilidad. Vallecillos y Batanero (1997) muestran las dificultades que tienen estudiantes universitarios para reconocer el promedio de las muestras como una variable aleatoria. Lo que es más, en Ruiz, Albert y Batanero (2006) se reporta la complejidad y dificultades asociadas para su aprendizaje de conceptos tan básicos como variable aleatoria. Otras investigaciones como Well, Pollatsek y Boyce (1990) dicen que las personas entienden que las medias de grandes muestras son más probables de parecerse a la media poblacional, pero no entienden las implicaciones sobre de la variabilidad de la media. Lipson (2000) sostiene que los estudiantes

necesitan valorar que la distribución de todas las muestras debe ser conocida o modelada para hacer una inferencia. Parte del problema reside, quizás, en, como dice el mismo Lipson, que la idea de distribuciones muestrales ha sido introducida en los cursos de estadística a través de un enfoque deductivo basado en la teoría de la probabilidad. Inzunsa (2006) aborda con los estudiantes las distribuciones muestrales apoyado de recursos computacionales y simulación. Al parecer eso les demandó un mayor esfuerzo cognitivo, pero les favoreció el desarrollo de ciertas ideas como el hecho de entender cuáles son los valores más probables de ciertos resultados muestrales y alrededor de que valor se encuentran, así como entender que cuando la variabilidad muestral rebasa ciertos límites, puede deberse a causas no atribuibles al azar. Pero también señala Inzunsa algunas dificultades tales como el establecimiento de expresiones matemáticas necesarias para calcular el estadístico muestral así como la presencia de cierta confusión de la distribución de una muestra con distribución muestral y tendencia a ver a las muestras en forma aislada, en lugar de verlas en forma de distribuciones.

Aunque en esta investigación se reconoce la complejidad sistémica e importancia de abordar el fenómeno didáctico desde sus distintas componentes: epistemológica, cognitiva, de enseñanza y de sus prácticas sociales, sólo se abordarán elementos de la componente cognitiva. Particularmente, la presente investigación hace un acercamiento cualitativo a algunos de los resultados hallados por Inzunsa (2006) con la intención de descubrir más elementos del proceso cognitivo que se da cuando los estudiantes responden a sus cuestionamientos relativos a algunas características de las distribuciones muestrales en contexto gráfico.

## **Metodología**

Este estudio se realizó con estudiantes de ingeniería de una universidad del Norte de México y constó de dos etapas: En la primera de ellas se aplicó la encuesta realizada por Inzunsa (2006) a un grupo de 22 estudiantes de Probabilidad y Estadística de

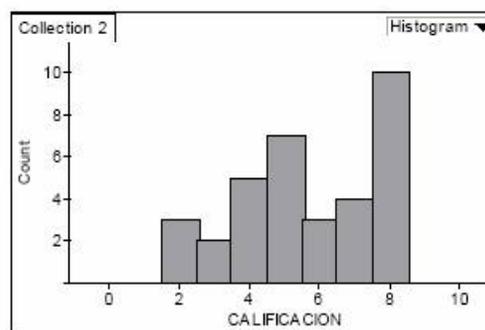
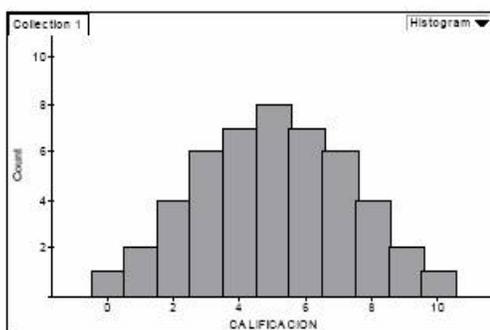
ingeniería. En la segunda etapa se escogieron 5 preguntas en contexto gráfico del mismo cuestionario relativas a variación, tamaño de la muestra y formas de la distribución y con ellas se hizo una entrevista exploratoria a dos estudiantes de ingeniería que ya habían cursado Probabilidad y Estadística para que las resolvieran en equipo y en voz alta. El entrevistador sólo intervino para esclarecer algunos de los procesos de solución de los estudiantes y se verbalizaran. Se videograbó la entrevista y se les pidió también sus escritos sobre los impresos de las actividades.

## Resultados y Discusión

El cuestionario aplicado en la primera etapa se detectó algunas concepciones erróneas acerca de la variabilidad de las distribuciones. En la pregunta 1 del cuestionario 77.27 % dijeron que la opción 2 es la distribución con más variabilidad.

### Pregunta 1

1. Marca con una X la distribución de datos que tiene más variabilidad.

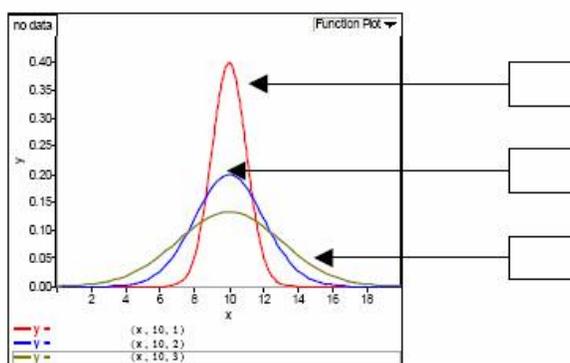


En la entrevista exploratoria (la etapa 2 del estudio) se halló que los dos estudiantes entrevistados coincidieron en decir que la segunda opción era la más variable, ellos argumentaron de manera similar a argumentos encontrados en el cuestionario

abierto, que los saltos en las columnas de la segunda gráfica eran mas evidentes que en la primera. Al cuestionarles, de acuerdo al protocolo diseñado, sobre la posibilidad de que exista dos variabilidades, una en la horizontal (variabilidad de la variable aleatoria) y la “variabilidad” en la vertical (la cantidad o “variabilidad “de las frecuencias) ellos aceptaron que, en ese sentido, la grafica uno seria la más variable, pero que la primera sería menos variable con respecto a la segunda ya que los saltos de las columnas eran mas evidentes. Se comprobó que la pregunta podía dar cabida a varias respuestas según el eje que observaban. También parece haber un efecto de lo estudiado en otros cursos de matemáticas o de otros cursos de su currículum, ya que es común que el estudio de la variación esté centrado en la ordenada y en el caso de los entrevistados, sus respuestas reiterativas de más o menos variación parecen sugerirlo.

### Pregunta 2

2. En las siguiente grafica aparecen tres distribuciones poblacionales cuya media es  $\mu = 10$  y sus desviaciones estándar son  $\sigma = 1$ ,  $\sigma = 2$  y  $\sigma = 3$  respectivamente. Coloca sobre cada una de ellas la desviación estándar que le corresponde.



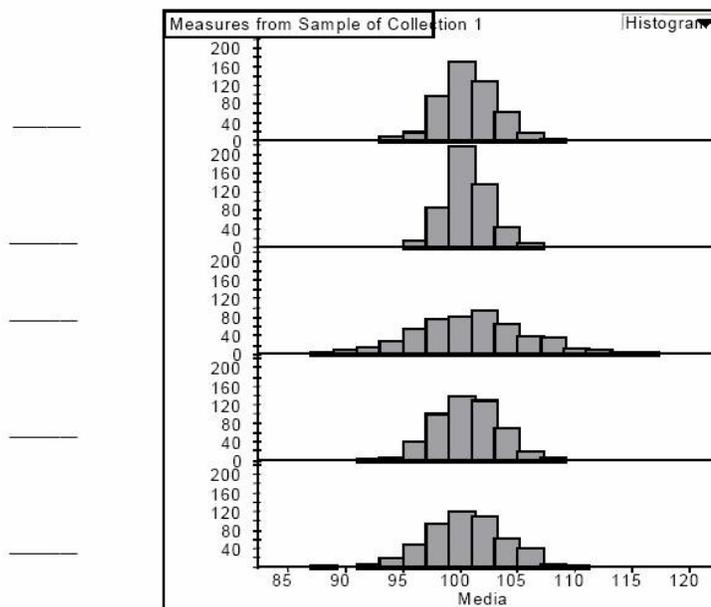
En este caso, el porcentaje de respuestas con secuencia 1, 2, 3 (de arriba abajo) consideradas correctas fue del 77.27 % (17 estudiantes de los 22 entrevistados en la etapa I) sólo el 18.18 % (4 de los 22) mantuvieron la congruencia con relación a la pregunta anterior. Esto es, la grafica con la cresta más alta tiene más variabilidad. Así,

la secuencia de asignación fue 3, 2, 1. Es de notarse que ante la presencia de lo continuo, uno de los estudiantes, en la entrevista exploratoria comentó: “sí pues son curvas de campana, entre más pequeña la desviación estándar el pico tiende a ser más alto y de eso estoy seguro. Es de acuerdo a lo que vimos en clase”. En este punto es importante mencionar que tanto en la etapa I como en la etapa II se encontró casos en los que se argumenta apoyándose en la ordenada de la graficas, por ejemplo, los estudiantes comentaban: “Entre más chica es la desviación más grande va hacer la grafica” o “Cuando la varianza es mayor hace que la gráfica crezca hacia arriba ya que la media sigue siendo de 10 en los 3 casos, por lo tanto la gráfica de la varianza = 3 tiene que crecer más que los demás “

### Pregunta 3

De una población con distribución normal se extrajeron 500 muestras aleatorias de cada tamaño (5, 10, 15, 20 y 25). Se calculó la media de cada muestra y los resultados se dibujaron en los histogramas que se muestran en la siguiente figura.

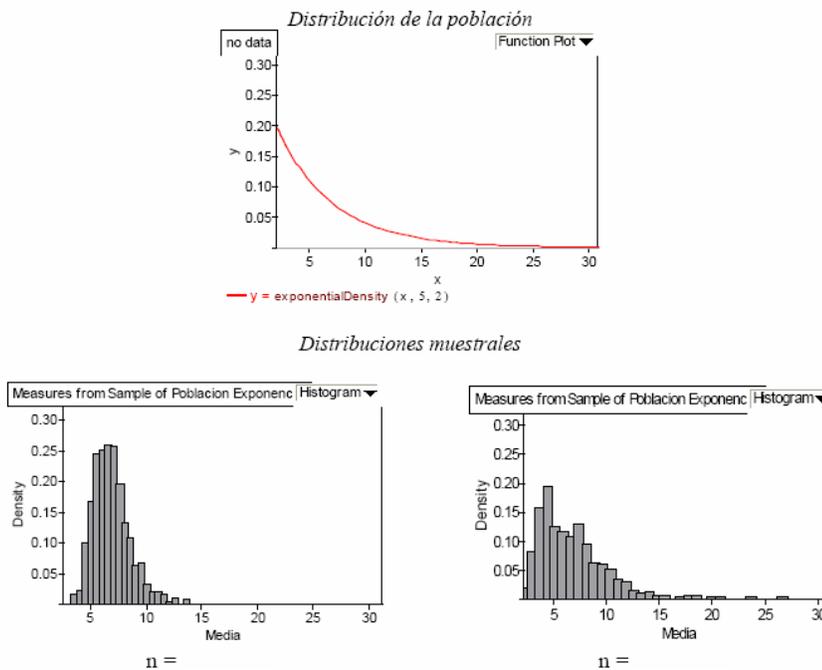
- a) Coloca a un lado de cada histograma, el tamaño de la muestra que corresponda.



Con relación a la pregunta 3, destaca el hecho de la existencia de dos posiciones contrarias. Por un lado los que dicen que las muestras de mayor tamaño mostraran graficas con menos variabilidad (en sus explicaciones evocan a la distribución normal) con los promedios más concentrados en la media (se debería marcar más la media) y, por otro lado, los que dicen que muestras pequeñas no reflejan bien a la población o, de otra manera, a muestras mayores mas dispersión. Los que asignaron con una secuencia correcta solo fueron 6 de los 22 y 12 asignaron en secuencia equivocada pero congruente con la creencia de que a mayor muestra más variabilidad, de hecho un estudiante escribió: “Con un tamaño mayor de la muestra puedes ver mejor como se distribuye, hay mas datos, en un tamaño menor de muestra la gráfica es mas pobre con menos columnas”.

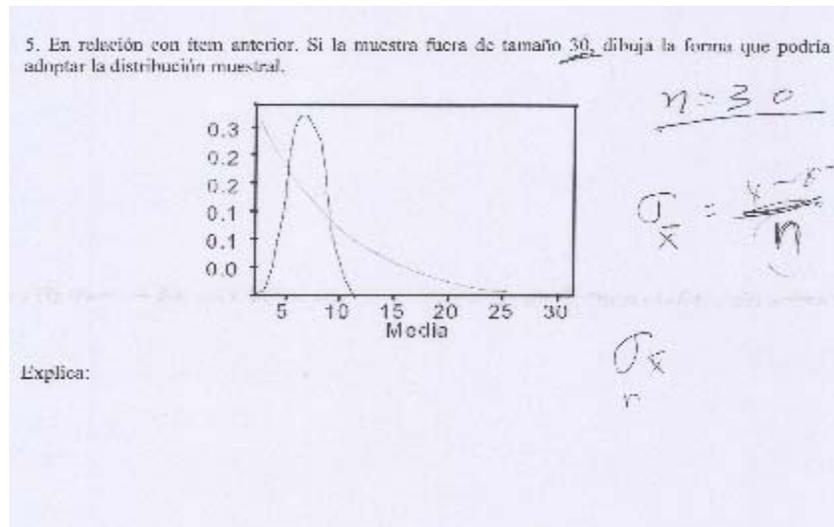
#### Pregunta 4

En la figura se muestra la distribución de una población y dos distribuciones muestrales para muestras de tamaño 2 y 15. Coloca el tamaño de muestra a la distribución que corresponda. Nota: se tomaron 500 muestras en cada caso.



En las respuestas a la pregunta 4 se presentan las posiciones contrarias ya comentadas en cuanto el efecto del tamaño de muestra. Sin embargo, al reducir el número de opciones e incluir la imagen visual de la población muestreada se destaca que el 50% del cuestionario abierto creen que a mayor número de muestra la grafica debe parecerse más a la poblacional (“A medida que el número de elementos en la muestra aumenta, ésta tiende acercarse o aproximarse a la gráfica de la distribución de la población”). Esta tendencia también se observó en la entrevista exploratoria, uno de los estudiantes insistía en que a mayor tamaño de muestra, los promedios deberían centrarse alrededor de la media (la media se debería marcar más) y el otro en que a mayor tamaño de muestra la grafica se debería parecerse más a la población. Esta postura se reafirmó al pedirles (pregunta 5) que dibujaran la forma que adopta la distribución muestral si el tamaño de la muestra fuera de 30. Los dibujos se muestran a continuación:

### Pregunta 5



Donde se puede apreciar el uso de la llamada *heurística de la representatividad* (Kahneman, D., y Tversky, A., 1972). Una de las conclusiones de estos investigadores es que la noción de que la varianza muestral decrece en proporción al tamaño de la muestra, aparentemente, no forma parte de la intuición del hombre.

## **Conclusiones**

Las distribuciones en el muestreo son un concepto fundamental para la probabilidad y estadística. Sin embargo, son muchos los conceptos antecedentes involucrados y cada uno de ellos tiene, a su vez, su complejidad y dificultades. Por otra parte, las distribuciones muestrales tienen por sí mismas sus propias dificultades. Una de ellas es la tendencia que tienen los estudiantes, por influjo de su formación en matemática determinista, a observar preponderantemente la variación de la ordenada en comparación a abscisa. Por otra parte, el uso de la intuición no resulta del todo exitoso. De hecho, es más frecuente que falle la intuición de los estudiantes acerca de la forma como se distribuyen las distribuciones muestrales dependiendo del tamaño de la muestra. Aunque la aplicación del cuestionario de Inzunza (2006) no hizo más que ratificar sus resultados, el acercamiento que se hizo en esta investigación con la entrevista exploratoria permitió identificar dos trayectorias cognitivas claramente distintas: una que se caracterizó por el uso de la heurística de la representatividad y que condujo, finalmente a errores. También se pudo identificar otra que consistió en tomar como criterio que a medida que el tamaño de la muestra crece, “la media de la distribución se debería remarcar más”. Esta estrategia, que le llevó finalmente a resolver con éxito las actividades planteadas, parece ser un elemento valioso a considerar en las innovaciones didácticas futuras dirigidas a mejorar el aprendizaje de los estudiantes relativo a las distribuciones muestrales.

## Referencias

- Chance, B. delMas, R. y Garfield, J. (2004). Reasoning about sampling distributions. *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*. Kluwer Academic Publishers.
- Inzunsa C. Santiago. (2006). *Significados que estudiantes universitarios atribuyen a las distribuciones muestrales en un ambiente de simulación computacional y estadística dinámica*. Tesis doctoral no publicada. CINVESTAV-IPN. México
- Kahneman, D., y Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3, 430-454
- Lipson, K. (2000). The role of the sampling distribution in developing understanding of statistical inference. Tesis doctoral no publicada. University of Technology of Swinburne, Australia
- Ruiz, B., Albert J. A. y Batanero, C. (2006). An exploratory study of student's difficulties with random variables. ICOTS-7
- Vallecillos, A. (1995). Sugerencias metodológicas para la introducción del teorema central del límite en la enseñanza secundaria. *Actas del I congreso nacional de bachillerato*. Universidad de Granada.
- Vallecillos, A. (1996). *Inferencia estadística y enseñanza: un análisis didáctico del contraste de hipótesis estadísticas*. Granada: COMARES.
- Vallecillos, A., y Batanero, C. (1997). Análisis del aprendizaje de conceptos clave en el contraste de hipótesis estadísticas mediante el estudio de casos. *Recherches en Didactique des Mathématiques*, 17(1), 29-48.
- Well, A. D., Pollatsek, A., y Boyce, S.J. (1990). Understanding the effects of sample size on the variability of the mean. *Journal of Organizational Behavior and Human Decision processes*, 47, 289-312