COMPRENSIÓN DE LA DISTRIBUCIÓN NORMAL EN BACHILLERATO

Yanet Karina González Arellano, Ana María Ojeda Salazar
Cinvestav. (México)
ygonzalez@cinvestav.mx, amojeda@cinvestav.mx

RESUMEN: Al terminar la enseñanza de la distribución normal, se planteó gráficamente un problema de tipificación a 26 estudiantes del sexto semestre de bachillerato. 80% del grupo no dio sentido a la variable aleatoria, su estandarización, la partición del espacio muestra de sus valores; emergieron conocimientos previos deficientes como producto cartesiano, área bajo la curva, función simétrica y notación de intervalo, así como no poder operar sin la presentación de las fórmulas respectivas. Un interrogatorio a dos estudiantes, una que contestó correctamente y otro incorrectamente, reveló su inadvertencia de la exclusión mutua de eventos determinada por intervalos disjuntos de los valores de la variable. En ambos casos no se reconoció el proceso de estandarización, ni contribuyó a comprender la gráfica mostrada en el planteamiento del problema. Estos precedentes imponen investigar resultados de la enseñanza correspondiente en nivel superior.

Palabras clave: distribución, normal, estandarización

ABSTRACT: After finishing teaching normal distribution, a classification problem was graphically posed to 26 students of the sixth semester of senior high school. An eighty per cent of the group did not understand the random variable neither its standardization, nor the space fragmentation that shows its values; deficient previous knowledge emerged, such as the Cartesian product, the area under the curve, the symmetric function and the interval notation, as well as the impossibility to work without the respective formulas. A questionnaire applied to two students (one of them answered incorrectly) showed their oversight of the mutual exclusion, determined by disjointed intervals of the variable values. Both students did not recognize the standardization process; it didn't help to understand the graph shown when posing the problem. Such precedents demand to investigate about the results of the correspondent teaching process at a higher level.

Key words: distribution, normal, standardization

Introducción

La enseñanza en el nivel superior de las distribuciones de probabilidad se centra en sus propiedades formales impuestas por su utilidad para determinar sus parámetros estadísticos, sin clarificar su significado en aplicaciones concretas (por ejemplo, Torres, 2013). Nuestra investigación, perfilada hacia la comprensión de estudiantes universitarios de las distribuciones de probabilidad, requiere identificar las características de su formación en estocásticos previa en el bachillerato. Así, examinamos las propuestas de dos programas de estudio de nivel medio superior (SEP, 2013; CCH-UNAM, 2003) para el tema de nuestro interés. También realizamos una indagación (Wittrock, 1986) concerniente al desempeño de estudiantes de bachillerato en la resolución de un problema relativo a la estandarización de la distribución normal, para caracterizar su comprensión de este tema.

Elementos teóricos

Con un enfoque epistemológico, Heitele (1975) ha propuesto diez ideas fundamentales de estocásticos como guía continua de un currículum en espiral, desde un plano intuitivo hasta uno formal: medida de probabilidad, espacio muestra, adición de probabilidades, regla del producto e independencia, equiprobabilidad y simetría, combinatoria, modelo de urna y simulación, variable estocástica, ley de los grandes números y muestra. La distribución normal implica las ideas de espacio muestra, medida de probabilidad, adición de probabilidades, equiprobabilidad y simetría, variable estocástica y muestra.

Steinbring (1991) ha propuesto el triángulo epistemológico para caracterizar la constitución del concepto matemático como una interrelación entre objeto, signo y concepto; lo cual indica que éstos no pueden ser tratados de forma independiente en la deducción del significado del conocimiento matemático.

Pollatsek, Lima y Well (1981) proponen tres tipos de conocimiento que deben considerarse para la comprensión de la media: conocimiento funcional, conocimiento de cálculo y conocimiento analógico. El conocimiento funcional de un concepto hace referencia a su comprensión como un concepto del mundo real significativo. El conocimiento computacional implica una fórmula de cálculo y el conocimiento analógico puede estar relacionado a imágenes visuales que permitan una interpretación razonable del resultado obtenido al aplicar el concepto a un problema.

Métodos e instrumentos

Planteamos a un grupo de 26 estudiantes (con edades de entre 17 a 19 años), del curso optativo de *Estadística y Probabilidad II* del sexto semestre de un bachillerato público, un problema relativo a la distribución normal de probabilidad, ya estudiada en el curso. Todos los estudiantes cursaban por primera vez la materia y sólo dos habían cursado y aprobado *Cálculo Diferencial e Integral I*, y en ese momento cursaban el correspondiente curso II, ambos propuestos también como optativos.

El problema, tomado del capítulo 16 de un solucionario para bachillerato (http://www.iessantvicent.com/web/images/departaments/ma tematiques/solucions/m1/1BAMA1_SO_ESB04U16.pdf), se les presentó impreso en hoja de papel e incluyó la gráfica con los datos de la media y las coordenadas x para los puntos de inflexión. También se les proporcionaron tablas para Z en hoja impresa.

Problema:

La longitud de cierto tipo de peces sigue una distribución normal con media $\mu = 100$ y varianza $\sigma^2 = 81$. ¿Cuál es la probabilidad de que uno de esos peces mida entre 82 y 91mm? (Ver Figura 1).

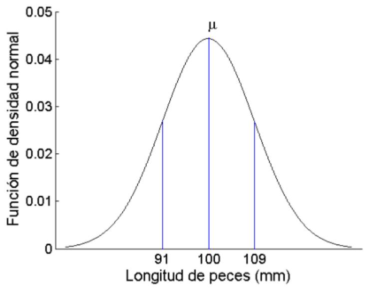


Figura 1. Distribución normal de la longitud de peces

Se pidió a los estudiantes que individualmente aplicaran lo estudiado: distribución normal estándar (identificada como la variable $Z = (X - \mu)/\sigma$), área bajo la curva normal, e identificar los valores de Z en las tablas. Cada estudiante dispuso del tiempo suficiente para dar su respuesta, pero se les dio la fórmula de la distribución normal estándar porque algunos argumentaron no recordarla y por eso no comenzaban a escribir.

La solución al problema requiere calcular $P(82 \leq X \leq 91)$, para la variable aleatoria continua X "longitud de un pez". Por lo tanto,

$$P(82 \leq X \leq 91) = \int_{82}^{91} f_X(x)dx = \int_{82}^{91} \frac{1}{9\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-100}{9}\right)^2} dx,$$

Con $f_X(x)$ como la función de densidad de la variable continua X y x como los valores que ella toma, lo cual relaciona la probabilidad con el área bajo la curva. El método de solución enseñado en el curso es el siguiente: de la Figura 1 se tiene que $\mu = 100$ y $\sigma = 9$. La medida de probabilidad en el problema de los peces se determina al estandarizar (tipificar) la variable aleatoria X. De aquí que, para $\mathrm{Z} = \frac{\mathrm{X}-100}{9}$ y mediante el uso de las tablas de la distribución normal estándar se tiene:

$$\begin{aligned} P(82 \leq X \leq 91) &= P\left(\frac{82-100}{9} \leq \frac{X-100}{9} \leq \frac{91-100}{9}\right) = P\left(\frac{-18}{9} \leq \frac{X-100}{9} \leq \frac{-9}{9}\right) \\ &= P\left(-2 \leq \frac{X-100}{9} \leq -1\right) = P(-2 \leq Z \leq -1) = P(Z \leq -1) - P(Z < -2) \\ &= 0.1587 - 0.0228 = 0.1359. \end{aligned}$$

A los datos recopilados se les aplicó la célula de análisis de la enseñanza (Ojeda, 2006). Según estos criterios, la caracterización del problema propuesto se resume en la Tabla 1.

Tabla 1. Criterios de análisis identificados en el problema planteado

Criterio	Solución al problema
Ideas fundamentales	
1) *Espacio muestra*	Conjunto Ω de todos los posibles resultados de la medición de longitudes de peces, que puede tomar distintos valores no negativos del intervalo $[0,\infty)$. El espacio muestra sería un subconjunto de $[0,\infty)$.
2) *Medida de probabilidad*	A cada evento (subconjunto de Ω), P asigna valores del intervalo real $[0,1]$.
3) *Adición de probabilidades*.	Dados dos subconjuntos $A, B \in [0,\infty)$, tal que $A \cap B = \emptyset$ se cumple que $P(A \cup B) = P(A) + P(B)$.
4) Equiprobabilidad y simetría	La función es simétrica respecto a $X = \mu$ por lo que respecto a este valor se tienen eventos equiprobables, por ejemplo, en $P(X \leq 91) = P(X \geq 109) = 1–P(X < 109)$.
5) Variable aleatoria X	La función X es la variable aleatoria continua $X: \Omega \to \mathbb{R}$, “longitud de peces”, que asocia a cada uno de los peces su medida de longitud.
6) Muestra	El enunciado supone una muestra de peces.
Otros concepto matemáticos	Plano cartesiano, orden en los números reales, intervalos reales, suma, resta, área, función simétrica y porcentaje.
Recursos semióticos	Gráfica, lengua natural escrita y simbología matemática, en particular, notación de intervalo real.
Términos para referirse a estocásticos	Distribución normal, media, varianza y probabilidad.
Referente	Probabilidad de que la longitud de peces, como variable aleatoria $X \sim N(\mu,\sigma)$, esté en un intervalo real dado $[a,b]$.

Luego de la revisión de las soluciones que dieron los estudiantes, seleccionamos a dos de ellos (*Pa* y *Lu*) para entrevistarlos individualmente en formato semiestructurado (Zazkis y Hazzan, 1999). *Pa* determinó de forma correcta la probabilidad solicitada en el problema y, de acuerdo al profesor titular del grupo, era su mejor estudiante, había aprobado *Cálculo Diferencial e Integral I* y ya se había inscrito al curso II. Al contrario, *Lu* dio una solución incorrecta al problema y no contaba con las bases del cálculo integral. Las entrevistas se efectuaron en el plantel respectivo seis semanas después de la aplicación del problema, con duración de 75 minutos aproximadamente cada una; y se les videograbó para su análisis.

Resultados del análisis

Planteamiento Institucional

En el bachillerato, la propuesta para la enseñanza de la distribución normal pretende que se identifique esta distribución como un "modelo continuo del comportamiento de una gran diversidad de fenómenos aleatorios de su entorno" (CCH-UNAM, 2003, p. 22) y subraya la estandarización de una variable aleatoria normal y su uso en la resolución de problemas.

Para la normalización, se espera que el estudiante comprenda "el significado de la estandarización de una variable aleatoria normal y las ventajas de efectuar este proceso" (loc. cit.), e identificar "el área bajo la curva normal estandarizada a partir de la distribución de probabilidad normal" (SEP, 2013, p. 22). Para lograr este objetivo se sugiere "explicar cómo se encuentra la curva normal estandarizada ubicando la media y la desviación estándar" (loc. cit.). El programa de estudios recomienda: "realizar en equipo el análisis de los ejemplos presentados y elaborar nuevos ejemplos donde se muestre el área bajo la curva normal estandarizada a partir de la tabla" (loc. cit.); y:

recordar al estudiante el papel que desempeñan las constantes a y c en la gráfica de una función de la forma $y = af(x - c)$ y asociarlo con μ y σ en la función de probabilidad Normal, mostrar al estudiante el uso de las propiedades geométricas de la Normal Estándar en la evaluación de probabilidades y en el cálculo de z. (CCH-UNAM, 2003, p. 22).

Algunas de las actividades para promover estos desempeños implican investigación documental, presentación en plenaria por parte de los estudiantes y el uso de las TIC's.

El problema. La Tabla 2 caracteriza las respuestas de acuerdo a la célula de análisis.

Tabla 2. Criterios de análisis identificados en la solución al problema planteado

Criterio	Problema
Espacio muestra	Un estudiante escribió 13.59, el porcentaje que asoció a la probabilidad, como un valor del eje longitud de peces.
Medida de probabilidad	Dos de los estudiantes dieron como respuesta un valor negativo.
Adición de	Los estudiantes no advirtieron que $P(Z \geq -2) - P(Z \geq -1) = P(-2 \leq Z \leq$

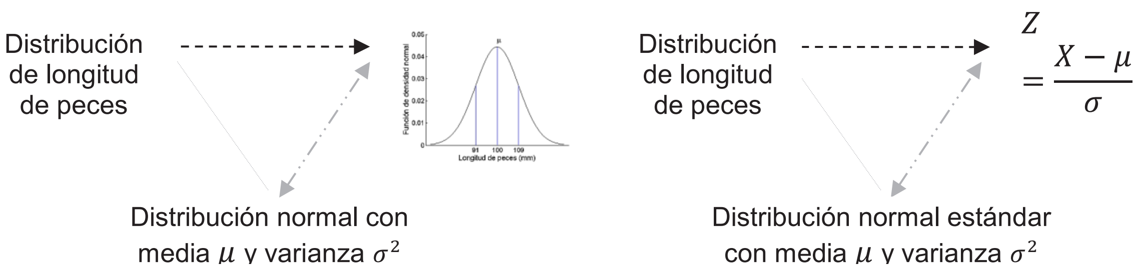
probabilidades	$-1)$, ni que $P(-2 \leq Z \leq -1) = P(Z \leq -1) - P(Z < -2)$.
Equiprobabilidad y simetría	Dos estudiantes determinaron al estandarizar: $P(Z \geq -2) - P(Z \geq -1) = 0.97725 - 0.84134 = 0.1359.$
Variable aleatoria X	Diecisiete estudiantes escribieron x (los valores que toma la variable X) como si fuera lo mismo que z o que X. Así, tres escribieron $P(x = 82)$ y $P(x = 91)$ y un estudiante escribió $P(x \geq 82)$ y $P(x \geq 91)$. Cinco estudiantes no incluyeron las variables en su solución, sino sólo sustituyeron los valores correspondientes.
6) Muestra	Aun al sombrear correctamente la región bajo la curva entre 82 y 91 no advirtieron que su respuesta, 0.8185 o 0.6826, distaba mucho de corresponder a la región que identificaron como solución al problema.
Otros conceptos matemáticos	Se manifestó insensibilidad al orden en los números. Por ejemplo, para referirse a la probabilidad de que uno de los peces midiera entre 82 y 91 mm, 23% (seis estudiantes) indicó que iban a determinar $P(82 \leq x \geq 91)$. El 12% (tres estudiantes) no se percató del orden de los números enteros al escribir que $91 \leq 82$ en $P(91 \leq x \geq 82)$. Dieciocho estudiantes concluyeron que la probabilidad de sacar un pez de entre 82 y 91 mm era el porcentaje asociado a su respuesta, aunque ésta no concordara con la región sombreada en su figura.
Recursos semióticos	Se identificó falta de conocimiento de la notación de intervalo como de la expresión algebraica para calcular la probabilidad de que un pez arbitrario midiera entre 82 y 91 mm. Ningún estudiante utilizó la información de la gráfica para solucionar el problema, sólo para sombrear la región que correspondía a la solución al problema, porque se les pidió lo hicieran.
Términos para referirse a estocásticos	Longitud, distribución normal, media, varianza, probabilidad, mida, entre 82 y 91mm.
Referente	Distribución normal de longitud de peces, con media μ y desviación σ.

En general, ocho estudiantes (31%) solucionaron correctamente el problema, 54% incorrectamente y 15% de ellos no lo resolvió. Mostraron dificultades para solucionarlo 14 de los 26 estudiantes. Al tipificar la variable (restar la media y dividir entre la desviación estándar los valores de *X* en el intervalo [82, 91]), los veintidós estudiantes sólo se limitaron a utilizar la fórmula proporcionada, $Z = \frac{X-\mu}{\sigma}$, ignorando el proceso implicado al estandarizar. Todos los estudiantes que resolvieron el problema (incorrecta o correctamente) realizaron por separado el cálculo de z_1 y z_2, buscaron en las tablas sus valores y finalmente realizaron una resta, aunque en dos casos invirtieron el orden y les dio negativo. Los estudiantes tendieron a seguir una secuencia de resultados en la misma línea, llegando a igualar números distintos. Por ejemplo, para el cálculo de z_1 obtuvieron -1 y a su derecha lo igualaron con

$P(z \leq -1)$, es decir, escribieron: $z_1 = \frac{91-100}{9} = -1 = 0.3413;$ aunque nadie distinguió z_1 de z_2 y solamente escribieron $z = \frac{91-100}{9}$ y $z = \frac{82-100}{9}$.

Triángulo epistemológico

Los resultados de este grupo de estudiantes indican que no se establecieron las interrelaciones en el triángulo epistemológico (Steinbring,1991), que corresponderían a la constitución del concepto de distribución normal (véase la Figura 3).



Entrevista

El guion de entrevista semiestructurada se basó en el análisis de las respuestas al problema planteado. A *Pa* y a *Lu* se les presentaron acetatos con las gráficas de la distribución normal, $N(100, 9^2)$ y $N(0,1)$, después de preguntarles: ¿Por qué estandarizan? ¿Es necesario estandarizar? ¿Qué significa estandarizar? ¿La gráfica de la distribución normal es igual a la de la normal estándar? También se les pidió que describieran gráficamente qué pasaba con la normal al estandarizar y que esbozaran su gráfica estandarizada. Ninguno logró explicar en qué consistía la estandarización y reconocieron operar sólo como una regla que les indicaba que debían evaluar la fórmula $Z = \frac{X-\mu}{\sigma}$ para 82 y 91 en este caso, buscar en las tablas para Z y restar:

I: ¿Gráficamente entiendes el proceso de estandarización? ¿Qué es lo que estás haciendo?

Pa: Mmmmm realmente… no, porque no vemos mucha teoría [ríe], entonces no le podría decir o sea, a mi igual, o sea, pónganme… esteee… se escucha como que un poco absurdo, perooo no vemos teoría. Vemos así como que un parrafito y ya, entonces a lo que nos vamos es a los problemas. Siento que por eso es que no puedo como que contestarle todo, porqueee… eh… a mi me pone un problema y le digo, va, lo hago porque ya me sé la fórmula, porque sé quee... qué tengo que sustituir con qué, más… sin embargo, no, tal vez no lo podría taaan… tan explicarlo. Por ejemplo, ahorita que me estaba diciendo de la… de la gráfica normal. Entonces, digoo, no [ríe]. Creo que no la podría hacer.

I: Si, ya. Mmm o sea lo que tú haces al estandarizar solamente se reduce a… ¿a aplicar esta fórmula? ¿Y buscar en las tablas y hasta ahí?

Pa: Y buscar en las tablas y hasta ahí. Ajá, es lo único.

El interrogatorio también dejó claro que la gráfica nada les sugirió para resolver el problema, sólo se le utilizó porque se les indicó sombrear la región que correspondía a la probabilidad solicitada. Sin embargo, ambos estudiantes reconocieron que pudieron haber resuelto el problema sin ella y que, por el contrario, de no proporcionarles los datos de media y varianza (o desviación estándar) no habrían podido determinarlos, sabiendo que 91 y 109 correspondían a las coordenadas x de los puntos de inflexión.

Una pregunta de la entrevista consideraba una respuesta incorrecta de otro estudiante, quien sombreó la región pedida erróneamente y daba como respuesta a la probabilidad el 97%. *Pa* indicó que no correspondía ese valor con lo sombreado, pero no logró indicar los posibles valores que tomaba la función de probabilidad ni que 97% era incorrecto.

I: ¿Qué con respecto a lo que puso? Él dice, “la probabilidad de que un, uno de los peces mida entre ochenta y dos y noventa y cinco es de noventa y siete punto, noventa y siete... punto setenta y dos por ciento”. ¿Es correcto decir que la probabilidad es de noventa y siete punto setenta y dos por ciento? ¿Es correcto decir eso?

Pa: Mmmmm. Pues se supone que debe de ser de cero para, deee... ¿no? cero punto uno, para arribaaaa… ¿no? o sea tiene que ser arriba de cero.

I: ¿Y puede seerr… mmm cinco, por ejemplo? Decir, la probabilidad de que… este… de que hoy llueva.

Pa: ¿es del cinco por ciento? Pueees…se supone que sí.

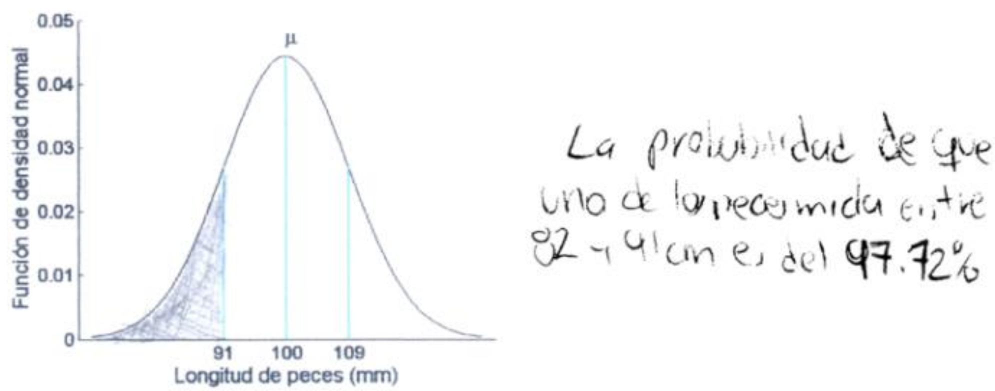


Figura 3. Ejemplo de respuesta de uno de los estudiantes

Tabla 3. Criterios de análisis identificados en las entrevistas

Criterio	**Entrevista**
Ideas fundamentales	
1) *Espacio muestra*	Ninguno de los dos estudiantes identificó los valores del eje de las ordenadas. *Lu*indicó que eran medidas de densidad de los peces.
2) *Medida de probabilidad*	Ninguno de los dos estudiantes entrevistados pudo indicar los valores que toma *P*, además de que expresaron su solución como un porcentaje.
3) *Adición de probabilidades*.	Reconocieron no saber en qué consistía la estandarización ni, por tanto, porqué efectuar $P(-2 \leq Z \leq -1) = P(Z \leq -1) - P(Z < -2)$.
4) Equiprobabilidad y simetría	Ninguno de los dos estudiantes describió la gráfica de la normal como simétrica. A *Lu* se le preguntó directamente si era simétrica y, aunque indicó que sí, su argumento fue incorrecto.
5) Variable aleatoria X	Ambos estudiantes reconocieron que los valores en el eje de las abscisas eran longitudes de peces, aunque ninguno dejó en claro que era una variable aleatoria. Incluso *Pa* preguntó: *¿una variable aleatoria?*
6) Muestra	Ambos estudiantes son conscientes de que la muestra utilizada en la gráfica corresponden a medidas de longitudes de peces.
Otros conceptos matemáticos	En la solución que *Lu* dio, la probabilidad de escoger un pez que midiera entre 82 y 91 mm era $P(x) = 3$; y en la entrevista, como $P(82 > x \leq 91)$. Ninguno de los dos estudiantes reconoció formalmente la relación entre área y probabilidad.
Recursos semióticos	Se desconoció la simbología matemática para representar la probabilidad de que la variable aleatoria *X* tome valores entre *a* y *b*. Ambos estudiantes reconocieron que la gráfica no les ayudó para resolver el problema. Ninguno pensó en una nueva gráfica al estandarizar. *Pa* presenta una falta de conocimiento por la notación y en la solución al problema lo plantea como $x = (82 \leq x \leq 91)$ y en la entrevista como $P(x82 \leq x91)$.
Términos para referirse a estocásticos	*Pa* señaló que no conocía la relación entre varianza ni desviación estándar. Ambos estudiantes utilizaron“media” como sinónimo de intermedio.
Referente	Distribución normal de longitud de peces, con media μ y varianza σ^2.

Comentarios

Las respuestas al problema planteado, así como respuestas a las entrevistas, revelaron que los estudiantes favorecen el pensamiento en el formato de frecuencia por sobre un formato de probabilidad, como lo señalan Gigerenzer y Hoffrage (1995). A pesar de obtener valores en decimales, todos ellos los expresaron en porcentajes y concluyeron así sus respuestas a la petición de la probabilidad y no con expresiones decimales entre cero y uno.

Los estudiantes manifestaron, a lo más, un conocimiento de cálculo (Pollatsek et al., 1981) de la distribución normal, pues interpretaron la resolución del problema planteado sólo como aplicación de un algoritmo aritmético para obtener un resultado, según ellos satisfactorio. No interrelacionaron el referente y la gráfica con el concepto de distribución normal; su respuesta del cálculo de la probabilidad pedida fue independiente de la región sombreada en su gráfica, por lo que no se puede deducir el conocimiento funcional ni analógico, en síntesis, matemático, enfoco. Un objetivo de los programas es que el estudiante “identifica el área bajo la curva normal estandarizada a partir de la distribución de probabilidad normal” (SEP, 2013, p. 22) o “comprende el significado de la estandarización de una variable aleatoria normal y las ventajas de efectuar este proceso” (CCH-UNAM, 2003). Pero del análisis de sus respuestas, estos objetivos no se cumplen.

Referencias bibliográficas

Colegio de Ciencias y Humanidades, UNAM (2003). *Programa de Estudios de Estadística y Probabilidad I y II*. Recuperado el 20/04/2014 de http://www.cch.unam.mx/sites /default/files/plan_estudio/mapa_estadistica.pdf.

Dirección General de Bachillerato, (2013). *Programa de Estudios. Probabilidad y Estadística II*. Recuperado el 19/04/2014 de http://www.dgb.sep.gob.mx/02-m1/03-iacademica/01programasdeestudio/cfp_6sem/PROBABILIDAD_ESTADISTICA_II.pdf

Gigerenzer, G. & Hoffrage, U. (1995). How to improve bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*, 684-704.

Heitele, D. (1975). An Epistemological View on Fundamental Stochastic Ideas. *EducationalStudies in Mathematics, 6* (2), 187-205.

(sf). *Solucionario de Matemáticas 1*, Bachillerato, Cap. 16, p. 84. Madrid: Ediciones SM.

Ojeda, A. M. (2006). Estrategia para un perfil nuevo de docencia: Un ensayo en la enseñanza de estocásticos. *Matemática educativa, treinta años.* México: Santillana-Cinvestav, pp. 195-214.

Pollatsek, A., Lima, S. & Well, A. (1981). Concept or Computation: Students’ Understanding of the Mean. *Educational Studies of Mathematics*, *12*(2), 191-204.

Steinbring, H. (1991). The Concept of Chance in Everyday Teaching: Aspects of a Social Epistemology of Mathematical Knowledge. *Educational Studies of Mathematics*, *22*, 503-522.

Torres, O. (2013). *Limitaciones en la adquisición de ideas fundamentales de estocásticos por estudiantes de ingeniería: El caso de un instituto tecnológico*. Tesis de maestría no publicada. México: Cinvestav.

Wittrock, M. (1986). *La investigación de la enseñanza II*. Barcelona: Paidós

Zazkis, R. & Hazzan, O. (1999). Interviewing in Mathematics Education Research: Choosing the Questions. *Journal of mathematical behavior*, 17(4), 429-439.