HISTOGRAMA: MUCHO MÁS QUE UNA REPRESENTACIÓN GRÁFICA

Roberto, Behar Gutiérrez
roberto.behar@correounivalle.edu.co
Escuela de Estadística, Universidad del Valle (Colombia)

Asunto: Formación de profesores
Temática: Estadística Descriptiva e Inferencial

RESUMEN

En los cursos generalmente introducimos el histograma como una forma más de representación gráfica, dejando pasar la oportunidad de convertirlo en un instrumento rico conceptualmente para comprender la idea central de la Estadística: La distribución de frecuencias y su relación con la función de densidad probabilística para las variables aleatorias continuas. En este taller, los participantes se enfrentarán al reto de construir un histograma con intervalos de clase de distinto ancho y podrán palpar el concepto de densidad empírica (o de frecuencia) y a partir de ella construir a mano la Función de Distribución acumulativa empírica, usando la intuición y el concepto de densidad y proporcionalidad y así, dar continuidad de manera informal con la idea de Función de Probabilidad Acumulada.

PALABRAS CLAVE

Histograma, Densidad empírica, Función de distribución acumulativa.

INTRODUCCIÓN

Tradicionalmente los profesores de Estadística de nivel medio y superior hemos mirado la Estadística Descriptiva como una temática divorciada de la probabilidad y de la Inferencia. Cuando llega el momento de explicar el histograma, generalmente se construyen intervalos de igual tamaño y el eje de las ordenadas representa directamente la frecuencia relativa. Sin embargo, cuando se trata la temática de las funciones de densidad en probabilidad, para calcular la probabilidad, en el marco de variables aleatorias continuas, que conceptualmente es el homólogo de la frecuencia relativa, cuando se mira como una extensión del concepto de la muestra a la población entera, debe calcularse un área, ya no son las ordenadas las que proporcionan esta información sobre la probabilidad, que es el límite de la frecuencia relativa.

La pregunta que surge es ¿Por qué si el concepto de probabilidad es una extensión a la población de la frecuencia relativa, en un caso se calcula un área y en el otro una altura? Esto parece conceptualmente incoherente. En este taller se plantea una estrategia para lograr coherencia entre el concepto muestral y el poblacional, definiendo el histograma como un gráfico de la densidad empírica. Esto tiene una doble función, por un lado ganar potencial intuitivo para dar sentido real a la idea de densidad, logrando que en el desarrollo del tema de probabilidad la definición de variable aleatoria continua no suene artificial para los estudiantes; y por otro resolver la mencionada incoherencia. Además, a partir del histograma, donde es el área y no altura la que representa la frecuencia, deja el eje Y automáticamente con el significado de densidad de frecuencia. Y a partir de allí, puede deducirse el concepto de distribución de frecuencia acumulada. En el espacio del taller se ilustra con un ejemplo la estrategia de cómo hacerlo.

MARCO DE REFERENCIA

La representación gráfica en la comunicación estadística es una pieza esencial para el proceso de transnumeración, el cual permite convertir una maraña de datos sobre un fenómeno, en representaciones simplificadas que revelan rasgos de los fenómenos ocultos en la complejidad de los números. Las características que se revelan a través de la trasformación que implica la transnumeración, potencializa la capacidad de dar respuesta a las preguntas de investigación, pero además facilita la comunicación de los resultados, al separar lo esencial de lo superfluo. La interpretación en contexto, la valoración crítica con base en referentes, el construir argumentos explicativos a partir de una buena representación gráfica, por ejemplo, es parte esencial de lo que se ha llamado Pensamiento Estadístico (Wild & Pfannkuch , 1999), y se constituye en factor determinante de la cultura estadística.

Sin embargo, hay muy serias evidencias empíricas en todos los niveles educativos, inclusive por parte de profesionales de todas las disciplinas, que revelan que hasta las representaciones más sencillas, tienen complejidades que hacen difícil encontrar significados. Estas afirmaciones se soportan en investigaciones como las de Batanero, Arteaga y Ruiz (2010), Kukliansky (2016), Lee y Meletiou (2003), Friel, Curcio y Brigh, (2001), Priti (1997), Wild & Pfannkuch (1999), Ruiz (2006) y Curcio (1987). Entre las principales deficiencias en la interpretación de gráficos, se encuentra la dificultad de leer más allá de los datos, nivel de interpretación de muy alto rango semiótico, en la escala de Curcio (1989). En particular en un gráfico de frecuencias de una variable estadística, interpretarlo como un diagrama de dispersión, en ocasiones sin poder asociar literalmente el gráfico con los datos. Interpretar que entre más alto es el histograma, hay mayor dispersión. Asociar mayor variabilidad con la irregularidad de las barras de un histograma. Identificar la idea de un histograma con la forma de campana de la normal. No diferenciación entre un histograma y un diagrama de barras. Uso de las frecuencias

(Eje *Y*) para determinar la media y el intervalo modal. Interpretar el eje *X*, con significado temporal y por lo tanto dar significado al gráfico como una serie o secuencia de tiempo. Este interesante tema, se presenta para dar contexto, pero su tratamiento no es el objetivo de este escrito, pues nuestra preocupación está específicamente en el histograma y su valor conceptual.

Histograma. ¿Qué es y cuál es su utilidad?

La palabra *Histograma* tan corriente hoy en día entre nosotros, para representar una variable estadística continua, fue acuñada por Karl Pearson, al parecer, en 1891. El origen etimológico de la palabra *Histograma,* tiene varias versiones, una de ellas muy difundida, es aquella que afirma que la palabra pretende sintetizar la expresión "*Historical Diagram*" (histo-gram) y daría la idea de un "cuadro histórico" de una distribución estadística (Schwartzman, 1994). Esta versión etimológica es compartida por Flood, Rice y Wilson, (2011) quien afirma que el "histograma [es] un diagrama similar a un diagrama de barras, pero que representa un conjunto de datos continuos, en lugar de discretos" (p. 12). Por esta razón, Pearson explicó que podría ser usado como una herramienta en el estudio de la historia, por ejemplo, un diagrama histórico en períodos de tiempo y acuñó el nombre de "histograma" para convenir su uso como un "*historical diagram".*

Riaño-Rufilanchas (2017) quien narra lo anterior, en su artículo "*On the Origin of Karl Pearson´s term "histogram*", discrepa profundamente de esta interpretación etimológica, pues a su juicio, Karl Pearson era un filólogo convertido en estadístico, por lo tanto muy riguroso y sistemático con la etimología. Como evidencia contundente para defender su punto de vista, destaca que durante sus conferencias de 1891, Pearson presentó los diez términos que describen los distintos tipos de gráficos, basándose en el elemento geométrico clave en la construcción del gráfico respectivo. Casi todos ellos, excepto el *euthigrama*, son inspiración de Levasseur (1885). Así por ejemplo se refirió a 1) *estigmogram* de "stigos" (punto); 2) *euthygramas* de "euthús" (recto) para designar el diagrama construido con líneas rectas, esta denominación fue sugerida por Pearson; 3) *Epipedogramas* de "epípedos" (planos), para aquellos diagramas donde la esencia de la representación son las áreas; 4) histogramas de "histós" (mástil) haciendo referencia a la representación por columnas; 5) Diagramas (curvas); 6) Radiogramas de "Radius" (rayo o radio) representación de diagramas de rayos (diagramas polares); 7) hormogramas de "hormós" (cadena, collar), representación por líneas rectas unidas, como sucede con el llamado "polígono de frecuencias"; 8) Cartogramas de "charta" (mapa); 9) Topogramas de "tópos" (lugar), mapas topográficos o mapas de relieve; y 10) estereogramas de "stereós" (firme, sólido).

Riaño-Rufilanchas (2017) precisa que en la actualidad, solo el histograma, el diagrama y el cartograma se usan en la terminología relacionada con la estadística (Porter, 2004)

El término *euthigrama* para designar “regla" es un arcaísmo en francés. Topograma hoy tiene un significado diferente, aunque de alguna manera relacionado, mientras que el uso actual de radiograma (o radiografía) y estereograma es completamente diferente, y su génesis parece no estar relacionada con el término de Pearson. Los hormogramas, estigmogramas y epipedogramas parecen haber muerto con su creador.

El tema que nos ocupa en este escrito y que será objeto del taller, es la definición de histograma, su utilidad práctica y conceptual. Desarrollaremos estos ítems a continuación.

Definición de histograma

El histograma es el resultado de un proceso de transnumeración (modelo) que convierte una masa de datos sobre una característica continua de un fenómeno, provenientes de una muestra (preferiblemente aleatoria) en una representación gráfica que pretende valorar la forma como están porcentualmente distribuidos los datos y observar algunas características que es imposible observar en la muestra cruda.

Como un proceso previo a la construcción de un histograma, se construyen intervalos (de clase) que constituyan categorías excluyentes y exhausticas con base en las cuales se clasificarán cada uno de los datos, de tal manera que un dato, pertenece a una y solo una categoría y todos los datos quedan en alguna de las categorías. Lo que sigue es la construcción de una tabla con las cuentas del número de datos en cada intervalo y su frecuencia relativa (porcentaje). Para después sobre un eje horizontal representar la escala de la variable en cuestión, marcar los intervalos de clase construidos y tomando como base cada intervalo, colocar sobre él, un rectángulo con área proporcional a la frecuencia de los datos que contiene. Ilustrémoslo con un ejemplo:

En el sector de la industria metalmecánica, se toma una muestra al azar de 500 obreros y se determina la antigüedad en su trabajo (Imagen 1). Por razones de índole administrativo, se quiere representar los datos por medio de un histograma que considere los siguientes intervalos de clase: 0-2 años, 2-3 años, 3-5 años, 5-10 años, 10-20 años. Después de contar el número de obreros que pertenecen a cada intervalo y expresarlo en porcentaje, se obtiene la Tabla 1.

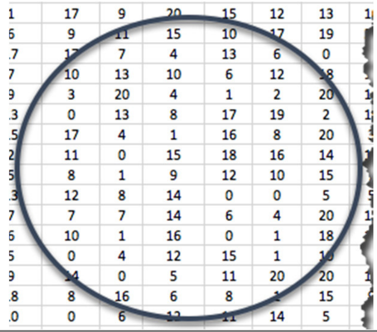

Imagen 1. Muestra aleatoria de la antigüedad de 500 trabajadores
Fuente. Creación propia

i	Intervalo (Años de antigüedad)	Conteo Frecuencia Absoluta (ni)	Frecuencia Relativa (%) (fi)
1	(0-2]	50	10%
2	(2-3]	25	5%
3	(3-5]	200	40%
4	(5-10]	200	40%
5	(10-20]	25	5%

Tabla 1. Cuadro de frecuencias de la antigüedad (años)
Fuente. Creación propia

Lamentablemente, la opción por defecto de los paquetes estadísticos, que nadie cuestiona, es construir intervalos de clase del mismo ancho colocados en el eje *X* y colocar la frecuencia relativa en el eje *Y*. En este caso, aunque la forma del histograma no se distorsiona, se pierde una valiosa oportunidad para que el tratamiento de una variable estadística, sea coherente con el que se dará posteriormente a una variable aleatoria continua, que se definirá con base en una función de densidad.

En algunas ocasiones, es muy conveniente el uso de intervalos de clase de distinto ancho. Por ejemplo, cuando se trata de la variable “salario”, sucede que para salarios bajos, por ejemplo para un salario mínimo de 700.000 pesos, una diferencia de 7.000 pesos, es bastante (10%), sin embargo esa misma diferencia, para salarios de siete millones, no es muy significativa. Posiblemente no será de interés construir para salarios elevados, intervalos de ancho 100 mil pesos. Además, en los salarios bajos, está concentrada la población, lo cual garantiza que aun para intervalos estrechos, no se presenten intervalos con cero datos. A medida que los salarios se hacen muy altos, la densidad baja, y para no obtener clases vacías, conviene construir intervalos más anchos. Si en el curso, no se trata la situación de intervalos de distinta anchura, el estudiante en esta situación, cometerá el error de hacer lo único que ha visto, colocar en el eje *Y* la frecuencia relativa, produciendo una distorsión en la forma del histograma. Si lo hiciere de esta forma, el histograma resultante, con los datos del ejemplo, es el de la Imagen 2.

Observe que el histograma de la Imagen 2, es incorrecto, pues de acuerdo con la definición de histograma, las áreas de los rectángulos que corresponden a cada intervalo, deben ser proporcionales a sus frecuencias relativas. De la Tabla 1, podemos observar que las frecuencias relativas de los dos primeros intervalos son 10% y 5% respectivamente, esperaríamos que el área del primer rectángulo fuera el doble de la del segundo. Observando la Imagen 2, vemos que el primer rectángulo es 4 veces el segundo. Igual ocurre si comparamos el tercer y cuarto intervalo que tienen cada uno 40% de los datos, deberían tener la misma área, sin embargo, el cuarto es 2,5 veces el tercero.

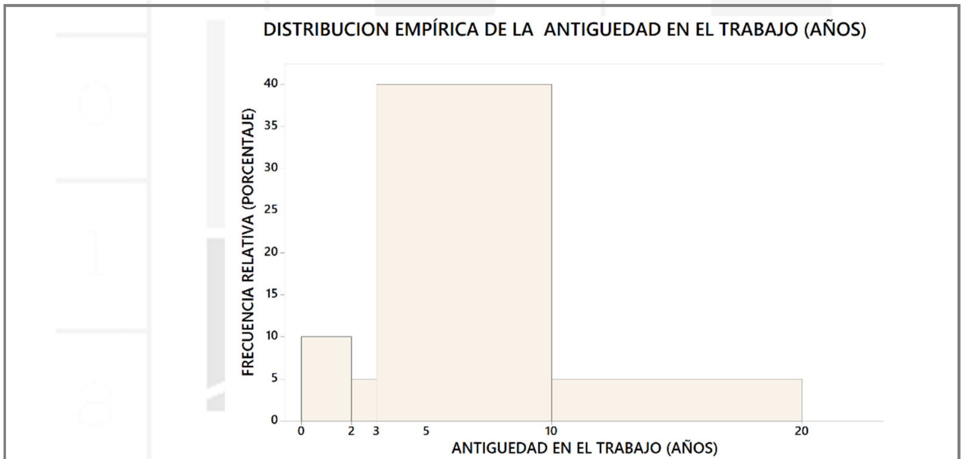

Imagen 2. Histograma con anchos de clase distintos y con el eje *Y*, como frecuencia relativa
Fuente. Creación propia

La moraleja es que haciendo lo que hacen los programas por defecto, cuando los intervalos son de diferente ancho se produce una distorsión.

Intentemos construirlos de manera correcta. Note que en esta situación los intervalos son de diferente ancho (Ci). Se debe ahora construir un conjunto de rectángulos cuya base sea el intervalo de clase correspondiente y cuya área (Ai) represente la frecuencia relativa (f_i) del intervalo de clase. La pregunta es ¿Qué alturas f_i^* deberá tener cada rectángulo para que el área coincida con la frecuencia relativa? De esta manera la ordenada, es decir las alturas, digamos f_i^*, del rectángulo construido sobre el i-ésimo intervalo, deberá ser tal que el área del rectángulo A_i coincida con su frecuencia f_i, es decir que: $A_i = f_i = (base).(altura) = C_i . f_i^*$ donde C_i es el ancho del intervalo. Así, despejando f_i^* se obtiene la altura (ordenada eje vertical) que debe tener cada rectángulo. $f_i^* = f_i / C_i$.

Observe que se divide la frecuencia relativa entre el número de unidades que tenga el intervalo correspondiente, entonces las unidades de f_i^* son (% de datos por cada unidad de la variable en dicho intervalo). Veamos por ejemplo para el primer intervalo: $f_1 = 10\%$ y $C_1 = 2$, así que la altura del primer rectángulo es: $f_1^* = \frac{f_1}{C_1} = \frac{10\%}{2\ años} = 5\% / año$

Es intuitivamente claro, que si el primer intervalo tiene el 10% de los datos y estos datos están distribuidos en un intervalo que tiene una longitud de dos (2) unidades, pues en promedio hay 5% por cada unidad ($f_1^* = 5\% / año \equiv 0.05 / año$). El cuarto intervalo, (5; 10], por ejemplo, en sus 5 unidades (5 años) contiene 40% de los datos. Así que, en promedio, hay 8% de los datos en cada unidad o lo que es lo mismo:

$$f_4^* = \frac{f_4}{C_4} = \frac{40\%}{5\ años} = 8\% / año \equiv 0{,}08 / año$$

Es decir, las unidades del eje *Y* en el gráfico del histograma son %/unidad de intervalo, por eso se le conoce como densidad de frecuencia (f_i^*) y en este caso, para tomar en consideración que se calcula con base en los datos de una muestra, se le llama función “empírica” de densidad de frecuencia. En la Tabla 2, se registra la densidad empírica de frecuencia para cada intervalo.

i	Intervalo (Años de Antigüedad)	Frecuencia Relativa f_i %	Densidad de Frecuencia f_i^* %/año
1	(0-2]	10%	5%/año
2	(2-3]	5%	5%/año
3	(3-5]	40%	20%/año
4	(5-10]	40%	8%/año
5	(10-20]	5%	0,5%/año
		100%	

Tabla 2. Densidad empírica de frecuencias para la variable Antigüedad (años)
Fuente. Creación propia

Construiremos un verdadero histograma, función de densidad empírica de frecuencia, construyendo los rectángulos con una altura igual a su densidad de frecuencia. De esta manera nos aseguramos que el área coincida con su frecuencia relativa. A partir de la Tabla 2, y asociando a cada intervalo, definido en el eje *X*, su correspondiente densidad empírica en el eje *Y*, se obtiene el histograma de la Imagen 3.

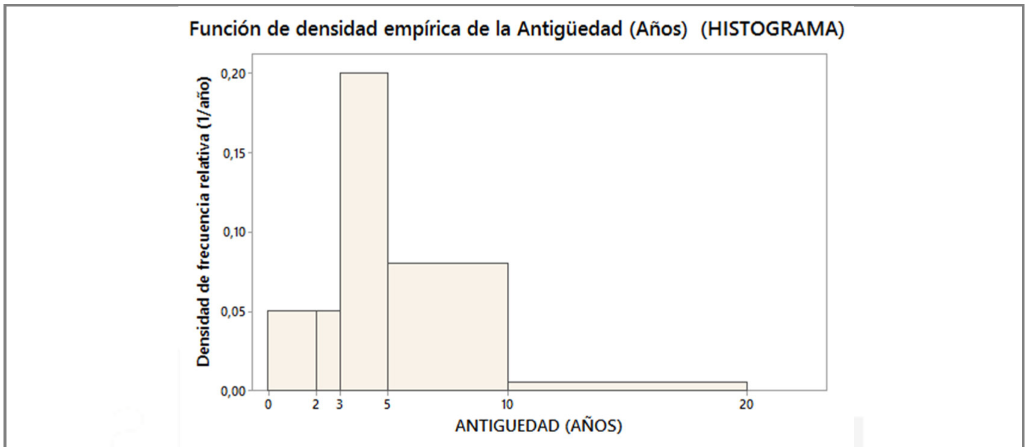

Imagen 3. Histograma para la variable "Antigüedad" cuando los intervalos son de anchura distinta
Fuente. Creación propia

Observe que escribir 5%/año es equivalente a expresarlo como decimal 0.05/año, en el eje *Y* están expresadas de esta manera. Note de la Imagen 3, que ahora el área del intervalo 1 es el doble del área del intervalo 2, conservando la relación que hay entre 10% y 5% que son sus frecuencias relativas respectivamente. Note además que las áreas de los rectángulos sobre los intervalos 3 y 4, son iguales, no obstante tienen diferente altura (haga la multiplicación de base por altura y compruébelo).

Usemos ahora el histograma para hacer una estimación (predicción). Así por ejemplo si se está interesado en estimar el porcentaje de obreros con antigüedad menor o igual a 4 años, digamos P(X≤4), bastará calcular el área del histograma comprendida entre cero (0) y cuatro (4), como se muestra en la Imagen 4. Note que encima de cada rectángulo de ha colocado la frecuencia relativa (porcentaje de datos).

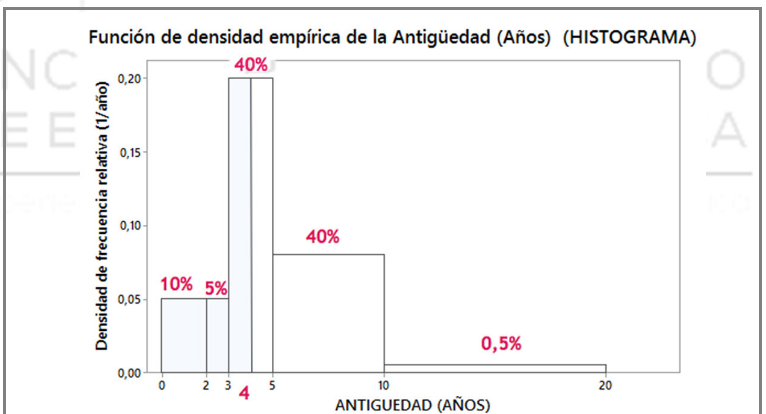

Imagen 4. El área sombreada representa el porcentaje de obreros con antigüedad menor de 4 años P(X<4)
Fuente. Creación propia

Observe que el área sombreada se calcula sumando por un lado las áreas de los primeros rectángulos (10%+5%) y por otro lado la parte del tercer rectángulo comprendida entre 3 y 4, como se conoce su densidad, que es 20%/año, y se requiere un año, así que el porcentaje de trabajadores con antigüedad de 4 años o menos se estima en: P(X≤4)=10%+5% +20%/año (1 año) = 35%. 35% de los obreros aproximadamente tiene antigüedad de 4 años o menos.

Hagamos una nueva pregunta: ¿Qué porcentaje de los obreros tienen antigüedad entre 4 y 7,5 años? Análogamente, si se desea estimar el porcentaje de obreros con antigüedad entre 4 y 7,5 años, es decir P(4≤X≤7,5). La respuesta será calcular el área del histograma entre dichos valores, como se muestra en la Imagen 5. Haciendo el cálculo, usando el concepto de densidad, se obtiene:

$$P\left(4 \le X \le 7,5\right) = f_3^* * \left(5-4\right) + f_4^* * (7,5-5) = 20\%/\,año * (1año) + 8\%/\,año * (2,5años) = 40\%$$

Aproximadamente, 40% de los obreros tiene antigüedad entre 4 y 7,5 años.

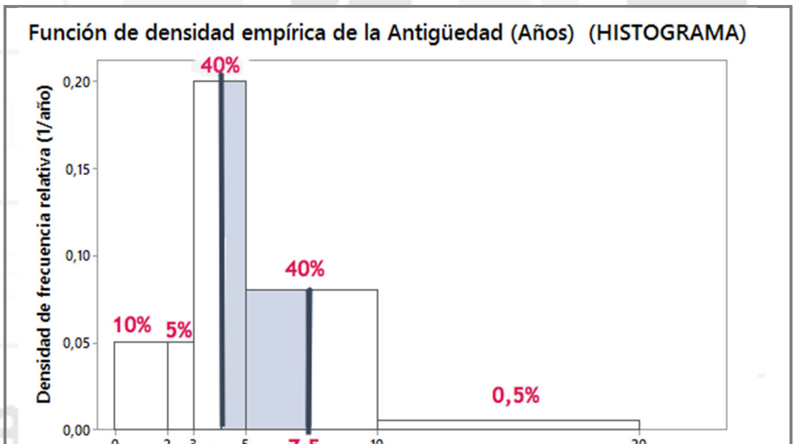

Imagen 5. Representación de P(4<X<7,5), porcentaje de obreros con antigüedad entre 4 y 7,5 años.
Fuente. Creación propia

Podríamos preguntar ahora por el primer cuartil o por la mediana (segundo cuartil). Sería hacer el proceso inverso, determinar cuál es el valor de la variable que deja atrás un área del 25% o el 50% respectivamente. Podríamos, con relativa facilidad, construir la Función Empírica de Frecuencias Acumuladas, más conocida como Función de distribución empírica. F(x).

Después de este recorrido, abordemos la definición de variable aleatoria continua, en el campo de la teoría de la probabilidad.

Variable aleatoria. Definición (Función de densidad de probabilidad)

Se dice que X es una variable Aleatoria Continua si existe una función $f(x)$, llamada función densidad de probabilidad (fdp) de X, que satisface las siguientes condiciones:

a) $f(x) \geq 0 \quad \forall x \in \Re$ — Es razonable que no tome valores negativos, si se asocia con la función empírica de densidad de frecuencia.

b) $\int_{-\infty}^{+\infty} f(x).dx = 1$ — Ya hemos dicho antes que el área del histograma y ahora el área bajo la función de densidad, debe ser 100%.

c) Para cualquier a, b se tiene que $P(a \leq X \leq b) = \int_{a}^{b} f(x).dx$ — El área atrapada entre los valores a y b es justamente el porcentaje de datos de la población que cumple con esas especificaciones, análogamente a lo observado en el histograma. Mirado como la experiencia aleatoria de sacar al azar un valor de X, esta área puede interpretarse como probabilidad.

Esta definición de variable aleatoria, sin haber recorrido con datos el proceso de construcción de la función de densidad empírica, suena muy artificial para el estudiante. Ahora se puede hacer una analogía perfecta. Existen los recursos conceptuales y operativos para comprender la definición de variable aleatoria.

METODOLOGÍA DE DESARROLLO DEL TALLER

En la primera parte del taller se explicará el contenido del marco de referencia antes expuesto, mismo que realizaremos con intervención de los participantes. En la segunda parte, los participantes, en grupo, abordarán desde el principio una situación nueva. Será necesario llevar calculadora, lápiz y regla. Conviene disponer de papel cuadriculado, para facilitar las labores de graficación.

DESCRIPCIÓN DE LAS ACTIVIDADES

Sesión 1. Explicación de la problemática que se pretende abordar y los referentes teóricos. Con la guía del profesor, los participantes, en grupo, podrán ir resolviendo el problema que se ha tratado en líneas anteriores.

Sesión 2. Construcción de la Función de distribución acumulativa, a partir del concepto de densidad empírica. Se da una situación nueva a los participantes con algunas preguntas de interés para que hagan uso del histograma y de la Función de Distribución empírica, para hacer algunos cálculos, entre ellos, los cuartiles de la distribución.

Ejercicio a desarrollar por parte de los participantes

Se dispone de la información sobre salarios, para 1000 empleados, expresados en “Número de Salarios Mínimos” para trabajadores de un cierto sector de la economía. Los resultados de la clasificación de los salarios en términos de categorías definidas por intervalos de clase, se presentan en la Tabla.

Salario	Numero de T. Frec. Absoluta. n_i
1 – 2	300
2 - 4	300
4 – 7	180
7 – 12	150
12 -19	70
	1000

Tabla 3. Salarios para 1000 empleados
Fuente. Creación propia

a) Construya un histograma y describa el comportamiento de los salarios en ese sector de la economía.
b) Se piensa dar un bono mensual por 300 mil pesos a cada uno de los trabajadores que ganen menos de 4,5 salarios mínimos. ¿Cuánto dinero debe presupuestar dicho sector, si el total de los empleados del sector es de 20.000?
c) Construya la Función de Distribución empírica y con base en ella obtenga los cuartiles.
d) Si se consideran puntos atípicos al 1% de los salarios más altos, ¿a partir de cuál salario empiezan los puntos atípicos?
e) ¿Se le ocurre alguna manera de estimar el salario promedio?

REFERENCIAS

Batanero, C., Arteaga, P., & Ruiz, B. (2010). Análisis de la Complejidad Semiótica de los Gráficos producidos por Futuros Profesores de Educación Primaria en una Tarea de Comparación de Dos Variables Estadísticas . *Enseñanza de las Ciencias, 28*(1), 141-154.

Curcio, F. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education , 18*(5), 382-293.

Curcio, F. (1989). *Developing graph comprehension.* Reston: VA: N.C.T.M.

Azcárate, P. (2006). ¿Por qué no nos gusta enseñar estadística y probabilidad? *Jornadas de Investigación en el Aula de Matemáticas: Estadística y Azar*, 1-22. Granada.

Barbosa, J. C. (2006). Mathematical modelling in classroom: A socio-critical and discursive perspective. *ZDM – The International Journal on Mathematics Education, 38*(3), 293-301.

Batanero, C., & Díaz, C. (2011). El Papel de los Proyectos en la Enseñanza y Aprendizaje de la Estadística. En J. P. Roya, *Aspectos didácticos de las matemáticas.* 125-164. Zaragoza: ICE.

Batanero, C., & Godino, J. (2002). Estocástica y su didáctica para maestros. En J. D. Godino, *Fundamentos para la enseñanza y el aprendizaje de las matemáticas para maestros.* 693-766. Granada: Facultad de Ciencias de la Educación. Universidad de Granada.

Batanero, C., & Godino, J. (2005). Perspectivas de la educación estadística como área de investigación. En R. Luengo (Ed.), *Líneas de investigación en Didáctica de las Matemáticas.* 203-226. Badajoz: Universidad de Extremadura.

Batanero, C., Arteaga, P., & Ruiz, B. (2010). Análisis de la Complejidad Semiótica de los Gráficos producidos por Futuros Profesores de Educación Primaria en una Tarea de Comparación de Dos Variables Estadísticas . *Enseñanza de las Ciencias, 28*(1), 141-154.

Cobb, P. (2000). *Handbook of Research Design in Mathematics and Science Education. Conducting teaching experiments in collaboration with teachers.* Lawrence erlbaum associates, publishers.

Curcio, F. (1987). Comprehension of mathematical relationships expressed in graphs. . *Journal for Research in Mathematics Education , 18*(5), 382-293.

Curcio, F. (1989). *Developing graph comprehension.* Reston: VA: N.C.T.M.

D'Ambrosio, U. (2009). Chapter 3.1 Some Reflections on Education, Mathematics, and Mathematics Education. En R. Even, & D. Ball (Edits.), *The Professional Education and Development of Teachers of Mathematics. The 15th ICMI Study.* 239-244. New York: Springer Science Business Media.

Del Pino, G., & Estrella, S. (2012). Educación estadística: relaciones con la matemática. *Pensamiento Educativo. Revista de Investigación Educacional Latinoamericana, 49*(1), 53-64.

delMas, R. (2002). Statistical Literacy , Reasoning , and Learning : A Commentary. *Journal of Statistics Education*, 1-10.

delMas, R. (2002). Statistical Literacy, Reasoning, and Learning: A Commentary. *Journal of Statistics Education, 10*(3). Recuperado el 11 de Abril de 2018, de https://ww2.amstat.org/publications/jse/v10n3/delmas_discussion.html

Flood, R., Rice, A., & Wilson, R. (. (2011). *Mathematics in Victorian Britain Oxford.* Oxford: University Press.

Flores, A., & Pinto, J. (2017). Características de la enseñanza de la estadística por proyectos. *Universidad Auntónoma de Yucatán*.

Friel, S., Curcio, F., & Brigh, G. (2001). Making Sense of Graphs: Critical Factors Influencing Comprehension and Instructional Implications. *Journal for Research in Mathematics Education , 32*(2), 124–158.

Gipps, C. (1994). *Beyond Testing: Towards a Theory of Educational Assessment.* London: The Falmer Press.

Hernández Sampiere, R., Baptista, M., & Fernández, C. (2014). *Metolodología de la investigación* (6 ed.). México: McGraw-Hill Interamericana Editores.

Holmes, P. (1997). Assessing project work by external examiners. En I. Gal, & G. J. B. (Eds.), *The assesment challenge in statistics education.* 153-164. Voorburg: IOS Press.

ICFES. (2016). *Las competencias ciudadanas en las pruebas Saber.* Bogotá.

James, M. (2006). Assessment, Teaching and Theories of Learning. En *Assessment and learning.* 47-60). London.

Kukliansky , I. (2016). Student's Conceptions in Statistical Graph's Interpretation. *International Conceptions in Statistical Graph¨s Interpretation, 5*(4), 262-267.

Lee, C., & Meletiou, M. (2003). Some difficulties of learning histograms in introductory statistics. En Nadaraya (Ed.), *en el Joint Statistical Meetings Section on Statistical Education. http://www.statlit.org/PDF/2003LeeASA.pdf* . E.A: ASA.

Levasseur, E. (1885). La Statistique graphique , Volume. pp. 218-250. *Journal of the Statistical Society of London, jubilee volume*, 218-250.

Meletieu-Mavrotheris, M., & Serradó, A. (2012). Formación a distancia para profesores de matemáticas: la experiencia de EarlyStatistics. *Revista de Universidad y Sociedad del Conocimiento (RUSC), 9*(1), 150-165.

Ministerio de Educacion Nacional. (1998). Lineamientos Curriculares de Matemáticas. Santa fé de Bogotá: Cooperativa Editorial Magisterio.

Ministerio de Educacion Nacional. (2006). Estandares básicos de competencias en lenguaje, matemáticas, ciencias y ciudadanas. Santa fé de Bogotá: Revolución Educativa Colombia Aprende.

Porter, T. (2004). *Karl Pearson. The scientific life in a statistical age* . Princeton: Princeton University Press.

Priti, S. (1997). *A Modelof the Cognitive andPerceptual Processes in GraphicalDisplay Comprehension.* AAAI Technical Report FS 97-03.

Pryor, J., & Crossouard, B. (2005). A socio-cultural theorisation of formative assessment.

Riaño-Rufilanchas, D. (2017). On the origin of Karl Pearson's term "histogram". *Estadística Española, 59*(192), 29-35.

Rocha, P. (2009). Una educación estadística: para una sociedad que tolere la incertidumbre . *Revista cientifica* (11), 6-14.

Rodríguez, M. (2013). *La educación matemática en la con-formación del ciudadano* (15).

Ruiz, B. (2006). *Un acercamiento cognitivo y epistemológico a la didáctica del concepto de variable aleatoria.* México: Tesis de Maestria. CICATA.

Sayago, S. (2014). El análisis del discurso como técnica de investigación cualitativa y cuantitativa en las ciencias sociales. *Cinta de Moebio*(49), 1-10.

Schwartzman, S. (1994). *The Words of Mathematics. An Etymological Dictionary of Mathematical Terms used in English».* Washington: The Mathematical Association of America.

Tagle, T. (2011). El enfoque reflexivo en la formación docente. *Calidad en la educación*(34), 203-215.

Valero, P. (2007). ¿De carne y hueso?: La vida social y política de las competencias matematicas. *Foro Educativo Nacional de Colombia–Competencias matemáticas*.

Vanegas, Y. M. (2013). *Competencias ciudadanas y desarrollo profesional en matemáticas.* (Tesis de doctorado). Uniersitat de Barcelona.

Wild , C., & Pfannkuch , M. (1999). Statistical Thinking in Empirical Enquiry. *International Statistical Review, 67*(3), 223-65.

Zapata-Cardona, L. (2011). Algunas reflexiones acerca del conocimiento pedagógico disciplinar del profesor de estadística. *Didac* (56-57), 9-14.

Zapata-Cardona, L., & González, D. (2017). Imágenes de los profesores sobre la estadística y su enseñanza. *Educación Matemática, 29*(1), 61-89.

Zapata-Cardona, L., & Rocha, P. (2011). Actitudes de profesores hacia la estadística y su enseñanza. *XIII CIAEM-IACME*, (pp. 1-12). Brasil.

Zapata-Cardona, L., & Rocha, P. (2013). La clase de estadística más allá del curriculo. En A. Salcedo (Ed.), *Educación Estadística en América Latina: Tendencias y Perspectivas* (pp. 153-165). Programa de Cooperación Interfacultades: Universidad Central de Venezuela .

Zeichner, K. M. (1993). El maestro como profesional reflexivo. En D. Liston, & K. M. Zeichner (Eds.), *La formación del profesorado y las condiciones sociales de la enseñanza* (pp. 1-9). Madrid: Morata.