Probabilidades y la curva normal

Las distribuciones reales y las distribuciones teóricas

Por Tevni Grajales Guerra

Tal cual estudiamos en nuestro tercer tema. Cuando registramos los valores de una variable observados en cierto número de casos, obtenemos diferentes valores de la variable repetidos. Y estos valores observados pueden ser agrupados de diversas maneras según sean estas variables discretas o continuas.

Tabla 1

Ejemplo de distribuciones de frecuencia de variables discretas y continuas

Variable discreta		Variable continua	
Hermanos	Frecuencia	Salario	Frecuencia
1	3	400 - 699	34
2	4	700 - 999	45
3	6	1 000 - 1 299	33
4	5	1 300 - 1 599	26
5	3	1 600 - 1 899	19

Los datos presentados en la tabla anterior, son datos reales que se refieren a una observación en particular. Los podemos representar en gráficos de frecuencia absoluta, frecuencia relativa, frecuencia acumulada "a más de" o "a menos de", etc.

Pero hay una diferencia muy importante que es necesario resaltar para nuestro nuevo tema. Si usted considera la variable Salario puede suponer que para cada una de las personas representadas, hay un valor específico. Es decir que en lugar de decir que tenemos 34 casos cuyo salario está entre 400 y 699 pesos, podríamos disponer del monto individual y así sucedería con los otros intervalos de clase. Porque se trata de una variable continua. Los valores a representar en la gráfica pueden ser unidades o subdivisiones de la unidad de medida hasta el infinito. En cambio para la variable hermanos, por tratarse de una variable discreta, los valores sólo serán números enteros.

Probabilidad teórica

Hasta aquí un repaso de nuestro tema No. 3. Ahora quisiera llamar su atención a otro tipo de distribución, en este caso se trata de valores que se supone puede asumir una variable, aunque no los hayamos observado en la realidad. Son valores teóricos y no reales. Se trata de las distribuciones de probabilidades (como si fuera una distribución teórica de frecuencias). Es la que describe el comportamiento que se espera tengan todos los resultados de un experimento aleatorio. Como tienen que ver con expectativas, se utilizan para obtener inferencias y decisiones bajo condiciones de incertidumbre.

En este caso se representa lo que se espera que suceda, no lo que en efecto ha sucedido. Por ejemplo: si usted lanza al aire dos veces una moneda perfecta, cuáles son los posibles resultados (note que se trata de posibles -probables- resultados, no que se puedan observar). Pues la respuesta es muy sencilla, si en la primera vez sale sol y si en la segunda sale sol entonces van dos de sol, si en la segunda sale águila apenas va uno y uno. En cambio si en la primera sale águila y en la segunda sol tendremos uno y uno, pero si en la segunda sale sol tendremos dos de sol.

Ahora representemos los resultados posibles de lanzar una moneda, lo cual se presenta en las primeras dos columnas, luego en la tercera se presentan en número de posibles águilas en los primeros dos lanzamientos y la última columna presenta la probabilidad de cada uno de los posibles resultados.

Cuadro 1

Resultados posibles al lanzar dos veces una moneda

Primer lanzamiento	Segundo lanzamiento	Número de águila en los dos lanzamientos	Probabilidad de los cuatro resultados posibles				
	A	2	0,5	*	0.5	=	0.25
Águila -A							
	S	1	0,5	*	0.5	=	0.25
	A	1	0,5	*	0.5	=	0.25
Sol -S							
	S	0	0,5	*	0.5	=	0.25

A partir de lo presentado en la columna tres se pueden anticipar las probabilidades del número de águilas en los dos lanzamientos según el cuadro 2.

Cuadro 2

Distribución de probabilidades del número de águilas en dos lanzamientos de una monda perfecta

Número de Aguilas (X)		Lanzamientos				Probabilidades de obtener X águilas. P(X)
0			(S,S)			0.25
1		(A,S)	+	(S,A)		0.50
2			(A,A)			0.25

Se puede presentar la frecuencia de probabilidades de la siguiente manera:

Probabilidad

0 1 2

Número de águilas

Gráfica 1

Probabilidad de la ocurrencia de águila después de lanzar una moneda perfecta dos veces

¿Puede usted imaginar cómo se representaría la gráfica 1 si se tratara de ilustrar la probabilidad de ocurrencia de cada uno de los posibles el número de águila si se lanzara la moneda 100 veces? Eso nos llevaría a ubicar en el eje de las abscisas una escala de que iría de 0 veces hasta 100 veces y de la misma manera como aparece en la gráfica anterior, los valores 0 y 100 tendrían el menor porcentaje de probabilidad y los valores alrededor de 50 serían los que obtendrían las más elevadas probabilidades. Pero lo importante en este momento es resaltar que se trata de frecuencias teóricas y no de observaciones reales.

La probabilidad puede ser considerada como una teoría referente a los resultados posibles de los experimentos. Estos deben estar en condiciones de ser repetidos. Como es el caso del lanzamiento de una moneda. Teóricamente la probabilidad de define como una proporción (*p*) y se trata del número de oportunidades en que un suceso puede ocurrir con respecto al total de sucesos posibles. Si usted tiene en una urna siete bolas de siete diferentes colores, la probabilidad (p) de que uno de los colores en particular salga al primer intento de retirar una bola será de uno en siete (1/7). Si tiene 52 cartas o barajas de póquer, la probabilidad de que al sacar una de ellas, ésta sea un rey de corazones es 1/52. Note que estamos refiriéndonos a la frecuencia relativa de que un suceso tenga lugar entre varios. Esto es conocido como el enfoque clásico en la teoría de probabilidad.

Probabilidad empírica

Cuando nos trasladamos al terreno de la vida diaria, encontramos que no es tan fácil determinar la frecuencia relativa esperada de un suceso. Por ejemplo, es posible que no conozcamos la proporción exacta de madres en una ciudad que tiene 3 hijos. Pero podemos estudiar una muestra elegida al azar de las madres de la ciudad y estimar la proporción de las que tendrán tres hijos. Si en una muestra aleatoria de 100 madres encontramos que 20 tienen un hijo, 25 tienen dos hijos, 27 tienen tres hijos, 13 tienen cuatro hijos, 5 tienen cinco hijos y las restantes 10 tiene seis o más hijos. Podríamos estimar que la proporción de madres con tres hijos es de .27, empleando la fórmula:

p (madre con tres hijos) = 27 / 100 = 0.21

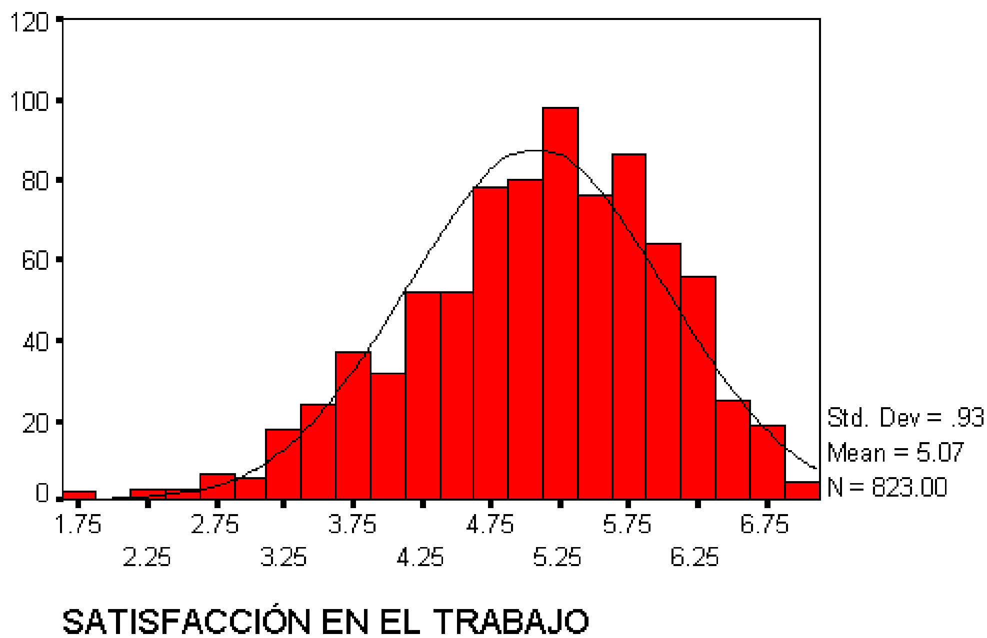
Como usted nota, hemos estado hablando de la probabilidad en función de frecuencia relativa y esto corresponde a variables discretas (número de hijos, Número de veces que sale un color en particular, número de veces que la moneda lanzada presenta águila). En el caso de variables continuas se representa la frecuencia en función de áreas situadas debajo de una curva de manera que la probabilidad se puede expresar en términos de proporción de cierta área por debajo de la curva respecto al área total bajo la curva.

La gráfica 2 representa la distribución de frecuencia de la satisfacción laboral de una muestra aleatoria (835 casos) seleccionada entre maestros de secundaria en el estado de Nuevo León, México durante 1998. Como puede observarse, se trata de una distribución en la que la frecuencia con que se presentan los valores en los extremos es menor y a medida que se mueve hacia el centro (media) la frecuencia es mayor. La totalidad de área sombreada representa la totalidad de los casos (el 100%), que en el de que fuese una distribución normal perfecta correspondería al área identificada por la curva sobre puesta.

En la gráfica 3 se presenta la misma distribución pero esta vez con sus valores estandarizados. Es dice con una media de 0 y desviación estándar de 1, lo que conocemos como puntuación z. Esa área debajo de la curva puede considerarse una distribución de probabilidades que en su totalidad es igual a 1.00 y cualquier porción de esa área debajo de la curva puede ser identificada como p. Con esta información podemos identificar la probabilidad que existía de encontrar entre los maestros de secundaria de Nuevo León en 1998 una cierta cantidad de satisfacción laboral. Por ejemplo cuál es la probabilidad de encontrar casos en una escala de 1 hasta 7 de encontrar casos por encima de 6.75 puntos de satisfacción. Siendo que la media muestral es de 5.07 y la desviación estándar es de 0.93 podemos identificar z para 6.75 restando este valor de 5.07 y dividiendo el resultado entre 0.93.

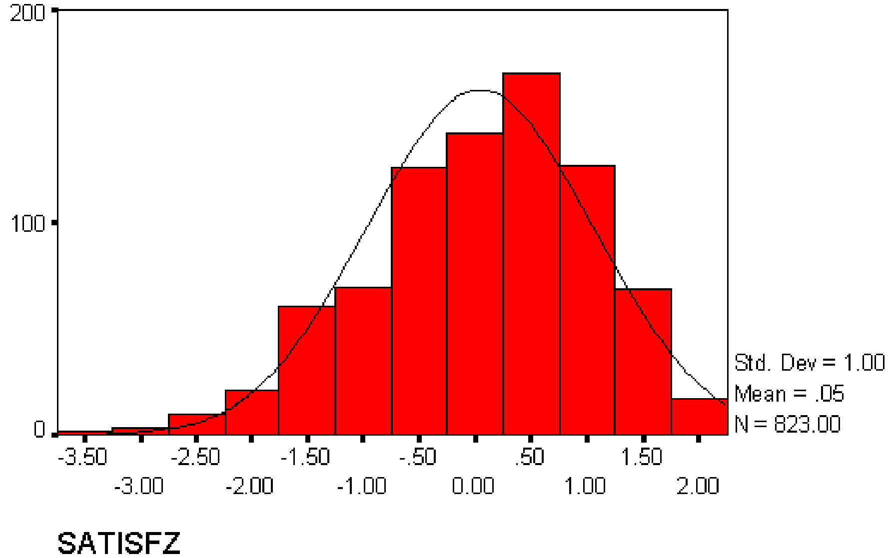
$Z = 6.75 - 5.07 / .93 = 1.68 / .93 = 1.80$

Es decir que el nivel de satisfacción 6.75 se ubica a 1.8 desviaciones estándar de la media y suponiendo que la distribución de la satisfacción laboral fuese una distribución normal podemos determinar el área de la curva que está más arriba de z = 1.80 (Esta área está representada en la Gráfica 3 por el espacio que está por debajo de la curva a la altura de la última columna a la derecha) haciendo uso de una tabla de proporciones de área bajo la curva normal. Lo cual nos dice que para una z = 1.80 el área ubicada a la izquierda de z =1.80 es igual a .9641 lo cual representa el 96 % porciento de toda el área debajo de la curva y el área que se ubica a la derecha de z = 1.80 es el restante .0359 lo que representa el 4%. Lo cual puede significar para la Gráfica 2 que el 4 % de los casos observado se ubica sobre 6.75 de satisfacción laboral y según la gráfica 3 que es probable que el 4 % de la totalidad de maestros de secundaria en Nuevo León en 1998 (12 525) valoren su satisfacción laboral, por encima de 6.75 puntos.



Gráfica 2

Distribución de frecuencias de la satisfacción laboral en una muestra de 835 maestros de secundaria en el estado de Nuevo León en 1998



Gráfica 3

Distribución estandarizada de la satisfacción laboral en una muestra de 835 maestros del estado de Nuevo León en 1998.

No olvide que para hablar de probabilidades es indispensable que los elementos de la muestra utilizada hayan tenido la oportunidad de ser elegidos para formar parte de la muestra. Aunque no es posible predecir un suceso cuando se considera aisladamente, las colecciones de sucesos aleatorios adquieren características y formas predecibles. En esta ocasión hemos visto la distribución que resultada del lanzamiento de una moneda perfecta y el número de veces que cae águila (una distribución binomial) y la distribución de una variable medida en una escala de intervalo cuyo comportamiento se asemeja a una distribución normal. En el caso de las frecuencias relativas idealizadas se determina que p es la razón entre el número de resultados favorables a un suceso y el número total de resultados posibles. Las probabilidades varían de 0 hasta 1. También tenemos probabilidades establecidas empíricamente por la determinación de frecuencias relativas. En el caso de las variables continuas, el área debajo de la curva representa el totalidad de casos y la probabilidad está definida por la razón que resulta de dividir el área bajo una parte de la curva entre el área total de la curva. Los valores z y la curva normal son muy útiles para calcular las probabilidades referentes a variables normalmente distribuidas.

Montemorelos, Setiembre de 1999

Facultad de Educación, Posgrado en Educación
Universidad de Montemorelos
tevgra@umontemorelos.edu.mx