Distribuciones de probabilidad-Variable aleatoria

Comenzamos a estudiar un tema muy importante en Estadística que nos va dirigiendo al tema de la inferencia estadística, tan utilizada en investigaciones de todas las disciplinas.

Las distribuciones de probabilidad representan “las expectativas de que algo suceda y resultan modelos útiles para hacer inferencias y tomar decisiones en condiciones de incertidumbre” (Levin y Rubin, 2012, p. 178).

En esta primera lectura, se introduce el concepto de *distribuciones de probabilidad* para después estudiar los modelos que nos ayudarán a resolver una gran cantidad de problemas. Podrás identificar una distribución de probabilidad y calcular algunas medidas de interés en esta lectura. ¡Comencemos entonces!

Caso: Hipermercado Indi

Se ha instalado hace pocos años en el país, un hipermercado de la cadena Indi. Con el objeto de conocer los gustos de los clientes, el gerente de marketing ha observado que la venta de un tipo de queso blanco en pote varía enormemente de mes a mes. Esto trajo problemas en el sector abastecimiento porque se vieron en la necesidad de encontrar una herramienta para poder predecir el comportamiento de los clientes. La distribución de probabilidad que se presenta en la tabla siguiente, basada en los datos de los dos últimos años, muestra la demanda mensual del producto.

Tabla 1: probabilidades de ventas de un producto del hipermercado Indi, según la cantidad de productos demandados

Descripción de la tabla 1: En la primera columna, se muestra la demanda unitaria mensual del producto queso blanco en pote, del hipermercado Indi. En la segunda columna, se escriben las correspondientes probabilidades de venta según la demanda del producto.

Demanda unitaria mensual (x)	**Probabilidad P(x)**
300	0,20
400	0,30
500	0,35
600	0,15

Fuente: elaboración propia.

En función de los datos, ¿podrías responder a estas cuestiones?

1. Si la empresa basa las órdenes mensuales en el valor esperado de la demanda mensual, ¿cuál será la cantidad ordenada mensualmente por la empresa para este producto?

2. Supón que cada unidad demandada genera $ 70 de ingresos y que cada unidad ordenada cuesta $ 50. ¿Cuánto ganará o perderá la empresa en un mes si coloca una orden con base en su respuesta al inciso a) y la demanda real de este artículo es de 300 unidades?

3. Determina la dispersión de la demanda del producto en los datos relevados.

Variable aleatoria

Distribuciones de probabilidad discretas

Referencias

Variable aleatoria

¿Qué es una variable aleatoria?

Al principio del módulo 2 definimos el *espacio muestral S* como el conjunto de todos los resultados de un experimento aleatorio.

Definimos, además, qué es un experimento aleatorio. Ahora nos detendremos en los resultados de ese experimento. Dijimos que cada resultado era un evento. “Una variable aleatoria es el resultado numérico de un experimento aleatorio” (Baronio y Vianco, 2015, https://bit.ly/2Y7WwRE).

En otras palabras, “una variable aleatoria asocia un valor numérico a cada uno de los resultados experimentales. El valor numérico de la variable aleatoria depende del resultado del experimento” (Anderson, Sweeney y Williams, 2008, p. 188).

Variables aleatorias discretas y continuas

VARIABLE ALEATORIA DISCRETA	VARIABLE ALEATORIA CONTINUA

Es la que asume un número finito de valores o una serie infinita de valores, como, por ejemplo, el de los números naturales. Si recuerdas, ya las definimos en la primera lectura y dijimos que surge de un conteo.

Tabla 2: ejemplos de variables aleatorias discretas
Descripción de la tabla 2: En esta tabla, se describe el experimento y qué variable aleatoria nos interesa. Luego, en la tercera columna se muestran los posibles resultados de ese experimento, que son siempre

valores numéricos.

Experimento	Variable aleatoria x	Valores posibles para la variable aleatoria x
Invitar 100 personas para un evento	Número de personas que asistirán al evento	0, 1, 2, 3, 4,..., 100
Se realiza un control de calidad en una fábrica de autopartes y se toma una muestra de 20 piezas	Número de piezas que tienen algún defecto	0, 1, 2, 3, 4, ..., 20
Calificar la atención de un profesional en un centro de salud.	Calificación del profesional	0: Malo 1: Regular 2: Bueno 3: Muy bueno 4: Excelente
Responder los llamados en un Call Center	Número de llamadas	1, 2, 3,...

Fuente: elaboración propia.

Observa que en todos estos casos los resultados arrojan un número entero de valores, aunque en algunos casos esos valores sean infinitos.

En los datos categóricos también las variables pueden adoptar un valor numérico establecido previamente, es decir que la variable aleatoria describe el experimento mediante una codificación con números enteros.

VARIABLE ALEATORIA DISCRETA

VARIABLE ALEATORIA CONTINUA

Las variables aleatorias continuas pueden adoptar cualquier valor numérico dentro de un intervalo de números reales. Cuando el experimento requiere resultados experimentales que están dentro de una escala de medición, como el tiempo, peso, distancia, entre otras magnitudes, en general, las variables aleatorias continuas están unidas al conjunto de números reales. Entre dos números reales hay infinitos valores reales.

Tabla 3: Ejemplos de variables continuas

Descripción de la tabla 3: En esta tabla se describe el experimento y qué variable aleatoria nos interesa. Luego, en la tercera columna, se muestran los posibles resultados de ese experimento, que son siempre valores numéricos.

Experimento	Variable aleatoria x	Valores posibles para la variable aleatoria x
Medir las llamadas telefónicas a un comercio	Tiempo en minutos entre dos llamadas consecutivas	$x \geq 0$
Construir una casa	Porcentaje del proyecto terminado en cuatro meses	$x \geq 0$

Fuente: elaboración propia.

Observa que en todos estos casos los resultados arrojan un intervalo real de valores.

Distribuciones de probabilidad discretas

Distribuciones de probabilidad

Diferencias entre distribuciones de frecuencias y distribución de probabilidad

Recuerda que, en Estadística Descriptiva, al estudiar frecuencias relativas, teníamos en la primera columna los valores que adopta la variable en estudio, que ahora llamamos *variable aleatoria*.

Cada frecuencia relativa puede asociarse al resultado de un experimento. La totalidad de frecuencias relativas tiene que sumar 1. Pero, ¡cuidado!, si bien pueden asociarse, no son lo mismo.

> … una distribución de frecuencias es un listado de las frecuencias observadas de todos los resultados de un experimento que se presentaron realmente cuando se efectuó este, mientras que una distribución de probabilidad es un listado de las probabilidades de todos los posibles resultados que podrían obtenerse si el experimento se *llevara* a cabo. (Levin y Rubin, 2012, p. 179).

- Las distribuciones de probabilidad pueden basarse, además, en consideraciones teóricas, como, por ejemplo, el lanzamiento de una moneda.

- También pueden surgir de estimaciones subjetivas, de la experiencia, como, por ejemplo, en las compañías de seguros pueden estimar los índices de mortalidad para asignar la probabilidad de muerte entre los diferentes grupos etarios.

- Inclusive, en el caso de las distribuciones de frecuencias podemos pensar una distribución de probabilidad como una distribución de frecuencias teórica, pues nos muestra cómo se espera que varíen los resultados.

Funciones de probabilidad

En el caso de una variable aleatoria discreta x, la forma que adopta la distribución de probabilidades está definida por una función de probabilidad *f(x)*, cuyo valor es cada una de las probabilidades de la variable aleatoria.

Ejemplo 1

Una compañía de seguros ha realizado una campaña de ventas de seguros para el hogar. Durante los tres últimos meses de la campaña (90 días), la información de las ventas muestra que hubo 10 días en los que no se vendieron seguros, 15 días en los que se vendió 1 seguro, 25 días en los que se vendieron 2 seguros, 30 días en los que se vendieron 3 seguros, 5 días en los que se vendieron 4 seguros, 3 días en los que se vendieron 5 seguros y 2 días en los que se vendieron 6 seguros.

Supongamos que seleccionamos un día al azar de esos tres meses (experimento) y definimos la variable aleatoria x como *número de seguros vendidos en un día*. Como *x* es una variable aleatoria discreta, según el experimento, puede tomar los valores 0, 1, 2, 3, 4, 5 y 6. La tabla de probabilidades de este ejemplo quedaría así:

Tabla 4: Distribución de probabilidad para el número de seguros vendidos en un día por la compañía de seguros

Descripción de la tabla 4: Por medio de esta tabla, se define la función de probabilidad que está dada por la distribución de probabilidad de la variable aleatoria *x: número de seguros vendidos* en un día. A cada variable aleatoria x le corresponde un valor de la función *f(x), es decir una probabilidad.*

x	**f(x)** **Probabilidad de que se dé la variable aleatoria *x***
1	10/90=0,1111
2	15/90=0,1667
3	25/90=0,2778
4	30/90=0,3333
5	5/90=0,0556
6	3/90=0,0333
7	2/90=0,0222
Sumatoria	1

Fuente: elaboración propia.

Es decir:

- *f(0*) es la probabilidad de que no se vendan seguros en un día, que es de 0,1111.
- La probabilidad de que se venda exactamente un seguro en un día es *f(1)=0,1667*, y así con las demás variables aleatorias. Todos estos valores constituyen una distribución de

probabilidades.

- Una ventaja importante de definir una variable aleatoria y su correspondiente distribución de probabilidad es que, una vez que se conoce la distribución de probabilidad, es relativamente fácil determinar la probabilidad de diversos eventos que pueden ser útiles para tomar decisiones.

- Por ejemplo, existe una mayor probabilidad de vender, en un día, tres seguros.

- También se observa que vender más de cuatro seguros en un día es una probabilidad muy baja:

f(5)+f(6) = 0,0333 + 0,0222 = 0,0555, solo un 5,55 %.

En el caso de la compañía de seguros, esta distribución de probabilidad se la puede tomar como referencia porque proporciona información para futuras decisiones.

Condiciones para que exista una función de probabilidad

Al elaborar una función de probabilidad para una variable aleatoria discreta, deben satisfacerse las dos condiciones siguientes:

$$1)\quad f(x) \geq 0$$

$$2)\quad \sum f(x) = 1$$

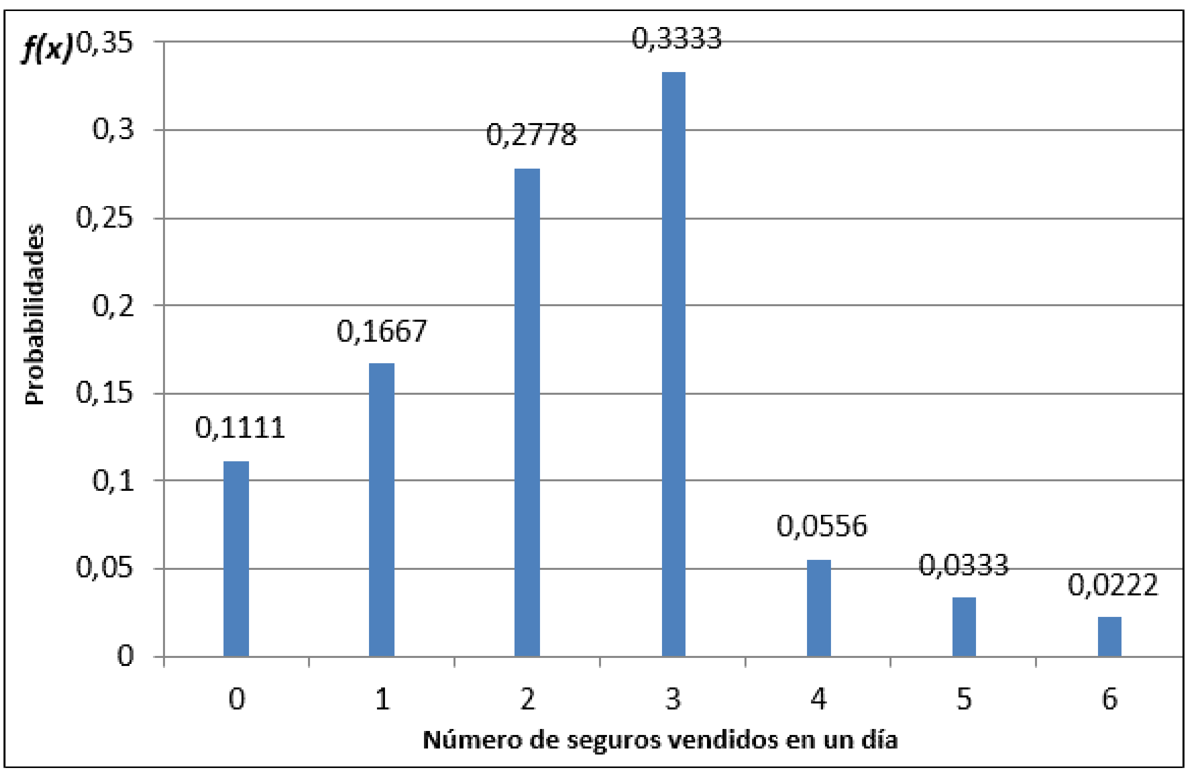
En la tabla 4, se observa que las probabilidades de la variable aleatoria x satisfacen las dos condiciones. Por tanto, la función de probabilidad de la compañía de seguros es una función de probabilidad discreta válida.

Representación gráfica de una función de probabilidad para variables aleatorias discretas

Las distribuciones de probabilidad también se representan gráficamente. En la *figura 1*, en el eje horizontal, aparecen los valores de la variable aleatoria x para el caso de la empresa de seguros y, en el eje vertical, aparecen las probabilidades correspondientes a estos valores.

Figura 1: Representación gráfica de la distribución de probabilidad del número de seguros vendidos en un día

Descripción de la figura 1: Es un gráfico de bastones, como el que se utiliza para variables aleatorias discretas. En el eje vertical está cada una de las probabilidades correspondientes a cada variable, también se llama eje de la función de probabilidad.



Fuente: elaboración propia.

Fórmulas para las funciones de probabilidad

Sabemos que las funciones pueden estar definidas por tablas y gráficos, entre otras formas. Pero la forma más utilizada en matemáticas es mediante una fórmula matemática que, para cada valor de x, se obtiene un valor de *f(x)*.

Análogamente, para describir las funciones de probabilidad, se suele usar una fórmula de la que se obtiene el valor de la función de probabilidad *f(x)* para cada valor x.

El ejemplo más sencillo es el de función de probabilidad uniforme discreta, donde:

$$f(x) = \frac{1}{x}$$

x es el número de valores que puede tomar la variable aleatoria

El típico caso de aplicación de esa función es lanzar un dado. Si definimos la variable aleatoria *x*, número de puntos en la cara del dado en la parte superior, observemos que el valor de probabilidad siempre será 1/6.

Las funciones de probabilidad discretas más empleadas suelen especificarse mediante fórmulas.

Tres de los modelos de este tipo de variables son las distribuciones binomial, hipergeométrica y de Poisson, que estudiaremos en la próxima lectura. Cada una de estas distribuciones tiene su fórmula de función de probabilidades, que permiten calcular las probabilidades para cada variable aleatoria.

También puede trabajarse mediante tablas de probabilidades para encontrar el valor de probabilidad buscado y no tener que aplicar la fórmula para cada variable aleatoria discreta que se busque.

Los modelos mencionados, repetimos, se utilizan para variables aleatorias discretas y a través de ellos podremos calcular las probabilidades para cada variable aleatoria discreta que necesitemos.

Aclaración: en términos prácticos, *P(x)=f(x)*, es decir que se utilizan indistintamente para indicar probabilidad de variable aleatoria y función de probabilidad de variable aleatoria.

Caso Hipermercado Indi

Por ahora, podemos analizar qué tipo de variable se presenta en el caso de la demanda del producto en cuestión. Además, investigar si existen las condiciones para que sea una distribución de probabilidad bien definida.

- Definimos la variable aleatoria *x: cantidad demandada de queso blanco en pote por mes*. Se trata de una distribución de variable discreta.
- Verificamos si se dan las condiciones para ser una distribución de probabilidad:
  1) P $(x) > 0$ se cumple. Todas las probabilidades son positivas.
  2) $\sum P(x) = 1$, verifiquemos 0,20 + 0,30 + 0,35 + 0,15 = 1

Entonces la distribución es válida, está bien definida.

Más adelante responderemos lo que se pregunta en el caso.

El director de admisión de una universidad calcula subjetivamente una distribución de probabilidad para x, el número de estudiantes que ingresarán en la universidad el próximo año. A continuación, se presenta esta distribución de probabilidad:

Tabla 5: Distribución de probabilidades de la cantidad de ingresantes a la universidad. Distribución de probabilidad de la variable aleatoria x: número de ingresantes a la Universidad. A cada variable aleatoria x, le corresponde un valor de la función f(x), es decir, una probabilidad.

Cantidad de ingresantes	Probabilidad
800	0,10
900	0,25
1000	0,45
1100	0,15
1200	0,05

Fuente: elaboración propia.

a) ¿Es válida esta distribución de probabilidades, por qué?

b) ¿Cuál es la probabilidad de que ingresen menos de mil estudiantes?

☐ a) Sí, es válida porque P(x)>0 y ∑P(x)=1. b) La probabilidad es de 0,35

☐ a) Sí, es válida porque f(x)>0 y ∑f(x)=1. b) La probabilidad es de 0,35

☐ a) Sí, es válida porque $P(x)>0$ y $\sum P(x)=1$. b) La probabilidad es de 0,75

☐ a) No es válida porque $P(x)\geq 0$ y $\sum P(x)=0$. b) La probabilidad es de 0,25

☐ a) no es válida porque $f(x)<0$ y $\sum f(x)=1$. b) La probabilidad es de 0,35

SUBMIT

Valor esperado y varianza

Valor esperado

El ***valor esperado***, o media, de una variable aleatoria es una medida de la posición central

El valor esperado es un promedio ponderado de los valores que toma la variable aleatoria. Esa ponderación se realiza multiplicando cada variable aleatoria por su probabilidad.

$$Ex=\mu=\sum x.f(x)$$

E(x): Esperanza o valor esperado de la variable aleatoria discreta x.

Tabla 6: Cálculo del valor esperado de la variable aleatoria para el ejemplo 1. Hemos agregado una columna auxiliar a la tabla 4 para poder realizar el cálculo de la esperanza matemática de la variable aleatoria, ponderándola con la probabilidad de ocurrencia de esta.

x	f(x)	x.f(x)
0	0,1111	0
1	0,1667	0,1667
2	0,2778	0,5556
3	0,3333	0,9999
4	0,0556	0,2224
5	0,0333	0,1665
6	0,0222	0,1332
Sumatoria	1	2,2443

Fuente: elaboración propia.

El valor esperado de ventas de seguros en un día es de 2,24.

- Esto se interpreta como que, a la larga, se venderán ese número de seguros en un día (recuerda que es un promedio; por eso, no es una cifra entera).
- ¡Cuidado! Esto no significa que en el día de mañana se vendan 2,24 seguros. Lo que se dice es que, en ausencia de más información, 2,24 seguros es el mejor número que se pudo obtener como base para el futuro. Si además uno puede distinguir patrones diarios, mensuales, estacionales en los datos, como, por ejemplo, los lunes se venden más seguros; entonces, deben incluirse en las decisiones.
- De aquí se puede hacer un pronóstico de lo que se venderá en un mes (30 días): 30 × 2,24 = 67,2 seguros por mes. **5,2**

Observa que este cálculo tiene mucha analogía con el cálculo de la media aritmética que estudiamos en el módulo 1. Sumábamos todos los productos de la variable en estudio multiplicados por la frecuencia. Luego a esa sumatoria la dividíamos por el total de observaciones. Esto es lo mismo solo que las divisiones por el total ya están hechas en el cálculo de cada probabilidad.

Varianza

Además de una medida de posición, necesitamos para una mejor descripción de la distribución de la variable aleatoria, una medida de dispersión, una medida que nos indique la dispersión de la variable aleatoria discreta.

Su fórmula es:

$$Var(x)=o^2=\sum(x\text{-}\mu)^2.f(x)$$

Nota: Recuerda que esta es la fórmula de la varianza, solo que como *f(x)* es la probabilidad de x, ya está dividida por el número total de ensayos (observaciones). Es decir, *f(x)* son las probabilidades o las frecuencias relativas.

La varianza es un promedio ponderado de los cuadrados de las desviaciones de una variable aleatoria con respecto a la media. Esa ponderación está dada por las probabilidades.

- la desviación (x - μ), mide cuán alejado del valor esperado μ o *E(X)*, se encuentra un valor determinado de la variable aleatoria.
- Para calcular la varianza de una variable aleatoria, estas desviaciones se elevan al cuadrado y después se ponderan con el correspondiente valor de la función de probabilidad.
- A la suma de estas desviaciones al cuadrado, ponderadas, se la conoce como varianza.

Tabla 7: Cálculo de la varianza para el número de seguros vendidos por día, del ejemplo 1. En esta tabla mostramos paso a paso, con columnas auxiliares, cómo se calcula la varianza para el caso del ejemplo 1.

x	f(x)	x-μ	$(x\text{-}\mu)^2$	$(x\text{-}\mu)^2.f(x)$
0	0,1111	-2,24	5,0176	0,5575
1	0,1667	-1,24	1,5376	0,2563
2	0,2778	-0,24	0,0576	0,0160

x	f(x)	x-µ	$(x-\mu)^2$	$(x-\mu)^2.f(x)$
3	0,3333	0,76	0,5776	0,1925
4	0,0556	1,76	3,0976	0,1722
5	0,0333	2,76	7,6176	0,2537
6	0,0222	3,76	14,1376	0,3139
Sumatoria	1			**1,7620**

Fuente: elaboración propia.

Entonces la varianza de la variable aleatoria x es 1,7620. Pero, al estar elevadas las desviaciones el cuadrado, nos están quedando —referido al ejemplo 1—, seguros al cuadrado.

Si extraemos la raíz cuadrada nos queda en la misma unidad del ejemplo: seguros. Entonces:

$$\sigma = \sqrt{1{,}7620} = 1{,}3274$$

Se prefiere utilizar esta medida para medir la variabilidad de una variable aleatoria. Pues, es más fácil de interpretar.

Recuerda que también se puede utilizar el coeficiente de variación para comparar esta variabilidad con respecto a la media o esperanza matemática:

$CV = \frac{\sigma}{\mu} = \frac{1{,}3274}{2{,}2443}$=0,5915 aproximadamente un 59 % de dispersión. Es elevada.

Esperanza y varianza para el caso del Hipermercado Indi

a) La primera pregunta que se hace el encargado del área de abastecimiento es, si la empresa basa las órdenes mensuales en el valor esperado de la demanda mensual, ¿cuál será la cantidad ordenada mensualmente por la empresa para este producto?

El valor esperado de la demanda mensual se calcula multiplicando cada demanda por su correspondiente probabilidad y sumando luego esos productos de la siguiente forma:

$Ex=\mu=\sum x.fx=300\times0,20+400\times0,30+500\times0,35+600\times0,15=445$

Es decir que se espera vender mensualmente 445 unidades de su producto. Además, tal como informa la empresa, tendrán que ordenar mensualmente 445 unidades del producto en cuestión.

b) Suponga que cada unidad demandada genera $ 70 de ingresos y que cada unidad ordenada cuesta $ 50. ¿Cuánto ganará o perderá la empresa en un mes si coloca una orden con base en su respuesta al inciso a) y la demanda real de este artículo es de 300 unidades?

Los ingresos por 300 unidades demandadas = 300 × 70 = $ 21 000

Los costos por 445 unidades compradas= 445 × 50 = $ 22 250

Se estima que la empresa perderá en ese mes $ 1250

c) $Varx=o2=\sum(x-\mu)^2.f(x)$

Realicemos los cálculos en la tabla para verlos mejor:

Tabla 8: Cálculo de la varianza para la demanda de un artículo del Hipermercado Indi

Se muestra paso a paso, con columnas auxiliares, cómo se calcula la varianza para el caso de demanda del pote de queso crema del Hipermercado Indi.

Demanda unitaria mensual (x)	**Probabilidad P(x)**	**$x-\mu$**	**$x-\mu^2$**	**$x-\mu^2.P(x)$**
300	0,20	-145	21025	4205
400	0,30	-45	2025	607,5
500	0,35	55	3025	1058,75
600	0,15	155	24025	3603,75
totales	1			9475

Fuente: elaboración propia.

La varianza es de 9475 unidades

La desviación estándar es

$$\sigma = \sqrt{9475} = 97{,}34 \text{ Artículos.}$$

El $\text{CV} = \frac{97{,}34}{445} = 0{,}2187$. En realidad no existe tanta dispersión de la variable aleatoria como suponían. Está casi al límite.

Valor esperado de una variable aleatoria

Es una medida de posición.

Es un promedio ponderado de los valores que toma la variable aleatoria.

Producto del valor de la variable aleatoria.

Varianza de una variable aleatoria

Es una medida de dispersión.

Su expresión matemática es ∑(x-μ)^2 .f(x)

Desviación estándar de una

variable aleatoria

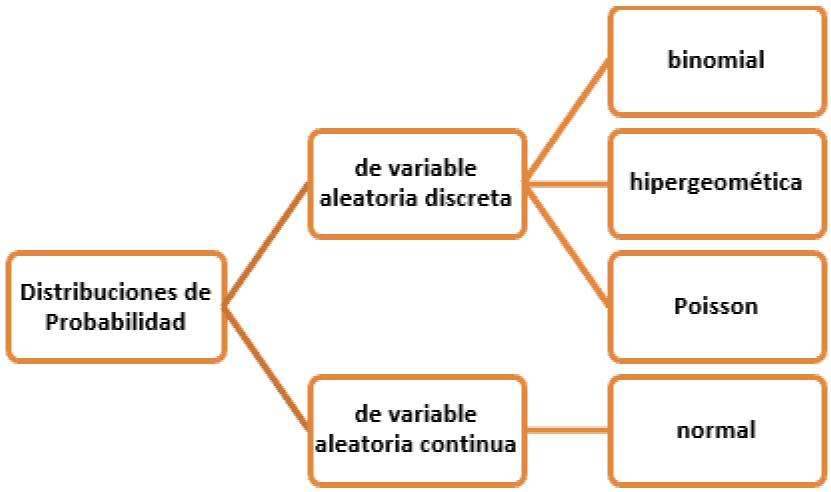
Es la raíz cuadrada de la varianza.	**Práctica para interpretar resultados, misma magnitud que la variable aleatoria.**

Modelos de distribuciones de probabilidad

Como venimos diciendo, las distribuciones de probabilidad son una herramienta poderosa en la toma de decisiones. También estudiamos las funciones de probabilidad, que son las que definen la distribución, por fórmula o por su gráfica. Las funciones de probabilidad modelan la distribución, pero pueden utilizarse dándoles el mismo sentido como verás entre distintos autores. Los modelos más utilizados están indicados en el cuadro de la figura 2.

Figura 2: Modelos de distribución de probabilidad

Descripción de la figura 2: En el cuadro se muestran los modelos de distribución de probabilidad que se estudiarán en las próximas lecturas del módulo, clasificados según el tipo de variable.



Fuente: elaboración propia.

Referencias

Anderson, D. R., Sweeney, D. J. y Williams, T. A. (2008). *Estadística para Administración y Economía.* Ciudad de México, MX: Cengage Learning Editores, S. A.

Baronio, A. y Vianco, A. M. (2015). La tabla de datos en el proceso de investigación econométrica. Recuperado de https://bit.ly/2Y7WwRE

Levin, R. y Rubin, D. (2012). *Estadística para Administración y Economía*. Ciudad de México, MX: Pearson.