

INTRODUCCIÓN AL VALOR ESPERADO Y VARIANZA (5 MINUTOS)

Cuando nos hablan del promedio con ocurre un evento, ¿cómo sabemos con certeza qué tan cerca estamos de alcanzar ese promedio? Esta pregunta nos la podemos hacer en distintas situaciones. Una de estas situaciones son los juegos de azar, ya que en ellos queremos calcular la probabilidad de que ocurra un evento con distintas posibilidades. Por ejemplo, en las últimas cuatro carreras podrían haber dos caballos que han salido en promedio en el 4° lugar: el primero con las posiciones 1, 2, 3 y 10, mientras que el segundo con las posiciones 3, 4, 6, 3. ¿A qué caballo le apostarían que llegará “colocado” (entre los tres primeros lugares)? La intuición nos lleva a apostarle al primer caballo, pero no necesitamos descansar en nuestra intuición. Comprender a qué caballo apostar y encontrar respuestas a muchas otras preguntas del tipo es algo que podemos hacer utilizando medidas de dispersión como la varianza y medidas de tendencia central como el valor esperado.

VALOR ESPERADO (30 MINUTOS)

Podríamos decir que el valor esperado nace en los juegos de azar, cuando los jugadores querían saber cuál era su esperanza de ganar o perder en un juego determinado. Por ejemplo, jugando a la ruleta y asignándole a cada casillero una probabilidad y un premio por apostar ahí, siendo el premio mayor en donde había poca esperanza de ganar.

El **valor esperado** de una variable aleatoria es una medida de la localización central de la variable aleatoria. Este se interpreta físicamente como el centro de gravedad de la distribución de probabilidad, que es igual a la media o promedio aritmético.

El valor esperado de una **variable aleatoria discreta** se representa de la siguiente manera:

$$E(x) = \mu = \sum xf(x)$$

Sea X una variable aleatoria discreta con función de probabilidades $f(x)$. El valor esperado se representa por $E(x)$ y μ (miu). En otras palabras, la esperanza de una variable aleatoria discreta es la suma de los posibles valores que puede tomar X multiplicado por su respectiva probabilidad.

El valor esperado de una **variable aleatoria continua** se representa con la siguiente fórmula:

$$E(x) = \int_{-\infty}^{+\infty} xf(x)dx$$

En este caso, el valor esperado de una variable aleatoria continua, es la suma de los posibles valores X por su probabilidad; al ser una integral es el área bajo la curva, es decir, bajo la función de distribución de la variable aleatoria.

Propiedades

1. La esperanza de una constante es la propia constante. Es decir, si k es una constante entonces: $E(k) = k$
Por ejemplo, si lanzamos un dado donde cada una de las 6 caras contiene el número 4, entonces la esperanza de obtener el número 4 es justamente 4.
2. Si una variable aleatoria X está acotada, es decir existen dos valores a y b tales que $a \leq X \leq b$ entonces se verifica que: $a \leq E(x) \leq b$

Por ejemplo, una cocinería que vende de 15 a 30 almuerzos todos los días de la semana, con distintas probabilidades por día, tiene un valor esperado del número de almuerzos que vende en un día a la semana que es mayor o igual a 15 y menor o igual a 30.

3. Sea X una variable aleatoria y sean $g(X)$ y $h(X)$ dos funciones de X , que a su vez son variables aleatorias, cuyos valores esperados existen y sean a y b dos constantes cualesquiera, entonces:

$$E[a * g(x) + b * h(x)] = a * E[g(x)] + b * E[h(x)]$$

La esperanza de una constante, en este caso “a”, por una función $g(x)$ más la constante “b” por $h(x)$, es igual a la constante “a” por la esperanza de $g(x)$ más “b” por la esperanza de $h(x)$

Por ejemplo, un vendedor de seguros de vida gana comisión por el número de seguros que vende al mes $g(x)$ y además tiene un segundo sueldo dependiendo de los seguros que se venden en total en el área de ventas de la empresa $h(x)$ (asumamos que es el único empleado de esa área). Durante el aniversario de la empresa el dueño después de largas horas de fiesta promete darles una prima sobre su función de ganancias, prima que está representada por las constantes a y b respectivamente. El valor esperado de la ganancia durante ese mes se puede calcular multiplicando la esperanza de la función o calcularla antes de sacar la esperanza. Entonces:

$$g(x) = 2x - 1$$

$$h(x) = x - 5$$

$$E((2xa - a) + (xb - 5b)) = a * E(2x - 1) + b * E(x - 5)$$

4. Sea X una variable aleatoria y sean $g(X)$ y $h(X)$ dos funciones de X que, a su vez son variables aleatorias cuyos valores esperados existen; si $g(x) \leq h(x)$ entonces: $E[g(x)] \leq E[h(x)]$ Si la función $g(x)$ es menor o igual a $h(x)$ entonces la esperanza de $g(x)$ es menor o igual a $h(x)$.

Utilizando el mismo ejemplo anterior:

$$g(x) = 2x - 1$$

$$h(x) = x - 5$$

$$(2x - 1) \leq (x - 5) \text{ entonces: } E[2x - 1] \leq E[x - 5]$$

5. Sea X una variable aleatoria y sea $g(X)$ una función de X que, a su vez es una variable aleatoria, cuyo valor esperado existe, entonces: $|E[g(x)]| \leq E[|g(x)|]$ El valor absoluto de la esperanza de una función $g(x)$ es menor o igual a la esperanza del valor absoluto de $g(x)$.

Utilizando el mismo ejemplo anterior:

$$|E[2x - 1]| \leq E[|2x - 1|]$$

6. Si X e Y son independientes entonces $E[x * y] = E[x] * E[y]$

Por ejemplo, el valor esperado de lanzar una moneda y un dado a la vez obteniendo cara y números pares se puede obtener de ambas maneras al ser la esperanza un operador lineal.

Para el caso de la moneda, la variable aleatoria $x=1$ si es cara y $x=0$ si es sello, y la probabilidad de obtener cara (o sello) es de 0.5, por lo tanto la esperanza de lanzar una moneda y obtener cara es de 0.5. Para el caso del dado es lo mismo: la variable aleatoria toma el valor de $x=1$ si es par y 0 lo contrario y la esperanza de lanzarlo y obtener un número par es de 0.5.

$$E[x * y] = (0 * 0)(0.5 * 0.5) + (0 * 1)(0.5 * 0.5) + (1 * 0)(0.5 * 0.5) + (1 * 1)(0.5 * 0.5) = 0.25$$

$$E[x] = 0 * 0.5 + 1 * 0.5 = 0.5$$

$$E[y] = 0 * 0.5 + 1 * 0.5 = 0.5$$

$$E[x] * E[y] = 0.25$$

Ejercicio de Valor Esperado 1

Una amiga generosa les ofrece jugar lanzando un dado y dependiendo lo que salga darles cierta cantidad de dinero. Pero su generosidad tiene un límite y ella tampoco quiere salir perdiendo, por lo que propone lo siguiente:

- Se lanza el dado.
- Si sale un 1, les da 10.000 pesos.

- Si sale un 2, 3, 4 o 5 ustedes le dan a ella 3.000 pesos.
- Si sale un 6, ella les da 2.000 pesos.

¿Será un juego justo? Esto quiere decir que en un juego donde se participa muchas veces un jugador no gana más que el otro.

Ejercicio de Valor Esperado 2

Sea una variable aleatoria X cuya función de densidad es la siguiente:

$$f(x) \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

¿Cuál es la esperanza de X ?

Ejercicio de Valor Esperado 3

Un agente de seguros de vida recibe un salario mensual de 800 mil pesos más una comisión de 20.000 pesos por cada seguro que vende. Si el número de seguros de vida que vende al mes es una variable aleatoria X con la siguiente función de probabilidad:

$$f(x) \begin{cases} \frac{1}{10} & x = 0 \\ \frac{1}{5} & x = 1 \\ \frac{2}{5} & x = 2 \\ \frac{1}{5} & x = 3 \\ \frac{1}{10} & x = 4 \\ 0 & \text{resto} \end{cases}$$

¿Cuál sería el salario mensual esperado?

VARIANZA (30 MINUTOS)

La varianza es usada como una medida de variabilidad o dispersión. La varianza permite resumir la variabilidad en los valores de la variable aleatoria.

La varianza de una **variable aleatoria discreta** se representa de la siguiente manera:

$$Var(X) = \sigma^2 = \sum (x - \mu)^2 f(x)$$

Sea X una variable aleatoria discreta con función de probabilidades $f(x)$. Siendo $x - \mu$ la desviación de x con respecto a la media, esta mide que tan alejado está la variable aleatoria determinada X del valor esperado μ . Denotado por $Var(x)$ y σ^2 (sigma al cuadrado)

Cabe destacar que la raíz cuadrada positiva de la varianza es la desviación estándar, denotada por σ . Está se mide en las mismas unidades que la variable aleatoria, siendo preferida a la hora de describir la variabilidad de una variable aleatoria. Como vimos anteriormente la varianza se mide en unidades al cuadrado, siendo más difícil de interpretar.

Propiedades

1. La varianza de una constante k es cero. Siguiendo con el mismo ejemplo anterior del dado con las 6 caras iguales, si lo lanzamos 1000 veces no habrá diferencia entre la variable aleatoria x y el valor esperado μ por lo que el resultado será 0.

2. Sea X una variable aleatoria cuya varianza existe, y k una constante cualquiera. Entonces:

$$Var(k * X) = k^2 * Var(X)$$

La varianza de una variable aleatoria multiplicada por una constante k es igual a la constante k al cuadrado por la varianza de la variable aleatoria.

3. Sea una variable aleatoria cuya varianza existe y a, b dos constantes cualesquiera. Entonces:

$$Var(a * X + b) = a^2 Var(x)$$

La varianza de una variable aleatoria X multiplicada por una constante cualquiera, "a", más otra constante, "b" es igual a la constante "a" al cuadrado por la varianza de X . Esto sucede porque la varianza de "b" en este caso es 0 por ser una constante.

4. Sea X e Y dos variables aleatorias independientes cuyas varianzas existen, entonces se verifica que la varianza de la suma o de la diferencia de ambas variables aleatorias independientes es igual a la suma de las varianzas. Es decir: $Var(X \pm Y) = Var(X) + Var(Y)$

Ejercicio de Varianza

La marca de autos Daewoo, debido a un boom comercial gatillado por un video viral en youtube, ha decidido como nueva estrategia de ventas rebajar los precios y así aumentar las ventas y disminuir su stock de autos. El director comercial estimó la siguiente función de probabilidad, utilizando como variable aleatoria el número de autos X que se venderán en el próximo mes al rebajar los precios:

$$\bullet \quad f(x) \begin{cases} 0,05 & x = 0 \\ 0,15 & x = 1 \\ 0,35 & x = 2 \\ 0,25 & x = 3 \\ 0,20 & x = 4 \end{cases}$$

Utilizando lo visto en clases para obtener el valor esperado μ :

$$\mu = E(X) = \sum xf(x) = 0 * 0.05 + 1 * 0.15 + 2 * 0.35 + 3 * 0.25 + 4 * 0.2 = 2,4$$

¿Cuál sería la varianza?

EJEMPLO DE ESPERANZA Y VARIANZA USANDO CASEN 2011 EN STATA (10 MINUTOS)

Al igual que en la primera clase utilizaremos un extracto de la base de datos CASEN 2011. En esta ocasión veremos cómo obtener la varianza y la desviación estándar de una variable en específico. Deben descargar la “Base de Datos Inferencia Estadística” de la sección “Inferencia Estadística” del siguiente enlace:

http://www.esteban-calvo.com/es/difusion_/recursos/. Luego guarden la base de datos en el disco C o en un pendrive.

```
.use "C:\casen2011_ie.dta", clear
```

Usaremos el comando <summarize> nuevamente. Como comentamos anteriormente esta herramienta es muy útil ya que nos arroja una descripción completa en una tabla resumen.

```
. sum edad
```

Variable	Obs	Mean	Std. Dev.	Min	Max
edad	4853	43.64517	18.75686	0	100

Pero no es toda la información necesaria, si queremos obtener la varianza hay dos opciones, tomar la desviación estándar y elevarla al cuadrado o utilizar opciones en Stata. Para especificar opciones hay que agregar una coma tras el comando y luego la opción necesaria. En este caso utilizaremos la opción <,detail>, donde entrega en detalle una serie de indicadores y medidas que procesa el comando <summarize> y no muestra de forma automática. La opción <detail> puede ser resumida en <d>.

```
. sum edad, d
```

```

                                edad (años cumplidos)
-----
Percentiles      Smallest
1%                3                0
5%                17               0
10%               22               0      Obs                4853
25%               29               0      Sum of Wgt.         4853

50%               43               Mean                43.64517
                                Largest             Std. Dev.           18.75686

```

75%	56	96		
90%	70	98	Variance	351.8197
95%	77	99	Skewness	.233991
99%	88	100	Kurtosis	2.635955

Esta tabla resumen, notoriamente más completa que la anterior, entrega el resultado de los percentiles, los números más pequeños dentro de la variable, los más grande de la variable, las medidas que ya vimos en la tabla simple como numero de observaciones, media y desviación estándar, más la varianza, el grado de asimetría de la distribución y la curtosis (medidas que veremos más adelante). En este caso la varianza es de 351.82, es decir, una desviación estándar de casi 19 años.

TAREA (5 MINUTOS)

Para seguir familiarizándose con Stata, la tarea que tienen para la siguiente clase es replicar las clases “Modifying Data” y “Managing Data”, disponibles en la siguiente página web de UCLA: <http://www.ats.ucla.edu/stat/stata/notes/default.htm>. Está permitido colaborar con compañeros, pero cada uno tiene que entregar individualmente su tarea. La Tarea se entrega por email en formato Word. Deben cortar y pegar los contenidos desde Stata hacia Word. Para que no haya problemas de formato, utilizar márgenes estrechos, letra Courier New tamaño 10, interlineado simple y sin espacio entre párrafos. Si es necesario pueden cambiar la orientación de la página para que sea horizontal y no vertical. No se olviden de agregar su nombre a la tarea. En el asunto escriban “Tarea 2.”