Variable Aleatoria I

Clase sincrónica

{desafío} latam_

Aplicar herramientas de Python para calcular indicadores estadísticos y probabilidades.

- **Unidad 1: Estadística descriptiva y probabilidades**
  (Parte I)

  (Parte II)

- **Unidad 2: Variable aleatoria**
  (Parte I)

  (Parte II)

- **Unidad 3:Estadística inferencial**

- **Unidad 4: Regresión**
  (Parte I)

  (Parte II)

Te encuentras aquí

{desafío} latam_

¿Qué aprenderás en esta sesión?

- *El concepto de variable aleatoria, discreta y continua, que permite introducir los conceptos de distribución normal y binomial.*

{desafío}
latam_

¿Cómo visualizamos estadísticamente eventos y la probabilidad de ocurrencia de los mismos?

{desafío} latam_

/*Variable aleatoria*/

Variable aleatoria
Definición

Se llama **variable aleatoria** a una función que asigna a cada resultado de un experimento aleatorio un valor numérico, y su distribución de probabilidad está determinada por las probabilidades de los posibles valores que puede tomar.

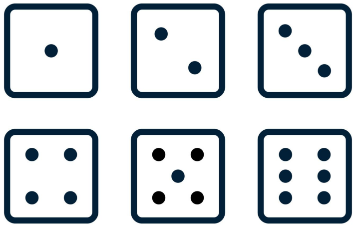
Es como una caja misteriosa que contiene diferentes valores posibles. ¡Cuando sacamos un valor de esa caja, no sabemos exactamente cuál será, pero sabemos que tiene una cierta probabilidad de ser uno u otro!

Variable aleatoria

Ejemplo

En el lanzamiento de un dado, si definimos la variable aleatoria X como el número obtenido al lanzar el dado, los posibles valores que puede tomar X son 1, 2, 3, 4, 5 y 6.

Cada valor tiene una probabilidad igual de 1/6, ya que el dado es justo. En este caso, X es una **variable aleatoria discreta**.

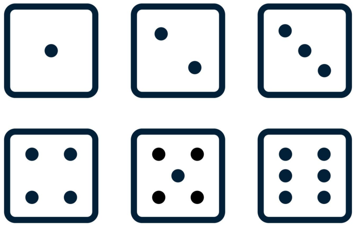


Variable aleatoria

Ejemplo

Pero, podemos fijarnos también en si el número obtenido es par o impar… en este caso, tenemos dos resultados posibles, cada uno con probabilidad ½.

Si lanzamos dos dados, ¿en qué resultados podemos fijarnos?



Variable aleatoria
Variable aleatoria continua

Un ejemplo de **variable aleatoria continua** es la duración de llamadas telefónicas. Si estamos estudiando las llamadas realizadas en un centro de atención al cliente y nos interesa la duración de cada llamada en minutos, la variable aleatoria continua sería el tiempo de duración de cada llamada.

La duración de una llamada puede tomar cualquier valor dentro de un rango continuo de números reales, desde fracciones de segundo hasta varios minutos. Por ejemplo, una llamada podría durar 1.5 minutos, 3.2 minutos, 4.8 minutos, etc.

Variable aleatoria
Lo esencial

Al observar un fenómeno, en realidad “elegimos” qué observar, o cuál va a ser el resultado que nos interesa. De un mismo conjunto de personas, por ejemplo, podemos fijarnos en su estatura, sexo, ingreso, nacionalidad, etc.

Cada una de estas características seguirá comportamientos diferentes, que son los que nos va a interesar observar y modelar.

{desafío} latam_

/* Función y distribución de probabilidad*/

Función de probabilidad
Ejemplo

Consideremos el lanzamiento de dos monedas, y contamos el número de caras obtenidas. ¿Cuáles son los resultados posibles, y sus probabilidades?

2 caras: cara - cara

1 cara: cara - sello   sello - cara

0 caras: sello - sello

Función de probabilidad
Ejemplo

Definimos entonces la variable aleatoria X: cantidad de caras obtenidas. Tenemos así:

$$P(X = x) = \begin{cases} 0,25 & \text{si } x = 0 \\ 0,5 & \text{si } x = 1 \\ 0,25 & \text{si } x = 2 \\ 0 & \text{en otro caso} \end{cases}$$

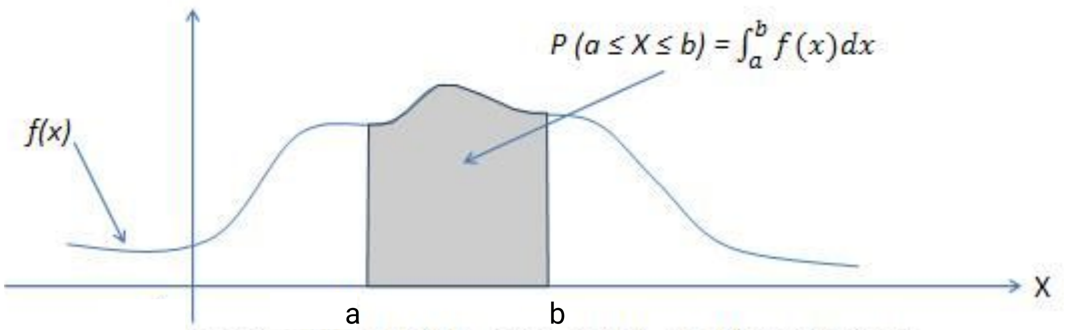
¿Qué podemos hacer si la variable aleatoria es continua?

Distribución de probabilidad

Variable Aleatoria Continua

Para una variable aleatoria continua, cada ocurrencia es un "punto", por lo que tiene dimensión cero. Esto hace que, en teoría, todo resultado tiene probabiidad cero.

En lugar de valores aislados, una variable aleatoria continua puede tomar cualquier valor dentro de un rango específico. Para definirlas usamos una función de probabilidad.



Muestra el área bajo la curva de *f(x)*, en el intervalo (a,b)

/* Distribución binomial*/

{desafío} latam_

Distribución binomial
Definición

Se llama **experimento de Bernouilli** a uno que solo tiene dos resultados posibles, éxito (al que asociamos valor 1) con probabilidad **p**, o fracaso (al que asociamos valor cero) con probabilidad **1 - p**.

Si repetimos este experimento n veces, ¿cuál es la probabilidad de obtener k éxitos? Podemos calcular esta probabilidad mediante la siguiente fórmula

$$P(k) = \frac{n!}{k!\,(n-k)!}p^k(1-p)^{n-k}$$

Distribución binomial
Ejemplo

Tenemos una caja con tres bolitas rojas y dos azules. Se extrae una y se anota su color. Si repetimos este experimento 8 veces, ¿cuál es la probabilidad de obtener 5 veces el color rojo?

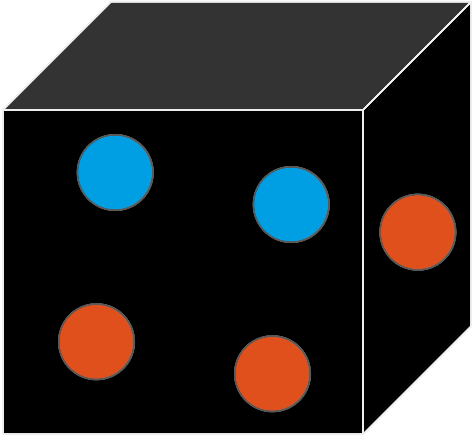
$$P = \frac{3}{5}$$

$$P(5) = \frac{8!}{5! \cdot (8-5)!} 0,6^5 \cdot 1 - 0,6^{8-5}$$

$$= \frac{6 \cdot 7 \cdot 8}{1 \cdot 2 \cdot 3} 0,6^5 \cdot 0,4^3$$

$$= 56 \cdot 0,07776 \cdot 0,064$$

$$0,27869184$$



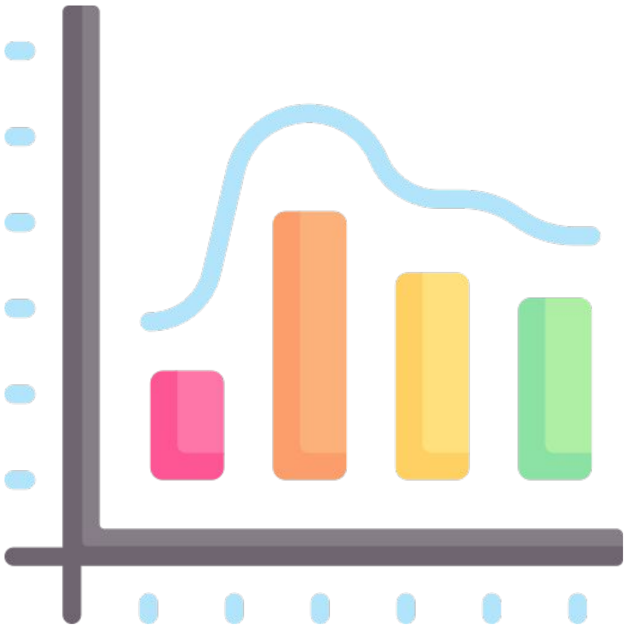
{desafío} latam_

/* Distribución Normal */

Distribución Normal
Definición

La distribución normal es un patrón o forma de distribución de datos que se asemeja a una campana simétrica. Es comúnmente observada en muchos fenómenos de la vida real, como la altura de las personas o las puntuaciones en exámenes.

En esta distribución, la mayoría de los valores se encuentran cerca del valor promedio, y a medida que nos alejamos de él, la frecuencia de los valores disminuye gradualmente. Es una distribución muy utilizada en estadística para modelar y analizar datos.



{desafío} latam_

Distribución Normal
Probabilidades

En la distribución normal, las probabilidades se relacionan con la probabilidad de que una variable aleatoria continua tome un valor dentro de un rango específico. La distribución normal es caracterizada por su función de densidad de probabilidad, que describe la forma de la curva y la probabilidad de que un valor particular ocurra.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Distribución Normal

¿Cómo la caracterizamos?

- Simetría
- Media, mediana y moda iguales
- Forma unimodal
- Tails o colas infinitas
- Parámetros de media y desviación estándar

{desafío} latam_

/* Ley de los grandes números */

Ley de los grandes números

Definición

La **Ley de los Grandes Números** es un principio fundamental en la teoría de la probabilidad y estadística que establece que, a medida que aumenta el tamaño de una muestra o el número de repeticiones de un experimento, la media o promedio de esos datos tiende a acercarse a la media teórica o esperada de la distribución de probabilidad.

En términos más sencillos, la Ley de los Grandes Números indica que, en el largo plazo, los resultados observados se acercarán cada vez más al resultado esperado o promedio.

Ley de los grandes números

Ley Débil y Ley Fuerte

Ley débil

la media muestral se acerca a la media poblacional cuando el tamaño de la muestra aumenta

Ley fuerte

la media muestral tiende, con probabilidad 1, a la media poblacional en la medida que el tamaño n de la muestra tiende a infinito

/* Teorema del límite central*/

{desafío}
latam_

Teorema del límite central

Definición

El **Teorema del Límite Central** es un resultado fundamental en la teoría de la probabilidad y estadística que establece las propiedades de la distribución de la suma o media de una gran cantidad de variables aleatorias independientes e idénticamente distribuidas.

Establece que, bajo ciertas condiciones, la distribución de la suma o media de un gran número de variables aleatorias tiende a aproximarse a una distribución normal, independientemente de la forma de distribución de las variables originales.

Teorema del límite central

Condiciones

1. **Independencia**

   Las variables aleatorias deben ser independientes entre sí.

2. **Identidad**

   Las variables aleatorias deben tener la misma distribución, es decir, la misma media y varianza.

3. **Muestra suficientemente grande**

   El tamaño de muestra (n) debe ser lo suficientemente grande, aunque no hay una regla específica y depende del contexto

Teorema del límite central

Interpretación y consecuencias

A medida que aumenta el tamaño de la muestra, la distribución de la suma o media se aproxima cada vez más a una distribución normal.

La distribución normal puede ser utilizada para estimar probabilidades o intervalos de confianza relacionados con la suma o media de una muestra grande.

Por ejemplo, permite realizar inferencias sobre la media de una población utilizando muestras grandes, ya que se puede confiar en que la distribución de las medias muestrales se aproxima a una distribución normal.

Desafío
Tabaquismo y gestación

{desafío} latam_

Desafío

"Tabaquismo y gestación"

- Descarga el archivo "Desafío".
- Tiempo de desarrollo asincrónico: desde 2 horas.
- Tipo de desafío: individual.

¡AHORA TE TOCA A TI! 💪

Ideas fuerza

A partir de un experimento u observación definimos ***variables aleatorias,*** *que dependen del aspecto que nos interesa analizar*

Las variables aleatorias pueden ser ***discretas o continuas,*** *presentando así distintas* ***distribuciones****.*

Las ***leyes de los grandes números*** *describen comportamientos de variables aleatorias al* ***aumentar*** *los tamaños de las* ***observaciones****.*

El ***teorema del límite central*** *nos permite modelar situaciones diversas a partir de la* ***distribución normal.***

¿Qué aspectos de la clase me parecen más relevantes?

{desafío} latam_

Recursos asincrónicos

¡No olvides revisarlos!

Para esta semana deberás revisar:

- Guía de estudio.
- Desafío “Tabaquismo y gestación”.

{desafío}
latam_

Próxima sesión...

- *Crea visualizaciones que permiten identificar la distribución de variables de un dataset*

{desafío} latam_

Academia de talentos digitales