Variable aleatoria

Tutoría - Parte I

{desafío} latam_

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

{desafío} latam_

Autoaprendizaje

Recursos asincrónicos

- ¿Revisaste los recursos de la semana 3 (Guía y desafío)?
- ¿Tienes dudas sobre alguno de ellos?

{desafío} latam_

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.***

/* Distribución binomial*/

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 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.

Distribución Normal

¿Cómo la caracterizamos?

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

Ejercicio:
La distribución normal en acción

{desafío} latam_

La distribución normal en acción

Distribución normal con Python

Veremos cómo la distribución normal nos puede ayudar a modelar algunas situaciones, para lo que utilizaremos Python, Puedes abrir tu propio archivo de Jupyter Notebook para replicar los pasos que te mostrará tu profesor, con los que aprenderemos:

1. Modelamiento con la distribución normal
2. Análisis de los parámetros de la distribución normal
3. Aplicación de la distribución normal

Distribución normal
Probabilidades y estandarización

Pese a su gran utilidad, la función de distribución normal tiene una gran dificultad: no tiene primitiva, por lo que la integral que permite calcular el área bajo la curva no puede determinarse algebraicamente.

Para resolver este problema, se debe recurrir a valores de tabla que se encuentran calculados para **valores estandarizados,** es decir, una función de media igual a cero y desviación estándar igual a 1.

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

Distribución normal

Probabilidades y estandarización

Así, por ejemplo, si la media de nuestro conjunto es μ, y su desviación estándar es σ, podemos **estandarizar** un valor cualquiera x de nuestro conjunto, restándole μ, y dividiendo el resultado por σ. Este valor estandarizado es el que puede ser buscado en una tabla.

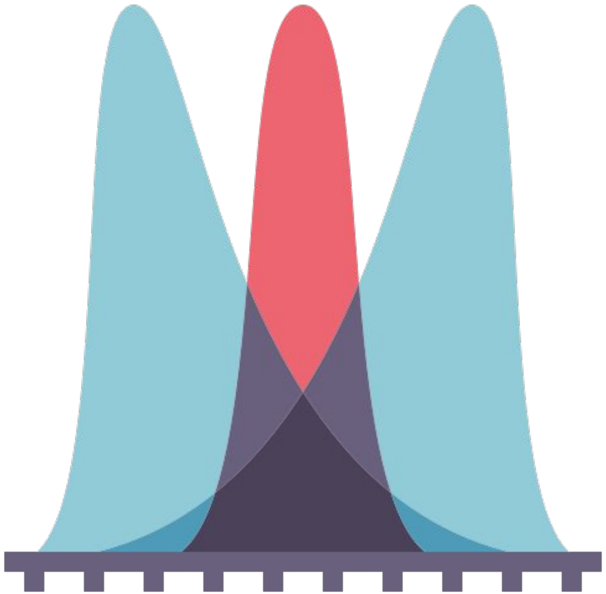
$$z = \frac{x - \mu}{\sigma}$$

Distribución normal

Probabilidades y estandarización

Un aspecto muy importante de la distribución normal es que nos permite establecer cuál es la probabilidad de que un valor dado se encuentre a una distancia dada de la media, en términos de la desviación estándar. Así, sabemos que si un conjunto de datos se distribuye en forma normal, entonces:

- Aproximadamente, un 68% de los datos se encuentran entre **x - µ** y **x + µ**
- Aproximadamente, un 95% de los datos se encuentran entre **x - 2µ** y **x + 2µ**
- Aproximadamente, un 99,7% de los datos se encuentran entre **x - 3µ** y **x + 3µ**



/* Ley de los grandes números */

{desafío}
latam_

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*/

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.

Ejercicio
Verifiquemos con Python

{desafío}
latam_

Ley de los grandes números y límite central

Verificando con Python

Verificaremos con Python los resultados anteriores, para lo que deberás seguir la presentación que te mostrará tu profesor. Puedes abrir tu propio archivo de Jupyter Notebook para replicar los pasos, con los que observaremos:

1. Una aplicación de la ley débil de los grandes números
2. Una aplicación de la ley fuerte de los grandes números
3. Una aplicación del teorema del límite central

Distribución binomial y ley de los grandes números

Esperanza y desviación teórica

Se llama **esperanza** al “promedio teórico” en un experimento. Para una variable aleatoria X con distribución binomial tenemos que

$$E(X) = n \cdot p$$

$$\sigma(X) = \sqrt{n \cdot p \cdot (1 - p)}$$

Desafío

“Estadística descriptiva y probabilidades (parte II)”

- ¿Leíste el desafío de esta semana? ¿Comprendes bien lo que se solicita en cada caso?
- ¿Hay contenidos que necesitas repasar antes de comenzar este desafío?
- ¿Necesitas algún ejemplo o indicación para alguna pregunta o requerimiento específico?

Próxima sesión...

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

{desafío}
latam_

{desafío}
latam_

Academia de
talentos digitales