

Estadística para negocios

Modulo 2: Fundamentos de cálculo de probabilidades de eventos

Dr. José Ignacio Hernández

Semana del 27 de mayo de 2024

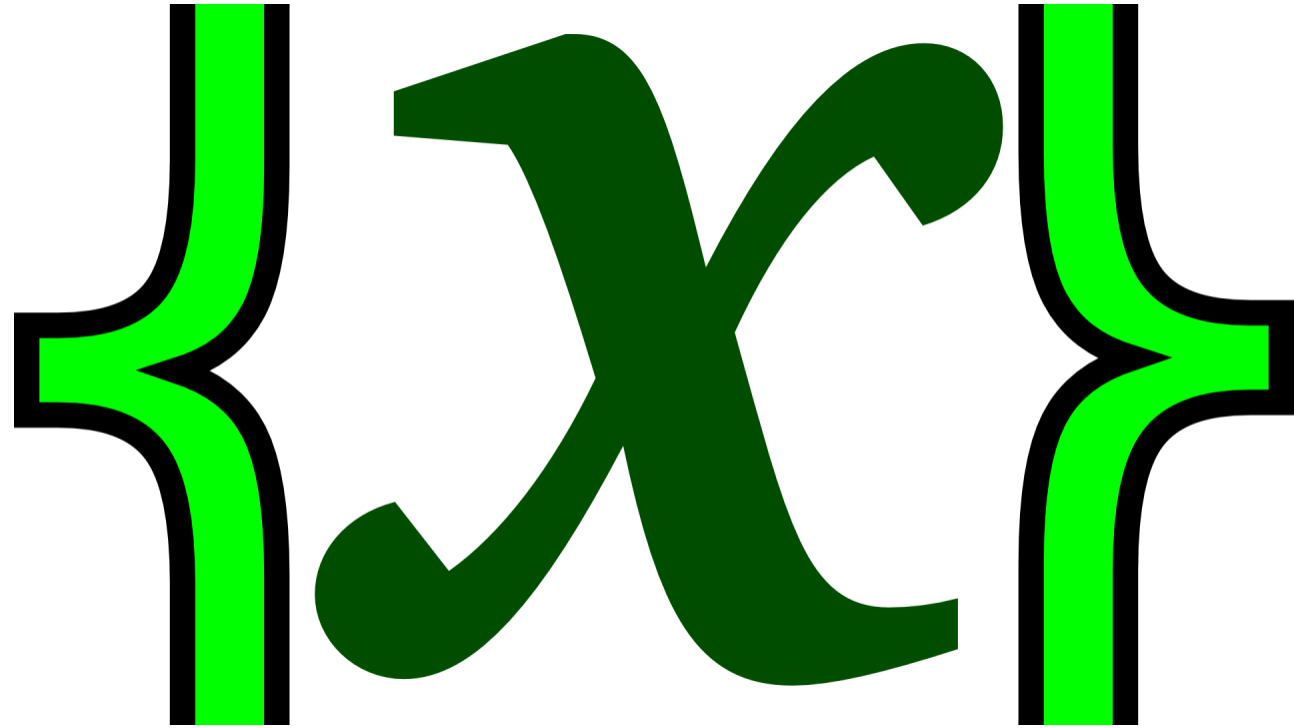
Resumen de la clase anterior

- En la clase anterior, vimos distintas formas de analizar datos de forma descriptiva
 - Frecuencias
 - Medidas de tendencia central, posición, dispersión
 - Correlación entre variables.
- Además, estudiamos cómo estas medidas pueden proveer de información a la empresa a través de un caso aplicado.
- **Sin embargo, lo que hemos visto solo describe a una parte de la población: la muestra con que contamos.**
- **Es posible hacer inferencia de la población usando la muestra. Para ello, requerimos el uso de la teoría de la probabilidad, lo cual veremos en este módulo**

Modulo II: Fundamentos de cálculo de probabilidades de eventos

Variables

- Una **variable** es una magnitud que puede tomar distintos valores en distintos puntos de observación
- La estadística permite describir variables en base a una estructura.



Universo, población y muestra

Universo: Conjunto de todos los posibles resultados de un experimento.

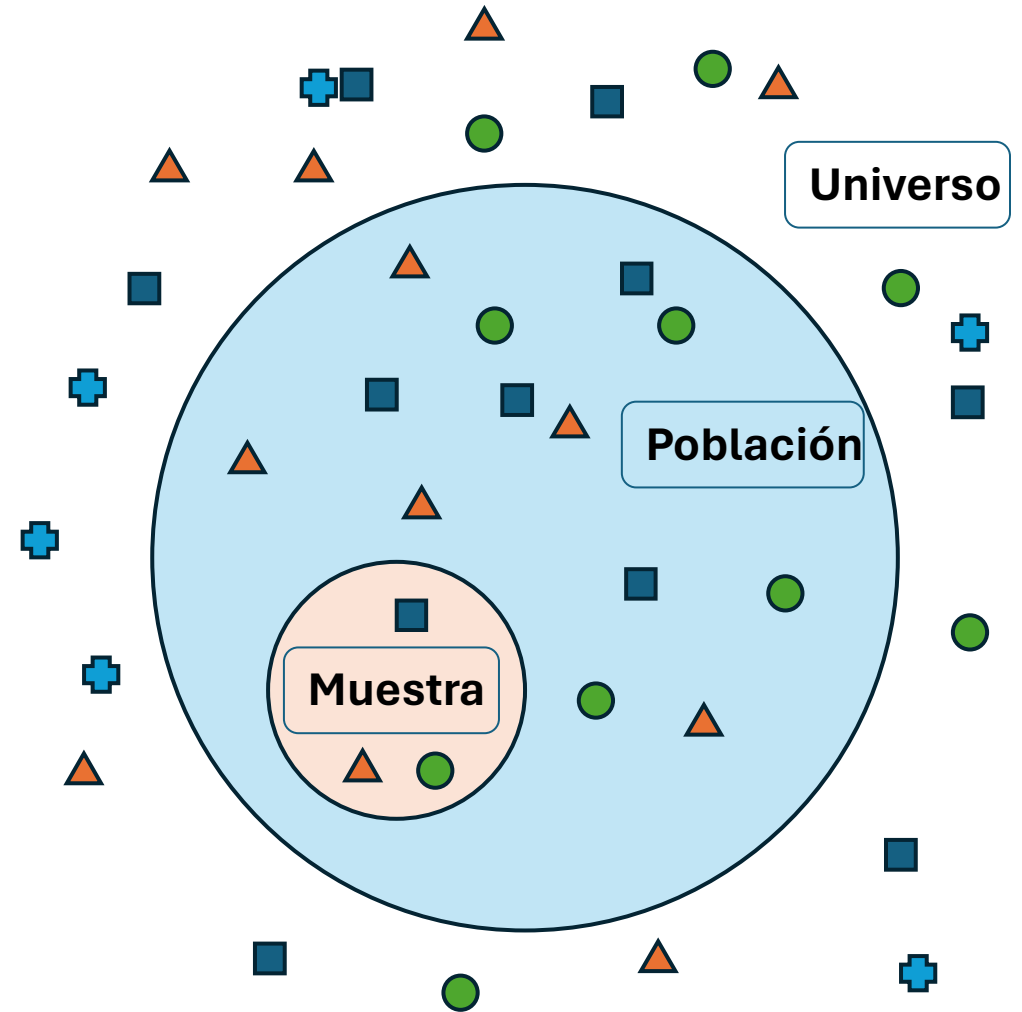
- Ejemplo: Todos los individuos habitantes de Chile

Población: Subconjunto del universo que engloba los resultados relevantes para el experimento.

- Ejemplo: Chilenos, mayores de 18 años.

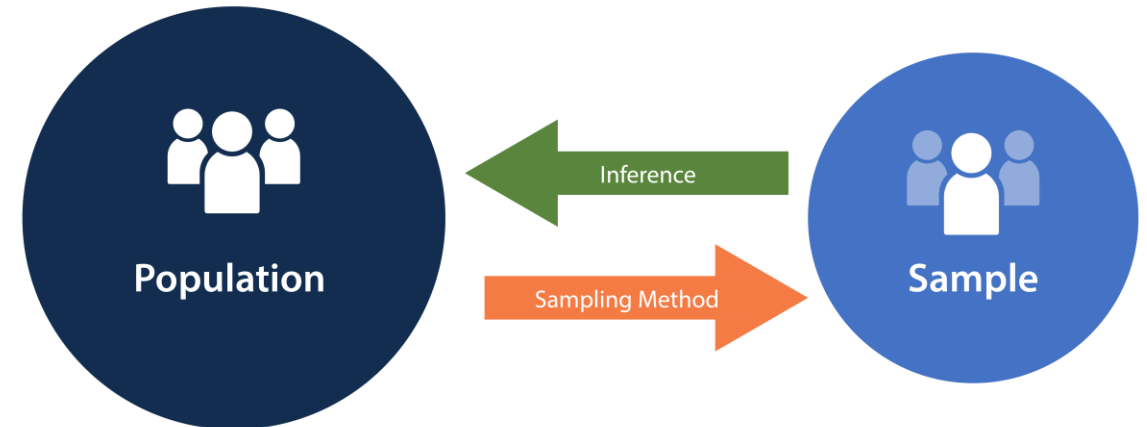
Muestra: Individuos o sucesos seleccionados de una población.

- Ejemplo: 1.000 chilenos, mayores de 18 años, tomados aleatoriamente.



Técnicas de conteo

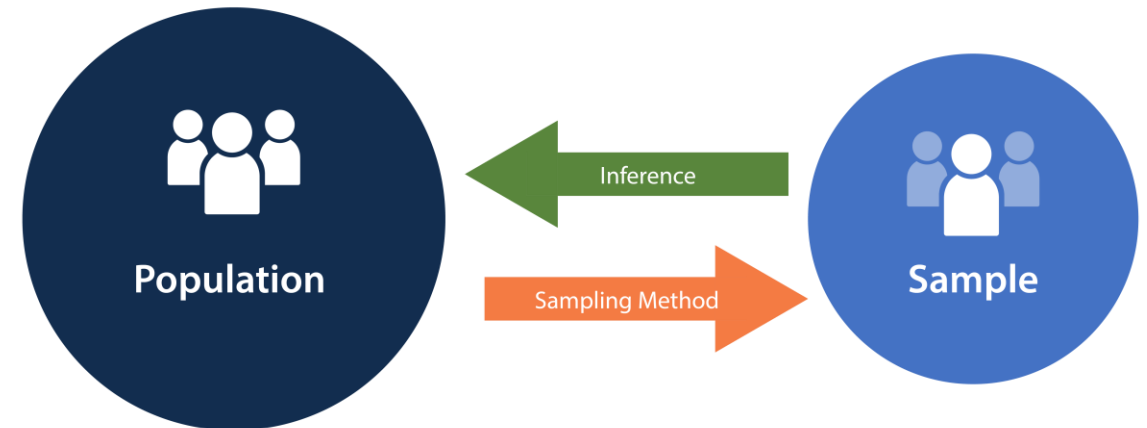
- Muestreo aleatorio: Tomar individuos de la muestra de forma aleatoria.
 - Ventajas: Simplicidad.
 - Limitaciones: Representatividad.
- Muestreo probabilístico: Tomar una muestra en base a probabilidades conocidas para cada individuo.
 - Ventajas: Representatividad
 - Limitaciones: Las probabilidades de población no son conocidas en la realidad



Técnicas de conteo

Muestreo estratificado: Es una versión moderada de los muestreos probabilísticos.

- Principio: La población se divide en grupos en base a variables (ej. Género, edad, etc.). Luego, se muestrea en cada grupo.



Técnicas de conteo

Manos a la obra: Realizaremos muestreos en nuestra base de datos.

- **Problema:** La dirección de la empresa requiere:
 - Una muestra aleatoria del 10% de los individuos.
 - Una muestra aleatoria de 1 persona por cada tramo de edad.
- **Posible estrategia:** Utilizar muestreos
- **Herramientas:** Base de datos + Python

¿Preguntas?

Probabilidades

Probabilidades

Antes, algunas definiciones:

- Experimento aleatorio: Acción en la cual no existe un resultado determinado.
 - Ejemplo: Lanzar una moneda, contratar un trabajador cuya productividad es conocida, etc.
- Evento: Corresponde a uno de los resultados



Variables aleatorias discretas y continuas

Variable aleatoria discreta:

- Se compone de un numero finito de numeros reales
- Ejemplos:
 - Numero de visitas al doctor
 - Genero
 - Numero de hijos

• Variable aleatoria continua:

- Es un intervalo o la recta real completa
- Ejemplos:
 - PIB, consumo
 - Salarios
 - Temperatura del día



Probabilidad

Si el experimento se compone de resultados mutuamente excluyentes, la probabilidad de un resultado es igual a:

$$\frac{1}{N}$$

Ejemplo: Si hay **10** bolitas iguales pero de distintos colores en una caja, la probabilidad de sacar una bolita de un color determinado es **$1/10 = 0,1$**

Si el experimento se repite **N** veces, y se observa que un evento **A** ocurre “**a**” veces, la frecuencia relativa del evento es:

$$\frac{a}{N}$$

La **probabilidad** de que ocurra el evento **A** es el limite de la frecuencia relativa, cuando **N** tiende a infinito.

En otras palabras, si el experimento se repite infinitas veces, la frecuencia relativa tiende a ser igual a la probabilidad de ocurrencia del evento

Probabilidad

- Uno de los teoremas básicos de la probabilidad es la **ley de los grandes números**
- A medida que un experimento aleatorio se repite, la probabilidad empírica de la suma de eventos tiende a su probabilidad real (o teórica).

¿Por qué esto es importante?

- En la realidad, solo contamos con una parte de la población (la muestra), pero queremos decir algo sobre la población completa.
- Las probabilidades nos permiten decir “algo” sobre la población completa, con una muestra.



Probabilidad

Manos a la obra: usando una hoja de cálculo, generaremos números aleatorios entre 0 y 1

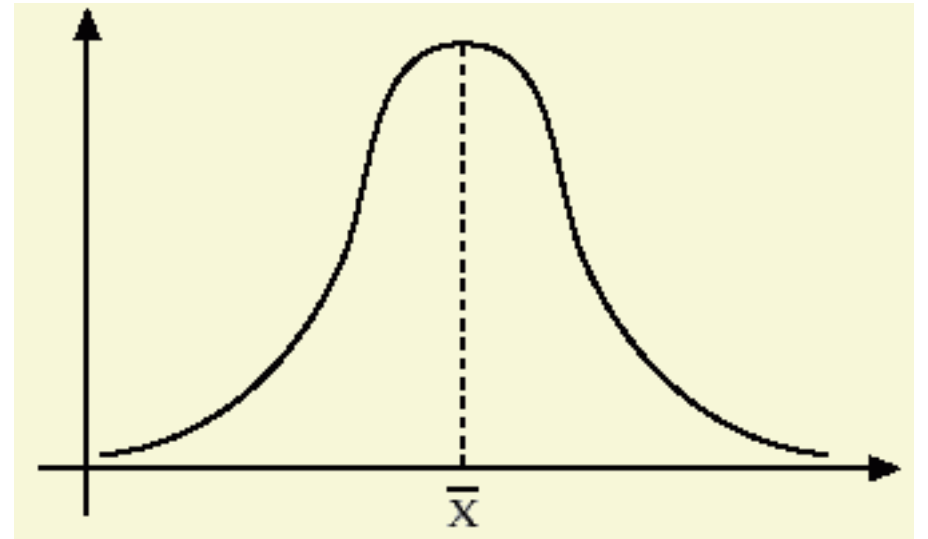
- Si el número es igual o está por sobre 0,5, se redondea a 1.
- Si el número está bajo 0,5, se redondea a 0.
- Generaremos:
 - 1 número aleatorio
 - 10, 100, 1000 números aleatorios
- Luego, calcularemos su media en cada caso. ¿A qué número se aproxima a medida que aumenta la muestra?

¿Preguntas?

Distribuciones de probabilidad

Una distribución de probabilidad es una abstracción teórica de una distribución de frecuencias

- Las distribuciones de probabilidad permiten caracterizar eventos probabilísticos a través de la matemática.
- Un ejemplo es la distribución normal (ver imagen)



Distribuciones de probabilidad

Propiedades:

- La probabilidad de un suceso solo puede tomar valores entre cero y uno

$$0 \leq f(x) \leq 1$$

- La probabilidad de un suceso no posible es cero
- La suma de las probabilidad de todos los sucesos posibles, mutuamente excluyentes, es igual a uno:

$$\sum_{i=1}^N f(x) = 1$$

Distribuciones de probabilidad

Definición

- Una distribución de probabilidad es una función matemática que asigna a cada evento posible, una probabilidad:

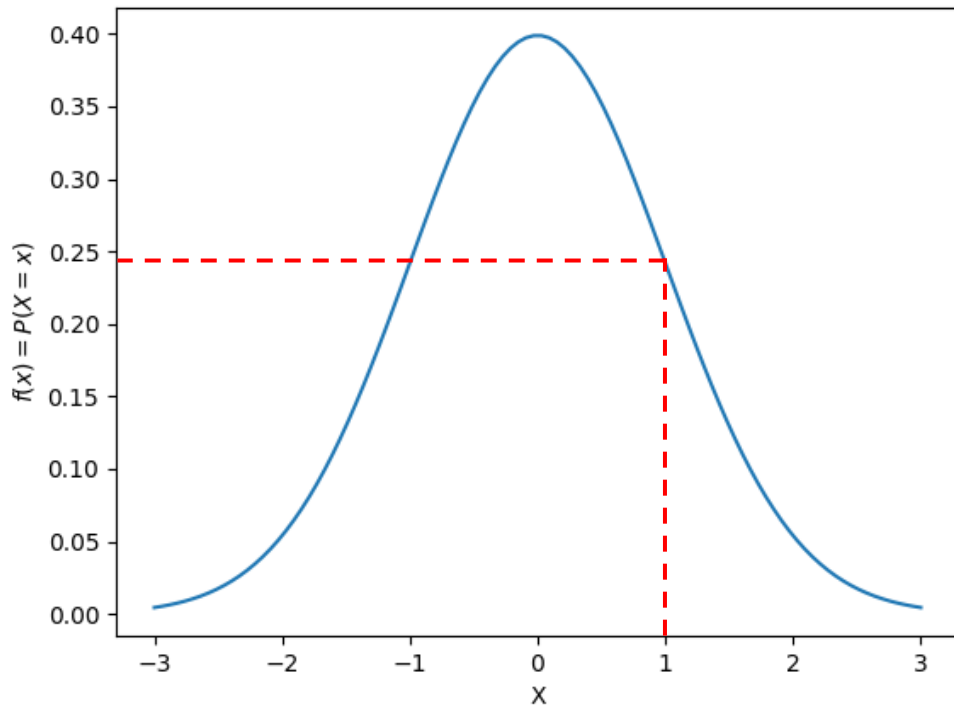
$$f: x \rightarrow f(x) = P(X = x)$$

Evento



Probabilidad de que la
variable aleatoria X sea igual
a x

Distribuciones de probabilidad



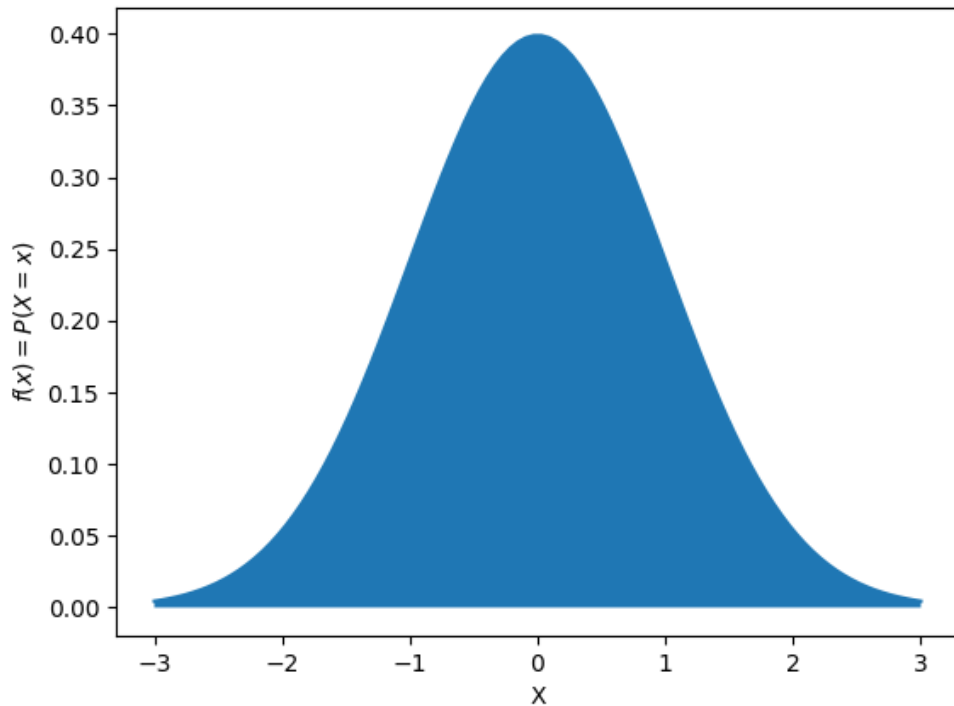
Tomemos un rato la distribución normal (ya hablaremos más adelante en detalle sobre ella).

- La función de distribución muestra la probabilidad de tomar un valor determinado. Por ejemplo:

$$f(x = 1) \approx 0.242$$

Además, la suma de todas las probabilidades, evaluadas en todos los puntos debe ser siempre igual a 1.

Distribuciones de probabilidad



Tomemos un rato la distribución normal (ya hablaremos más adelante en detalle sobre ella).

- La función de distribución muestra la probabilidad de tomar un valor determinado. Por ejemplo:

$$f(x = 1) \approx 0.242$$

La suma de todas las probabilidades, evaluadas en todos los puntos debe ser siempre igual a 1.

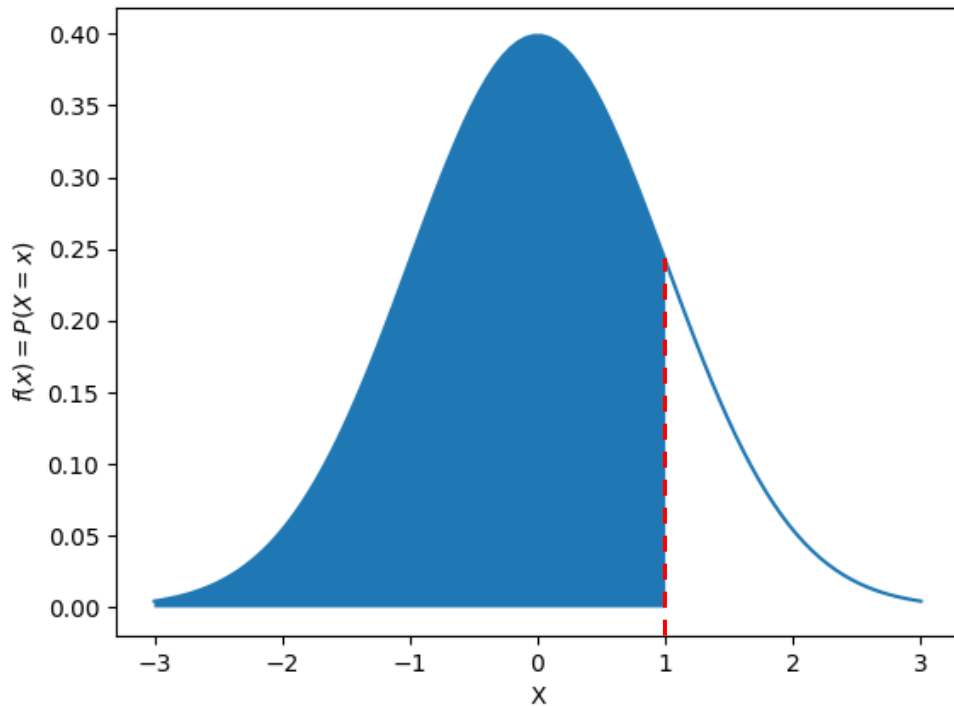
Distribuciones de probabilidad

Definición

- Además de la función de distribución, se define la función de probabilidad acumulada, que describe la probabilidad de que el valor de un evento sea menor a un determinado valor:

$$F: x \rightarrow F(x) = P(X \leq x)$$

Distribuciones de probabilidad

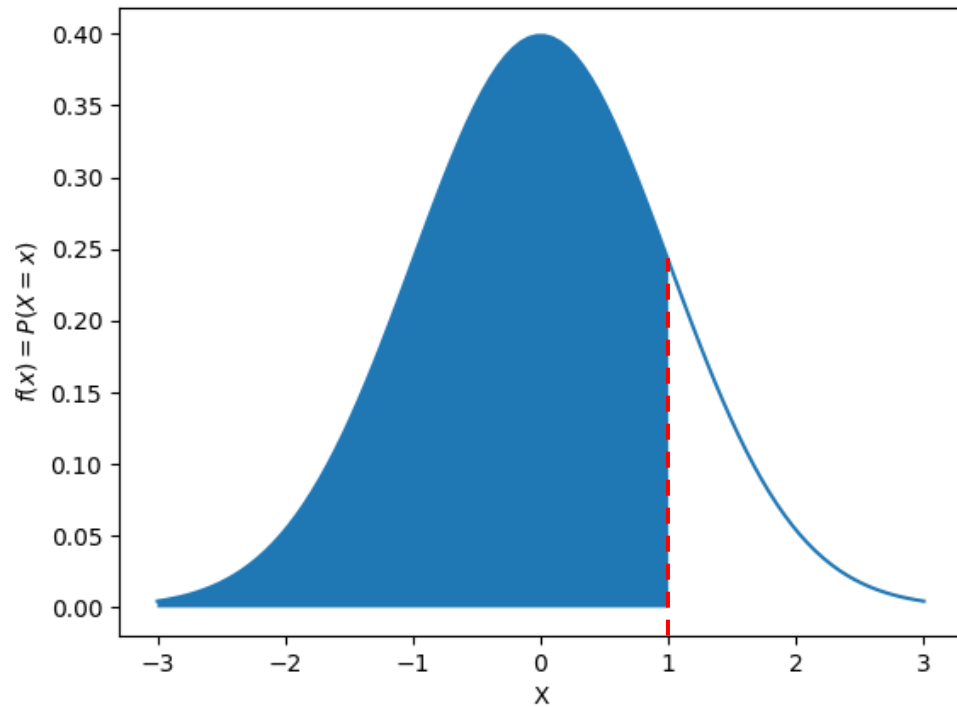


Volvamos a nuestro ejemplo de la distribución normal

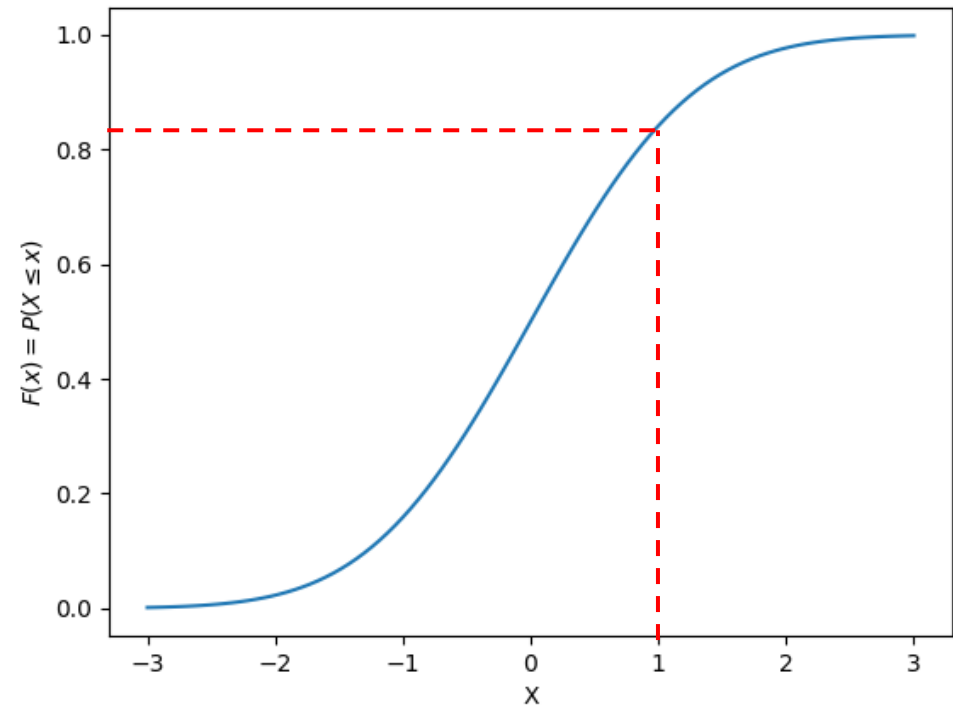
- La función de probabilidad acumulada es la probabilidad de que una variable sea menor que un valor determinado. Por ejemplo:

$$F(x = 1) = f(x \leq 1) \approx 0.841$$

Distribuciones de probabilidad



Función de distribución



Función de distribución acumulada

Distribuciones de probabilidad

Tipos de distribución según su dimensión:

- Distribución unidimensional: Describe la probabilidad de ocurrencia de un solo evento:
 - Ejemplo: Probabilidad de contratar a una persona productiva
- Distribución bidimensional: Describe la probabilidad de ocurrencia de dos eventos:
 - Ejemplo: Probabilidad de contratar una persona productiva y que sea activa físicamente.

Distribuciones de probabilidad

Tipos de distribución según la naturaleza de los datos:

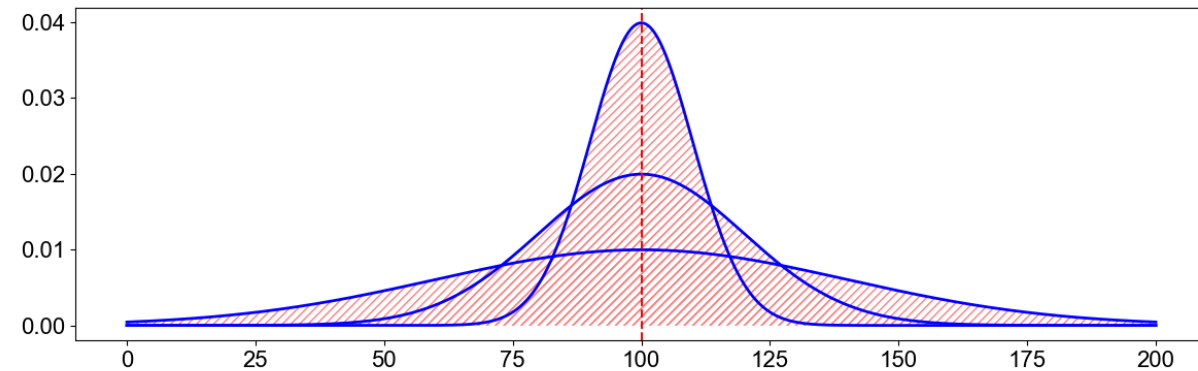
- Distribución discreta: Los eventos son “contables” (ej. números enteros)
 - Ejemplo: Probabilidad de cerrar dos días en el periodo por fuerza mayor.
- Distribución continua: Los eventos son un número de la recta real.
 - Ejemplo: Probabilidad de que las ventas sean mayores a un monto determinado.

¿Preguntas?

Momentos de la distribución

Los momentos permiten describir y explicar diferencias en una distribución

- Los momentos son un equivalente a la estadística descriptiva (media, varianza, etc.).
- En el caso de las distribuciones, estas descripciones son una función matemática.
- Veremos dos momentos en particular:
 - Esperanza o valor esperado
 - Varianza



Momentos de la distribución

Esperanza o valor esperado: Corresponde al valor promedio de una distribución de probabilidad.

- Para distribuciones discretas:

$$E(X) = \sum_{i=1}^N x_i \cdot f(x_i)$$

Probabilidad
de ocurrencia

- Para distribuciones continuas:

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$$



Momentos de la distribución

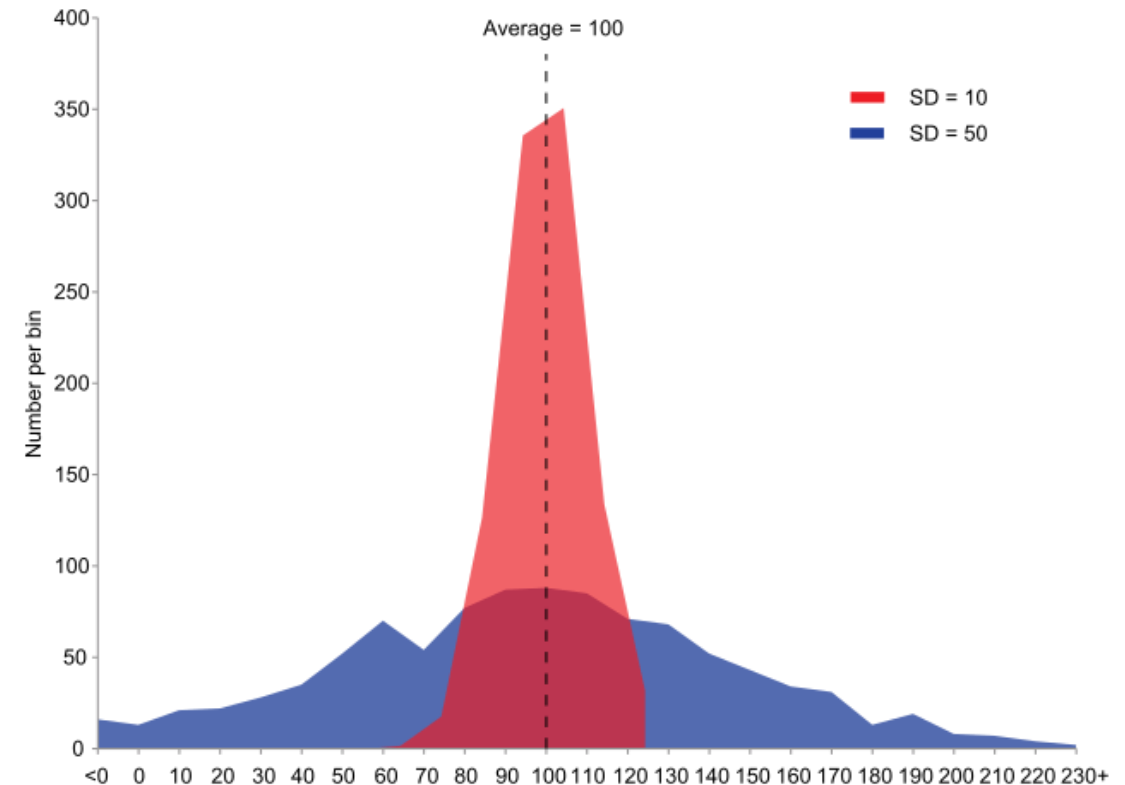
Varianza: Corresponde a la dispersión promedio de los datos de una distribución, con respecto a la media.

- Para distribuciones discretas:

$$E(X) = \sum_{i=1}^N (x_i - E(X))^2 \cdot f(x_i)$$

- Para distribuciones continuas:

$$E(X) = \int_{-\infty}^{\infty} (x - E(X))^2 \cdot f(x) dx$$



¿Preguntas?

Distribuciones clásicas de probabilidad

Distribuciones clásicas de probabilidad

Distribución normal

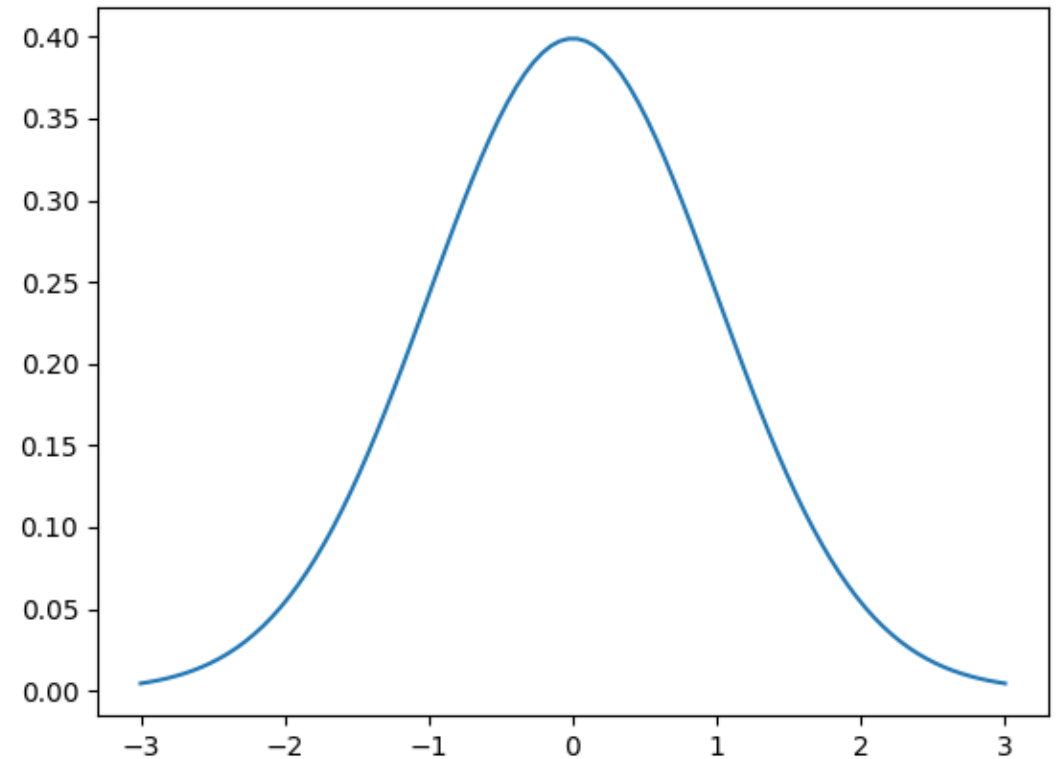
- Es una de las distribuciones continuas más conocidas. Presente frecuentemente en fenómenos de la naturaleza.
- Para describir una variable aleatoria \mathbf{X} con distribución normal, se usa la notación:

$$X \sim N(\mu, \sigma^2)$$

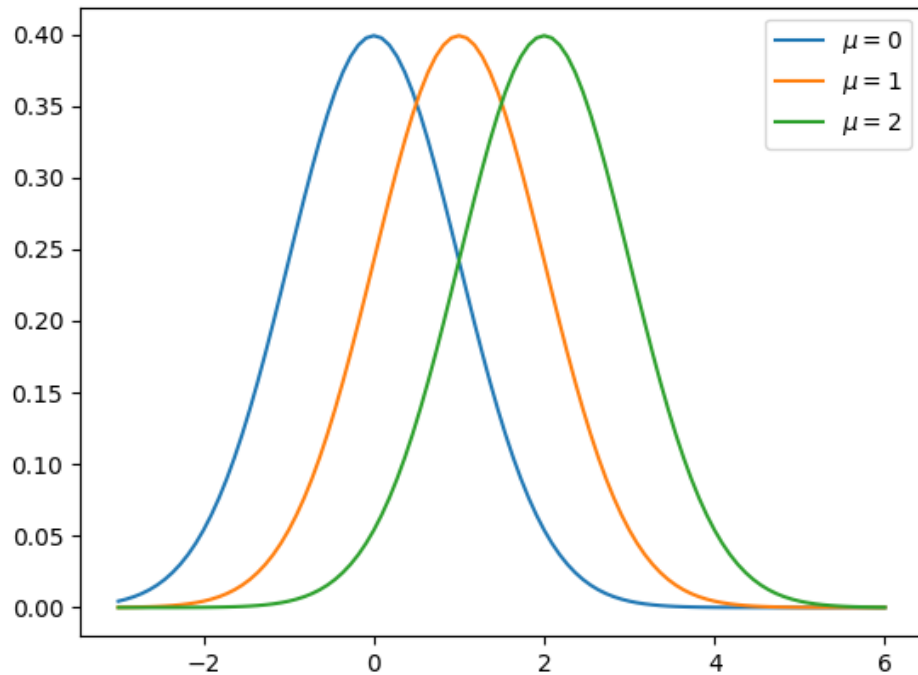
Diagram illustrating the notation $X \sim N(\mu, \sigma^2)$. The parameter μ is labeled "Media" (Mean) and the parameter σ^2 is labeled "Varianza" (Variance).

- Su función de densidad es:

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



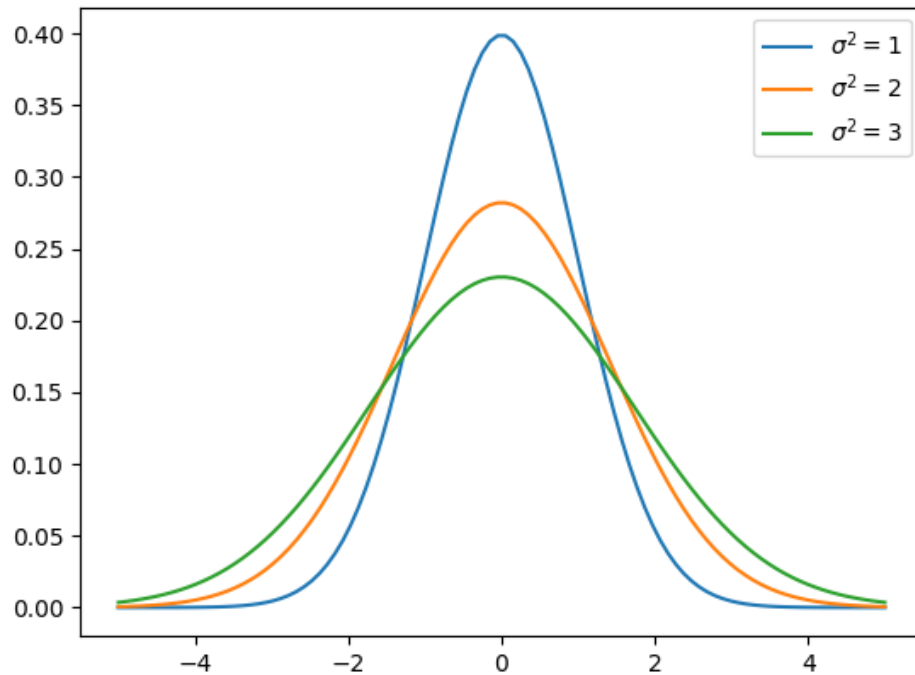
Distribuciones clásicas de probabilidad



Los parámetros de la distribución permiten modificar su comportamiento.

En el caso de la distribución normal, al cambiar el parámetro de la media (μ), la distribución normal se “mueve” en el eje, manteniendo la anchura de su campana.

Distribuciones clásicas de probabilidad



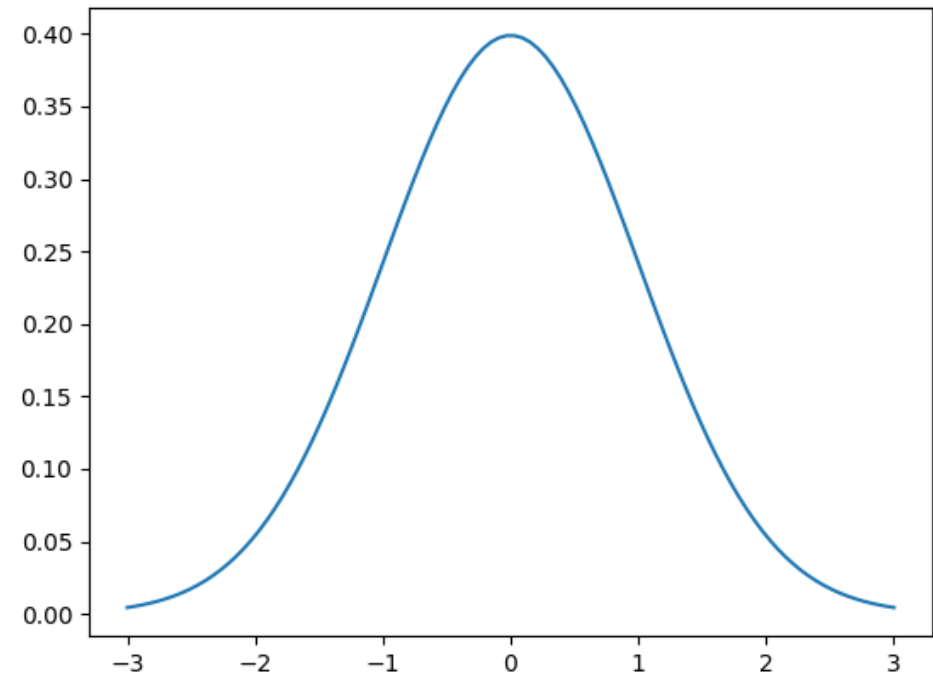
Los parámetros de la distribución permiten modificar su comportamiento.

Cuando cambia la varianza (σ^2), es la “anchura” de la distribución la que cambia, sin cambiar el punto donde está concentrada (es decir, la media).

Distribuciones clásicas de probabilidad

¿Por qué la distribución normal es tan relevante?

1. Teorema del límite central: postula que, una variable aleatoria tiende a una distribución normal a medida que la muestra tiende a un número muy grande.
2. ... (lo veremos en un momento)



Distribuciones clásicas de probabilidad

Manos a la obra: Veremos visualmente si nuestros datos tienen un comportamiento de una distribución normal

- **Problema:** La dirección de la empresa quiere conocer:
 - Si la edad de las personas tiene un comportamiento similar a una distribución normal
- **Posible estrategia:** Usar gráficas y un histograma
- **Herramientas:** Base de datos + Python

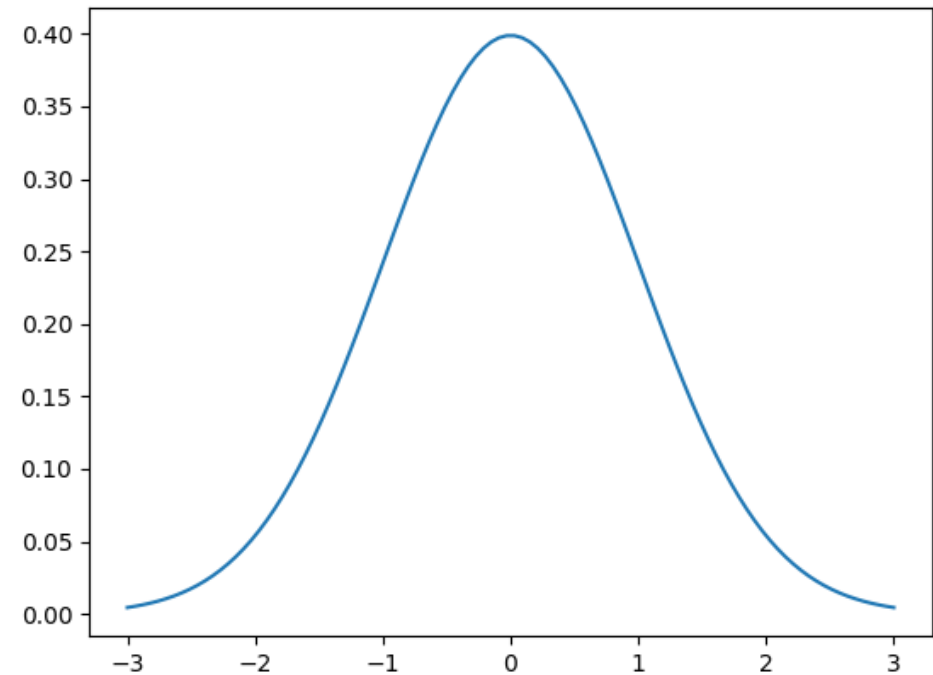
Distribuciones clásicas de probabilidad

¿Por qué la distribución normal es tan relevante?

2. Distribución normal estándar:
Es un caso particular de la distribución normal, con media igual a cero, y varianza igual a 1:

$$Z \sim N(0,1)$$

¿Por qué es relevante esta distribución?



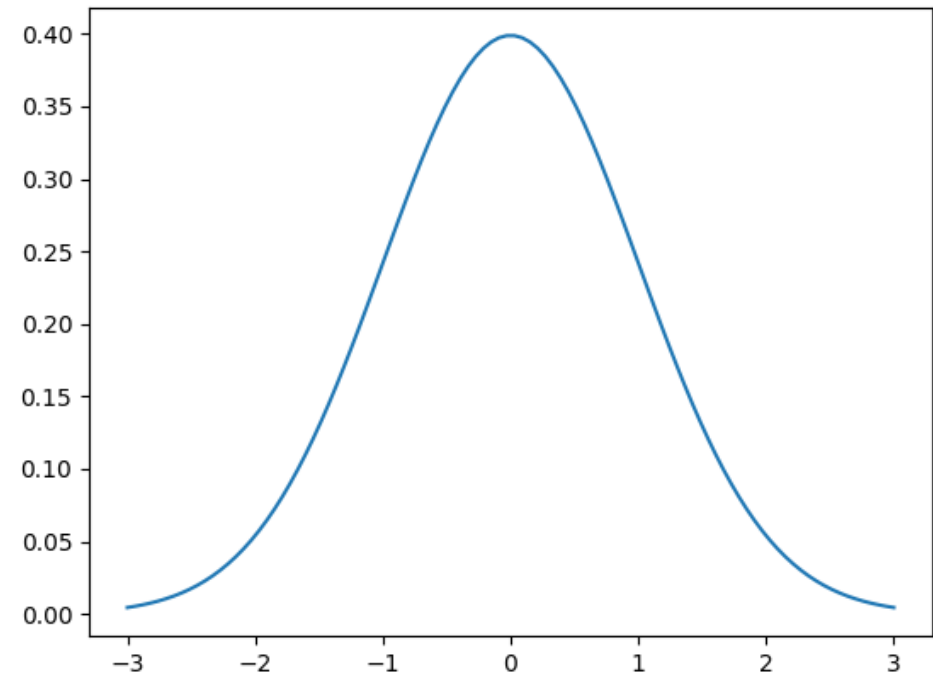
Distribuciones clásicas de probabilidad

¿Por qué es relevante esta distribución?

Estandarización: Cualquier variable con distribución normal puede ser estandarizada, calculando:

$$Z_i = \frac{x_i - \mu}{\sqrt{\sigma^2}}$$

Relevancia: Permite traer distintas distribuciones a una escala común.



Distribuciones clásicas de probabilidad

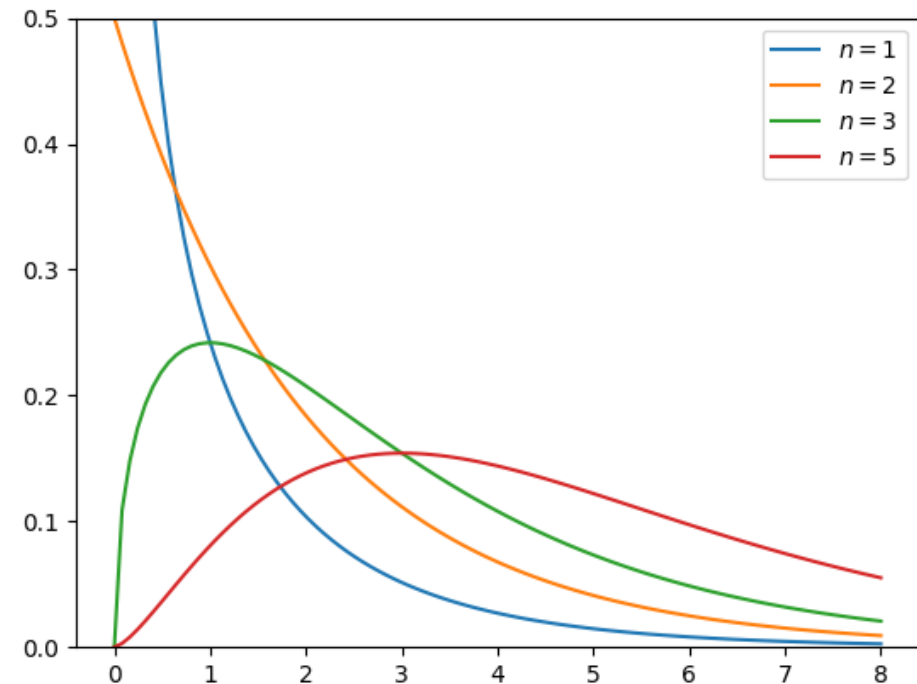
Distribución Chi-cuadrada (χ^2)

- Se construye a partir de la suma de variables con una distribución normal estándar elevadas al cuadrado:

$$X = \sum_{i=1}^n Z_i^2 \sim \chi^2(n)$$

- La distribución depende del **número de grados de libertad**

Uso: Pruebas de hipótesis de proporciones (lo veremos en los próximos módulos)



Distribuciones clásicas de probabilidad

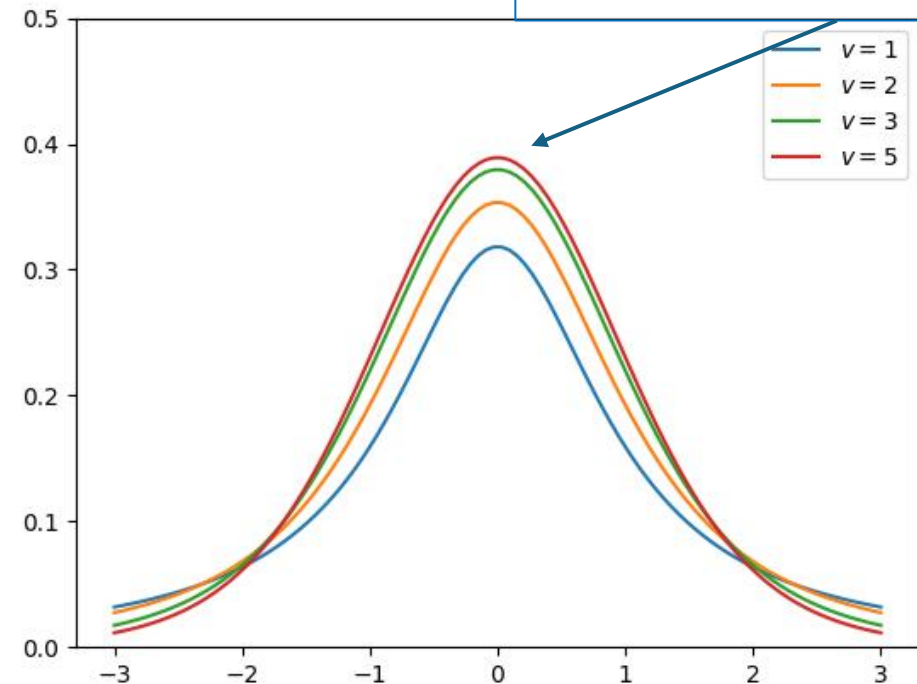
Distribución T de Student (T)

- Se construye a partir de una variable X_1 con distribución normal estándar, y una variable X_2 con distribución chi-cuadrada con ν grados de libertad:

$$T = \frac{X_1}{\sqrt{\frac{X_2}{\nu}}}$$

Uso: Pruebas de hipótesis de media (lo veremos en los próximos módulos)

Dato relevante: a medida que los grados de libertad de la distribución T aumentan, esta converge a una distribución normal estándar



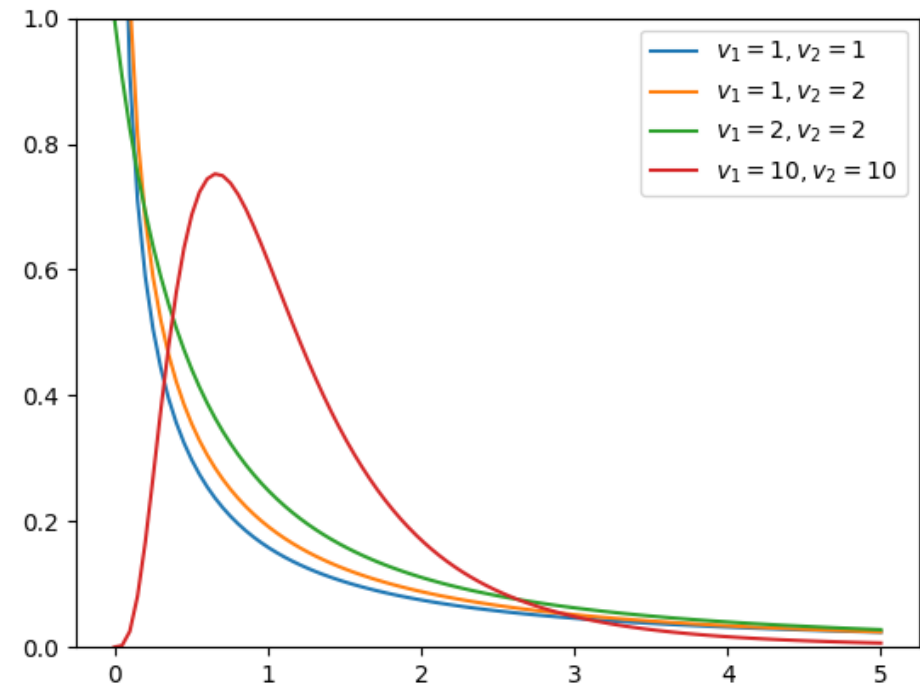
Distribuciones clásicas de probabilidad

Distribución F de Fisher:

- Se construye a partir de dos variables X_1 y X_2 con distribución chi-cuadrada, con grados de libertad v_1 y v_2 , respectivamente:

$$F = \frac{X_1/v_1}{X_2/v_2}$$

Uso: Pruebas de hipótesis sobre múltiples estadísticos.



Distribuciones clásicas de probabilidad

Manos a la obra: Graficaremos distintas distribuciones de probabilidad en Python.

- **Problema:** La dirección de la empresa quiere conocer:
 - Si como analistas estamos en condiciones de realizar pruebas de hipótesis en una sesión futura
- **Posible estrategia:** Usar programación
- **Herramientas:** Python

¿Preguntas?

Para la sesión PM de hoy

- **Realizaremos un caso aplicado:** Formarán grupos de 4-5 personas, y deberán resolver el caso aplicado provisto en el material.
- **Cada grupo debe tener, al menos:**
 - Coordinador: Se encarga de hacer llegar el reporte al profesor.
 - Redactor: Se encarga de escribir el reporte.
- **El reporte se entrega a las 17:20 del día de hoy.**

¡Gracias!