HENRY

Distribución de la Probabilidad

clase 4

Data Science

Agenda

- ➜ Distribuciones de probabilidad
- ➜ Variables discretas
- ➜ Variables continuas

OBJETIVOS DE LA CLASE

Al finalizar esta lecture estarás en la capacidad de…

- **Entender** el concepto de distribución de probabilidad y sus características.

Al **finalizar** cada uno de los temas, tendremos un **espacio de consultas**.

Hay un **mentor** asignado para responder el **Q&A**.

¡Pregunta, pregunta, pregunta! :D

Distribuciones de Probabilidad

Cuando a todos los posibles **valores numéricos** de una variable aleatoria se le asignan valores de probabilidad, ya sea mediante un listado o una **función matemática** el resultado es una distribución de probabilidad. La suma de las probabilidades de todos los resultados numéricos posibles debe ser igual a 1.

Una **función de probabilidad** puede asignar valores de probabilidad a cada estado del espacio muestral.

Características

Las **características generales** de las distribuciones de probabilidad **difieren según el tipo de variable** aleatoria, discreta o continua, que se encuentre bajo estudio.

Si la **variable aleatoria** es continua, no pueden listarse todos los posibles valores de la variable, motivo por el cual las probabilidades que se determinan por medio de una función matemática son gráficamente representadas por una función de densidad de probabilidad, o curva de probabilidad.

Variables discretas

- Puede tomar solamente algunos valores dentro de un intervalo definido.
- Las probabilidades se representan con los símbolos pi o p(xi).
- El gráfico de la distribución de probabilidad se denomina gráfico de bastones.

- La condición de cierre se verifica realizando la sumatoria de las **probabilidades p = 1.**

- La **distribución de probabilidad** en el caso de una variable aleatoria discreta se denomina genéricamente función de probabilidad.

- Las **probabilidades** se calculan mediante la aplicación de las reglas provenientes de la **teoría clásica** de probabilidad así como de fórmulas específicas.

Distribuciones

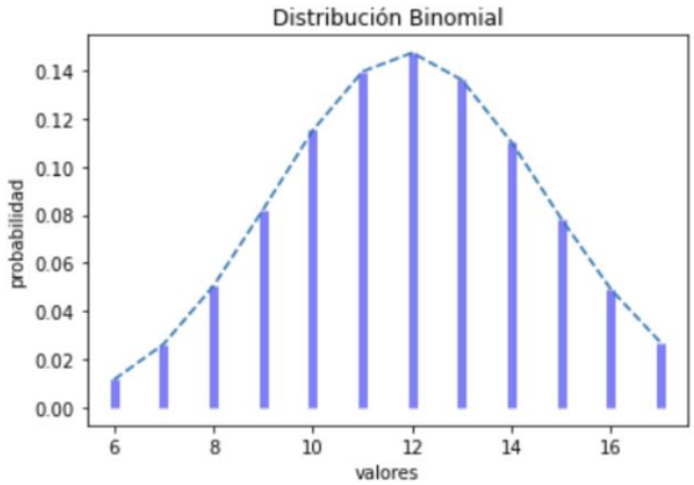
Distribución Binomial

Se utilizan para **variables del tipo binario** (lanzar una moneda) para describir el resultado del experimento.

p= probabilidad de éxito.
q= Probabilidad de fracaso.
n= espaci
k= número de éxitos.

$$P_{(k)} = \frac{n!}{k!\,(n-k)!} p^k q^{(n-k)}$$

```python
import matplotlib.pyplot as plt
import numpy as np
from scipy import stats
import seaborn as sns
# Graficando Binomial
N, p = 30, 0.4 # parametros de forma
binomial = stats.binom(N, p) # Distribución
x = np.arange(binomial.ppf(0.01),
              binomial.ppf(0.99))
fmp = binomial.pmf(x) # Función de Masa de Probabilidad
plt.plot(x, fmp, '--') #Esta función recibe un conjunto de valores x e y y los muestra en el plano como puntos unidos por línea.
plt.vlines(x, 0, fmp, colors='b', lw=5, alpha=0.5) #Esta función da formato a las figuras.
plt.title('Distribución Binomial') #Esta función asigna un título.
plt.ylabel('probabilidad') #Esta función etiqueta el eje Y.
plt.xlabel('valores') #Esta función etiqueta el eje X.
plt.show() #Esta función muestra las figuras
```



Ejemplo

Una novela ha tenido un gran éxito, y se estima que el 80% de un grupo de lectores ya la han leído.
En un grupo de 4 amigos aficionados a la lectura:

¿Cuál es la probabilidad de que en el grupo hayan leído la novela 2 personas?

Caso binario :

Haber leído la novela - No haber leído la novela.

```python
# Construir una función Binomial
from math import factorial

def funcion_binomial(k,n,p):
  num_exitos = factorial(n) #Factorial del espacio muestral.
  num_eventos = factorial (k) * factorial(n-k) #Factorial de la cantidad de casos de éxito buscados.
  exitos_fracaso=pow(p,k) * pow(1-p,(n-k)) # Probabilidad de exitos y fracasos.

  binomial = (num_exitos / num_eventos) * exitos_fracaso #Aplicación de la función binomial.

  return binomial
```

```python
#Probabilidad de que 2 integrantes del grupo hayan leido la novela con una probabilidad de éxito del 0.8.

print(funcion_binomial(2,4,0.8))
```

Distribución Poisson

Se utiliza para describir sucesos en donde se considera que la probabilidad del suceso es muy pequeña. La **variable aleatoria** es el número de veces que ocurre un evento en un intervalo de tiempo, distancia, área, volumen u otra similar.

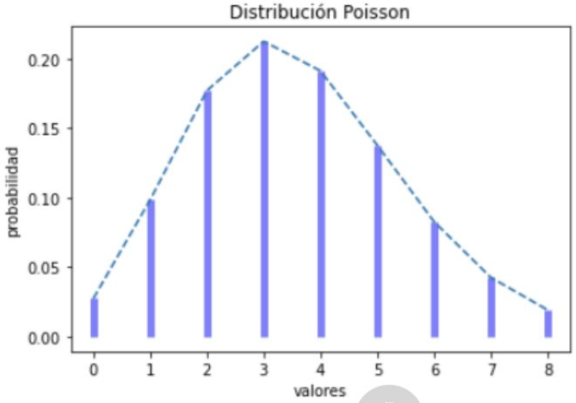
lamda= espacio

k= número de éxitos.

$$P(X = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}$$

```python
import matplotlib.pyplot as plt
import numpy as np
from scipy import stats
import seaborn as sns
mu =  3.6 # parametro de forma
poisson = stats.poisson(mu) # Distribución
x = np.arange(poisson.ppf(0.01),
              poisson.ppf(0.99))
fmp = poisson.pmf(x) # Función de Masa de Probabilidad
plt.plot(x, fmp, '--')
plt.vlines(x, 0, fmp, colors='b', lw=5, alpha=0.5)
plt.title('Distribución Poisson')
plt.ylabel('probabilidad')
plt.xlabel('valores')
plt.show()

# histograma
aleatorios = poisson.rvs(1000)  # genera aleatorios
cuenta, cajas, ignorar = plt.hist(aleatorios, 20)
plt.ylabel('frequencia')
plt.xlabel('valores')
plt.title('Histograma Poisson')
plt.show()
```



Ejemplo

La probabilidad de que en el lapso de una semana en el taller de la concesionaria uno de los autos vendidos tenga problemas cubiertos por la garantía es 0,02. Suponiendo que en el taller se atienden 450 autos semanalmente.

¿Cuál es la probabilidad de que se presenten 5 autos con problemas por semana?

```
from math import e,factorial

def probabilidad_poisson(lamba_np,x):
    probabilidad = (pow(e,-lamba_np) * pow(lamba_np,x))/factorial(x)
    return probabilidad

print(probabilidad_poisson((450*0.02),5))
```

Distribución Hipergeométrica

En un experimento de características **hipergeométricas** el resultado de una observación es afectado por los resultados de las observaciones previas, por tanto las probabilidades son condicionales.

$$p(X = x) = \frac{\binom{k}{x} \cdot \binom{N-k}{n-x}}{\binom{N}{n}}$$

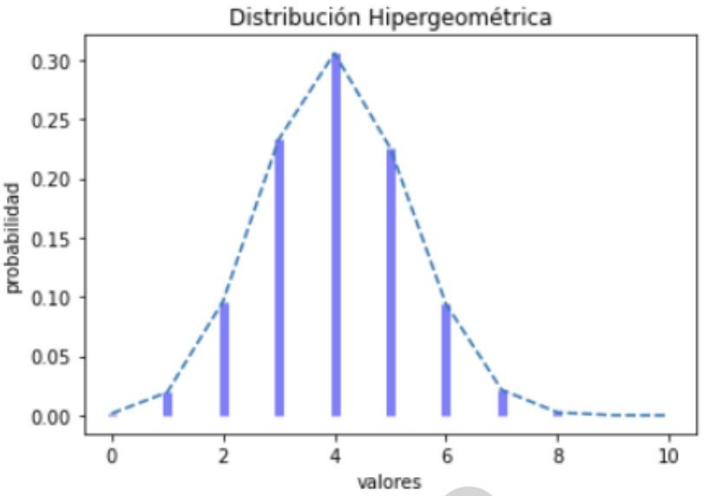
$N = \textit{tamaño de población}$

$K = n^{o} \textit{ individuos que...}$

$n = \textit{tamaño de la muestra}$

$x = \textit{valor que toma la variable}$

```python
import matplotlib.pyplot as plt
import numpy as np
from scipy import stats
import seaborn as sns
# Graficando Hipergeométrica
M, n, N = 30, 10, 12 # parametros de forma
hipergeometrica = stats.hypergeom(M, n, N) # Distribución
x = np.arange(0, n+1)
fmp = hipergeometrica.pmf(x) # Función de Masa de Probabilidad
plt.plot(x, fmp, '--')
plt.vlines(x, 0, fmp, colors='b', lw=5, alpha=0.5)
plt.title('Distribución Hipergeométrica')
plt.ylabel('probabilidad')
plt.xlabel('valores')
plt.show()
```



Ejemplo

Una empresa que importa los autos que vende una concesionaria, desea hacer una encuesta de satisfacción a los compradores de estos autos. De una muestra de 80 autos, 30 son importados.

Si se seleccionan 9 clientes. **¿Cuál es la probabilidad de que haya 2 que compraron autos importados?.**

```python
from math import e,factorial

N,X,n,x= 80,30,9,2

Xx = factorial(X)/(factorial(x)*factorial(X-x))
NX_nx= factorial(N-X)/(factorial(n-x)*factorial((N-X)-(n-x)))
Nn = factorial(N)/(factorial(n)*factorial(N-n))

probabilidad_hipergeometrica = (Xx * NX_nx)/Nn

print(probabilidad_hipergeometrica)
```

Variables continuas

- Puede tomar cualquier valor en un determinado campo de variación.
- La **probabilidad** se representa con los símbolos fi o f(x).
- En un punto la probabilidad no tiene sentido. Sólo tiene sentido en un intervalo particular de la variable aleatoria xi, por más pequeño que éste sea.
- En el **gráfico**, se ve como una función continua f(x), y la probabilidad en sí misma, denominada A, se representa como un área entre los puntos x1 y x2.

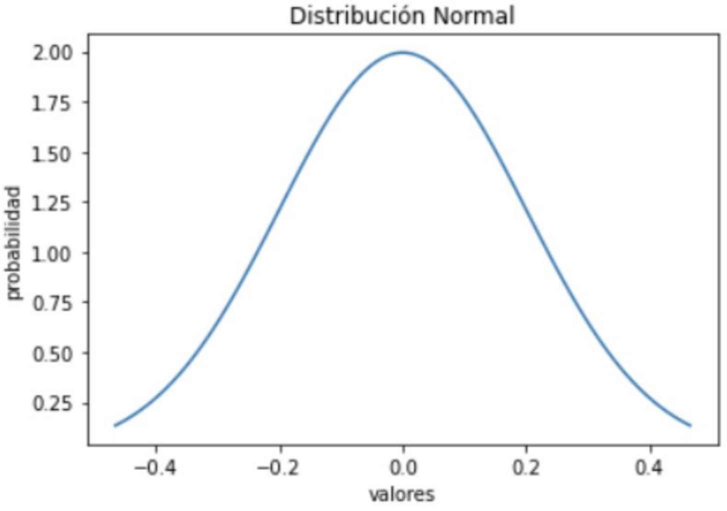
Distribución Normal

Dada por una función de **densidad** y la **probabilidad** se obtiene en base a una variable aleatoria xi que se encuentra entre dos valores arbitrarios de x1 y x2, la cual está dada por el área A bajo la curva cuyo valor se encuentra integrando la función f(x) entre ambos valores, es decir que en tanto la probabilidad en un punto cualquiera no tiene sentido.

Distribución Normal

La solución práctica para obtener esas probabilidades consiste en utilizar la **Tabla de Probabilidades** apropiada para calcular cualquier probabilidad en el caso normal, sin que importe cuáles son los valores particulares de la variable aleatoria ni los parámetros de la distribución.

```python
import matplotlib.pyplot as plt
import numpy as np
from scipy import stats
import seaborn as sns
# Graficando Normal
mu, sigma = 0, 0.2 # media y desvio estandar
normal = stats.norm(mu, sigma)
x = np.linspace(normal.ppf(0.01),
                normal.ppf(0.99), 100)
fp = normal.pdf(x) # Función de Probabilidad
plt.plot(x, fp)
plt.title('Distribución Normal')
plt.ylabel('probabilidad')
plt.xlabel('valores')
plt.show()
```



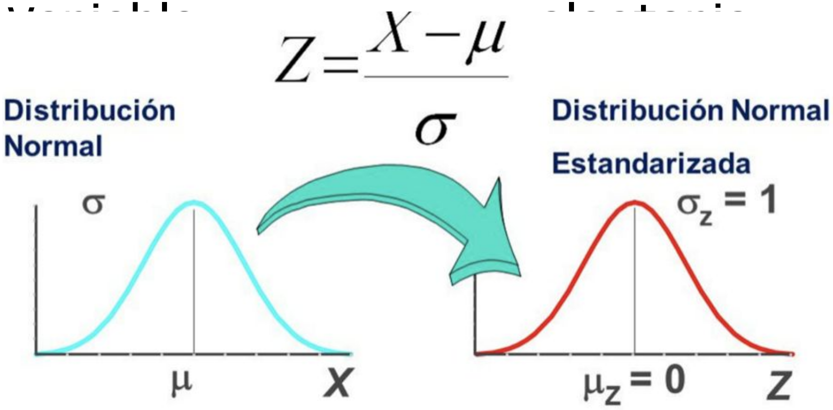
Estandarización

La **distribución normal** requiere la estandarización de las variables mediante la siguiente fórmula:

X =

mu = Media.

sigma = Desvío estándar.

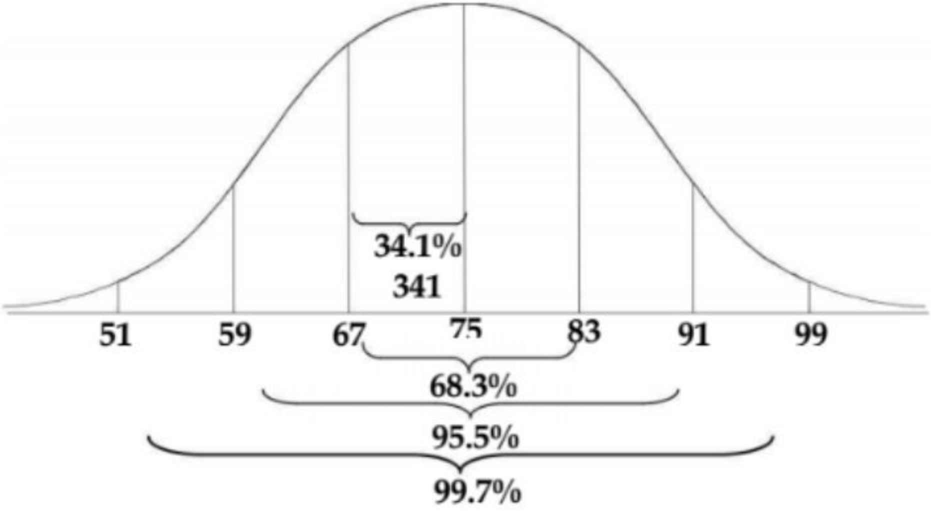


Ejemplo

Luego de estandarizar las variables, se debe buscar el **valor de Z** en la tabla de distribución normal y determinar la probabilidad con base al área delimitada por el experimento.

Ejemplo.

Una población de 1,000 personas tiene una media de edad de 75 años y una desviación estándar calculada de 8, ¿cuántas tienen entre 67 y 75 años?

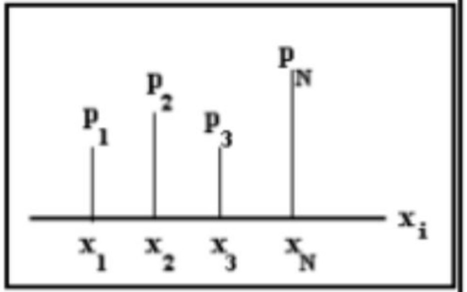
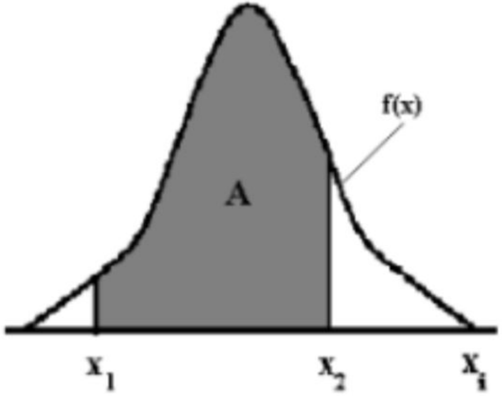


Sustituyendo la media en el centro y las desviaciones estándar (σ) a la izquierda (-8) y a la derecha (+8), la respuesta será: 68.3/2 = 34.1% de 1,000 → 341 personas tienen entre 67 y 75 años.

Resumen

Características de las Distribuciones de Probabilidad

Tipo de variable	***Discreta***	***Continua***
Dentro de un intervalo	Puede tomar sólo algunos valores	Puede tomar cualquier valor
Simbología de la probabilidad	p_i o $p(x)$	f_i o $f(x)$
Concepto de probabilidad	En un punto	En un intervalo (en un punto no tiene sentido)
Cálculo de la Probabilidad	$P(x_i = x_u) = p_u$	$P(x_1 \leq x_i \leq x_2) = \int_{x_1}^{x_2} f(x)dx = A$

Tipo de variable	*Discreta*	*Continua*
Gráfico	De bastones	De áreas
		
Valor de la probabilidad	Vale p_i (bastón) o cero	Vale A (área)
Condición de cierre	$\sum_{i=1}^{N} p_i = 1$	$\int_{-\infty}^{+\infty} f(x)dx = 1$
Denominación genérica	Función de probabilidad	Función de Densidad

¿PREGUNTAS?

Resumen

- **TEMA |** .....
- **TEMA |** .....
- **TEMA |** .....

¿Alguien dijo Homework?

HENRY

Proxíma lecture

Introducción a las bases de datos

¡Feedback!

Click on me

Dispones de un formulario en:

- Homeworks
- Guías de clase
- Slack

HENRY