

MAGNA

INSTITUCIÓN DE ESPECIALIZACIÓN PROFESIONAL



Estadística descriptiva

Ing. Breynner Chávez

Conceptos básicos de estadística descriptiva:

La estadística descriptiva se refiere al análisis y resumen de datos mediante herramientas estadísticas para describir y comprender las características de un conjunto de datos.

Nos permite analizar datos relacionados con el rendimiento, fallas y tiempos de reparación de equipos o sistemas, buscando mejorar la confiabilidad y reducir los tiempos de paradas.

Conceptos:

Población: Es el conjunto completo de elementos o unidades de análisis que se desean estudiar (por ejemplo, todos los equipos de una planta industrial).

Muestra: Es un subconjunto de la población que se selecciona para su análisis (por ejemplo, un grupo de máquinas de una planta)

Conceptos básicos de estadística descriptiva:

Media:

La media es el promedio aritmético de un conjunto de datos. Se calcula sumando todos los valores y dividiendo entre el número total de observaciones.

Donde X_i son los valores individuales y n es el número total de observaciones.

$$\text{Media}(\mu) = \frac{\sum X_i}{n}$$

Aplicaciones:

- El tiempo medio entre fallas (**MTBF**, Mean Time Between Failures) o el tiempo medio de reparación (**MTTR**, Mean Time To Repair).
- Por ejemplo, la media de los tiempos entre fallas de un equipo ayuda a predecir su comportamiento futuro y a planificar el mantenimiento preventivo.

Conceptos básicos de estadística descriptiva:

Mediana:

Es una medida de tendencia central que divide un conjunto de datos en dos partes iguales, con el 50% de los datos por debajo de la mediana y el otro 50% por encima. A diferencia de la media, la mediana no se ve afectada por valores extremos o atípicos (outliers), lo que la hace especialmente útil en situaciones donde los datos pueden estar sesgados o contener valores fuera de lo común.

Aplicaciones:

- MTBF, si hay valores extremos (por ejemplo, un equipo que funciona durante un tiempo excepcionalmente largo o corto antes de fallar), la mediana puede proporcionar una mejor representación del comportamiento.
- Si un equipo tiene tiempos entre fallas de 50, 60, 70, 80, 400 horas, la media será elevada por el valor extremo de 400 horas. La mediana, en cambio, es 70 horas, lo que refleja mejor el comportamiento común.

Conceptos básicos de estadística descriptiva:

Ejemplo:

Si tenemos los tiempos entre fallas de un grupo de máquinas en horas: 100, 120, 150, 800, 180, 200, 150.

Media: El promedio sería:

$$\frac{100 + 120 + 150 + 800 + 180 + 200 + 150}{7} = 242.86 \text{ horas}$$

Debido al valor extremo de 800 horas, la media es mucho mayor que el valor que típicamente se observa

Mediana: Ordenamos los datos de menor a mayor (100, 120, 150, 150, 180, 200, 800), el valor central es 150 horas, lo que refleja de manera más precisa el tiempo que comúnmente pasa entre fallas.

En lugar de depender solo de la media, que puede estar sesgada por fallas o tiempos de reparación inusualmente largos o cortos, la mediana ofrece una mejor visión de lo que se puede esperar en términos de tiempo entre fallas y tiempo de reparación.

Conceptos básicos de estadística descriptiva:

Moda:

Es una medida de tendencia central que identifica el valor que aparece con mayor frecuencia en un conjunto de datos. A diferencia de la media y la mediana, la moda se centra en el valor más común, lo que la hace especialmente útil cuando se busca identificar patrones recurrentes o eventos repetitivos.

Aplicaciones:

- Se utiliza para identificar valores o comportamientos que ocurren con mayor frecuencia en aspectos como fallas, reparaciones y tiempos de operación de equipos. Es útil para resaltar patrones comunes que pueden ayudar a optimizar las estrategias de mantenimiento

Conceptos básicos de estadística descriptiva:

Ejemplo:

MTTR, en el análisis del tiempo medio de reparación, si varía mucho, la moda te dará una idea de cuánto tiempo tiende a durar una reparación típica.

Si los tiempos de reparación son 2, 3, 5, 3 y 4 horas, la moda es 3 horas, lo que indica que la mayoría de las reparaciones duran ese tiempo. Esto ayuda a planificar mejor los recursos y tiempos de parada para el mantenimiento.

La moda identifica el valor más común en datos de fallas y reparaciones, ayudando a mejorar la eficiencia operativa, optimizar el mantenimiento y reducir los tiempos de parada.

Conceptos básicos de estadística descriptiva:

Varianza

Es una medida que indica cómo se dispersan o distribuyen los datos alrededor de la media.

La varianza evalúa la consistencia y variabilidad los datos sobre el comportamiento de sistemas o equipos, como los tiempos entre fallas y de reparación. Una mayor varianza indica mayor dispersión de los datos y, por lo tanto, mayor variabilidad en el rendimiento o confiabilidad del equipo.

Baja varianza: Indica que los equipos o sistemas tienen un comportamiento predecible y constante. Esto es ideal para planificar mantenimientos y prever tiempos de parada.

Alta varianza: Señala una mayor imprevisibilidad en el desempeño, lo que puede requerir un análisis más detallado para mejorar la confiabilidad mediante ajustes en las operaciones o en el mantenimiento preventivo

Conceptos básicos de estadística descriptiva:

Ejemplo:

$$\sigma^2 = \frac{\sum(X_i - \mu)^2}{n}$$

X_i son los valores individuales

μ es la media

n es el número de observaciones

MTTR, Si los tiempos entre fallas (en horas) 50, 60, 70, 90, 100

Primero, calculamos la media (promedio) de los tiempos entre fallas:

$$\text{Media} = \frac{50 + 60 + 70 + 90 + 100}{5} = \frac{370}{5} = 74 \text{ horas}$$

$$(50 - 74)^2 = (-24)^2 = 576$$

$$(60 - 74)^2 = (-14)^2 = 196$$

$$(70 - 74)^2 = (-4)^2 = 16$$

$$(90 - 74)^2 = (16)^2 = 256$$

$$(100 - 74)^2 = (26)^2 = 676$$

$$576 + 196 + 16 + 256 + 676 = 1720$$

Ahora, dividimos por el número de observaciones ($n = 5$):

$$\text{Varianza} = \frac{1720}{5} = 344$$

La varianza de 344 horas² sugiere una moderada variabilidad en los tiempos entre fallas del equipo. Una varianza más baja indicaría tiempos más consistentes y predecibles. Por el contrario, una varianza muy alta señalaría que el equipo es menos confiable, dificultando la predicción de fallas y llevando a un enfoque proactivo en el mantenimiento para prevenir fallas inesperadas.

Conceptos básicos de estadística descriptiva:

Desviación estándar

Mide la dispersión o variabilidad en un conjunto de datos, se utiliza para evaluar la consistencia del rendimiento de un equipo, proporcionando información sobre la variabilidad en los tiempos entre fallas y de reparación. Una desviación estándar baja indica que los datos están más agrupados alrededor de la media, mientras que una alta señala mayor dispersión.

Ejemplo

Si un equipo tiene un tiempo medio entre fallas de 100 horas y una desviación estándar de 5 horas, la mayoría de las fallas ocurrirán cerca de las 100 horas. Sin embargo, si la desviación estándar es de 30 horas, los tiempos entre fallas son más variables, lo que dificulta la planificación del mantenimiento.

Conceptos básicos de estadística descriptiva:

La varianza y la desviación estándar son dos medidas estadísticas que indican la dispersión de un conjunto de datos.

DEFINICIÓN

Varianza: Es el promedio de las diferencias al cuadrado entre cada dato y la media del conjunto. Mide la dispersión en unidades al cuadrado.

$$\text{Varianza} = \sigma^2 = \frac{\sum(X_i - \mu)^2}{n}$$

Desviación estándar: Es la raíz cuadrada de la varianza. Mide la dispersión en las mismas unidades que los datos originales.

$$\text{Desviación Estándar} = \sigma = \sqrt{\sigma^2}$$

Conceptos básicos de estadística descriptiva:

UNIDADES

Varianza: Se expresa en unidades al cuadrado (si los datos están en horas, la varianza se expresa en horas²).

Desviación estándar: Se expresa en las mismas unidades que los datos originales (por ejemplo, horas).

INTERPRETACIÓN

Varianza: Puede ser menos intuitiva para interpretar debido a su unidad al cuadrado. Sin embargo, es útil para cálculos matemáticos y estadísticos.

Desviación estándar: Es más fácil de interpretar porque se expresa en las mismas unidades que los datos. Indica, en promedio, cuánto se desvían los datos de la media.

Conceptos básicos de estadística descriptiva:

```
import numpy as np
from scipy import stats

# Datos de ejemplo
data = [23, 20, 21, 22, 25, 20, 23, 24, 22, 21]

# Cálculo de estadísticas
media = np.mean(data)
mediana = np.median(data)
moda = stats.mode(data, keepdims=True).mode[0] # Asegura que la moda sea un array
varianza = np.var(data, ddof=1) # Varianza muestral
desviacion_estandar = np.std(data, ddof=1) # Desviación estándar muestral

# Resultados
print(f"Promedio (Media): {media}")
print(f"Mediana: {mediana}")
print(f"Moda: {moda}")
print(f"Varianza: {varianza}")
print(f"Desviación estándar: {desviacion_estandar}")
```

```
Promedio (Media): 22.1
Mediana: 22.0
Moda: 20
Varianza: 2.7666666666666667
Desviación estándar: 1.66332999331662
```

Conceptos básicos de estadística descriptiva:

Boxplots (Diagramas de Cajas)

Son gráficos que resumen la distribución de un conjunto de datos mediante sus cuartiles. Muestran la mediana, los cuartiles (Q1 y Q3), y los valores atípicos.

APLICACIÓN

Visualización de Tiempos entre Fallas: Permiten identificar la mediana, el rango intercuartílico y los valores atípicos en los tiempos entre fallas de un equipo.

Comparación de Equipos: Facilitan la comparación de la variabilidad y la consistencia entre diferentes equipos o componentes.

Conceptos básicos de estadística descriptiva:

Datos: 2, 3, 5, 3, 4

Paso 1: Ordenar los Datos

- Orden ascendente: 2, 3, 3, 4, 5

Paso 2: Calcular Q1

- Número de datos (n) = 5
- $Q1 = \left(\frac{5+1}{4}\right) = 1.5$
- Esto significa que Q1 está entre el 1er y el 2do dato (2 y 3).
- $Q1 = \frac{2+3}{2} = 2.5$

Paso 3: Calcular Q2 (Mediana)

- Para un conjunto de 5 datos, la mediana es el 3er dato.
- $Q2 = 3$

Paso 4: Calcular Q3

- $Q3 = \left(\frac{3(5+1)}{4}\right) = 4.5$
- Esto significa que Q3 está entre el 4to y el 5to dato (4 y 5).
- $Q3 = \frac{4+5}{2} = 4.5$

Resultados:

Q1: 2.5

Q2 (Mediana): 3

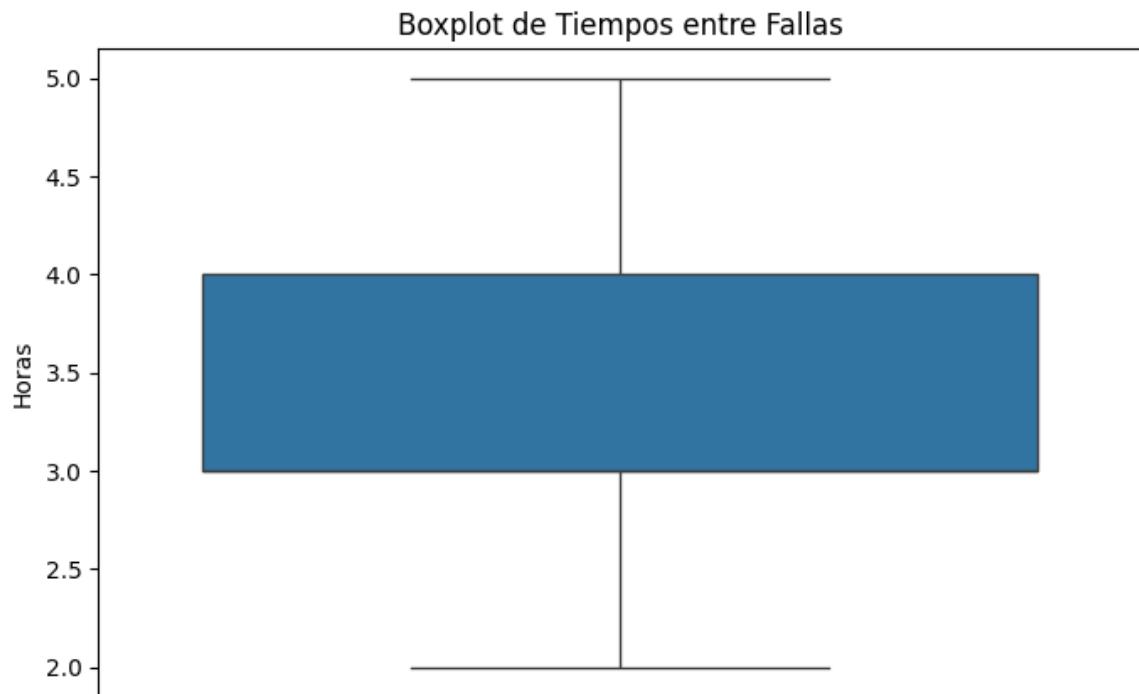
Q3: 4.5

Conceptos básicos de estadística descriptiva:

```
import matplotlib.pyplot as plt
import seaborn as sns

# Datos de ejemplo
data = [2, 3, 5, 3, 4]

# Crear el boxplot
plt.figure(figsize=(8, 5))
sns.boxplot(data=data)
plt.title('Boxplot de Tiempos entre Fallas')
plt.ylabel('Horas')
plt.show()
```



Conceptos básicos de estadística descriptiva:

Gráficos de Barras

Los gráficos de barras son herramientas visuales efectivas para mostrar y comparar la frecuencia de diferentes categorías de datos.

APLICACIÓN

Tipos de Fallas: Los gráficos de barras pueden mostrar la cantidad de fallas clasificadas por tipo (mecánica, eléctrica, etc.). Esto ayuda a identificar las áreas que requieren más atención en el mantenimiento.

Tipos de Mantenimiento: Se pueden utilizar para visualizar la cantidad de mantenimiento realizado por tipo (preventivo, correctivo, etc.) o por equipo. Esto ayuda a planificar recursos y optimizar programas de mantenimiento.

Conceptos básicos de estadística descriptiva:

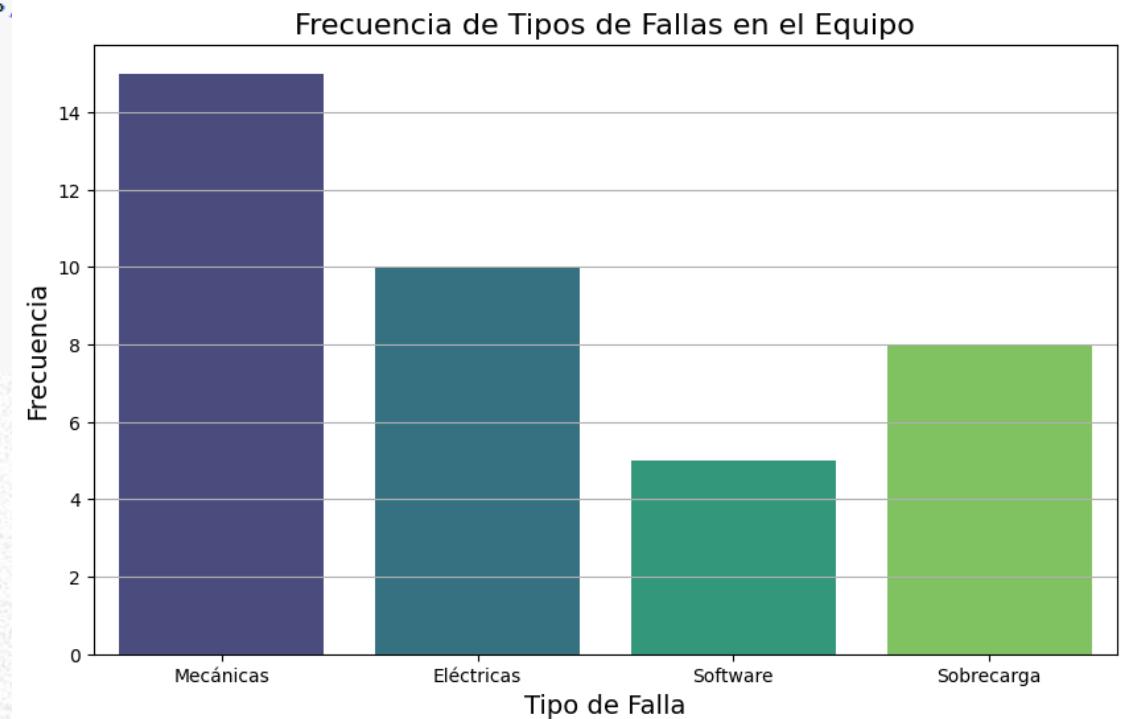
```
import matplotlib.pyplot as plt
import seaborn as sns

# Datos de ejemplo
tipos_fallas = ['Mecánicas', 'Eléctricas', 'Software', 'Sobrecarga']
frecuencias = [15, 10, 5, 8]

# Crear el gráfico de barras
plt.figure(figsize=(10, 6))
# Asignar los tipos de fallas como un eje categórico
sns.barplot(x=tipos_fallas, y=frecuencias, palette='viridis', hue=tipos_fallas)

# Títulos y etiquetas
plt.title('Frecuencia de Tipos de Fallas en el Equipo', fontsize=16)
plt.xlabel('Tipo de Falla', fontsize=14)
plt.ylabel('Frecuencia', fontsize=14)
plt.grid(axis='y')

# Mostrar el gráfico
plt.show()
```



Conceptos básicos de estadística descriptiva:

Gráficos de Líneas

Son herramientas efectivas para visualizar datos continuos en el tiempo, ayudan a mostrar tendencias en el rendimiento de un equipo, la frecuencia de fallas y los tiempos de reparación, lo que permite a los responsables de mantenimiento identificar patrones y tomar decisiones informadas.

APLICACIÓN

Tendencias de Fallas a lo Largo del Tiempo: Muestra cómo varía la frecuencia de fallas en un equipo a lo largo de un período determinado. Esto puede ayudar a identificar si un equipo se está volviendo menos confiable.

Conceptos básicos de estadística descriptiva:

```
import matplotlib.pyplot as plt
import numpy as np

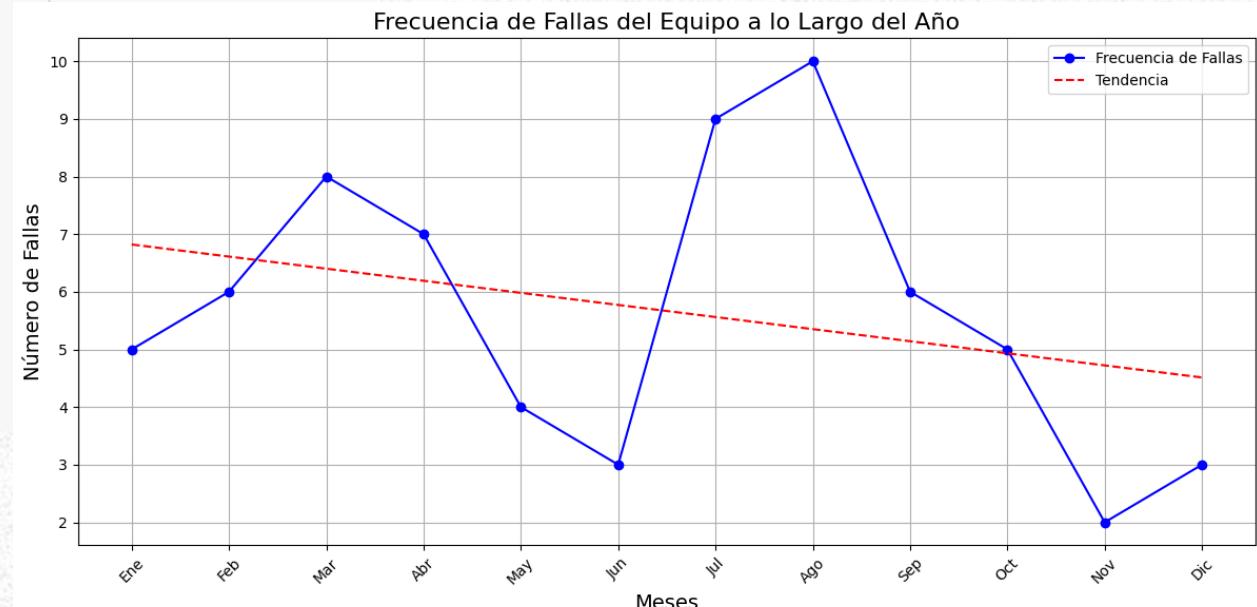
# Datos de ejemplo: Frecuencia de fallas por mes
meses = np.array(['Ene', 'Feb', 'Mar', 'Abr', 'May', 'Jun', 'Jul', 'Ago', 'Sep', 'Oct', 'Nov', 'Dic'])
fallas = np.array([5, 6, 8, 7, 4, 3, 9, 10, 6, 5, 2, 3])

# Crear el gráfico de líneas
plt.figure(figsize=(12, 6))
plt.plot(meses, fallas, marker='o', linestyle='-', color='b', label='Frecuencia de Fallas')

# Calcular y agregar línea de tendencia
z = np.polyfit(range(len(meses)), fallas, 1) # Ajustar una línea de primer grado
p = np.poly1d(z)
plt.plot(meses, p(range(len(meses))), linestyle='--', color='r', label='Tendencia')

# Títulos y etiquetas
plt.title('Frecuencia de Fallas del Equipo a lo Largo del Año', fontsize=16)
plt.xlabel('Meses', fontsize=14)
plt.ylabel('Número de Fallas', fontsize=14)
plt.xticks(rotation=45)
plt.grid()
plt.legend()

# Mostrar el gráfico
plt.tight_layout() # Ajustar el layout para que no se corten los textos
plt.show()
```



HISTOGRAMAS

Los gráficos de barras son herramientas visuales efectivas para mostrar y comparar la frecuencia de diferentes categorías de datos.

APLICACIÓN

Distribución de Tiempos entre Fallas: Permiten observar cómo se distribuyen los tiempos entre fallas en un equipo, ayudando a identificar si hay tendencias en los tiempos de funcionamiento antes de que ocurran fallas.

Análisis de Modas: Un histograma puede ayudar a identificar el rango de tiempo que es más común entre las fallas, lo que puede ser crucial para la planificación del mantenimiento.

HISTOGRAMAS

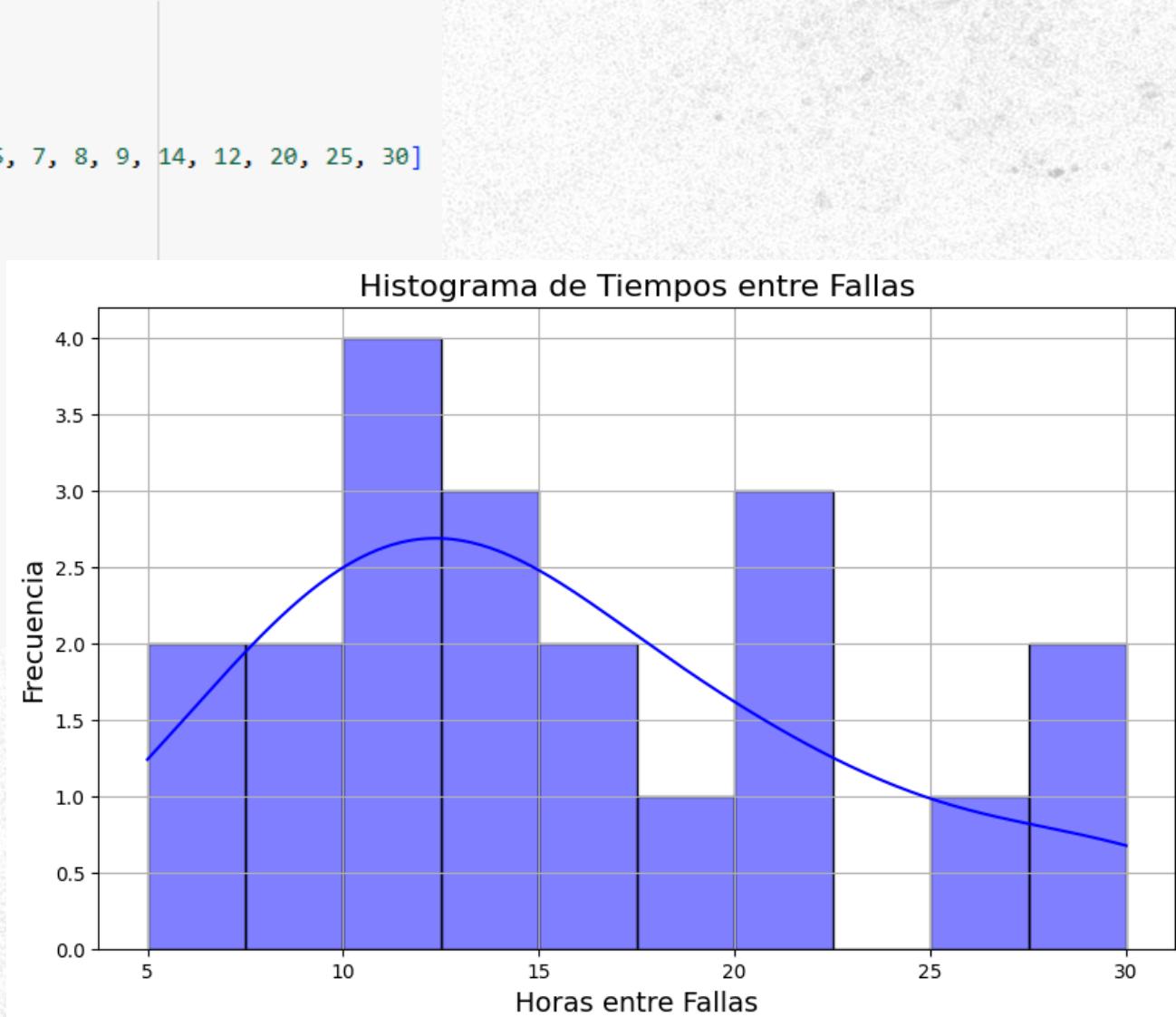
```
import matplotlib.pyplot as plt
import seaborn as sns

# Datos de ejemplo
tiempos_entre_fallas = [10, 12, 15, 10, 13, 14, 18, 20, 15, 22, 30, 5, 7, 8, 9, 14, 12, 20, 25, 30]

# Crear el histograma
plt.figure(figsize=(10, 6))
sns.histplot(tiempos_entre_fallas, bins=10, kde=True, color='blue')

# Títulos y etiquetas
plt.title('Histograma de Tiempos entre Fallas', fontsize=16)
plt.xlabel('Horas entre Fallas', fontsize=14)
plt.ylabel('Frecuencia', fontsize=14)

# Mostrar el gráfico
plt.grid()
plt.show()
```



DISTRIBUCION EXPONENCIAL

La distribución exponencial se utiliza para modelar el tiempo que transcurre entre eventos, donde los eventos ocurren de manera continua y a una tasa constante.

Tiempos entre Fallas: Se utiliza para modelar el tiempo que transcurre entre fallas de un sistema o equipo, donde es importante predecir cuándo puede ocurrir una falla

Vida Útil de Productos: se puede utilizar para estimar la vida útil de productos y componentes, ayudando a las empresas a diseñar programas de mantenimiento

DISTRIBUCION EXPONENCIAL

Tasa de Fallas La tasa de fallas (λ) se define como el inverso del tiempo promedio entre fallas (media).

$$\lambda = \frac{1}{\text{Media}}$$

Función de Densidad de Probabilidad (PDF)

La función de densidad de probabilidad de la distribución exponencial se define como:

$$f(t; \lambda) = \lambda e^{-\lambda t} \quad \text{para } t \geq 0$$

Función de Distribución Acumulativa (CDF)

La función de distribución acumulativa se define como:

$$F(t; \lambda) = 1 - e^{-\lambda t}$$

Esto permite calcular la probabilidad de que un evento ocurra antes de un tiempo específico t .

DISTRIBUCION EXPONENCIAL

Ejemplo

Los técnicos han recopilado datos sobre el tiempo que les toma reparar una máquina. Se ha observado que, en promedio, el tiempo de reparación es de 3 horas. Esto significa que la distribución de los tiempos de reparación se puede modelar utilizando una distribución exponencial.

Definición de parámetros

1. Media (μ): 3 horas

2. Tasa de fallas (λ):

$$\lambda = \frac{1}{\text{Media}} = \frac{1}{3} \approx 0.333 \text{ reparaciones/hora}$$

3. Calcular la probabilidad de que el tiempo de reparación sea **menos de 2 horas**.

Usamos la **función de distribución acumulativa (CDF)** de la distribución exponencial

$$F(t; \lambda) = 1 - e^{-\lambda t}$$

Para $t = 2$ horas, la probabilidad es:

$$F(2; 0.333) = 1 - e^{-0.333 \cdot 2}$$

DISTRIBUCION EXPONENCIAL

```
import numpy as np
import matplotlib.pyplot as plt

# Parámetros
media = 3 # horas
lambda_rate = 1 / media # tasa de reparaciones

# Simular 100 tiempos de reparación
num_simulaciones = 100
tiempos_reparacion = np.random.exponential(scale=media, size=num_simulaciones)

# Calcular la probabilidad de que el tiempo de reparación sea menor a 2 horas
t = 2 # horas
probabilidad = 1 - np.exp(-lambda_rate * t)
print(f"Probabilidad de que el tiempo de reparación sea menor a {t} horas: {probabilidad:.4f}")

# Crear el histograma
plt.hist(tiempos_reparacion, bins=10, density=True, alpha=0.7, color='blue', label='Histograma de')

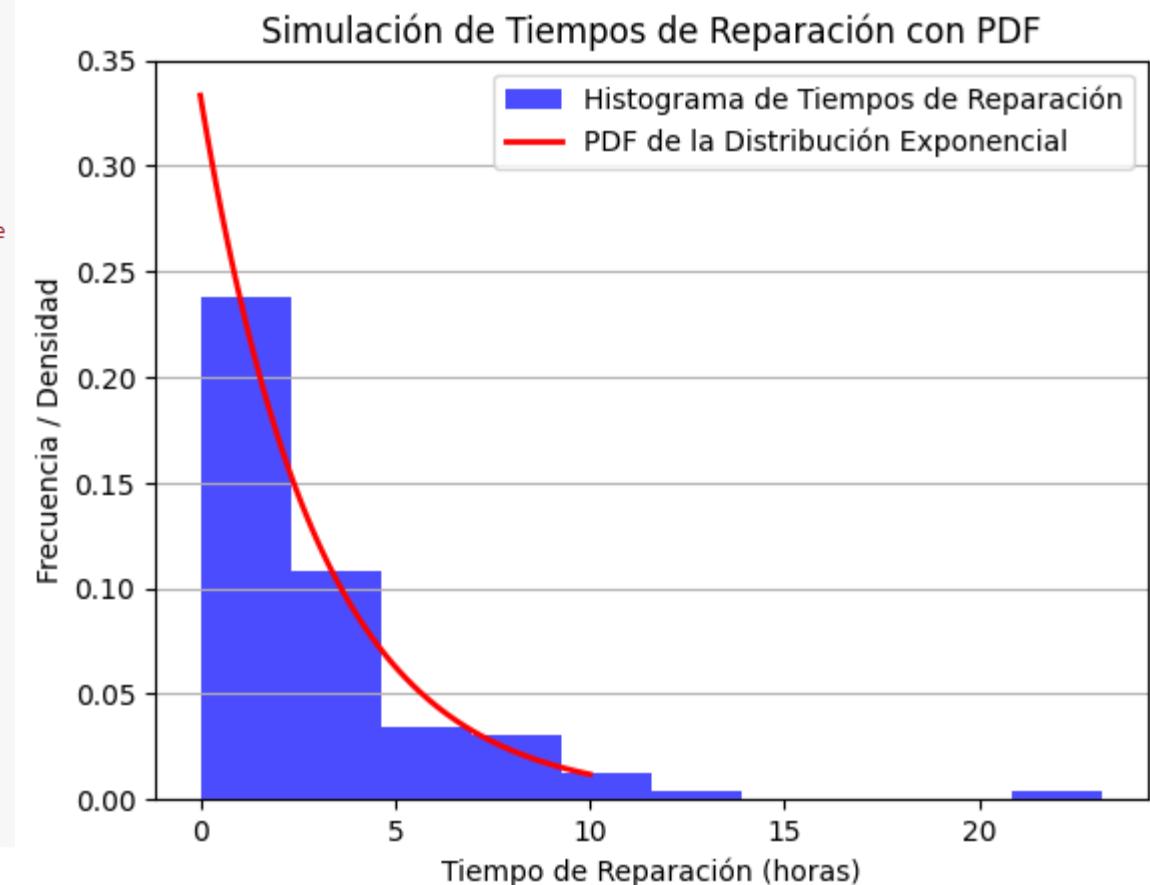
# Crear la función de densidad de probabilidad (PDF)
x = np.linspace(0, 10, 100) # Rango de valores
pdf = lambda_rate * np.exp(-lambda_rate * x) # PDF de la distribución exponencial

# Graficar la PDF
plt.plot(x, pdf, color='red', lw=2, label='PDF de la Distribución Exponencial')

# Títulos y etiquetas
plt.title('Simulación de Tiempos de Reparación con PDF')
plt.xlabel('Tiempo de Reparación (horas)')
plt.ylabel('Frecuencia / Densidad')
plt.legend()
plt.grid(axis='y')

# Mostrar el gráfico
plt.show()
```

Probabilidad de que el tiempo de reparación sea menor a 2 horas: 0.4866



DISTRIBUCION NORMAL

La distribución normal es una distribución continua que se caracteriza por su forma de campana, simétrica alrededor de su media.

Tiempos de Reparación: Si los tiempos de reparación de un equipo son normalmente distribuidos, se pueden estimar las probabilidades de que una reparación tome menos o más tiempo del promedio. Esto ayuda en la planificación de recursos y en la gestión del tiempo

Variabilidad en el Rendimiento: En el análisis del rendimiento de los equipos, se pueden usar distribuciones normales para evaluar si las variaciones en el rendimiento son aceptables o si requieren atención.

DISTRIBUCION NORMAL

Media (μ): El valor promedio o central de la distribución.

Desviación Estándar (σ): Una medida de la dispersión o variabilidad de los datos alrededor de la media

La función de densidad de probabilidad (PDF) de la distribución normal se expresa como:

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

Propiedades

Simetría: La curva es simétrica en torno a la media. Esto significa que la mitad de los datos se encuentra a la izquierda y la otra mitad a la derecha.

Campana: La mayoría de los datos se agrupan alrededor de la media, y la probabilidad de encontrar valores extremos disminuye a medida que nos alejamos de la media.

DISTRIBUCION NORMAL

Ejemplo

Teniendo un tiempo medio de reparación de una máquina es de 6 horas con una desviación estándar de 1.5 horas. Vamos a calcular y visualizar la probabilidad de que el tiempo de reparación sea menor a 5 horas.

Definición de parámetros

Z es la variable normal estándar.

μ es la media.

σ es la desviación estándar.

$$P(X < t) = P\left(Z < \frac{t - \mu}{\sigma}\right)$$

$\mu = 6$ (media de tiempos de reparación)

$\sigma = 1.5$ (desviación estándar)

$t = 5$ (tiempo específico)

$$Z = \frac{t - \mu}{\sigma} = \frac{5 - 6}{1.5} = \frac{-1}{1.5} \approx -0.6667$$

$$P(Z < -0.6667)$$

DISTRIBUCION NORMAL

Tabla de la Distribución Normal Estándar $N(0,1)$

	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.50000	0.49601	0.49202	0.48803	0.48405	0.48006	0.47608	0.47210	0.46812	0.46414
0.1	0.46017	0.45620	0.45224	0.44828	0.44433	0.44038	0.43644	0.43251	0.42858	0.42465
0.2	0.42074	0.41683	0.41294	0.40905	0.40517	0.40129	0.39743	0.39358	0.38974	0.38591
0.3	0.38209	0.37828	0.37448	0.37070	0.36693	0.36317	0.35942	0.35569	0.35197	0.34827
0.4	0.34458	0.34090	0.33724	0.33360	0.32997	0.32636	0.32276	0.31918	0.31561	0.31207
0.5	0.30854	0.30503	0.30153	0.29806	0.29460	0.29116	0.28774	0.28434	0.28096	0.27760
0.6	0.27425	0.27093	0.26763	0.26435	0.26109	0.25785	0.25463	0.25143	0.24825	0.24510
0.7	0.24196	0.23885	0.23576	0.23270	0.22965	0.22663	0.22363	0.22065	0.21770	0.21476
0.8	0.21186	0.20897	0.20611	0.20327	0.20045	0.19766	0.19489	0.19215	0.18943	0.18673
0.9	0.18406	0.18141	0.17879	0.17619	0.17361	0.17106	0.16853	0.16602	0.16354	0.16109
1	0.15866	0.15625	0.15386	0.15151	0.14917	0.14686	0.14457	0.14231	0.14007	0.13786
1.1	0.13567	0.13350	0.13136	0.12924	0.12714	0.12507	0.12302	0.12100	0.11900	0.11702
1.2	0.11507	0.11314	0.11123	0.10935	0.10749	0.10565	0.10383	0.10204	0.10027	0.09853
1.3	0.09680	0.09510	0.09342	0.09176	0.09012	0.08851	0.08692	0.08534	0.08379	0.08226
1.4	0.08076	0.07927	0.07780	0.07636	0.07493	0.07353	0.07215	0.07078	0.06944	0.06811
1.5	0.06681	0.06552	0.06426	0.06301	0.06178	0.06057	0.05938	0.05821	0.05705	0.05592
1.6	0.05480	0.05370	0.05262	0.05155	0.05050	0.04947	0.04846	0.04746	0.04648	0.04551
1.7	0.04457	0.04363	0.04272	0.04182	0.04093	0.04006	0.03920	0.03836	0.03754	0.03673
1.8	0.03593	0.03515	0.03438	0.03362	0.03288	0.03216	0.03144	0.03074	0.03005	0.02938
1.9	0.02872	0.02807	0.02743	0.02680	0.02619	0.02559	0.02500	0.02442	0.02385	0.02330
2	0.02275	0.02222	0.02169	0.02118	0.02068	0.02018	0.01970	0.01923	0.01876	0.01831
2.1	0.01786	0.01743	0.01700	0.01659	0.01618	0.01578	0.01539	0.01500	0.01463	0.01426
2.2	0.01390	0.01355	0.01321	0.01287	0.01255	0.01222	0.01191	0.01160	0.01130	0.01101
2.3	0.01072	0.01044	0.01017	0.00990	0.00964	0.00939	0.00914	0.00889	0.00866	0.00842
2.4	0.00820	0.00798	0.00776	0.00755	0.00734	0.00714	0.00695	0.00676	0.00657	0.00639
2.5	0.00621	0.00604	0.00587	0.00570	0.00554	0.00539	0.00523	0.00508	0.00494	0.00480
2.6	0.00466	0.00453	0.00440	0.00427	0.00415	0.00402	0.00391	0.00379	0.00368	0.00357
2.7	0.00347	0.00336	0.00326	0.00317	0.00307	0.00298	0.00289	0.00280	0.00272	0.00264
2.8	0.00256	0.00248	0.00240	0.00233	0.00226	0.00219	0.00212	0.00205	0.00199	0.00193
2.9	0.00187	0.00181	0.00175	0.00169	0.00164	0.00159	0.00154	0.00149	0.00144	0.00139
3	0.0013500	0.0013063	0.0012639	0.0012228	0.0011830	0.0011443	0.0011068	0.0010704	0.0010351	0.0010009
3.1	0.0009677	0.0009355	0.0009043	0.0008741	0.0008448	0.0008164	0.0007889	0.0007623	0.0007364	0.0007114
3.2	0.0006872	0.0006637	0.0006410	0.0006190	0.0005977	0.0005771	0.0005571	0.0005378	0.0005191	0.0005010
3.3	0.0004835	0.0004665	0.0004501	0.0004343	0.0004189	0.0004041	0.0003898	0.0003759	0.0003625	0.0003495
3.4	0.0003370	0.0003249	0.0003132	0.0003018	0.0002909	0.0002803	0.0002701	0.0002603	0.0002508	0.0002416
3.5	0.0002327	0.0002241	0.0002158	0.0002078	0.0002001	0.0001927	0.0001855	0.0001785	0.0001718	0.0001654
3.6	0.0001591	0.0001531	0.0001473	0.0001417	0.0001364	0.0001312	0.0001261	0.0001213	0.0001166	0.0001122
3.7	0.0001078	0.0001037	0.0000996	0.0000958	0.0000920	0.0000884	0.0000850	0.0000816	0.0000784	0.0000753
3.8	0.0000724	0.0000695	0.0000667	0.0000641	0.0000615	0.0000591	0.0000567	0.0000544	0.0000522	0.0000501
3.9	0.0000481	0.0000462	0.0000443	0.0000425	0.0000408	0.0000391	0.0000375	0.0000360	0.0000345	0.0000331
4	0.0000317	0.0000304	0.0000291	0.0000279	0.0000267	0.0000256	0.0000245	0.0000235	0.0000225	0.0000216

DISTRIBUCION NORMAL

```
# Parámetros de la distribución normal
mu = 6 # media (horas)
sigma = 1.5 # desviación estándar (horas)
t = 5 # tiempo específico

# Cálculo manual de la variable Z
Z = (t - mu) / sigma
print(f"Valor de Z para t = {t}: {Z:.4f}")

# Cálculo manual de la probabilidad usando la CDF
# Este valor se puede obtener usando una tabla de la distribución normal estándar
probabilidad_manual = stats.norm.cdf(Z)
print(f"Probabilidad (manual) de que el tiempo de reparación sea menor a {t} horas: {probabilidad_manual:.4f}")

# Cálculo de la probabilidad usando scipy
probabilidad_scipy = stats.norm.cdf(t, mu, sigma)
print(f"Probabilidad (scipy) de que el tiempo de reparación sea menor a {t} horas: {probabilidad_scipy:.4f}")

# Generar datos de la distribución normal
x = np.linspace(mu - 4*sigma, mu + 4*sigma, 1000) # Rango de valores
y = stats.norm.pdf(x, mu, sigma) # PDF

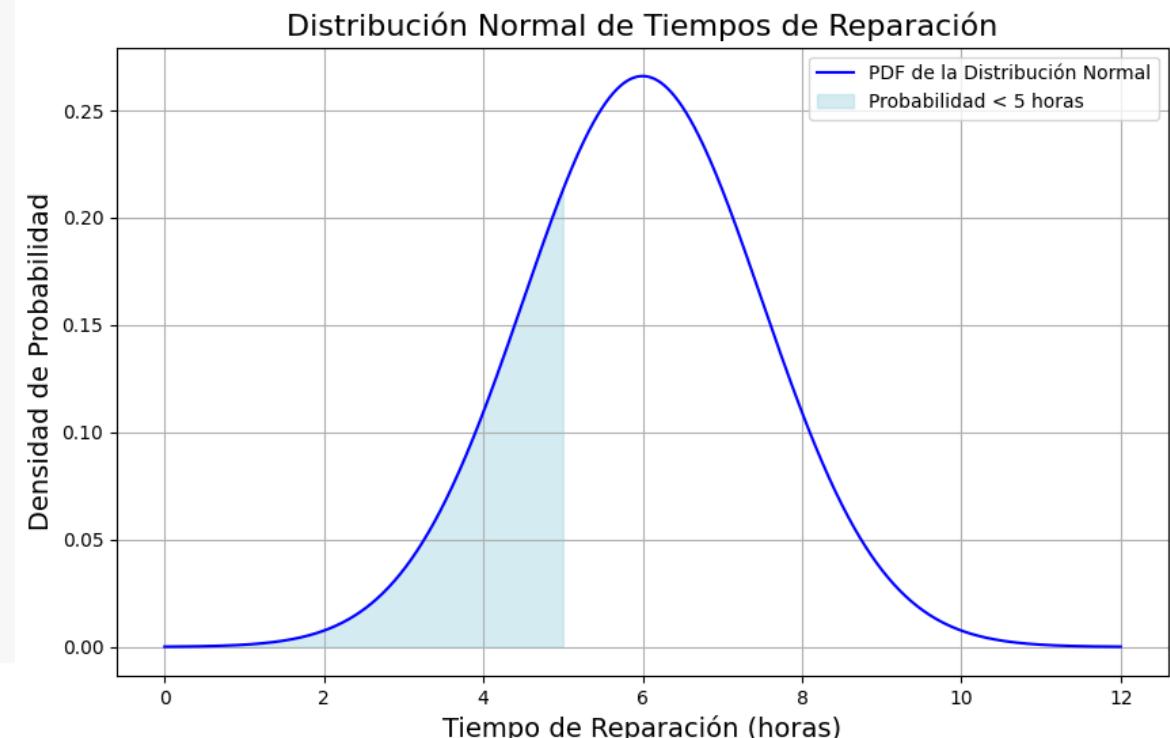
# Crear el gráfico
plt.figure(figsize=(10, 6))
plt.plot(x, y, label='PDF de la Distribución Normal', color='blue')

# Marcar la región de probabilidad
x_fill = np.linspace(mu - 4*sigma, t, 100)
y_fill = stats.norm.pdf(x_fill, mu, sigma)
plt.fill_between(x_fill, y_fill, alpha=0.5, color='lightblue', label=f'Probabilidad < {t} horas')

# Títulos y etiquetas
plt.title('Distribución Normal de Tiempos de Reparación', fontsize=16)
plt.xlabel('Tiempo de Reparación (horas)', fontsize=14)
plt.ylabel('Densidad de Probabilidad', fontsize=14)
plt.legend()
plt.grid()

# Mostrar el gráfico
plt.show()
```

Valor de Z para t = 5: -0.6667
Probabilidad (manual) de que el tiempo de reparación sea menor a 5 horas: 0.2525
Probabilidad (scipy) de que el tiempo de reparación sea menor a 5 horas: 0.2525



DISTRIBUCION LOG-NORMAL

Se utiliza para modelar tiempos de vida de equipos o componentes, cuando los datos son positivos y asimétricos. En esta distribución, el logaritmo de la variable sigue una distribución normal, lo que implica que la variable original está sesgada hacia la derecha. Es útil para representar situaciones en las que los tiempos de fallo, tiempos de reparación, o la duración de eventos tienen una gran variabilidad y no siguen una distribución simétrica.

Modelar tiempos entre fallos (MTTF): cuando los fallos se deben a degradación.

Tiempos de reparación: cuando hay gran variabilidad en la duración de las reparaciones.

Duración de eventos o procesos de mantenimiento: donde los tiempos varían de manera asimétrica

DISTRIBUCION LOG-NORMAL

Una variable aleatoria X sigue una distribución log-normal si su logaritmo natural $\ln(X)$ sigue una distribución normal.

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

Entonces, X tiene una distribución log-normal, y su densidad de probabilidad está dada por:

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

Donde:

μ y σ son la media y la desviación estándar del logaritmo natural de X , respectivamente.

$x > 0$, porque el valor de X en una distribución log-normal debe ser positivo.

DISTRIBUCION NORMAL

Ejemplo

Tiempo hasta la falla en horas: [20,30,25,50,40,80,60,70,65,90]

$\mu \approx 3.4012$ (logaritmo natural de la media)

$\sigma \approx 0.2610$ (desviación estándar del logaritmo)

$$F(x) = P(X \leq x) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{\ln(x) - \mu}{\sigma\sqrt{2}} \right) \right]$$

erf es la función error.

$\ln(x)$ es el logaritmo natural del valor x .

Cálculo para $x = 60$

1. Calcular el logaritmo natural de 60:

$$\ln(60) \approx 4.0943$$

2. Calcular el valor del argumento de la función error:

$$\frac{\ln(60) - \mu}{\sigma\sqrt{2}} = \frac{4.0943 - 3.4012}{0.2610\sqrt{2}}$$

Primero, calculemos $\sigma\sqrt{2}$:

$$\sigma\sqrt{2} \approx 0.2610 \times 1.4142 \approx 0.3696$$

Ahora, sustituimos:

$$\frac{4.0943 - 3.4012}{0.3696} \approx \frac{0.6931}{0.3696} \approx 1.8772$$

3. Calcular $\operatorname{erf}(1.8772)$:

Para calcular $\operatorname{erf}(1.8772)$, normalmente se consultaría una tabla de la función error o una calculadora científica. Supongamos que:

$$\operatorname{erf}(1.8772) \approx 0.9662$$

4. Sustituyendo en la fórmula de la CDF:

Ahora, sustituimos este valor en la fórmula de la CDF:

$$F(60) = \frac{1}{2} [1 + 0.9662] = \frac{1}{2} \times 1.9662 \approx 0.9831$$

Esto significa que hay aproximadamente un 98.31% de probabilidad de que el tiempo hasta la falla sea 60 horas o menos.]

DISTRIBUCION NORMAL

```
# Datos de tiempo hasta la falla (en horas)
tiempos_falla = np.array([20, 30, 25, 50, 40, 80, 60, 70, 65, 90])

# Calcular la media y la desviación estándar de los logaritmos
log_tiempos_falla = np.log(tiempos_falla)
mu = np.mean(log_tiempos_falla)
sigma = np.std(log_tiempos_falla)

print(f"Media logarítmica ( $\mu$ ): {mu:.4f}")
print(f"Desviación estándar logarítmica ( $\sigma$ ): {sigma:.4f}")

# Crear la distribución log-normal
dist = lognorm(sigma, scale=np.exp(mu))

# Calcular la CDF para un tiempo de falla específico
tiempo_falla_especifico = 60
cdf_valor = dist.cdf(tiempo_falla_especifico)
print(f"Probabilidad acumulada (CDF) para que el tiempo hasta la falla sea  $\leq$  {tiempo_falla_especifico} horas: {cdf_valor:.4f}")

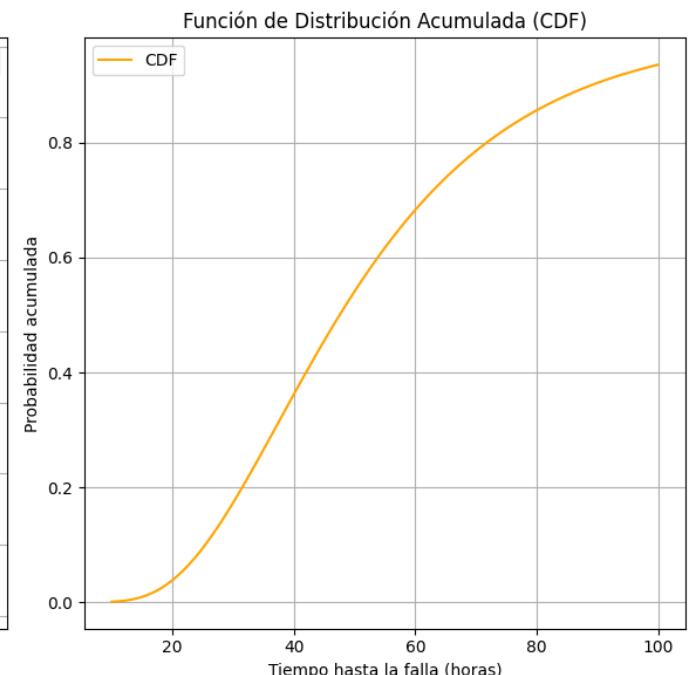
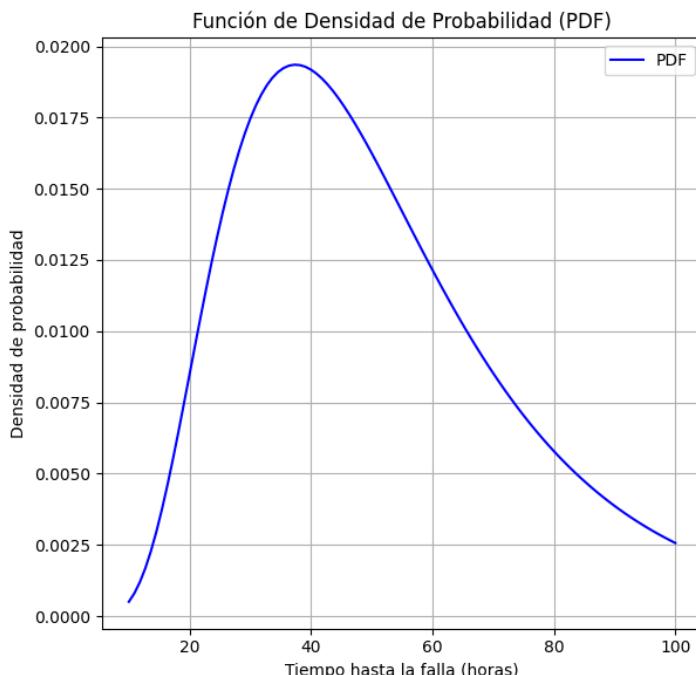
# Graficar la PDF y la CDF
x = np.linspace(10, 100, 100)
pdf = dist.pdf(x)
cdf = dist.cdf(x)

plt.figure(figsize=(12, 6))

# Graficar PDF
plt.subplot(1, 2, 1)
plt.plot(x, pdf, label='PDF', color='blue')
plt.title('Función de Densidad de Probabilidad (PDF)')
plt.xlabel('Tiempo hasta la falla (horas)')
plt.ylabel('Densidad de probabilidad')
plt.grid(True)
plt.legend()

# Graficar CDF
plt.subplot(1, 2, 2)
plt.plot(x, cdf, label='CDF', color='orange')
plt.title('Función de Distribución Acumulada (CDF)')
plt.xlabel('Tiempo hasta la falla (horas)')
plt.ylabel('Probabilidad acumulada')
plt.grid(True)
plt.legend()

plt.tight_layout()
plt.show()
```



DISTRIBUCION WEIBULL

Es especialmente útil para modelar el tiempo hasta el fallo de componentes y sistemas, ya que puede representar diversas formas de comportamiento del tiempo hasta el fallo dependiendo de sus parámetros

Modelar el tiempo hasta la falla: Ayuda a predecir cuándo es probable que ocurran las fallas.

Evaluar la confiabilidad de productos: Permite estimar la vida útil de productos y sistemas.

Análisis de mantenibilidad: Ayuda a determinar el momento óptimo para realizar mantenimiento preventivo.

DISTRIBUCION WEIBULL

Parámetro de Forma (k)

Determina la forma de la distribución e influye en la tasa de fallas a lo largo del tiempo.

$k < 1$, La tasa de fallas disminuye con el tiempo.

Ejemplo: Fenómeno conocido como "fallas iniciales", donde los componentes tienden a fallar rápidamente al principio debido a defectos de fabricación o errores de diseño.

$k = 1$, La distribución se convierte en una distribución exponencial.

Ejemplo: Esto implica que la tasa de fallas es constante a lo largo del tiempo, componentes que se utilizan bajo condiciones estables, como bombillas que fallan aleatoriamente.

$k > 1$, La tasa de fallas aumenta con el tiempo.

Ejemplo: Esto es típico en el caso de "desgaste", donde los componentes se vuelven más propensos a fallar a medida que envejecen, piezas mecánicas.

DISTRIBUCION WEIBULL

Parámetro de Escala (λ)

Establece la escala de la distribución y está relacionado con la vida útil esperada del componente o sistema.

Este parámetro no afecta la forma de la distribución, pero sí afecta la ubicación de la distribución en el eje x.

λ representa el "tiempo característico" para la tasa de fallas. A medida que λ aumenta, la distribución se desplaza hacia la derecha, lo que indica que se espera que el componente funcione por más tiempo antes de fallar.

Ejemplo: Si $\lambda=100$ horas y $k=2$, eso significa que los componentes están diseñados para fallar en promedio alrededor de 100 horas. Si λ aumenta a 150 horas, se espera que la vida útil media del componente también aumente.

DISTRIBUCION WEIBULL

Función de Densidad de Probabilidad (PDF)

La función de densidad de probabilidad de la distribución weibull se define como:

$$f(x; k, \lambda) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} \quad \text{para } x \geq 0$$

Función de Distribución Acumulativa (CDF)

La función de distribución acumulativa se define como:

$$F(x; k, \lambda) = 1 - e^{-(x/\lambda)^k} \quad \text{para } x \geq 0$$

Esto permite calcular la probabilidad de que un evento ocurra antes de un tiempo específico t.

DISTRIBUCION WEIBULL

Ejemplo

Tiempo hasta la falla en horas: [20,25,30,40,50,60,65,70,80,90]

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import weibull_min

# Datos de tiempo hasta la falla
tiempos_falla = np.array([20, 25, 30, 40, 50, 60, 65, 70, 80, 90])

# Ajustar la distribución Weibull
params = weibull_min.fit(tiempos_falla, floc=0) # floc=0 para fijar la localización en 0
k, loc, scale = params # k = forma, loc = localización, scale = escala

print(f"Parámetro de forma (k): {k:.4f}")
print(f"Parámetro de escala (λ): {scale:.4f}")

# Calcular la probabilidad de falla para un tiempo específico (ejemplo: 50 horas)
tiempo_especifico = 50
probabilidad_falla = weibull_min.pdf(tiempo_especifico, k, loc, scale)
probabilidad_falla_acumulada = weibull_min.cdf(tiempo_especifico, k, loc, scale)

print(f"Probabilidad de falla en {tiempo_especifico} horas (PDF): {probabilidad_falla:.4f}")
print(f"Probabilidad acumulada de falla hasta {tiempo_especifico} horas (CDF): {probabilidad_falla_acumulada:.4f}")
```

Parámetro de forma (k): 2.5737
Parámetro de escala (λ): 59.9106

DISTRIBUCION WEIBULL

Cálculo de la Probabilidad de Falla (PDF)

$$f(t; k, \lambda) = \frac{k}{\lambda} \left(\frac{t}{\lambda} \right)^{k-1} e^{-\left(\frac{t}{\lambda}\right)^k}$$

$$k = 2.5737$$

$$\lambda = 59.9106$$

$$t = 50$$

Ahora, calculamos $f(50; 2.5737, 59.9106)$:

$$f(50; 2.5737, 59.9106) = \frac{2.5737}{59.9106} (0.834)^{2.5737} e^{-0.5963}$$

Calculando paso a paso:

- $\frac{2.5737}{59.9106} \approx 0.0429$

Por lo tanto:

$$f(50; 2.5737, 59.9106) = 0.0429 \cdot 0.5963 \cdot 0.5512 \approx 0.0143$$

Cálculo de la Probabilidad de Falla (PDF)

$$F(t; k, \lambda) = 1 - e^{-\left(\frac{t}{\lambda}\right)^k}$$

$$F(50; 2.5737, 59.9106) = 1 - e^{-(0.834)^{2.5737}} = 1 - e^{-0.5963}$$

$$F(50; 2.5737, 59.9106) = 1 - 0.5512 \approx 0.4488$$

Resultados Manuales

Probabilidad de falla en 50 horas (PDF): Aproximadamente **0.0143** (o **1.43%**).

Probabilidad acumulada de falla hasta 50 horas (CDF): Aproximadamente **0.4488** (o **44.88%**)

DISTRIBUCION WEIBULL

```
t = 50

# Calculo de PDF y CDF para el tiempo específico
pdf_at_t = stats.weibull_min.pdf(t, k, loc, scale)
cdf_at_t = stats.weibull_min.cdf(t, k, loc, scale)

# Resultados
print(f'Parámetro de forma (k): {k:.4f}')
print(f'Parámetro de escala (λ): {scale:.4f}')
print(f'Probabilidad de falla en 50 horas (PDF): {pdf_at_t:.4f}')
print(f'Probabilidad acumulada de falla hasta 50 horas (CDF): {cdf_at_t:.4f}')

# Graficar la distribución
x = np.linspace(0, 100, 100)

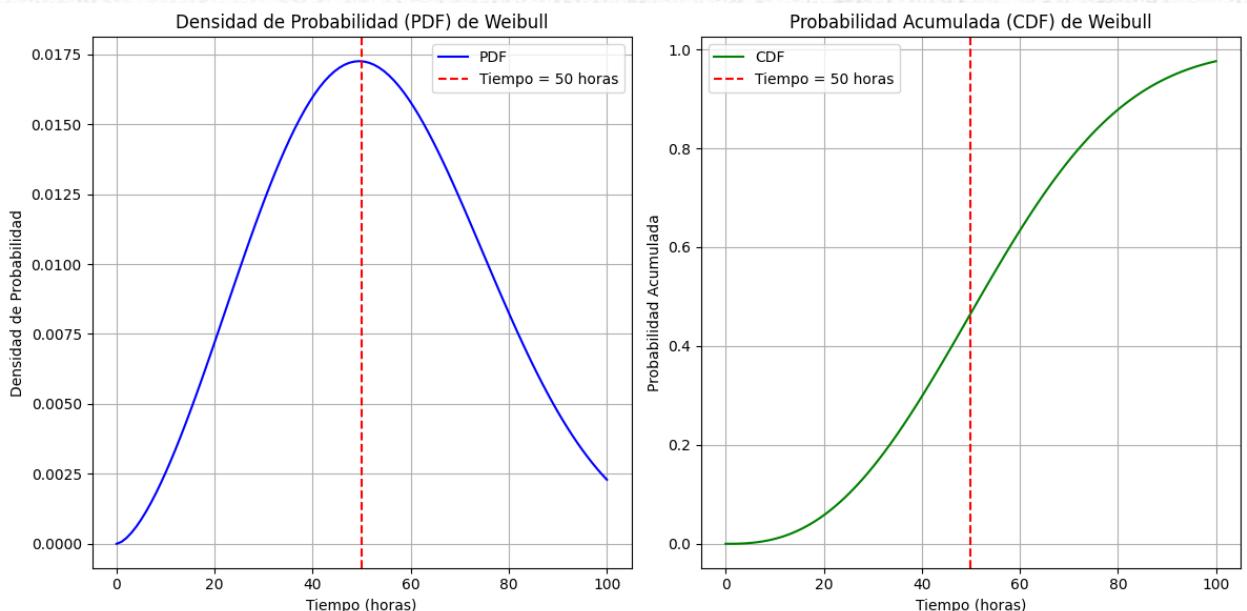
# Calcular PDF y CDF para todo el rango de x
pdf = stats.weibull_min.pdf(x, k, loc, scale)
cdf = stats.weibull_min.cdf(x, k, loc, scale)

plt.figure(figsize=(12, 6))

# Gráfico de PDF
plt.subplot(1, 2, 1)
plt.plot(x, pdf, label='PDF', color='blue')
plt.axvline(x=t, color='red', linestyle='--', label=f'Tiempo = {t} horas')
plt.title('Densidad de Probabilidad (PDF) de Weibull')
plt.xlabel('Tiempo (horas)')
plt.ylabel('Densidad de Probabilidad')
plt.legend()
plt.grid()
```

```
# Gráfico de CDF
plt.subplot(1, 2, 2)
plt.plot(x, cdf, label='CDF', color='green')
plt.axvline(x=t, color='red', linestyle='--', label=f'Tiempo = {t} horas')
plt.title('Probabilidad Acumulada (CDF) de Weibull')
plt.xlabel('Tiempo (horas)')
plt.ylabel('Probabilidad Acumulada')
plt.legend()
plt.grid()
```

Parámetro de forma (k): 2.5737
Parámetro de escala (λ): 59.9106
Probabilidad de falla en 50 horas (PDF): 0.0172
Probabilidad acumulada de falla hasta 50 horas (CDF): 0.4663



Distribución	Uso	Ejemplo en Confiabilidad	Características Clave
Normal	Datos simétricos alrededor de una media	Tiempos de reparación de equipos de mantenimiento, variaciones en la producción de piezas que cumplen especificaciones	Simétrica, media y desviación estándar, 68% dentro de $\pm 1\sigma$
Exponencial	Tiempos entre eventos independientes	Tiempos entre fallas de maquinaria que funcionan de manera continua, tiempos hasta el fallo de sistemas electrónicos	"Sin memoria", un solo parámetro λ
Lognormal	Datos positivos y sesgados	Tiempos de vida útil de componentes mecánicos, duración de la vida útil de productos que no siguen un ciclo de vida regular	Distribución de la variable logarítmica
Weibull	Análisis de confiabilidad y comportamiento variable	Tiempos hasta la falla de rodamientos en maquinaria, durabilidad de materiales compuestos bajo estrés	Parámetro de forma y escala, adaptable a diferentes patrones de falla

