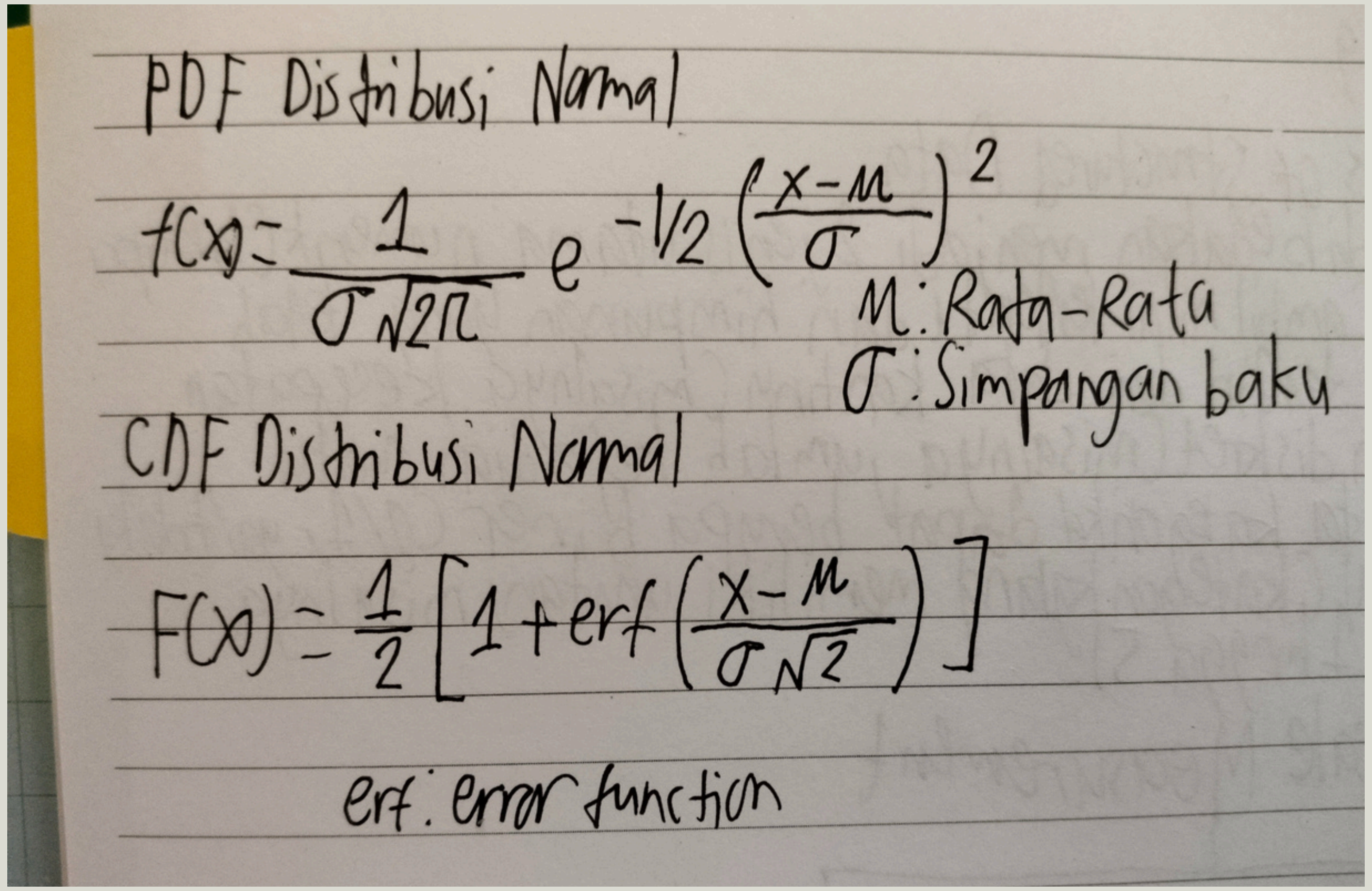
DATA DAN SAMPLING DISTRIBUTION

Benhard Leroy Situmorang/Statistika/20244920002

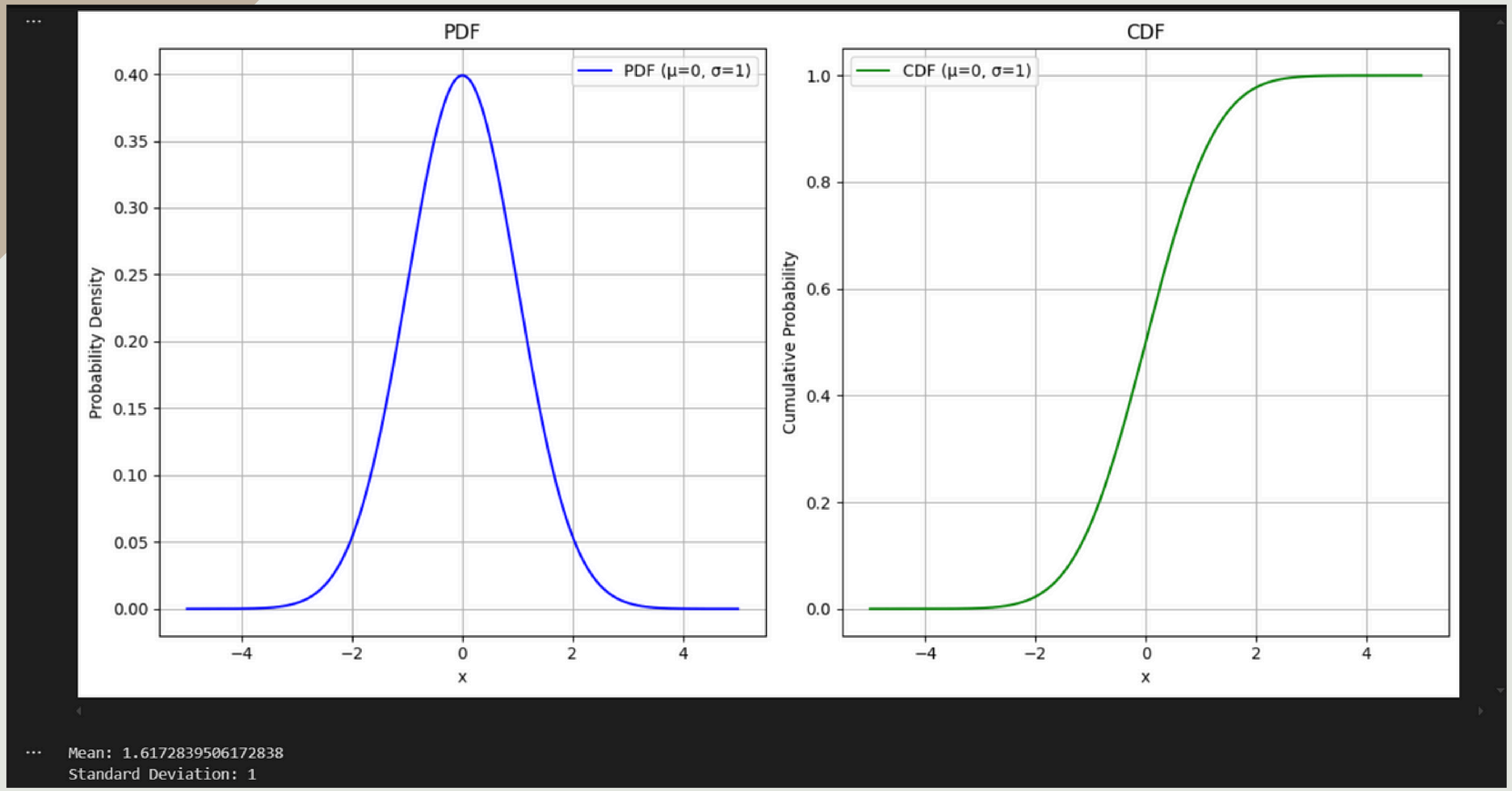
DISTRIBUSI NORMAL

Distribusi normal adalah sebuah distribusi probabilitas kontinu yang bentuknya itu digambarkan dengan kurva lonceng yang simetris. Distribusi ini biasanya digunakan untuk menggambarkan sebaran data yang ada dari variabel kontinu seperti tinggi badan, berat badan, atau tekanan darah.

PDF dan CDF:



VISUALISASI DATA




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

mean = 0
std_dev = 1

x = np.linspace(-5, 5, 1000)

pdf = norm.pdf(x, mean, std_dev)
cdf = norm.cdf(x, mean, std_dev)

plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.plot(x, pdf, label=f'PDF (μ={mean}, σ={std_dev})', color='blue')
plt.title('PDF')
plt.xlabel('x')
plt.ylabel('Probability Density')
plt.legend()
plt.grid()

plt.subplot(1, 2, 2)
plt.plot(x, cdf, label=f'CDF (μ={mean}, σ={std_dev})', color='green')
plt.title('CDF')
plt.xlabel('x')
plt.ylabel('Cumulative Probability')
plt.legend()
plt.grid()
plt.tight_layout()
plt.show()

print(f"Mean: {mean_value}")
print(f"Standard Deviation: {std_dev}")
```

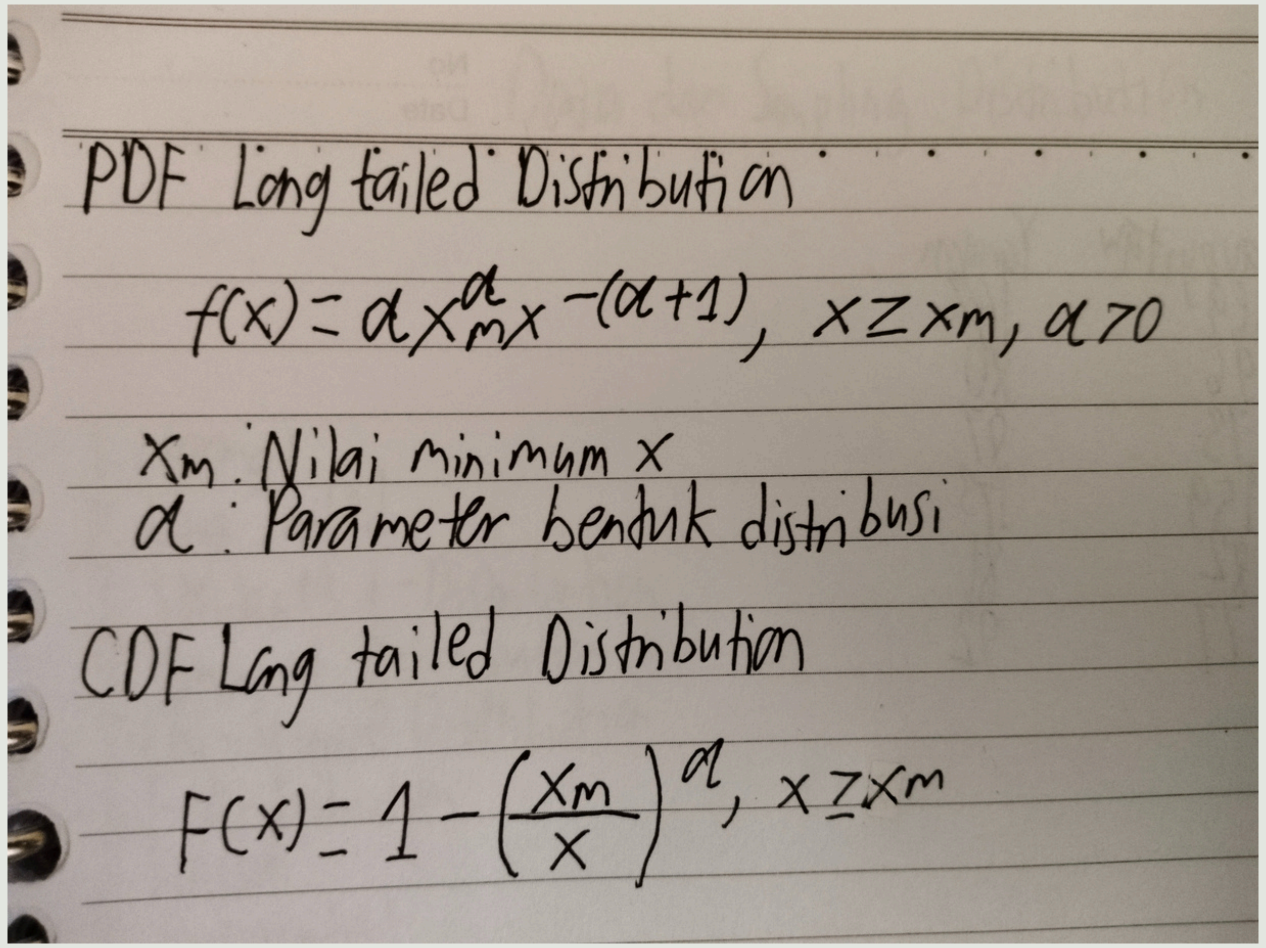
Mean = 1.617

Standar Deviasi = 1

DISTRIBUSI LONG-TAILED

Distribusi long-tailed adalah distribusi yang mempunyai sebuah ekor panjang yang perlahan-lahan mengecil menjelang akhir distribusinya. Ini mewakili beberapa jaush sebaran data dari nilai tengah atau kepala. Distribusinya dapat melibatkan popularitas, jumlah kemunculan peristiwa secara acak dengan berbagai probabilitas, dll.

PDF dan CDF:



PDF Long tailed Distribution

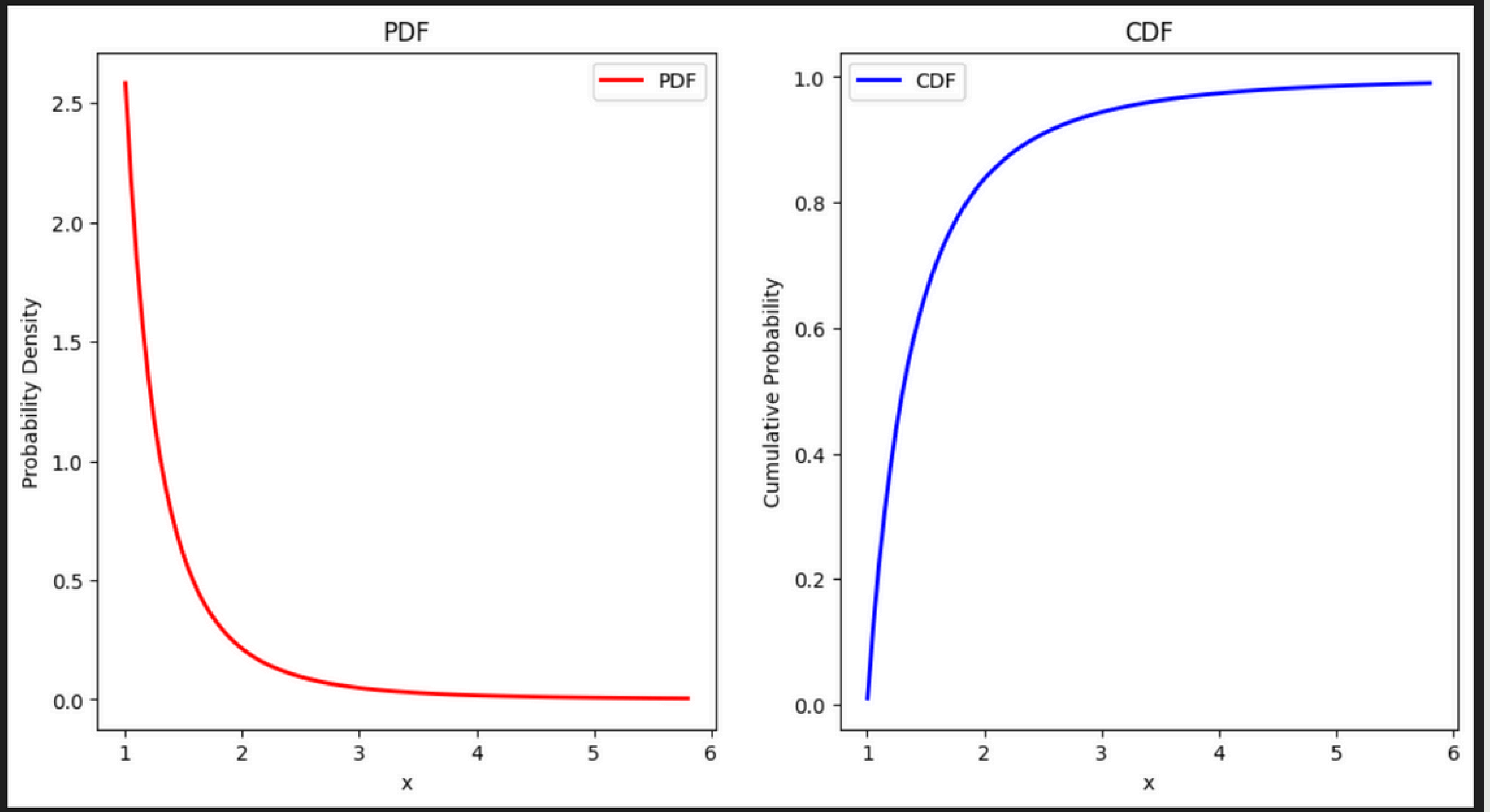
$$f(x) = \alpha x_m^{\alpha} x^{-(\alpha+1)}, \quad x \geq x_m, \alpha > 0$$

x_m : Nilai minimum x
α : Parameter bentuk distribusi

CDF Long tailed Distribution

$$F(x) = 1 - \left(\frac{x_m}{x}\right)^{\alpha}, \quad x \geq x_m$$

VISUALISASI DATA




```
Mean: 1.6172839506172838
Standard Deviation: 1.2689361980370224
```

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

b = 2.62
mean, var, skew, kurt = pareto.stats(b, moments='mvsk')
x = np.linspace(pareto.ppf(0.01, b), pareto.ppf(0.99, b), 100)

pdf = pareto.pdf(x, b)
cdf = pareto.cdf(x, b)

mean_value = pareto.mean(b)
std_dev = pareto.std(b)

plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.plot(x, pdf, 'r-', lw=2, label='PDF')
plt.title('PDF')
plt.xlabel('x')
plt.ylabel('Probability Density')
plt.legend()

plt.subplot(1, 2, 2)
plt.plot(x, cdf, 'b-', lw=2, label='CDF')
plt.title('CDF')
plt.xlabel('x')
plt.ylabel('Cumulative Probability')
plt.legend()
plt.show()

print(f"Mean: {mean_value}")
print(f"Standard Deviation: {std_dev}")
```

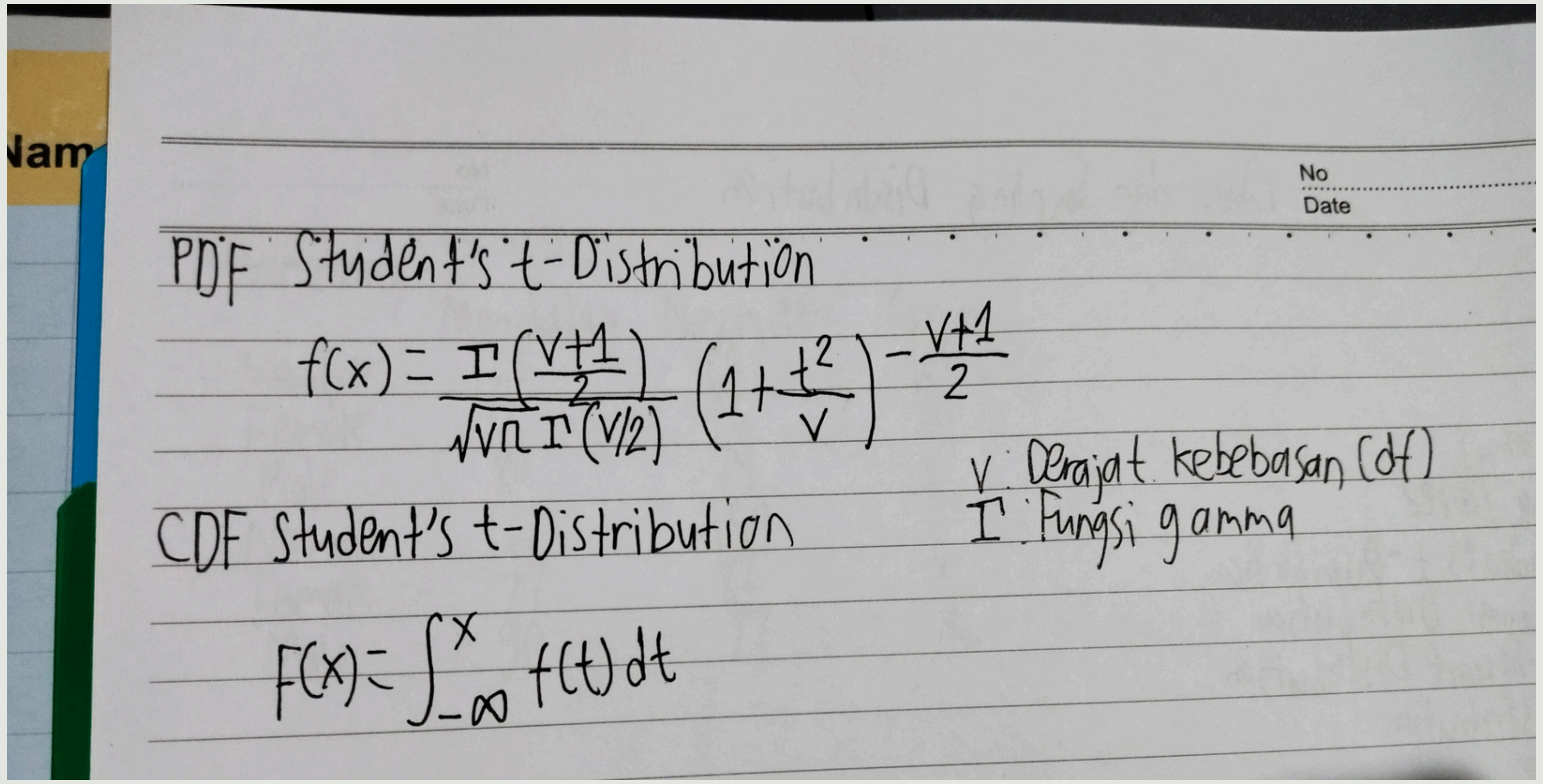
Mean = 1.617

Standar deviasi = 1.268

DISTRIBUSI STUDENT'S T

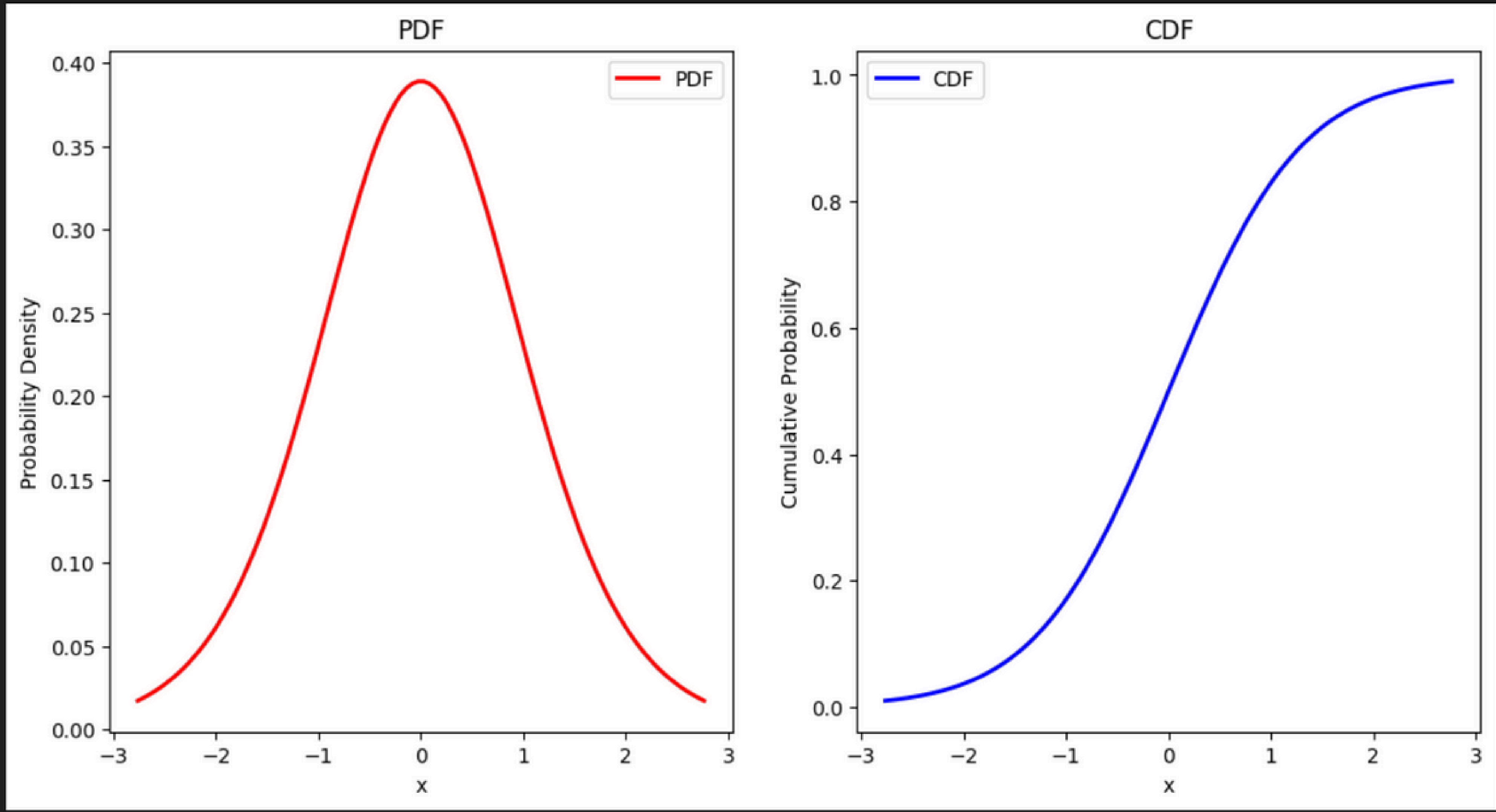
Distribusi Student's t adalah sebuah distribusi probabilitas yang sering digunakan dalam inferensi statistik, terutama ketika ukuran sampel kecil dan standar deviasi populasi tidak diketahui. Student's t-distribution bisa memperkirakan parameter yang tidak diketahui dengan menggunakan distribusi sampel untuk mengubah standar deviasi populasi.

PDF dan CDF:



VISUALISASI DATA

```
Mean: 0.0
Standar Deviasi: 1.118033988749895
```




```python
import numpy as np
import scipy.stats as stats
import matplotlib.pyplot as plt

df = 10
mean, var, skew, kurt = stats.t.stats(df, moments='mvsk')

x = np.linspace(stats.t.ppf(0.01, df), stats.t.ppf(0.99, df), 100)
pdf = stats.t.pdf(x, df)
cdf = stats.t.cdf(x, df)

std_dev = np.sqrt(var)
print(f"Mean: {mean}")
print(f"Standar Deviasi: {std_dev}")

plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.plot(x, pdf, 'r-', lw=2, label='PDF')
plt.title('PDF')
plt.xlabel('x')
plt.ylabel('Probability Density')
plt.legend()

plt.subplot(1, 2, 2)
plt.plot(x, cdf, 'b-', lw=2, label='CDF')
plt.title('CDF')
plt.xlabel('x')
plt.ylabel('Cumulative Probability')
plt.legend()

plt.show()
```

Mean = 0

Standar deviasi = 1.118

DISTRIBUSI BINOMIAL

Distribusi Binomial adalah sebuah distribusi probabilitas yang menggambarkan jumlah keberhasilan dalam beberapa n percobaan secara independen. Distribusi ini digunakan dengan 2 kemungkinan saja seperti koin/dadu.

PDF dan CDF:

PDF Binomial Distribution

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$$

k: Jumlah keberhasilan

n: Jumlah percobaan

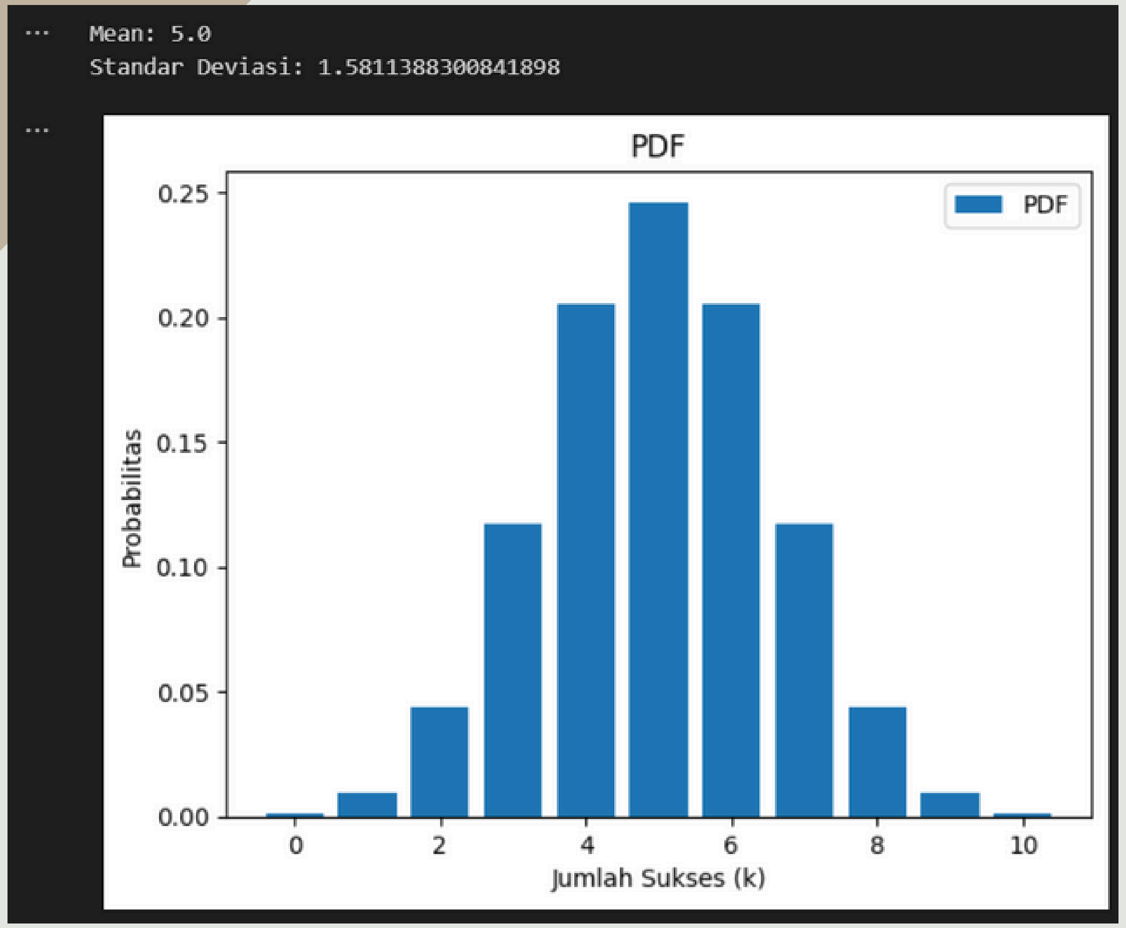
p: Probabilitas keberhasilan

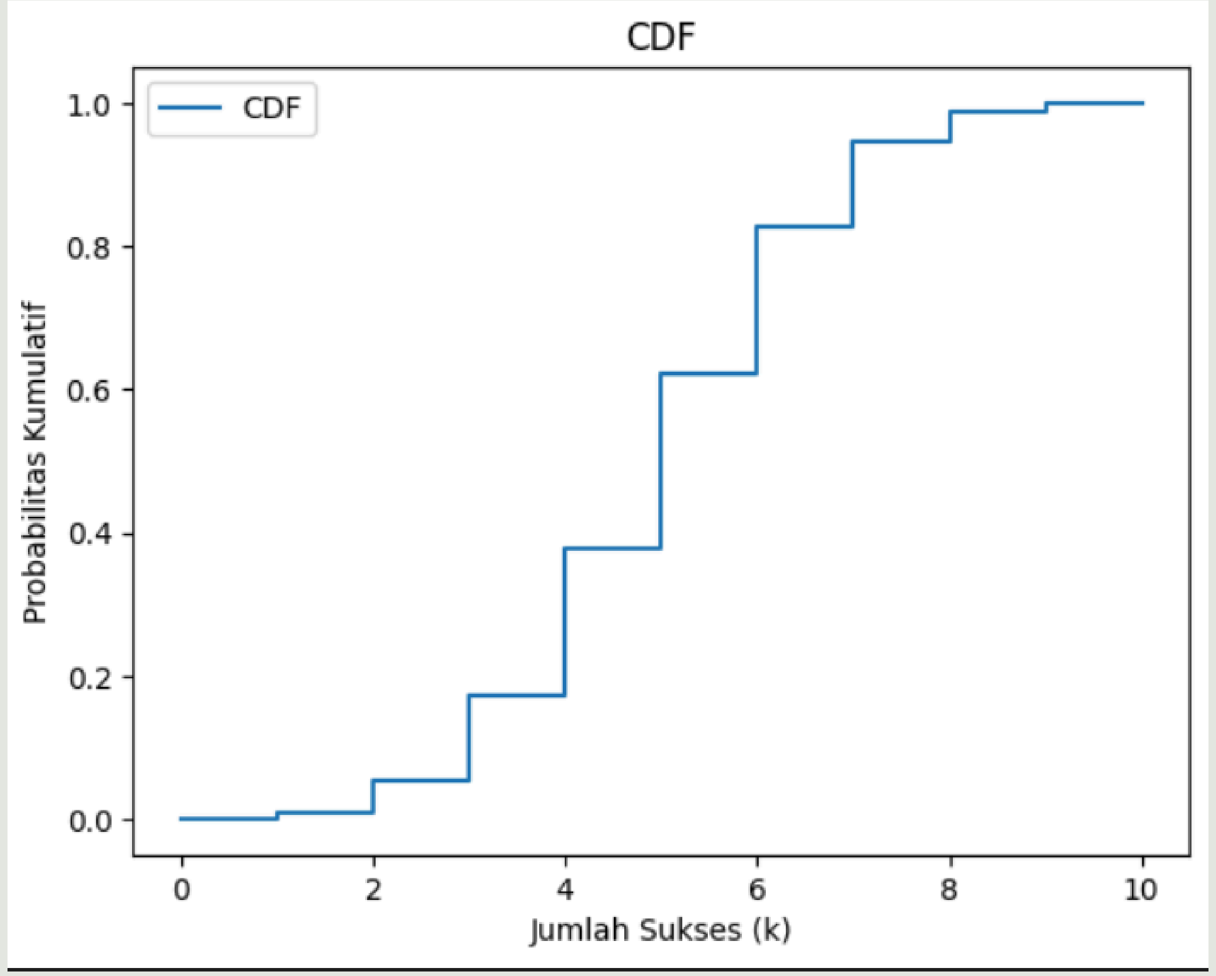
$\binom{n}{k}$: Kombinasi $= \frac{n!}{k!(n-k)!}$

CDF Binomial Distribution

$$F(x)=P(X\leq k)=\sum_{i=0}^{k} P(X=i)$$

VISUALISASI DATA






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

n = 10
p = 0.5

k_values = np.arange(0, n+1)
pdf_values = binom.pmf(k_values, n, p)
cdf_values = binom.cdf(k_values, n, p)

mean = binom.mean(n, p)
std_dev = binom.std(n, p)

print(f"Mean: {mean}")
print(f"Standar Deviasi: {std_dev}")

plt.bar(k_values, pdf_values, label='PDF')
plt.title('PDF')
plt.xlabel('Jumlah Sukses (k)')
plt.ylabel('Probabilitas')
plt.legend()
plt.show()

plt.step(k_values, cdf_values, label='CDF', where='post')
plt.title('CDF')
plt.xlabel('Jumlah Sukses (k)')
plt.ylabel('Probabilitas Kumulatif')
plt.legend()
plt.show()
```

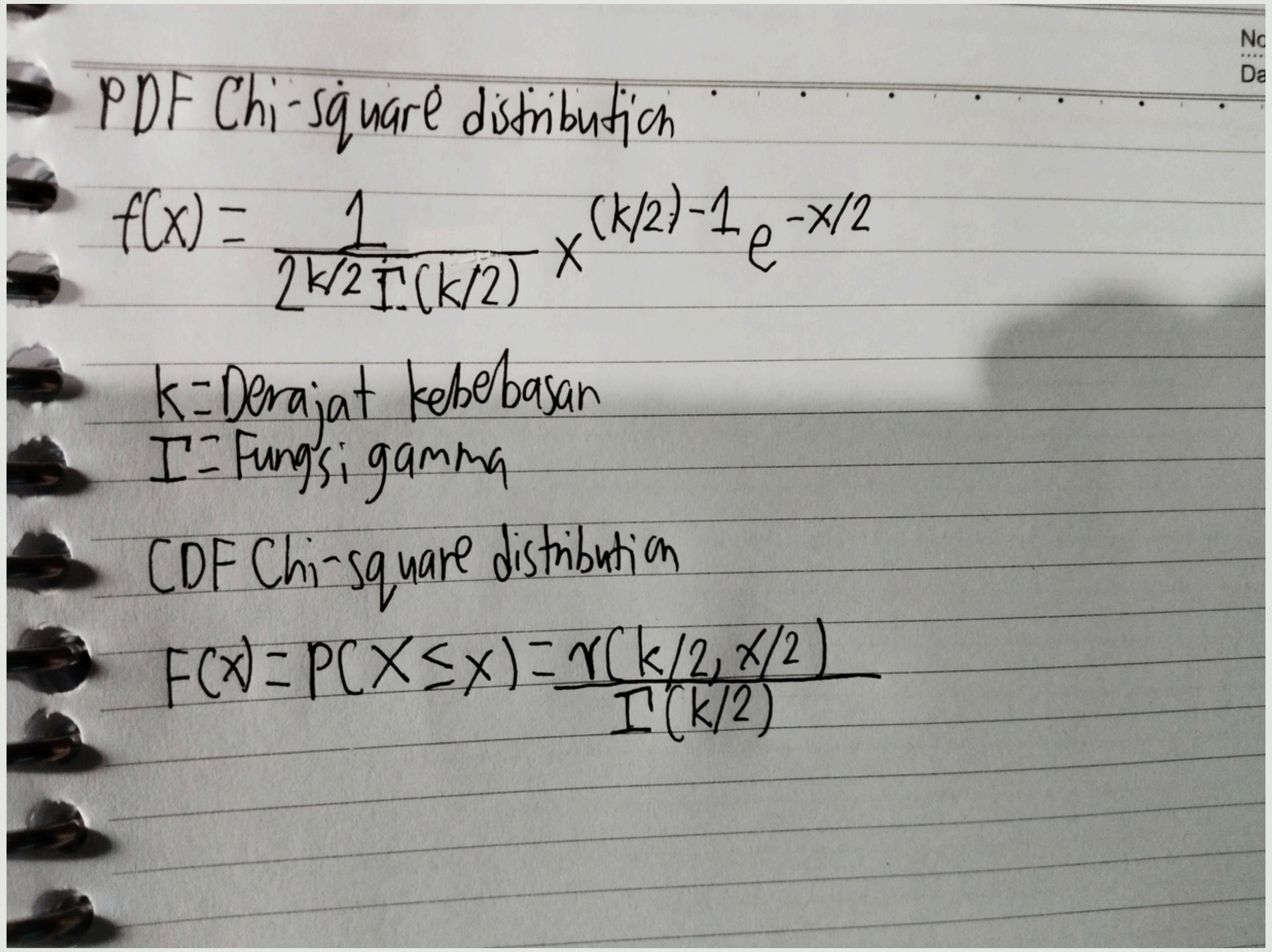
Mean = 5

Standar deviasi = 1.581

DISTRIBUSI CHI-SQUARE

Distribusi Chi-square adalah distribusi jumlah kuadrat k variabel acak normal standar independen. Distribusi ini menjadi solusi pengujian sampel terutama dalam kasus uji kesesuaian.

PDF dan CDF:



PDF Chi-square distribution

$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2}$$

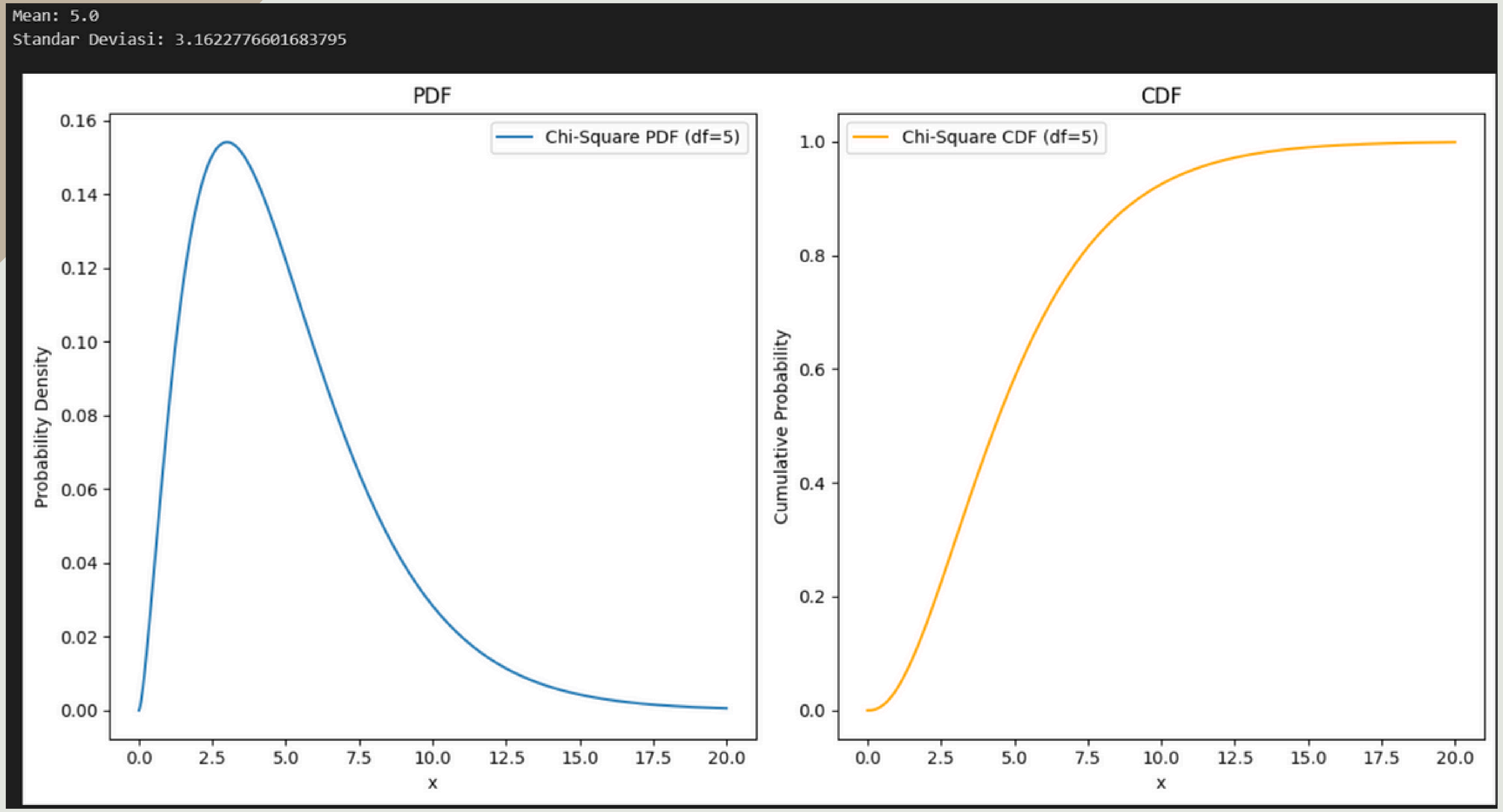
k = Derajat kebebasan

Γ = Fungsi gamma

CDF Chi-square distribution

$$F(x) = P(X \le x) = \frac{\gamma(k/2, x/2)}{\Gamma(k/2)}$$

VISUALISASI DATA




```python
import numpy as np
import scipy.stats as stats
import matplotlib.pyplot as plt

df = 5

mean, var = stats.chi2.stats(df, moments='mv')
std_dev = np.sqrt(var)

print(f"Mean: {mean}")
print(f"Standar Deviasi: {std_dev}")

x = np.linspace(0, 20, 1000)

pdf = stats.chi2.pdf(x, df)
cdf = stats.chi2.cdf(x, df)

plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.plot(x, pdf, label=f'Chi-Square PDF (df={df})')
plt.title('Probability Density Function (PDF)')
plt.xlabel('x')
plt.ylabel('Probability Density')
plt.legend()

plt.subplot(1, 2, 2)
plt.plot(x, cdf, label=f'Chi-Square CDF (df={df})', color='orange')
plt.title('Cumulative Distribution Function (CDF)')
plt.xlabel('x')
plt.ylabel('Cumulative Probability')
plt.legend()
plt.tight_layout()
plt.show()
```

Mean = 5

Standar deviasi = 3.162

DISTRIBUSI F

Distribusi F adalah sebuah distribusi yang memiliki peran penting dalam statistika, khususnya dalam konteks perancangan eksperimen. Distribusi F digunakan untuk menentukan apakah varians antar kelompok data tersebut signifikan atau hanya hasil dari variasi yang wajar.

PDF dan CDF:

PDF F-Distribution

$$f(x) = \frac{\sqrt{\dfrac{(d1x)^{d1}\, d_2^{\,d2}}{(d1x+d2)^{d1+d2}}}}{x\,B\left(\frac{d1}{2}, \frac{d2}{2}\right)}$$

d1: Derajat kebebasan pembilang
d2: Derajat kebebasan penyebut
B: Beta Function

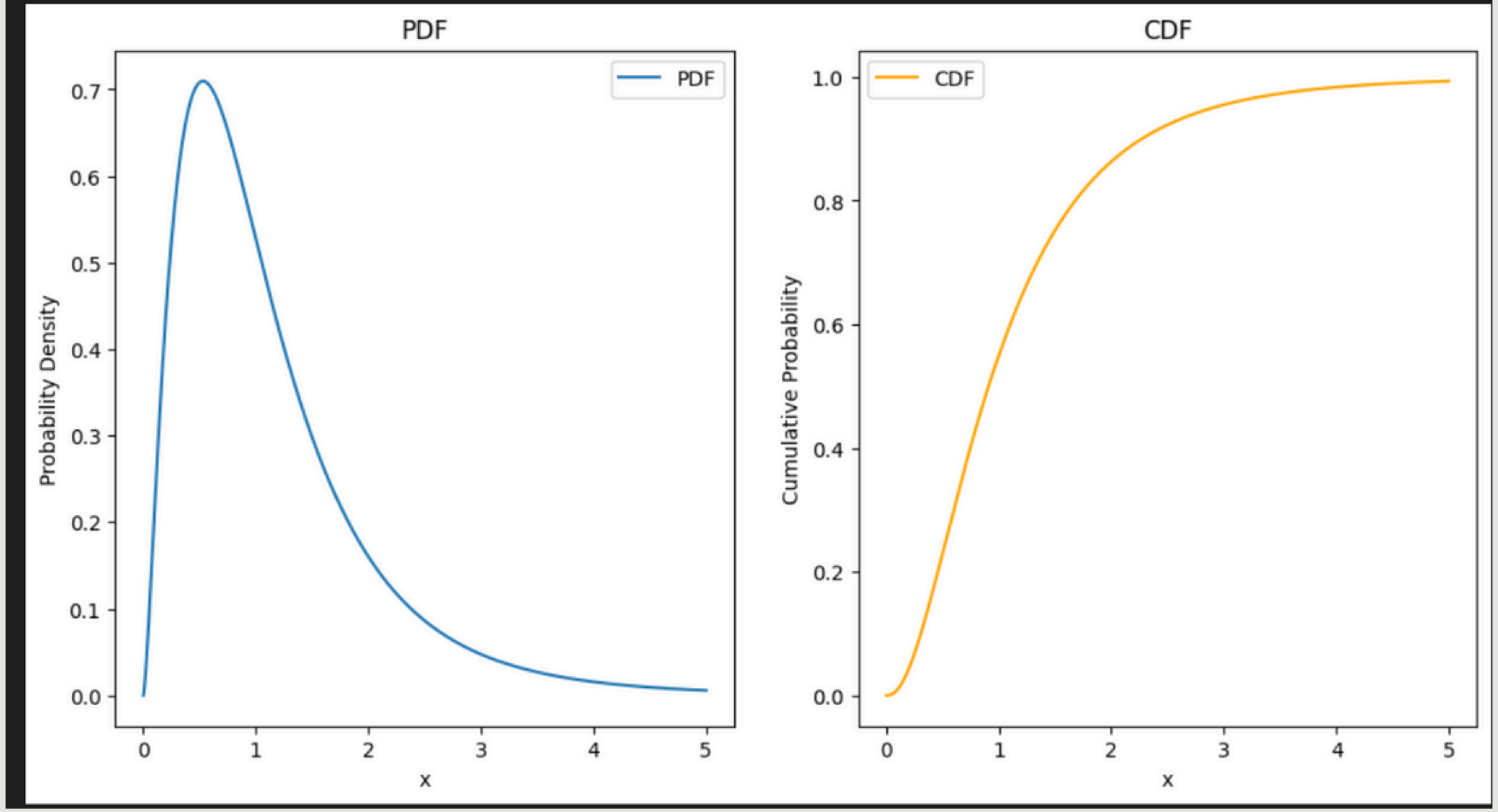
CDF F-Distribution

$$F(x) = I_{\frac{d1x}{d1x+d2}}\left(\frac{d1}{2}, \frac{d2}{2}\right)$$

I: Fungsi Beta tidak lengkap yg diatur

VISUALISASI DATA

```
Mean: 1.1538461538461537
Standar Deviasi: 0.9335075018721813
```




```python
import numpy as np
import scipy.stats as stats
import matplotlib.pyplot as plt

dfn = 5
dfd = 15

x = np.linspace(0, 5, 1000)

pdf = stats.f.pdf(x, dfn, dfd)
cdf = stats.f.cdf(x, dfn, dfd)

mean, var = stats.f.stats(dfn, dfd, moments='mv')
std_dev = np.sqrt(var)

print(f"Mean: {mean}")
print(f"Standar Deviasi: {std_dev}")

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

plt.subplot(1, 2, 1)
plt.plot(x, pdf, label='PDF')
plt.title('PDF')
plt.xlabel('x')
plt.ylabel('Probability Density')
plt.legend()

plt.subplot(1, 2, 2)
plt.plot(x, cdf, label='CDF', color='orange')
plt.title('CDF')
plt.xlabel('x')
plt.ylabel('Cumulative Probability')
plt.legend()
plt.show()
```

Mean= 1.153

Standar Deviasi= 0.933

DISTRIBUSI POISSON

Distribusi Poisson adalah distribusi probabilitas yang digunakan untuk memodelkan jumlah kejadian dalam interval waktu atau ruang tertentu, dengan asumsi bahwa kejadian tersebut terjadi secara independen dengan laju rata-rata tertentu.

PDF dan CDF:

No

Date

PDF Poisson Distribution

$$f(x) = \frac{\lambda^k e^{-\lambda}}{k!}$$

CDF Poisson Distribution

$$F(x) = \sum_{x=0}^{|k|} \frac{e^{-\lambda} \lambda^x}{x!}$$

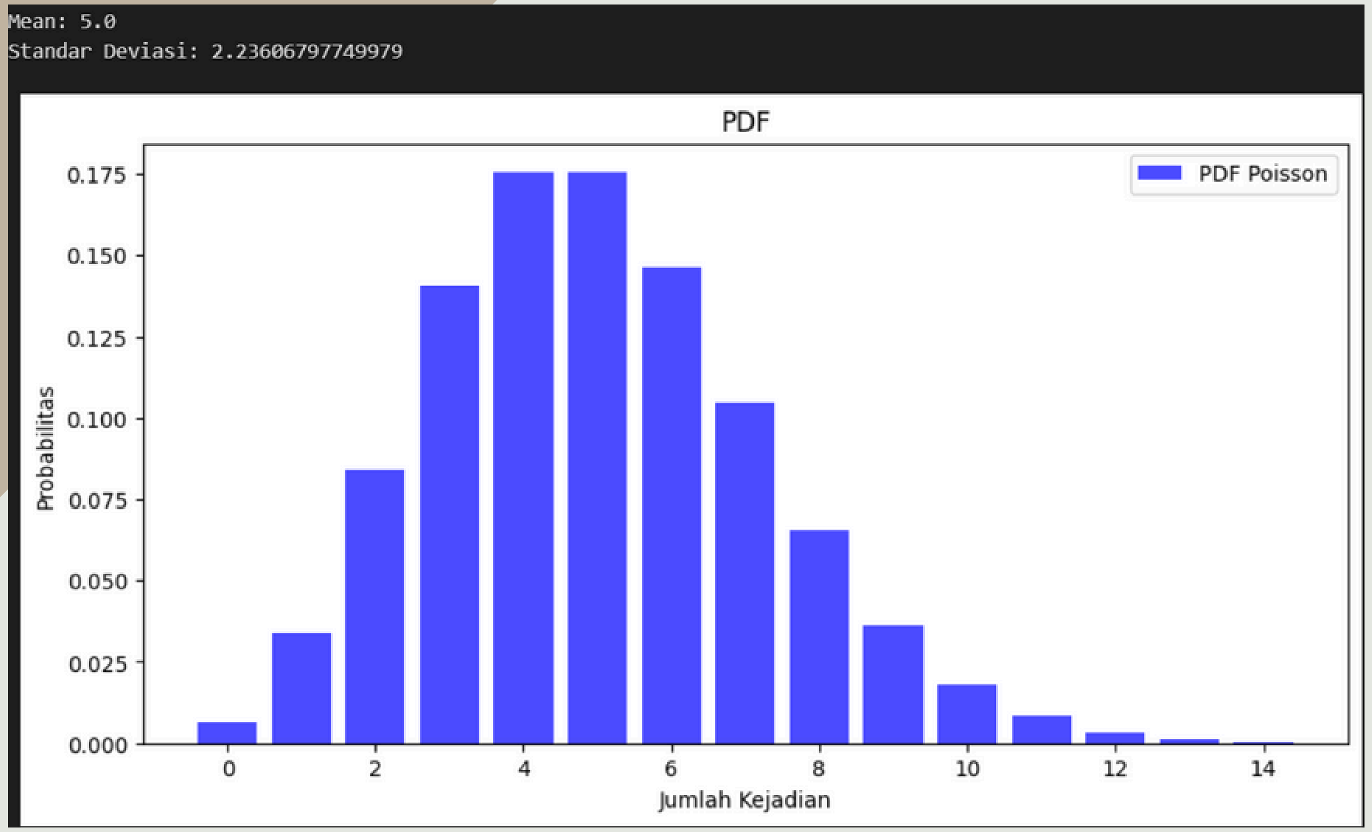
k: Jumlah kejadian yang diamati

λ: Rata-rata kejadian/interval

e: Bilangan Euler

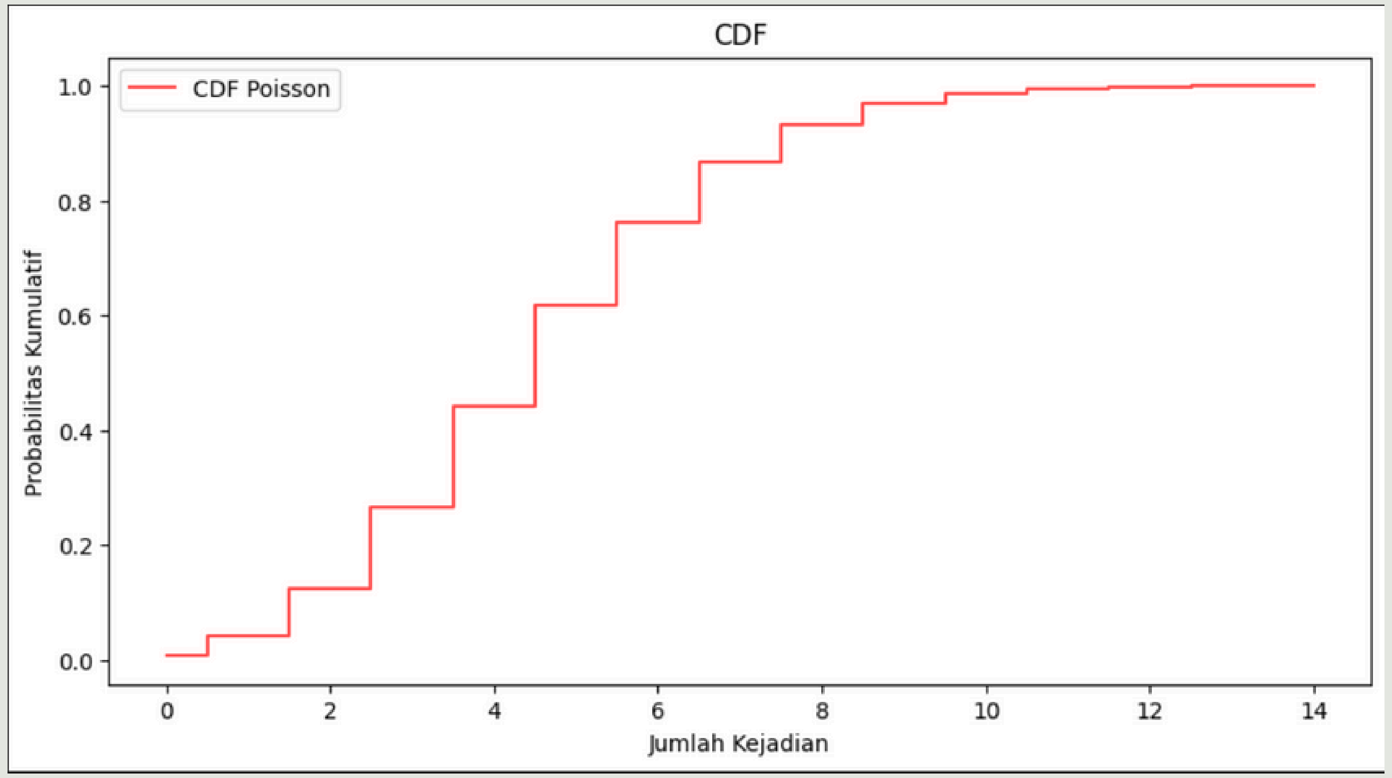
k!: Faktorial k

VISUALISASI DATA



Mean= 5

Standar Deviasi= 2.236




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

lambda_ = 5

x = np.arange(0, 15)

pdf = poisson.pmf(x, lambda_)
cdf = poisson.cdf(x, lambda_)

mean, var = poisson.stats(lambda_, moments='mv')
std_dev = np.sqrt(var)

print(f"Mean: {mean}")
print(f"Standar Deviasi: {std_dev}")

plt.figure(figsize=(10, 5))
plt.bar(x, pdf, color='blue', alpha=0.7, label='PDF Poisson')
plt.title('PDF')
plt.xlabel('Jumlah Kejadian')
plt.ylabel('Probabilitas')
plt.legend()
plt.show()

plt.figure(figsize=(10, 5))
plt.step(x, cdf, where='mid', color='red', alpha=0.7, label='CDF Poisson')
plt.title('CDF')
plt.xlabel('Jumlah Kejadian')
plt.ylabel('Probabilitas Kumulatif')
plt.legend()
plt.show()
```

DISTRIBUSI EKSPONENSIAL

Distribusi Eksponensial adalah sebuah distribution yang memodelkan waktu antara kejadian dalam suatu proses yang terjadi secara kontinu dan acak. Distribusi eksponensial umumnya digunakan oleh para analisis dalam berbagai kebutuhan. Penggunaannya juga tepat untuk mengetahui hasil pasti dari sebuah analisis data.

PDF dan CDF:

PDF Exponential Distribution

$f(x) = \lambda e^{\lambda x}$

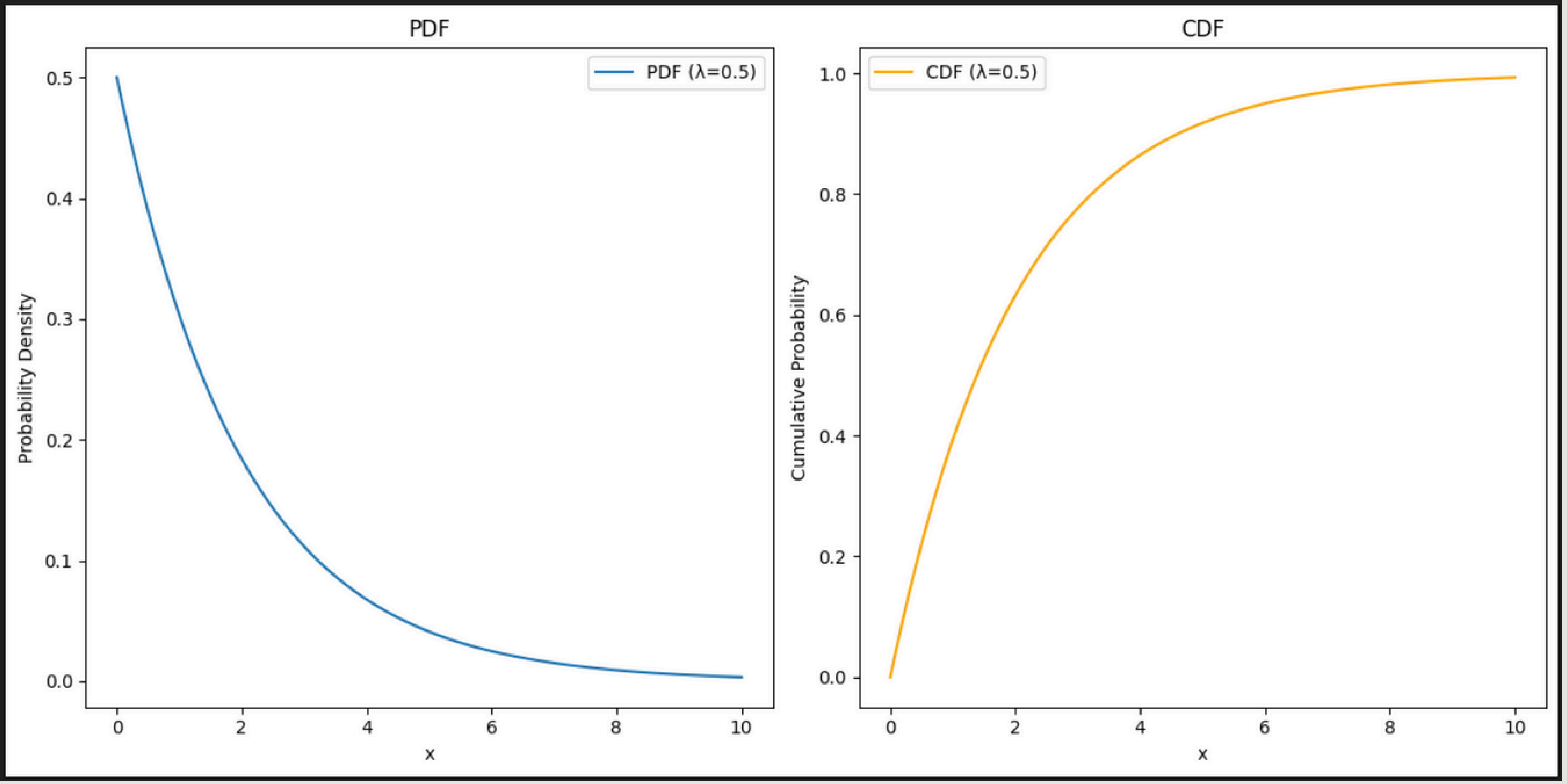
λ = Parameter tingkat kejadian
e = Bilangan Euler
x = waktu kejadian

CDF Exponential Distribution

$1 - e^{-\lambda x}$

VISUALISASI DATA

```
Mean: 2.0
Standar Deviasi: 2.0
```




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

# Parameter distribusi eksponensial
lambda_ = 0.5  # Parameter rate (lambda)
mean, var = expon.stats(scale=1/lambda_, moments='mv')
std_dev = np.sqrt(var)

print(f"Mean: {mean}")
print(f"Standar Deviasi: {std_dev}")

x = np.linspace(0, 10, 1000)

plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.plot(x, pdf, label=f'PDF (λ={lambda_})')
plt.title('PDF')
plt.xlabel('x')
plt.ylabel('Probability Density')
plt.legend()

plt.subplot(1, 2, 2)
plt.plot(x, cdf, label=f'CDF (λ={lambda_})', color='orange')
plt.title('CDF')
plt.xlabel('x')
plt.ylabel('Cumulative Probability')
plt.legend()
plt.tight_layout()
plt.show()
```

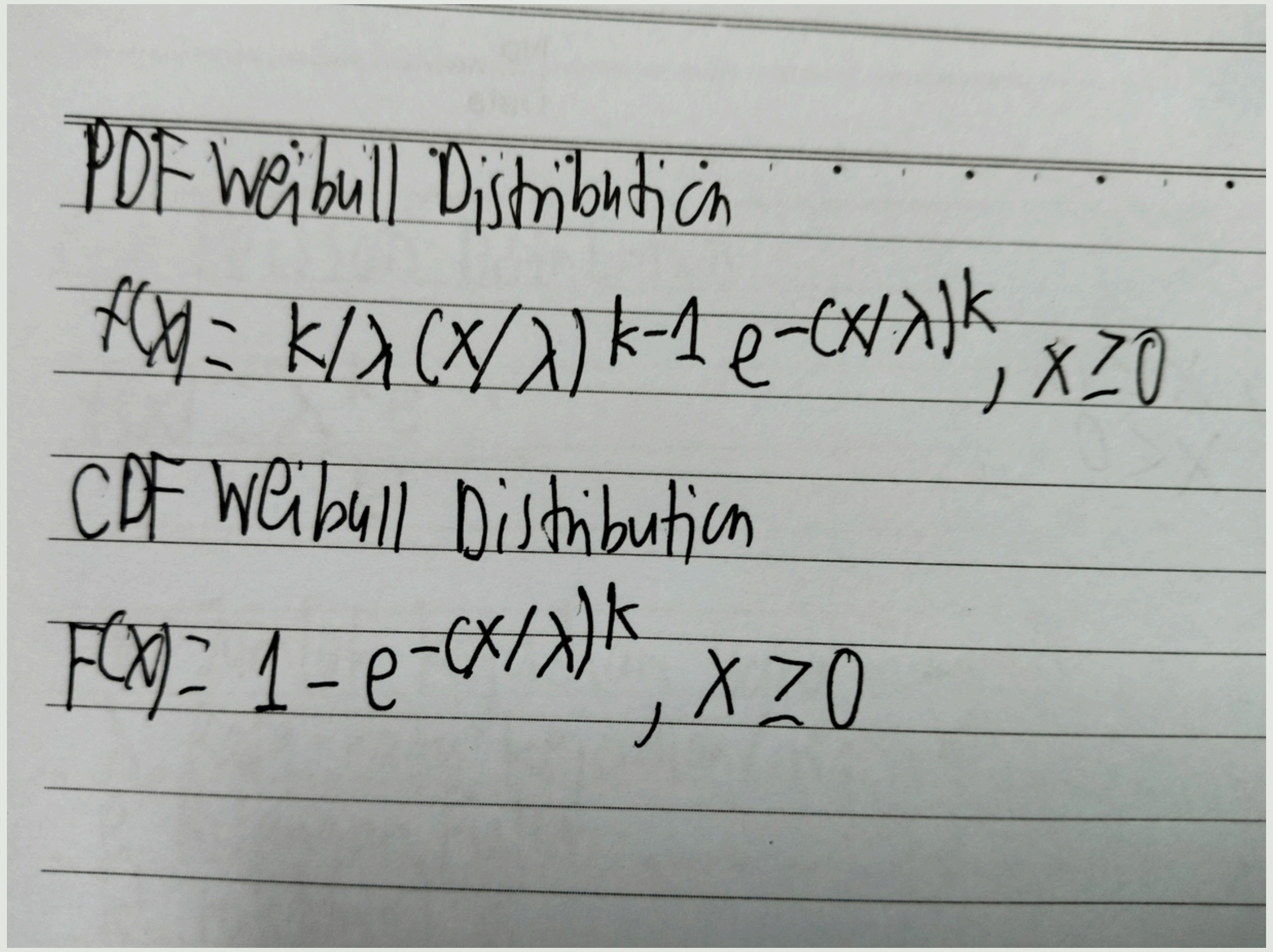
Mean= 2

Standar Deviasi= 2

DISTRIBUSI WEIBULL

Distribusi weibull merupakan salah satu metode stasistika yang sering digunakan untuk menjelaskan data kecepatan angin. Distribusi ini adalah metode pencocokan yang tepat untuk data ekperimen.

PDF dan CDF:



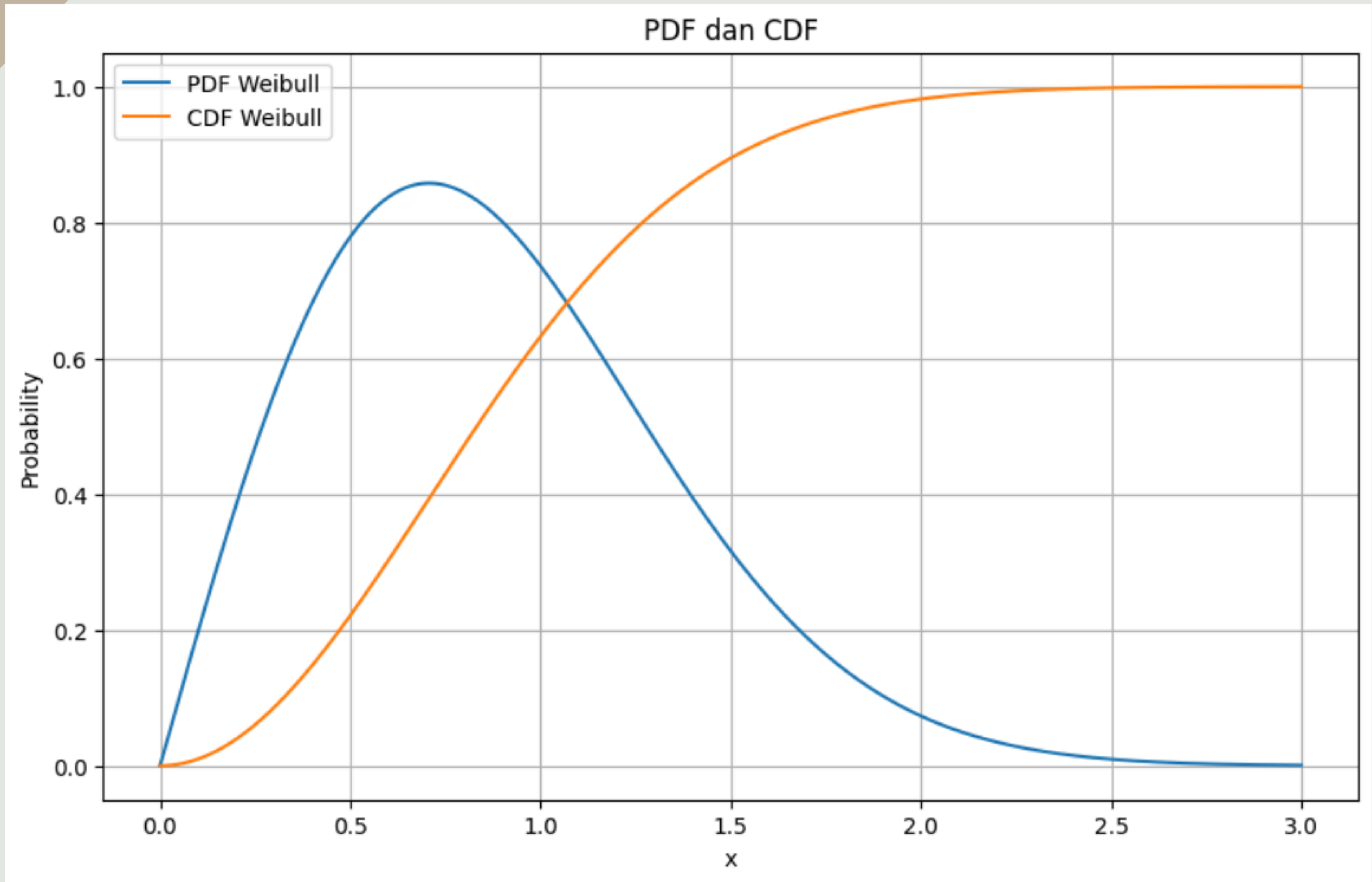
PDF Weibull Distribution

$$f(x) = k/\lambda \, (x/\lambda)^{k-1} e^{-(x/\lambda)^k}, \; x \geq 0$$

CDF Weibull Distribution

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

VISUALISASI DATA




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

shape = 2.0
scale = 1.0

x = np.linspace(0, 3, 1000)

pdf = weibull_min.pdf(x, shape, scale=scale)
cdf = weibull_min.cdf(x, shape, scale=scale)

mean = weibull_min.mean(shape, scale=scale)
std_dev = weibull_min.std(shape, scale=scale)

print(f"Mean: {mean}")
print(f"Standard Deviation: {std_dev}")

plt.figure(figsize=(10, 6))
plt.plot(x, pdf, label='PDF Weibull')
plt.plot(x, cdf, label='CDF Weibull')
plt.title('PDF dan CDF')
plt.xlabel('x')
plt.ylabel('Probability')
plt.legend()
plt.grid(True)
plt.show()
```

Mean= 0.886

Standar Deviasi= 0.463

TABEL PERBANDINGAN

Distribution	Mean	Standar deviasi
Normal	1.617	1
Long-tailed	1.617	1.268
Student's T	0	1.118
Binomial	5	1.581
Chi-Square	5	3.162
F	1.153	0.933
Poisson	5	2.236
Exponential	2	2
Weibull	0.886	0.463

Kesimpulan dari tabel ini memberikan sebuah gambaran tentang karakteristik sentral dan dispersi dari berbagai distribusi probabilitas dari mean dan standar deviasi. Ada yang sama seperti distribusi Eksponensial dan ada juga yang lebih jauh beda seperti distribusi binomial dan poisson. Apa yang dapat dianalisis data distribusi itu membantu untuk memilih distribusi yang sesuai untuk analisis statistik tertentu, terutama dengan perbandingan seperti di tabel itu.

Terima Kasih