

Estadística_descriptiva

May 14, 2020

```
[2]: import numpy as np
from scipy import stats
import pandas as pd

import os
from pandas import ExcelWriter

import matplotlib.pyplot as plt
import seaborn as sns

%matplotlib inline

[3]: tablaN01_df = pd.read_excel('input_file/Tabla_05.xlsx', 'Hoja7')
df = pd.DataFrame(tablaN01_df) # df es nuestra variable a utilizar
a = type(tablaN01_df)
b = type(df)
print(df.head())
print("\n" )
print("La variable 'tablaN01_df' es de tipo: "+str(a))
print("La variable 'df' es de tipo: "+str(b))
if a == b:
    print("'tablaN01_df' y 'df' son el mismo tipo de dato.")
print("\n")
print("Las dimensiones de la tabla son: ")
print("Total de filas: "+str(len(df.index)))
print("Total de columnas: "+str(len(df.columns)))
print("\n")
print("El tipo de dato de la variable 'df' es: "+str(type(df)))
print("\n")
print("ESTADÍSTICA DESCRIPTIVA")
print("Media aritmetica:")
media_aritmetica_01 = df.mean()
print(media_aritmetica_01)
print("Media aritmética para cada columna:")
media_aritmetica_02 = df.mean(axis=1)
print(media_aritmetica_02)
#print("Media aritmética para cada fila:")
```

```

#media_aritmetica_02 = df.mean(axis=1)
#print(media_aritmetica_02)
#print("Mediana:")
#print("Mediana ")
#mediana_01 = np.median(df)
#print(mediana_01)
print("Mediana por cada columna:")
mediana_02 = np.median(df, 0)
print(mediana_02)
#print("Desviación:")
#desviacion_01 = np.std(df)
#print(desviacion_01)
print("Desviación típica de cada columna:")
desviacion_02 = np.std(df, 0)
print(desviacion_02)
#print("Varianza")
#varianza_01 = np.var(df)
#print(varianza_01)
print("Varianza de cada columna:")
varianza_02 = np.var(df, 0)
print(varianza_02)
print("Moda:")
moda_01 = stats.mode(df)
print(moda_01)
#print("Correlación:")
#correlacion_01 = np.corrcoef(df)
#print(correlacion_01)
#print("Co
#print("Covarianza:")
#covarianza_01 = np.cov(df)
#print(covarianza_01)

resumen = df.describe()
print("Resumen estadístico:")
print(resumen)

print("Sumando el valor de cada columna:")
suma = df.sum()
print(suma)

print("Acumulado de las sumas por columnas")
acumulado = df.cumsum()
print(acumulado)

#print("Media aritmetica:")
#media_aritmetica = df.
#pri

```

```

print("\n")
print("HISTOGRAMA DE DISTRIBUCIÓN")

sns.set_palette("deep", desat=.6)
sns.set_context(rc={"figure.figsize": (8,4)})

mu, sigma = 0, 0.1 # media y desvío estándar
s = np.random.normal(mu, sigma, 100)

cuenta, cajas, ignorar = plt.hist(s, 30, density=True)
normal = plt.plot(cajas, 1/(sigma * np.sqrt(2*np.pi)) *
                  np.exp( - (cajas -mu)**2 / ( 2*sigma**2)),
                  linewidth=2, color='r')

plt.show()

```

	SEDAPAL	H&O - ADR
0	4.794392	3.718465
1	4.635618	3.652710
2	5.010507	3.879566
3	4.804975	4.106422
4	4.763774	3.771070

La variable 'tablaN01_df' es de tipo: <class 'pandas.core.frame.DataFrame'>
 La variable 'df' es de tipo: <class 'pandas.core.frame.DataFrame'>
 'tablaN01_df' y 'df' son el mismo tipo de dato.

Las dimensiones de la tabla son:

Total de filas: 433
 Total de columnas: 2

El tipo de dato de la variable 'df' es: <class 'pandas.core.frame.DataFrame'>

ESTADÍSTICA DESCRIPTIVA

Media aritmetica:

SEDAPAL	4.891346
H&O - ADR	4.381972

dtype: float64

Media aritmética para cada columna:

0	4.256429
1	4.144164
2	4.445037
3	4.455698

```

4      4.267422
...
428    4.491630
429    5.071717
430    4.653496
431    4.540694
432    4.607437
Length: 433, dtype: float64
Mediana por cada columna:
[4.96175473 4.21163034]
Desviación típica de cada columna:
SEDAPAL      0.314227
H&O - ADR    0.530033
dtype: float64
Varianza de cada columna:
SEDAPAL      0.098738
H&O - ADR    0.280935
dtype: float64
Moda:
ModeResult(mode=array([[3.79160687, 3.87956595]]), count=array([[1, 5]]))
Resumen estadístico:

```

	SEDAPAL	H&O - ADR
count	433.000000	433.000000
mean	4.891346	4.381972
std	0.314590	0.530646
min	3.791607	3.350235
25%	4.763774	3.915731
50%	4.961755	4.211630
75%	5.042730	4.960474
max	5.431105	5.431105

```

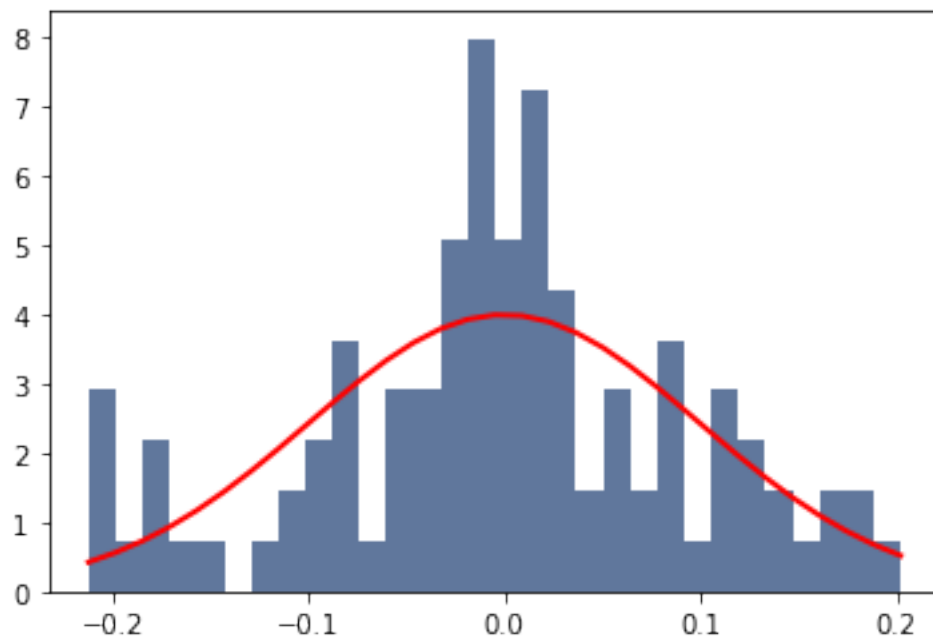
Sumando el valor de cada columna:
SEDAPAL      2117.952986
H&O - ADR    1897.393997
dtype: float64
Acumulado de las sumas por columnas

```

	SEDAPAL	H&O - ADR
0	4.794392	3.718465
1	9.430010	7.371175
2	14.440517	11.250741
3	19.245492	15.357163
4	24.009266	19.128233
..
428	2097.886605	1879.713691
429	2102.958322	1884.785407
430	2107.938612	1889.112110
431	2112.933304	1893.198805
432	2117.952986	1897.393997

[433 rows x 2 columns]

HISTOGRAMA DE DISTRIBUCIÓN



[]: