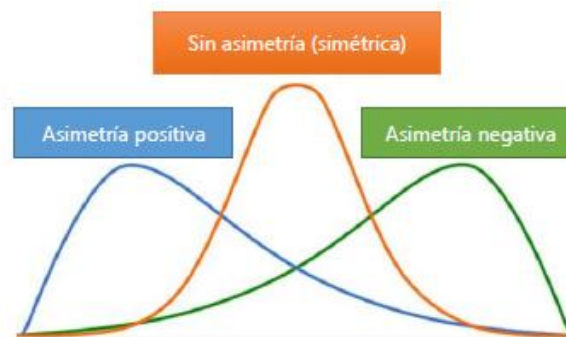


3.3. Asimetría e histogramas.

- **Asimetría (skewness)**
 - El coeficiente de asimetría es muy útil para detectar casos en los que las variables no se comportan como variables normales (distribución gaussiana).
 - La asimetría puede ser positiva o negativa. Valores mayores de +1 o menores de -1 indican una asimetría pronunciada. Valores cercanos a cero indican una distribución bastante simétrica.
 - Muchos algoritmos de aprendizaje automático asumen una distribución gaussiana.
 - Conocer la asimetría nos permite preparar los datos para poder corregir las variables necesarias, de forma que aumente la precisión de los modelos.
 - Transformaciones comunes para reducir la asimetría son la aplicación de la raíz cuadrada, transformaciones logarítmicas o transformaciones Box-Cox.




Ejercicio 90_01. Cálculo de la asimetría en Python

```
import pandas as pd
from matplotlib import pyplot
nomFichero = 'winequality-red.csv'
# No es necesario añadir nombres de columnas, ya que el csv los trae.
# Se usa ; como separador, y no la coma por defecto.
dfVinos = pd.read_csv(nomFichero, sep=";")
print('Asimetría de las variables: ')
print(dfVinos.skew())
```

Asimetría de las variables:

fixedAcidity	0.982751
volatileAcidity	0.671593
citricAcid	0.318337
residualSugar	4.540655
chlorides	5.680347
freeSulfurDioxide	1.250567
totalSulfurDioxide	1.515531
density	0.071288
pH	0.193683
sulphates	2.428672

Autores: María Inmaculada Rodríguez García , María Gema Carrasco García, Javier González Enrique, Juan Jesús Ruiz Aguilar, Ignacio J. Turias Domínguez. [Universidad de Cádiz](#)

```
alcohol          0.860829
quality          0.217802
```

Ejercicio 90_02. Cálculo de la asimetría en MATLAB


```
% Importar datos desde un archivo CSV en MATLAB
nomFichero = 'winequality-red.csv';
dfVinos = readtable(nomFichero, 'Delimiter', ',');
% Calcular la asimetría de las variables
asimetria = skewness(table2array(dfVinos));
% Mostrar la asimetría de las variables
nombres_variables = dfVinos.Properties.VariableNames;
asimetria_tabla = table(asimetria, 'RowNames', nombres_variables, 'VariableNames',
{'Asimetría'});
% Mostrar la asimetría en formato de tabla
disp('Asimetría de las variables:');
disp(asimetria_tabla);
```

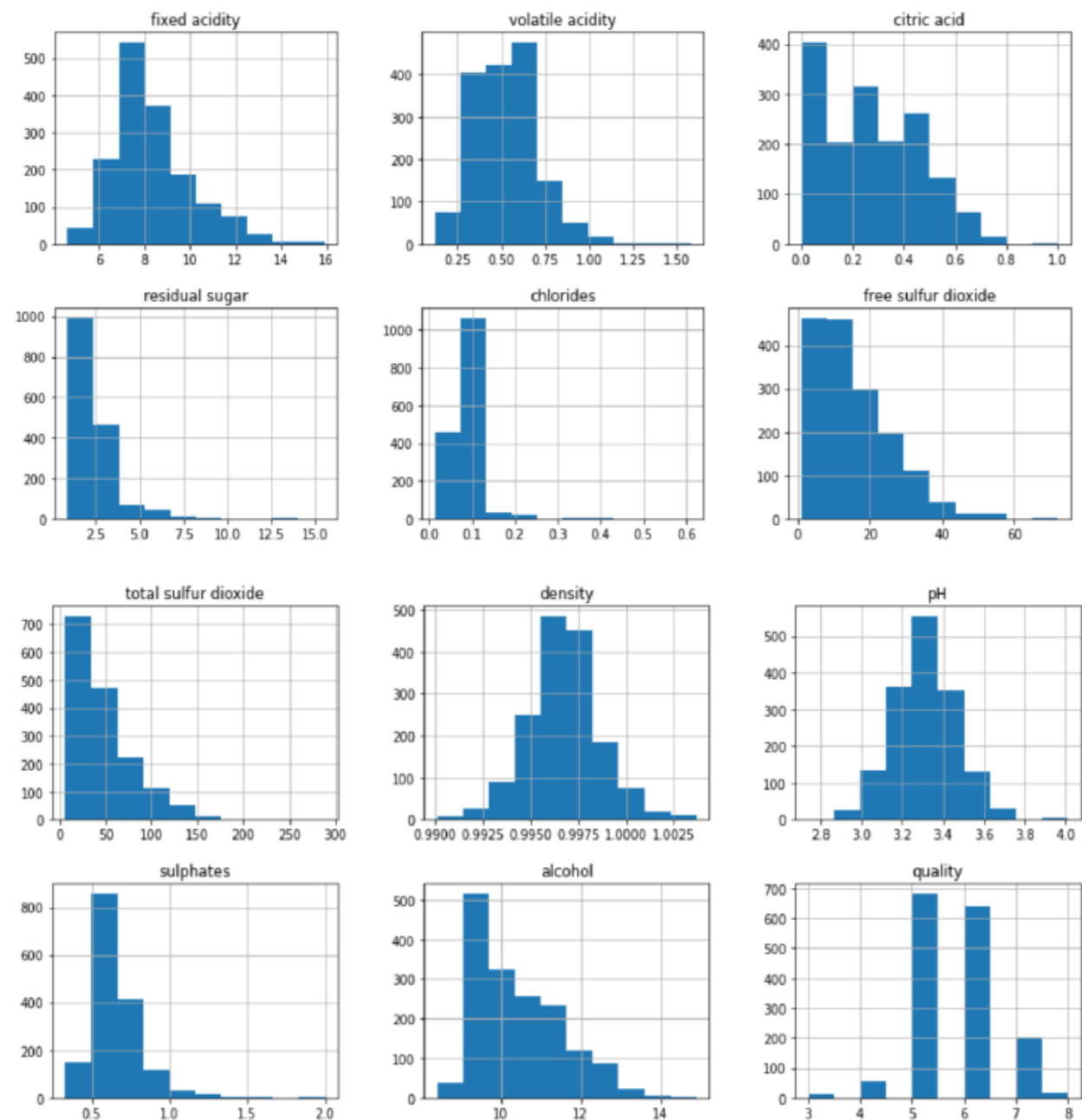
Asimetría de las variables:

	Asimetria
fixedAcidity	0.98183
volatileAcidity	0.67096
citricAcid	0.31804
residualSugar	4.5364
chlorides	5.675
freeSulfurDioxide	1.2494
totalSulfurDioxide	1.5141
density	0.071221
pH	0.1935
sulphates	2.4264
alcohol	0.86002
quality	0.2176

Ejercicio 91_01. Mostramos los histogramas de las variables en Python


```
# Mostramos los histogramas de las variables
# Necesario para corregir la proporción de los histogramas
plt.rcParams['figure.figsize'] = (15,15)
# Histogramas de las variables
dfVinos.hist()
plt.show()
```

Autores: María Inmaculada Rodríguez García , María Gema Carrasco García, Javier González Enrique, Juan Jesús Ruiz Aguilar, Ignacio J. Turias Domínguez. [Universidad de Cádiz](https://www.usc.es/)

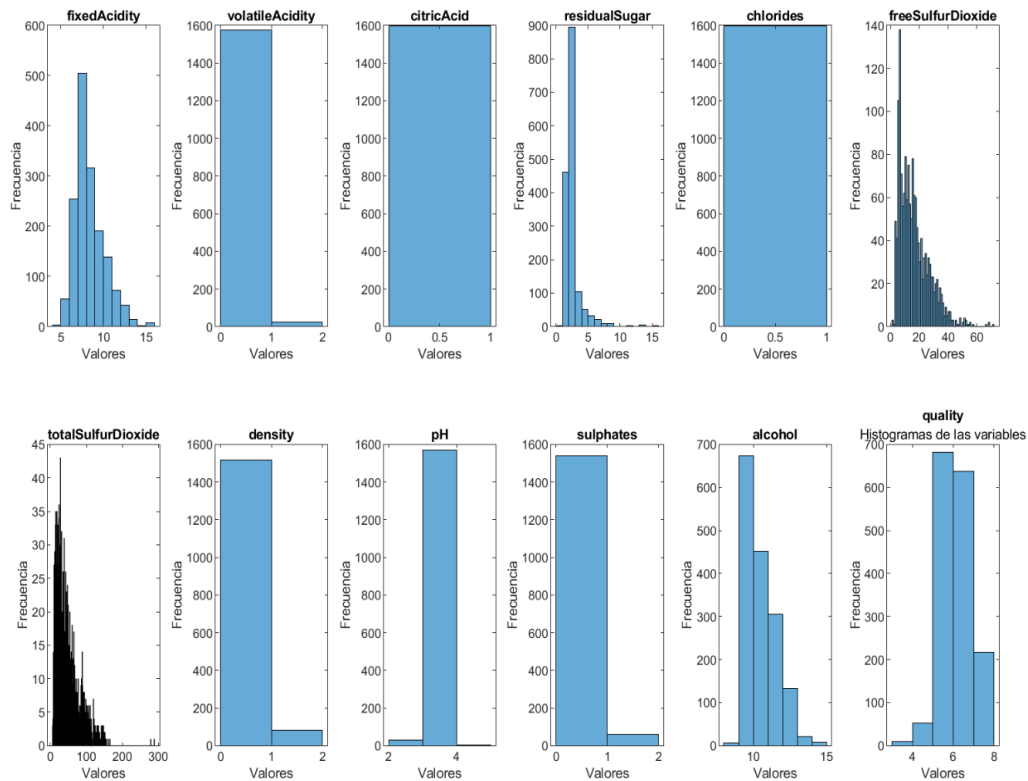


Ejercicio 91_02. Mostramos los histogramas de las variables en MATLAB

```
% Importar datos desde un archivo CSV en MATLAB
nomFichero = 'winequality-red.csv';
dfVinos = readtable(nomFichero, 'Delimiter', ',');
% Configurar el tamaño del gráfico
figure('Position', [100, 100, 1200, 800]);
% Iterar a través de las columnas y mostrar histogramas
numColumnas = width(dfVinos);
for i = 1:numColumnas
    subplot(2, 6, i); % Ajusta el número de filas y columnas según sea necesario
    histogram(dfVinos.(i), 'BinWidth', 1); % Puedes ajustar el ancho de los bins
end
```

Autores: María Inmaculada Rodríguez García , María Gema Carrasco García, Javier González Enrique, Juan Jesús Ruiz Aguilar, Ignacio J. Turias Domínguez. [Universidad de Cádiz](#)


```
title(dfVinos.Properties.VariableNames{i});  
xlabel('Valores');  
ylabel('Frecuencia');  
end  
% Mostrar el gráfico  
filename = sprintf('Gráficas de asimetría en MATLAB.tif');  
subtitle('Histogramas de las variables');  
saveas(gcf, filename, 'tif');
```



Ejercicio 92_01. Reducción de asimetría usando Box-Cox y Yeo-Johnson

```
import pandas as pd  
import numpy as np  
from matplotlib import pyplot  
from sklearn.preprocessing import PowerTransformer  
nomFichero = 'winequality-red.csv'  
# No es necesario añadir nombres de columnas, ya que el csv los trae  
# Se usa ; como separador, y no la coma por defecto  
dfDatosVinos = pd.read_csv(nomFichero, sep=";")  
# Nos quedamos con un par de variables para hacer el ejemplo:  
dfDatos = dfDatosVinos.loc[:,["residual sugar", "chlorides"]]  
# Mostramos las 3 primeras filas  
print(dfDatos.head(3))
```

```
      residual sugar    chlorides  
0                1.9         0.076
```

Autores: María Inmaculada Rodríguez García , María Gema Carrasco García, Javier González Enrique, Juan Jesús Ruiz Aguilar, Ignacio J. Turias Domínguez. [Universidad de Cádiz](#)


1	2.6	0.098
2	2.3	0.092

Ejercicio 92_02. Reducción de asimetría usando Box-Cox y Yeo-Johnson en MATLAB

```
% Importar datos desde un archivo CSV en MATLAB
nomFichero = 'winequality-red.csv';
dfDatosVinos = readtable(nomFichero, 'Delimiter', ',');
% Nos quedamos con un par de variables para hacer el ejemplo
columnas_seleccionadas = {'residualSugar', 'chlorides'};
dfDatos = dfDatosVinos(:, columnas_seleccionadas);
% Mostrar las 3 primeras filas
disp('Las 3 primeras filas de la tabla: ');
disp(dfDatos(1:3, :));
```

```
Las 3 primeras filas de la tabla:
    residualSugar    chlorides
    _____    _____
           1.9           0.076
           2.6           0.098
           2.3           0.092
```

- **Reducción de asimetría mediante transformaciones Box Cox y Yeo Johnson: Realizamos la transformación Box-Cox y Yeo-Johnson:**
 - Importante: La transformación Box-Cox necesita que los valores sean estrictamente positivos, a diferencia de la transformación Yeo-Johnson, que admite valores positivos y negativos.
 - Las citadas transformaciones permiten estandarizar los valores, lo que puede ser muy aconsejable. Pero eso lo veremos en el ejemplo 45.

Autores: María Inmaculada Rodríguez García , María Gema Carrasco García, Javier González Enrique, Juan Jesús Ruiz Aguilar, Ignacio J. Turias Domínguez. [Universidad de Cádiz](#)