

LAPORAN TUGAS BESAR
IF2220 PROBABILITAS DAN STATISTIKA



Disusun oleh

Rachel Gabriela Chen 13521044

Jeffrey Chow 13521046

PROGRAM STUDI TEKNIK INFORMATIKA
SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA
INSTITUT TEKNOLOGI BANDUNG

2023

NOMOR 1

Menulis deskripsi statistika (Descriptive Statistics) dari semua kolom pada data yang bersifat numerik, terdiri dari mean, median, modus, standar deviasi, variansi, range, nilai minimum, maksimum, kuartil, IQR, skewness dan kurtosis. Boleh juga ditambahkan deskripsi lain.

```
In [ ]: import pandas as pd
import statistics

# All columns contain numeric data
df = pd.read_csv('../dataset/anggur.csv')
```

```
In [ ]: # Show first 5 datas
df.head()
```

```
Out[ ]:
```

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0	5.90	0.4451	0.1813	2.049401	0.070574	16.593818	42.27	0.9982	3.27	0.71	8.64	7
1	8.40	0.5768	0.2099	3.109590	0.101681	22.555519	16.01	0.9960	3.35	0.57	10.03	8
2	7.54	0.5918	0.3248	3.673744	0.072416	9.316866	35.52	0.9990	3.31	0.64	9.23	8
3	5.39	0.4201	0.3131	3.371815	0.072755	18.212300	41.97	0.9945	3.34	0.55	14.07	9
4	6.51	0.5675	0.1940	4.404723	0.066379	9.360591	46.27	0.9925	3.27	0.45	11.49	8

```
In [ ]: # Descriptive Statistics
def desc_stat(col_data):

    # Mean
    print('Mean          : ' + str(col_data.mean()))

    # Median
    print('Median         : ' + str(col_data.median()))

    # Modus
    print('Modus          : ', end='')
    modes = statistics.multimode(col_data)
    sorted_modes = sorted(modes)
    if (len(sorted_modes) != len(df)):
        print(sorted_modes[0]) # Jika ada beberapa nilai modus, akan diambil yang terkecil
    else :
        print("Semua data pada kolom ini unik")

    # Standar Deviasi
    print('Standar Deviasi  : ' + str(col_data.std()))

    # Variansi
    print('Variansi         : ' + str(col_data.var()))

    # Range
    print('Range           : ' + str(col_data.max() - col_data.min()))

    # Nilai minimum
    print('Nilai minimum    : ' + str(col_data.min()))

    # Nilai maksimum
    print('Nilai Maksimum   : ' + str(col_data.max()))

    # Kuartil I
    print('Kuartil I        : ' + str(col_data.quantile(0.25)))

    # Kuartil II
    print('Kuartil II       : ' + str(col_data.quantile(0.5)))

    # Kuartil III
    print('Kuartil III      : ' + str(col_data.quantile(0.75)))

    # Interquartile
    print('Interquartile    : ' + str(col_data.quantile(0.75) - col_data.quantile(0.25)))

    # Skewness
    print('Skewness         : ' + str(col_data.skew()))
```

```
# Kurtosis
print('Kurtosis      : ' + str(col_data.kurtosis()))
```

```
In [ ]: for column in df.columns:
        print("Descriptive Statistics kolom " + column)
        desc_stat(df[column])
        print()
```

Descriptive Statistics kolom fixed acidity
Mean : 7.15253
Median : 7.15
Modus : 6.54
Standar Deviasi : 1.2015975764938258
Variansi : 1.4438367358358357
Range : 8.17
Nilai minimum : 3.32
Nilai Maksimum : 11.49
Kuartil I : 6.3774999999999995
Kuartil II : 7.15
Kuartil III : 8.0
Interquartile : 1.6225000000000005
Skewness : -0.028878575532660055
Kurtosis : -0.019292120932933532

Descriptive Statistics kolom volatile acidity
Mean : 0.5208385000000001
Median : 0.52485
Modus : 0.5546
Standar Deviasi : 0.09584827405534951
Variansi : 0.009186891639389388
Range : 0.6652
Nilai minimum : 0.1399
Nilai Maksimum : 0.8051
Kuartil I : 0.4561
Kuartil II : 0.52485
Kuartil III : 0.585375
Interquartile : 0.12927499999999997
Skewness : -0.1976986986092083
Kurtosis : 0.16185290336961788

Descriptive Statistics kolom citric acid
Mean : 0.27051699999999995
Median : 0.2722
Modus : 0.3019
Standar Deviasi : 0.04909837147076348
Variansi : 0.0024106500810810814
Range : 0.29290000000000005
Nilai minimum : 0.1167
Nilai Maksimum : 0.4096
Kuartil I : 0.2378
Kuartil II : 0.2722
Kuartil III : 0.302325
Interquartile : 0.064525
Skewness : -0.045576058685017296
Kurtosis : -0.1046792495951605

Descriptive Statistics kolom residual sugar
Mean : 2.5671036825067595
Median : 2.519430272865794
Modus : Semua data pada kolom ini unik
Standar Deviasi : 0.9879154365046929
Variansi : 0.9759769096842579
Range : 5.5182004097078625
Nilai minimum : 0.032554525015195
Nilai Maksimum : 5.550754934723058
Kuartil I : 1.896329943488683
Kuartil II : 2.519430272865794
Kuartil III : 3.220873482829786
Interquartile : 1.3245435393411031
Skewness : 0.13263808618992312
Kurtosis : -0.04298003436476261

Descriptive Statistics kolom chlorides
Mean : 0.08119515250784977
Median : 0.0821669021645236
Modus : Semua data pada kolom ini unik
Standar Deviasi : 0.020110647243996742
Variansi : 0.00040443813257247374
Range : 0.1256351302653488
Nilai minimum : 0.0151224391657095
Nilai Maksimum : 0.1407575694310583
Kuartil I : 0.06657363190977357
Kuartil II : 0.0821669021645236
Kuartil III : 0.09531150148556258
Interquartile : 0.028737869575789013
Skewness : -0.05131929742072573
Kurtosis : -0.2465081359240382

Descriptive Statistics kolom free sulfur dioxide
Mean : 14.907679251029796
Median : 14.860346236568924
Modus : Semua data pada kolom ini unik
Standar Deviasi : 4.888099705756562
Variansi : 23.893518733417388
Range : 27.26784690109891
Nilai minimum : 0.194678523326937
Nilai Maksimum : 27.462525424425845
Kuartil I : 11.426716949457617
Kuartil II : 14.860346236568924
Kuartil III : 18.313097915395005
Interquartile : 6.886380965937388
Skewness : 0.007130415991143398
Kurtosis : -0.36496364342685306

Descriptive Statistics kolom total sulfur dioxide
Mean : 40.290150000000004
Median : 40.19
Modus : 35.2
Standar Deviasi : 9.965767376218295
Variansi : 99.3165193968969
Range : 66.809999999999999
Nilai minimum : 3.15
Nilai Maksimum : 69.96
Kuartil I : 33.785
Kuartil II : 40.19
Kuartil III : 47.0225
Interquartile : 13.237500000000004
Skewness : -0.024060026812269975
Kurtosis : 0.06394978916172311

Descriptive Statistics kolom density
Mean : 0.9959253000000001
Median : 0.996
Modus : 0.9959
Standar Deviasi : 0.0020201809426487133
Variansi : 4.081131041041045e-06
Range : 0.013799999999999923
Nilai minimum : 0.9888
Nilai Maksimum : 1.0026
Kuartil I : 0.9946
Kuartil II : 0.996
Kuartil III : 0.9972
Interquartile : 0.002599999999999357
Skewness : -0.07688278915513917
Kurtosis : 0.01636562128503849

Descriptive Statistics kolom pH
Mean : 3.30361
Median : 3.3
Modus : 3.34
Standar Deviasi : 0.10487548220040166
Variansi : 0.010998866766766766
Range : 0.7399999999999998
Nilai minimum : 2.97
Nilai Maksimum : 3.71
Kuartil I : 3.23
Kuartil II : 3.3
Kuartil III : 3.37
Interquartile : 0.14000000000000012
Skewness : 0.14767259510827038
Kurtosis : 0.0809095518741838

Descriptive Statistics kolom sulphates
Mean : 0.5983899999999999
Median : 0.595
Modus : 0.59
Standar Deviasi : 0.10081900799141187
Variansi : 0.010164472372372372
Range : 0.6699999999999999
Nilai minimum : 0.29
Nilai Maksimum : 0.96
Kuartil I : 0.53
Kuartil II : 0.595
Kuartil III : 0.67
Interquartile : 0.14
Skewness : 0.1491989008699043
Kurtosis : 0.06481928180859686

Descriptive Statistics kolom alcohol
Mean : 10.592279999999999
Median : 10.61
Modus : 9.86
Standar Deviasi : 1.5107060052287586
Variansi : 2.282232634234234
Range : 8.989999999999998
Nilai minimum : 6.03
Nilai Maksimum : 15.02
Kuartil I : 9.56
Kuartil II : 10.61
Kuartil III : 11.622499999999999
Interquartile : 2.062499999999982
Skewness : -0.01899140432111647
Kurtosis : -0.13173155932281988

Descriptive Statistics kolom quality
Mean : 7.958
Median : 8.0
Modus : 8
Standar Deviasi : 0.9028017783827471
Variansi : 0.8150510510510509
Range : 5
Nilai minimum : 5
Nilai Maksimum : 10
Kuartil I : 7.0
Kuartil II : 8.0
Kuartil III : 9.0
Interquartile : 2.0
Skewness : -0.08905409122491781
Kurtosis : 0.10829100232871003

NOMOR 2

Membuat Visualisasi plot distribusi, dalam bentuk histogram dan boxplot untuk setiap kolom numerik. Berikan uraian penjelasan kondisi setiap kolom berdasarkan kedua plot tersebut.

```
In [ ]: import pandas as pd
import matplotlib.pyplot as plt
import scipy.stats as st
import numpy as np

df = pd.read_csv('../dataset/anggur.csv')

df.head()
```

```
Out[ ]:
```

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0	5.90	0.4451	0.1813	2.049401	0.070574	16.593818	42.27	0.9982	3.27	0.71	8.64	7
1	8.40	0.5768	0.2099	3.109590	0.101681	22.555519	16.01	0.9960	3.35	0.57	10.03	8
2	7.54	0.5918	0.3248	3.673744	0.072416	9.316866	35.52	0.9990	3.31	0.64	9.23	8
3	5.39	0.4201	0.3131	3.371815	0.072755	18.212300	41.97	0.9945	3.34	0.55	14.07	9
4	6.51	0.5675	0.1940	4.404723	0.066379	9.360591	46.27	0.9925	3.27	0.45	11.49	8

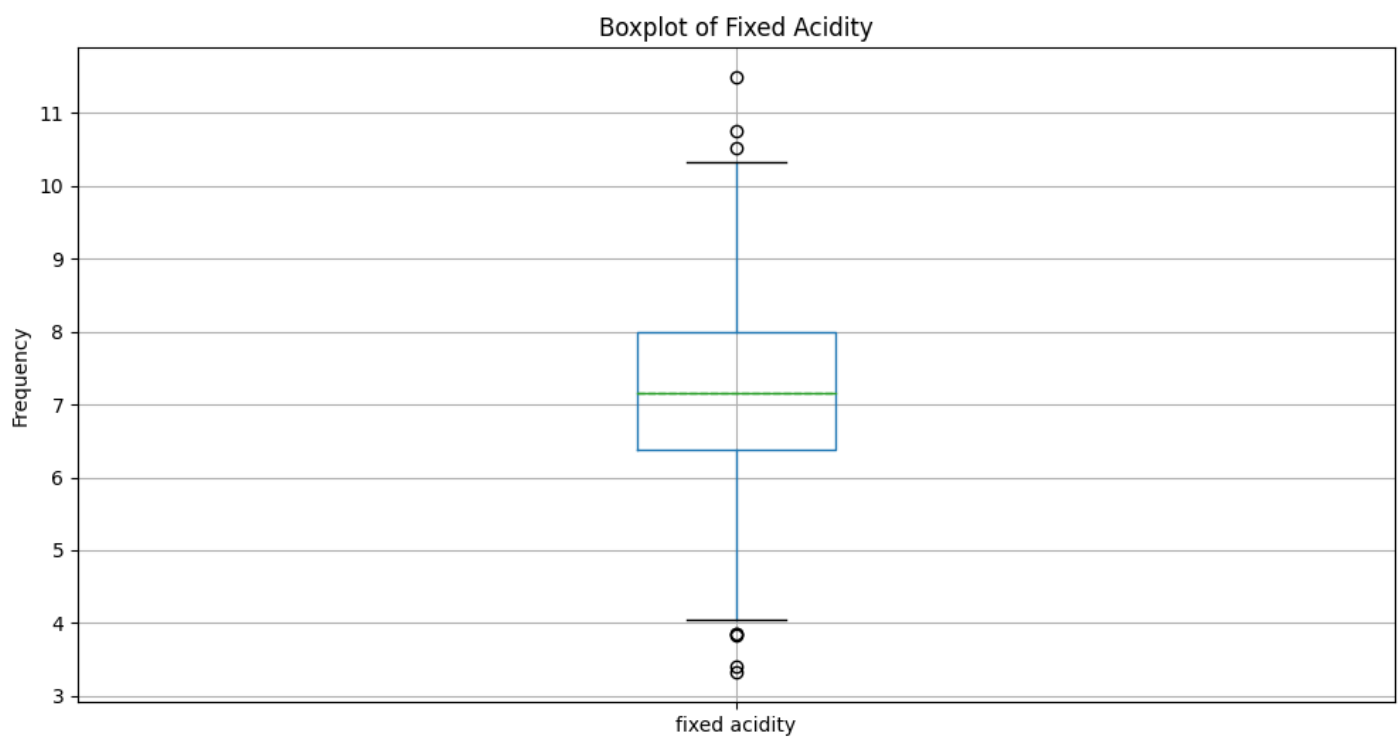
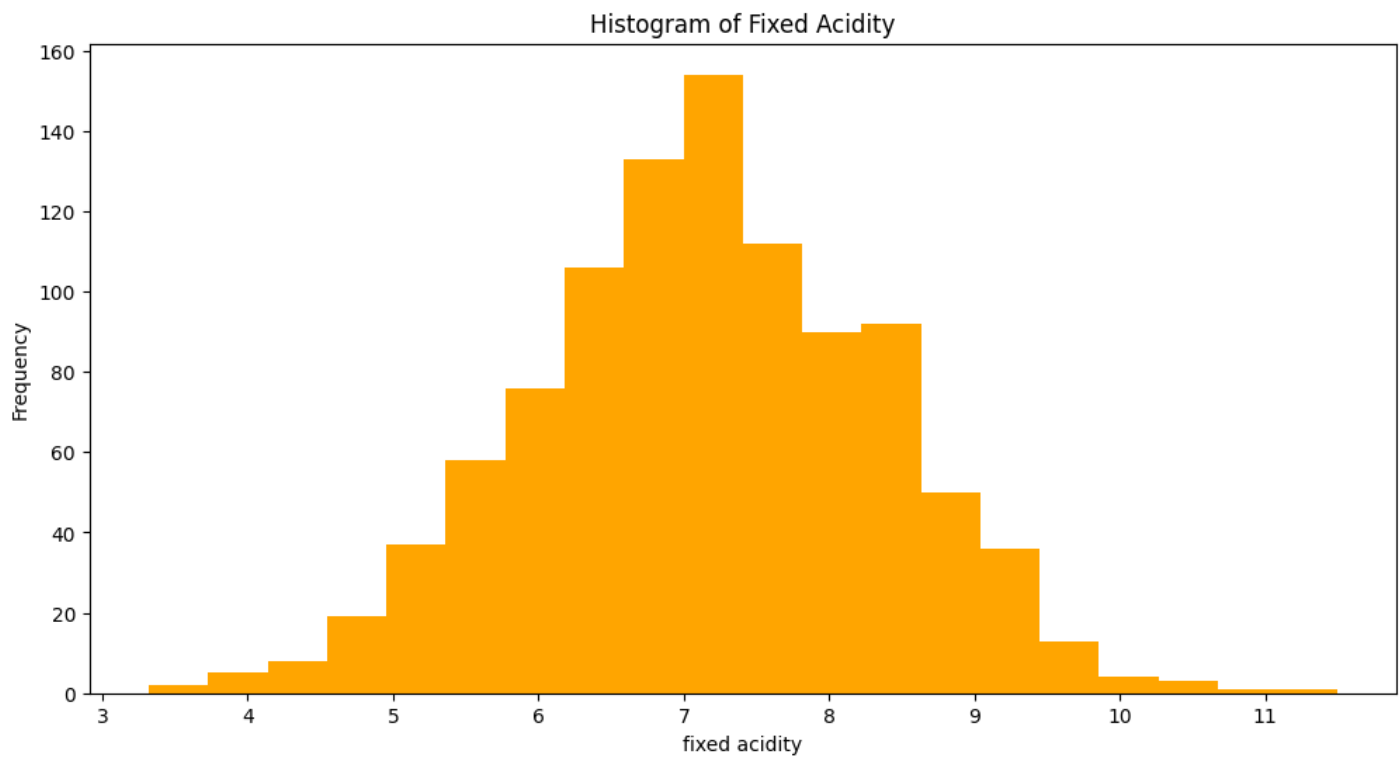
Function to Generate Histogram

```
In [ ]: # create histogram
def generate_hist(col):
    df[col].plot(kind="hist", figsize=(12, 6), bins=20, color='orange')
    # set title, x-label, and y-label for each histogram
    plt.title('Histogram of ' + col.title())
    plt.xlabel(col)
    plt.ylabel('Frequency')
    # show the histogram
    plt.show()

def generate_boxplot(col):
    df.boxplot(column=col, figsize = (12,6), meanline = True, showmeans = True)
    # set title, x-label, and y-label for each histogram
    plt.title('Boxplot of ' + col.title())
    plt.ylabel('Frequency')
    # show the boxplot
    plt.show()
```

Fixed Acidity

```
In [ ]: generate_hist("fixed acidity")
generate_boxplot("fixed acidity")
```

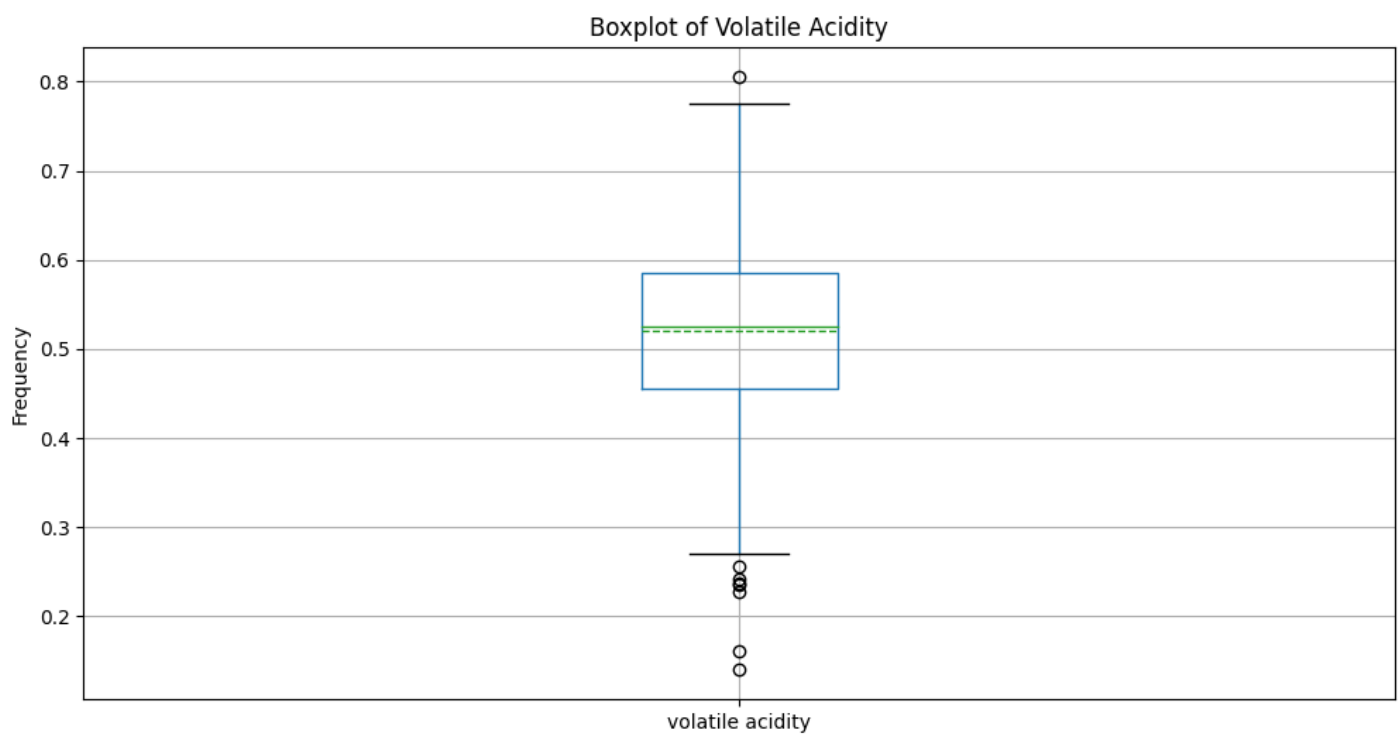
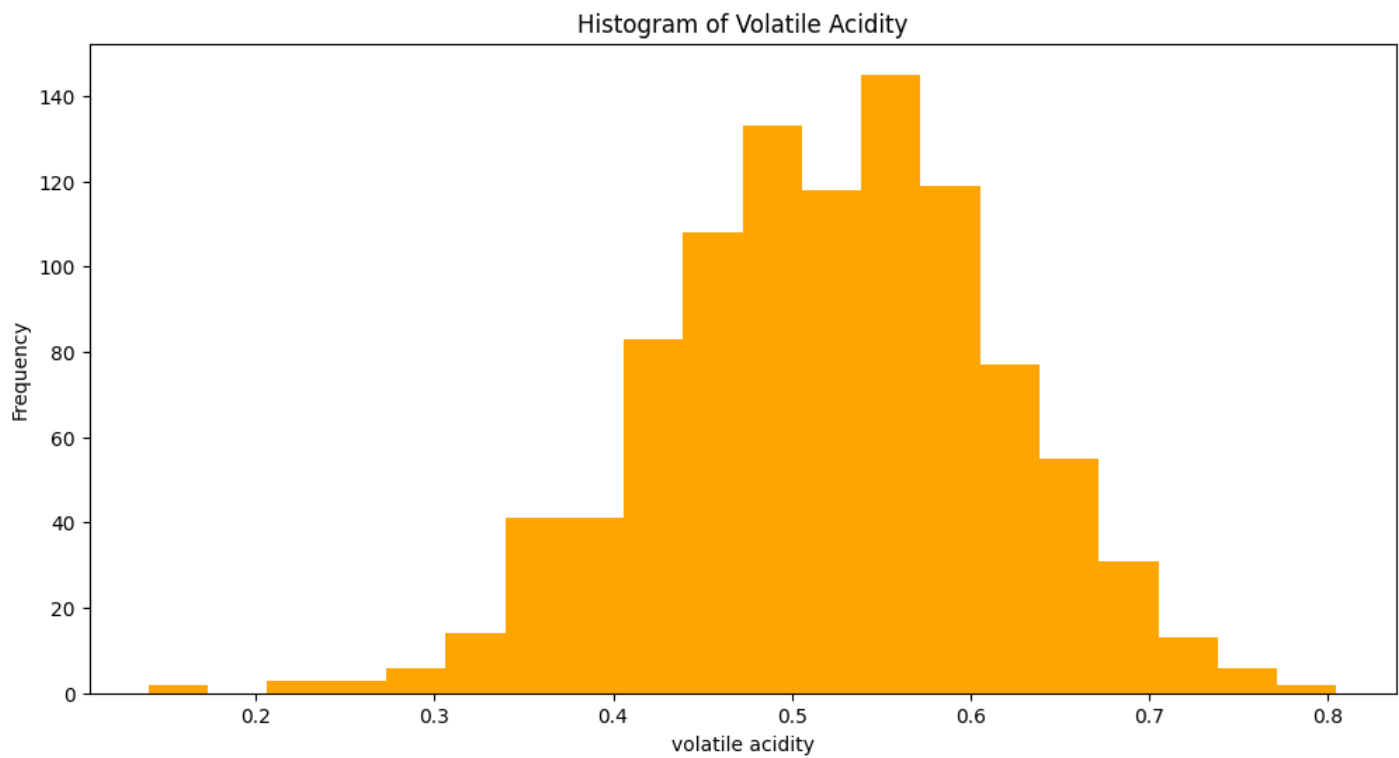


Berdasarkan kedua plot di atas, dapat disimpulkan bahwa:

1. Distribusi fixed acidity memiliki *skewness* normal.
2. Terdapat beberapa data *outlier* yang berada pada rentang 3 - 4 dan 10 - 12. Terlihat juga median berada di sekitar 7.

Volatile Acidity

```
In [ ]: generate_hist("volatile acidity")
generate_boxplot("volatile acidity")
```

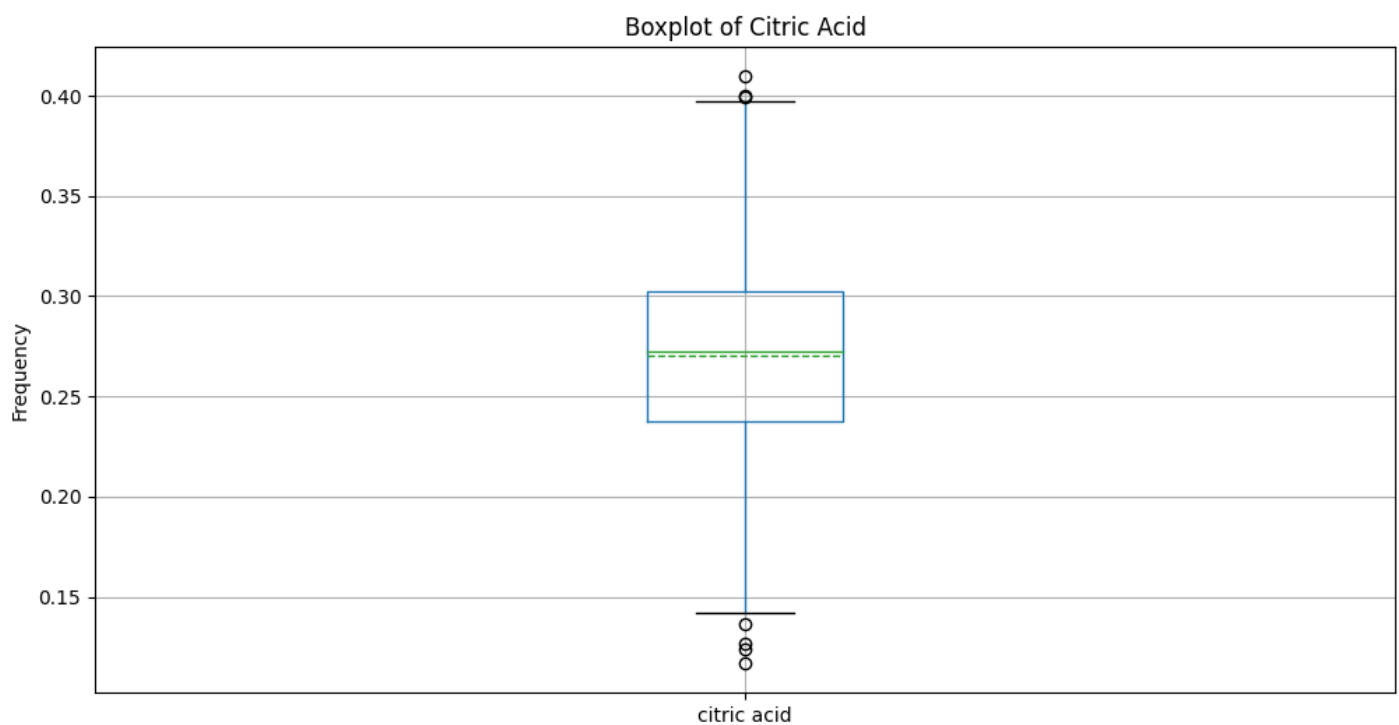
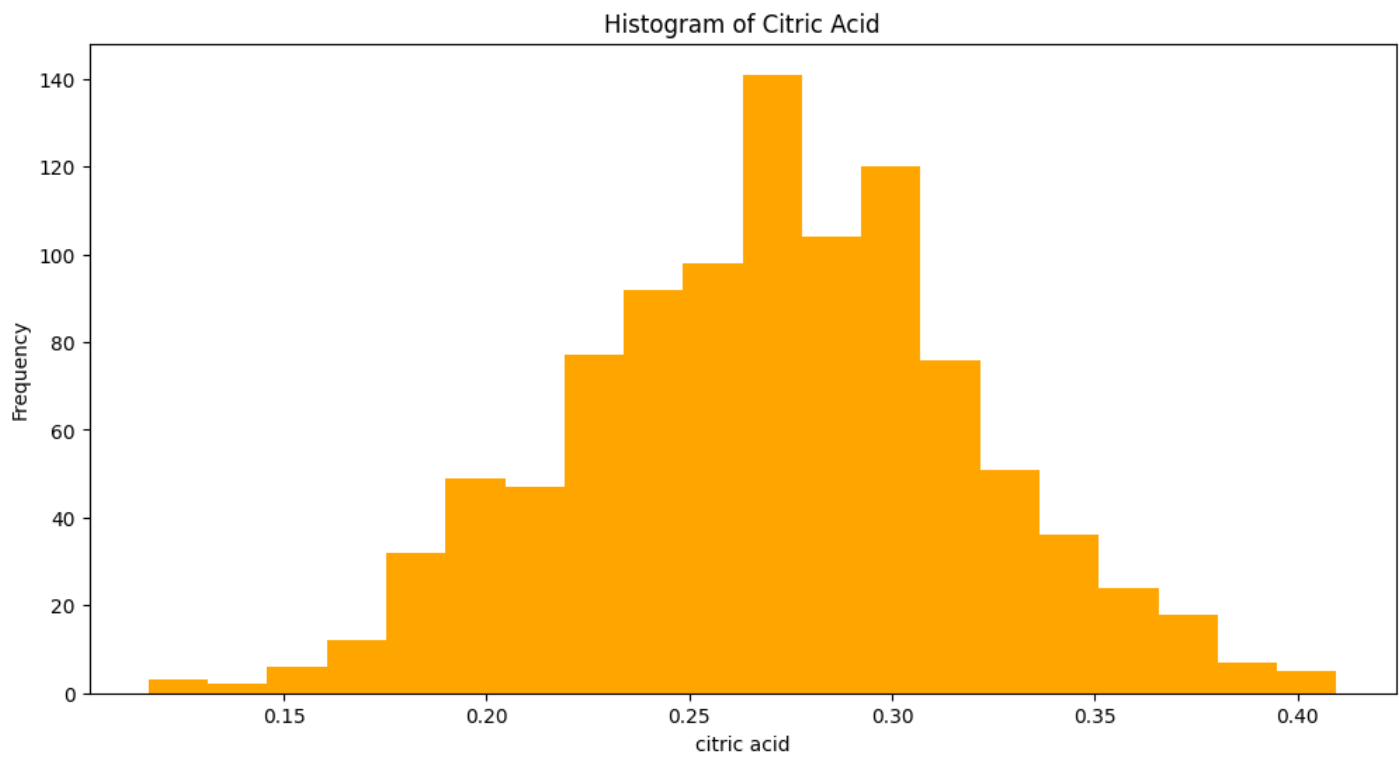



Berdasarkan kedua plot di atas, dapat disimpulkan bahwa:

1. Distribusi volatile acidity bersifat *mildly left skewed*.
2. Terdapat beberapa data *outlier* yang berada pada rentang 0 - 0.3 dan sekitar 0.8. Terlihat juga median berada di sekitar 0.5 - 0.6.

Citric Acid

```
In [ ]: generate_hist("citric acid")
generate_boxplot("citric acid")
```

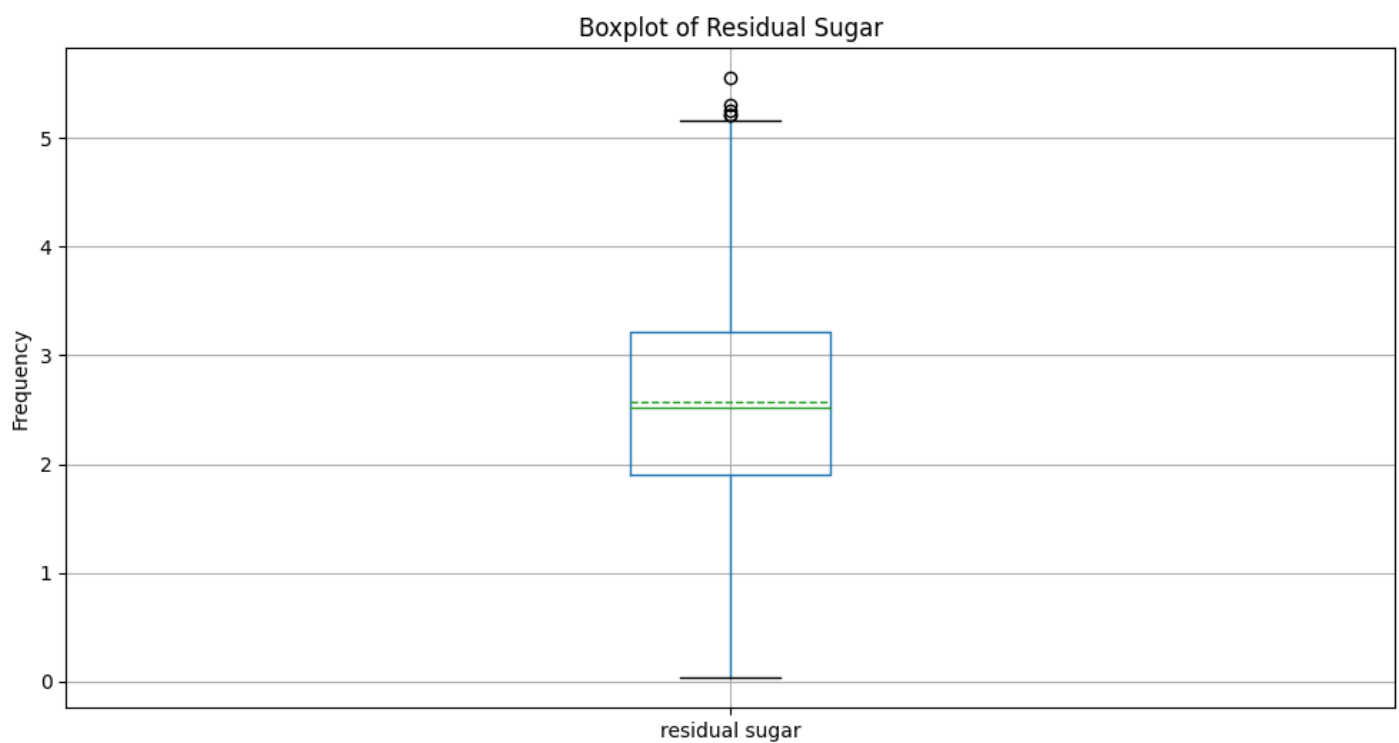
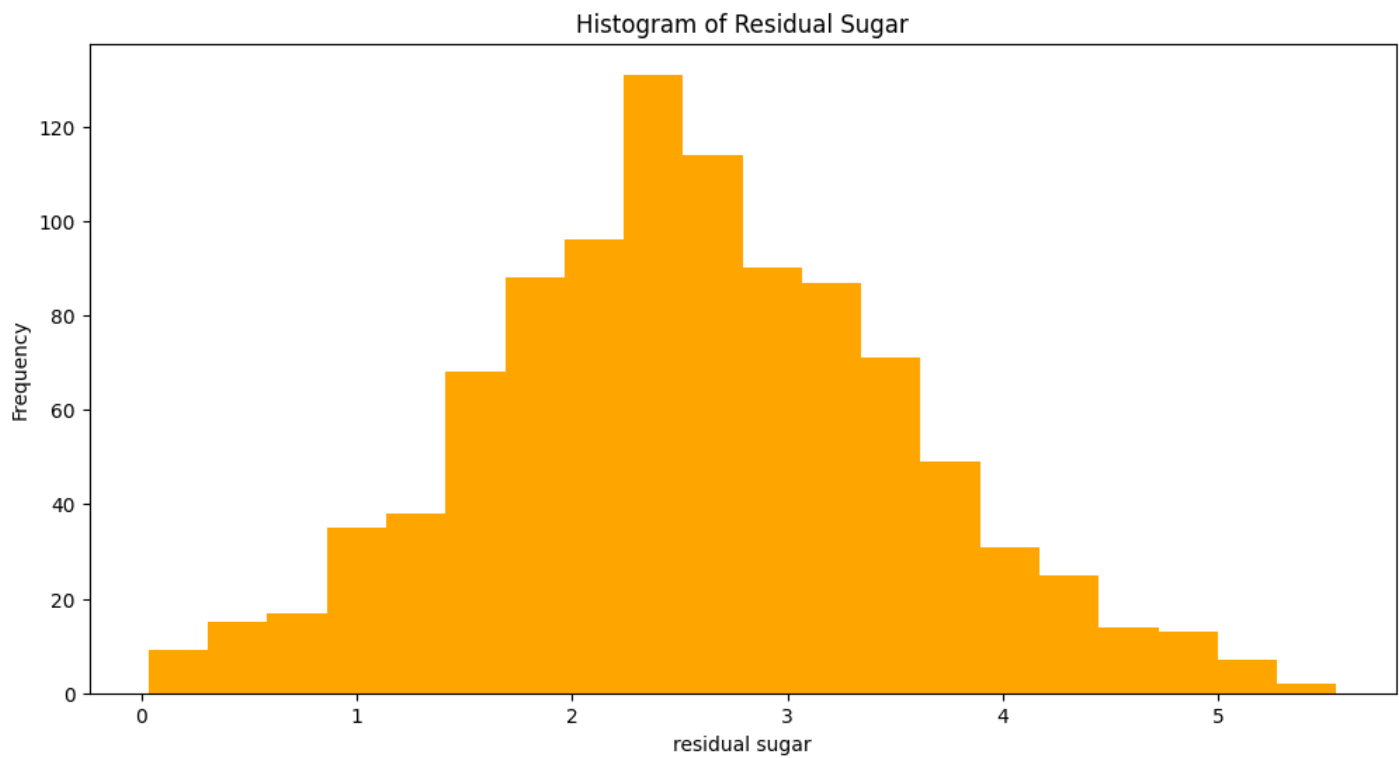


Berdasarkan kedua plot di atas, dapat disimpulkan bahwa:

1. Distribusi citric acid memiliki *skewness* normal.
2. Terdapat beberapa data *outlier* yang berada pada rentang 0 - 0.15 dan sekitar 0.4 - 0.45. Terlihat juga median berada di sekitar 0.275.

Residual Sugar

```
In [ ]: generate_hist("residual sugar")
generate_boxplot("residual sugar")
```

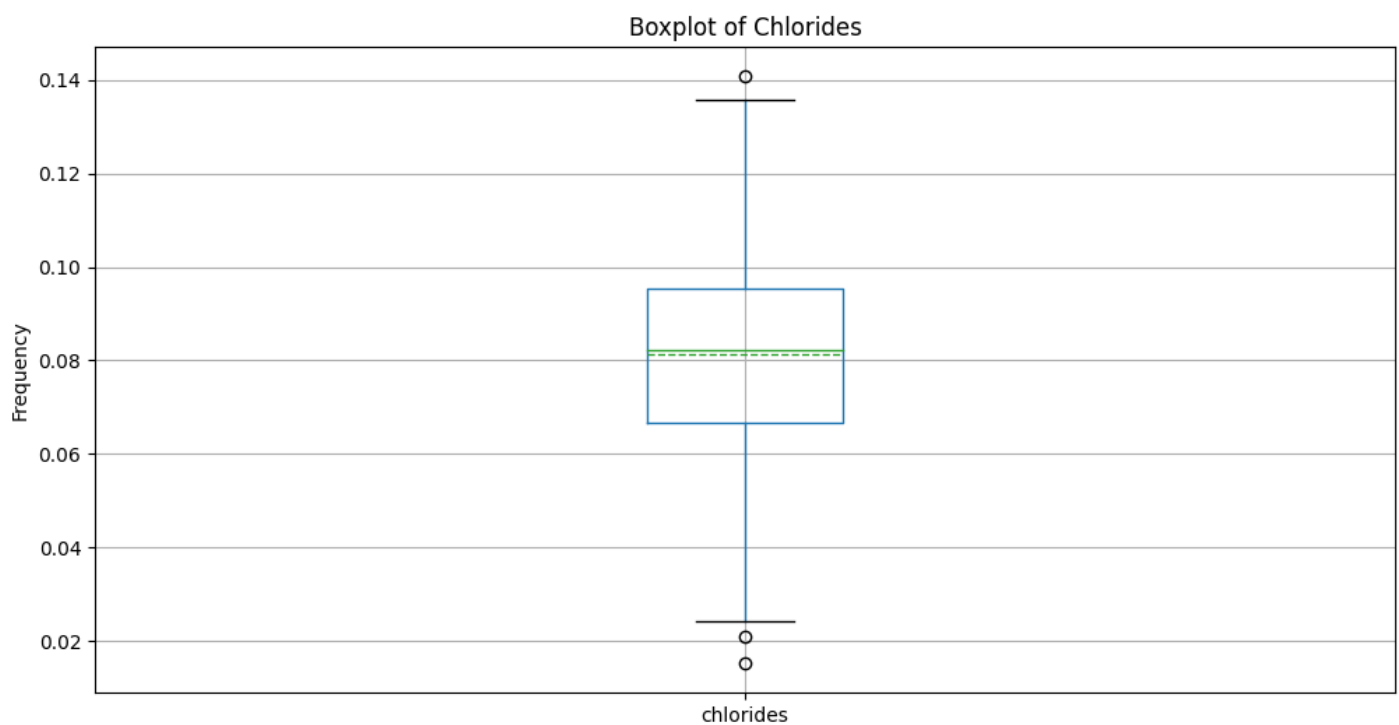
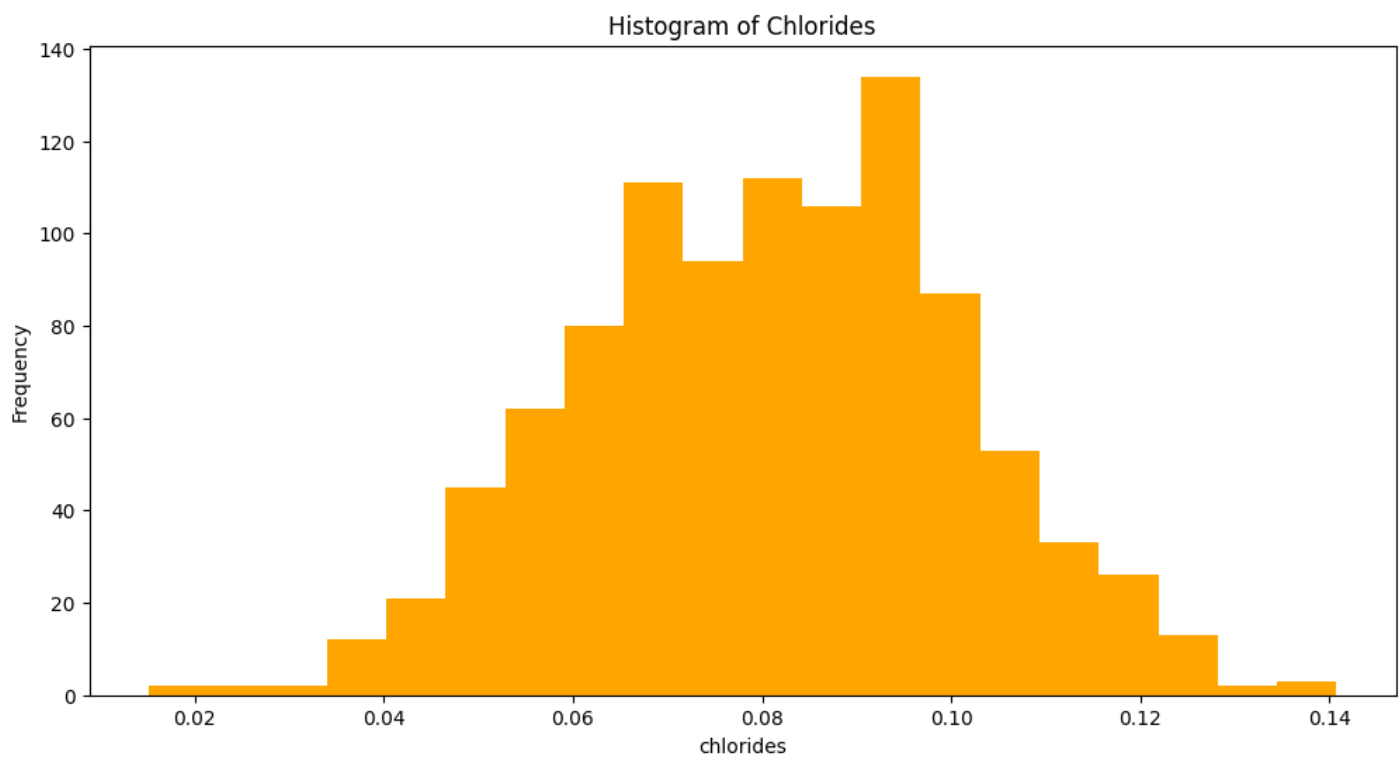


Berdasarkan kedua plot di atas, dapat disimpulkan bahwa:

1. Distribusi residual memiliki *skewness* normal.
2. Terdapat beberapa data *outlier* yang berada pada rentang 5 - 6. Terlihat juga median berada di sekitar 0.25.

Chlorides

```
In [ ]: generate_hist("chlorides")
generate_boxplot("chlorides")
```

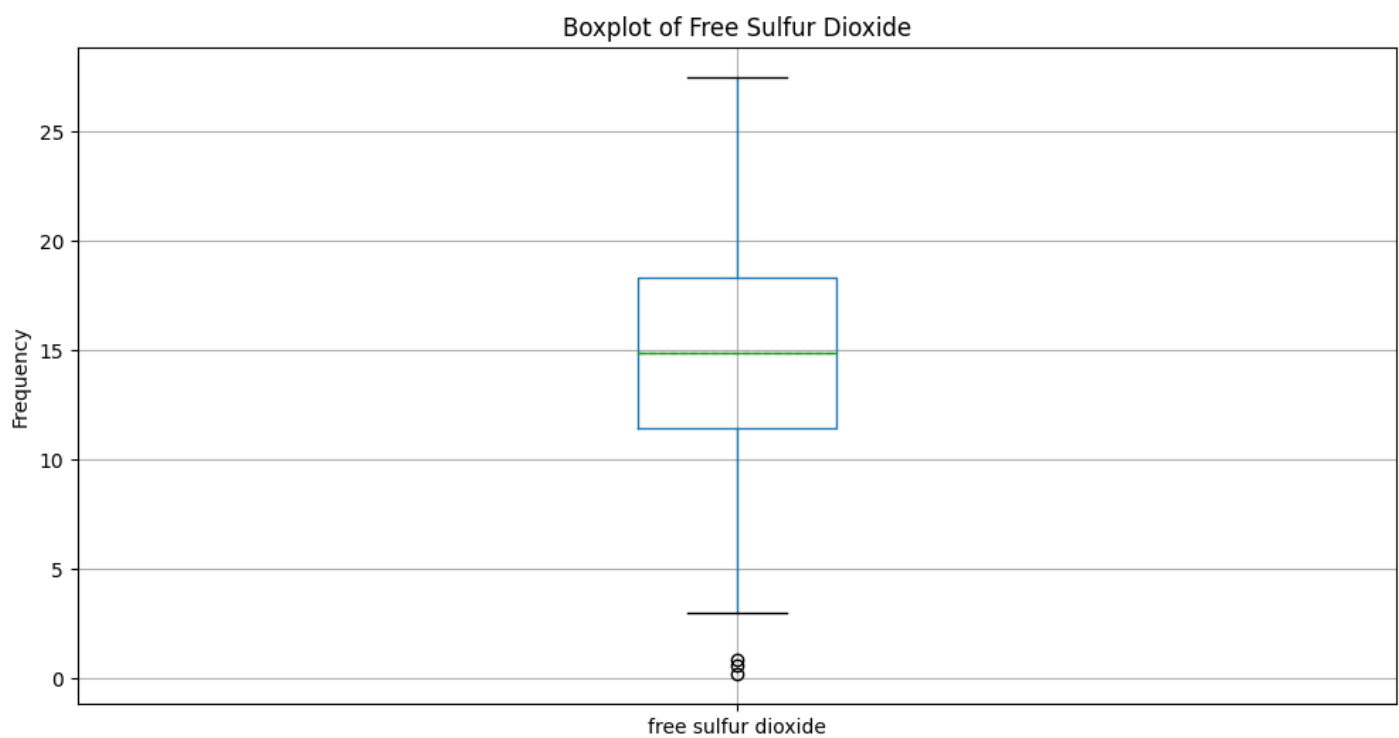
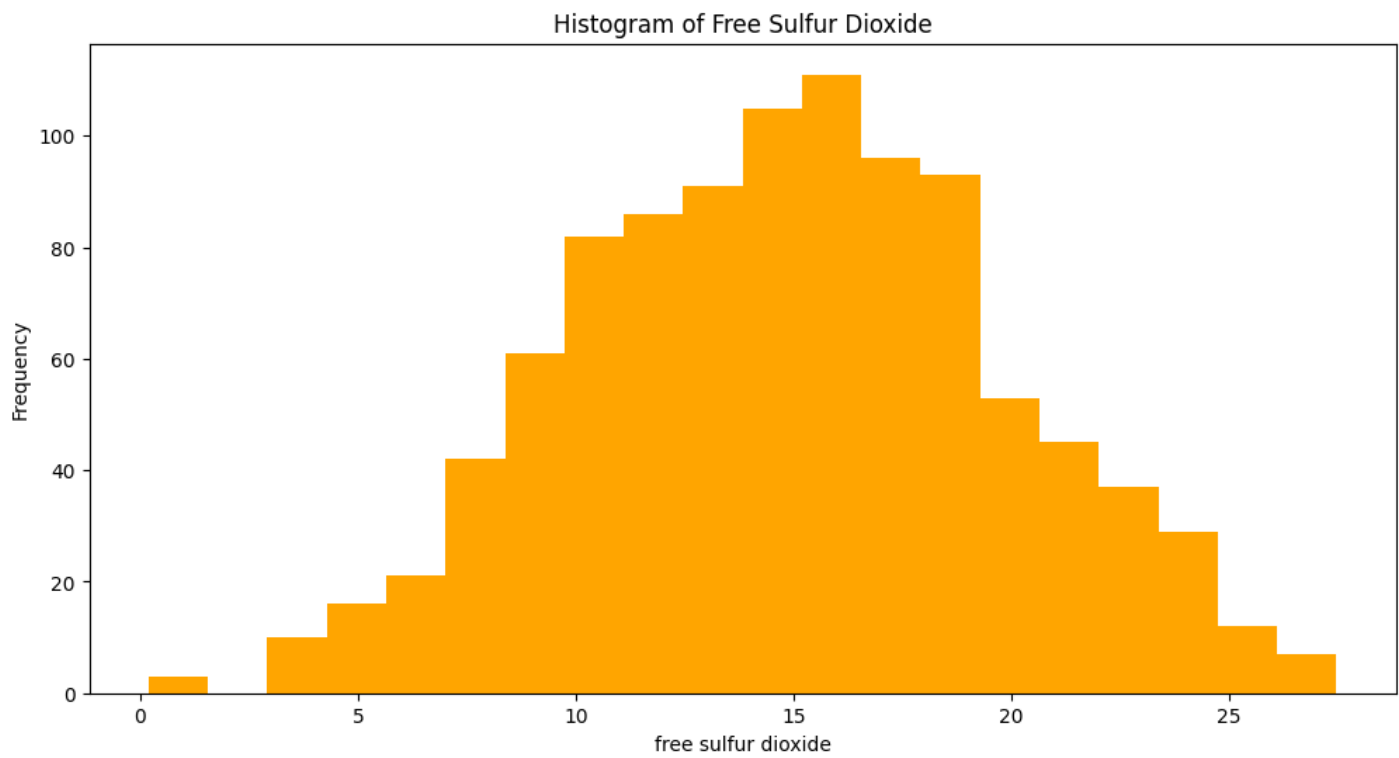


Berdasarkan kedua plot di atas, dapat disimpulkan bahwa:

1. Distribusi chlorides memiliki *skewness* normal.
2. Terdapat beberapa data *outlier* yang berada pada rentang 0 - 0.03 dan sekitar 0.14. Terlihat juga median berada di sekitar 0.08.

Free Sulfur Dioxide

```
In [ ]: generate_hist("free sulfur dioxide")
generate_boxplot("free sulfur dioxide")
```

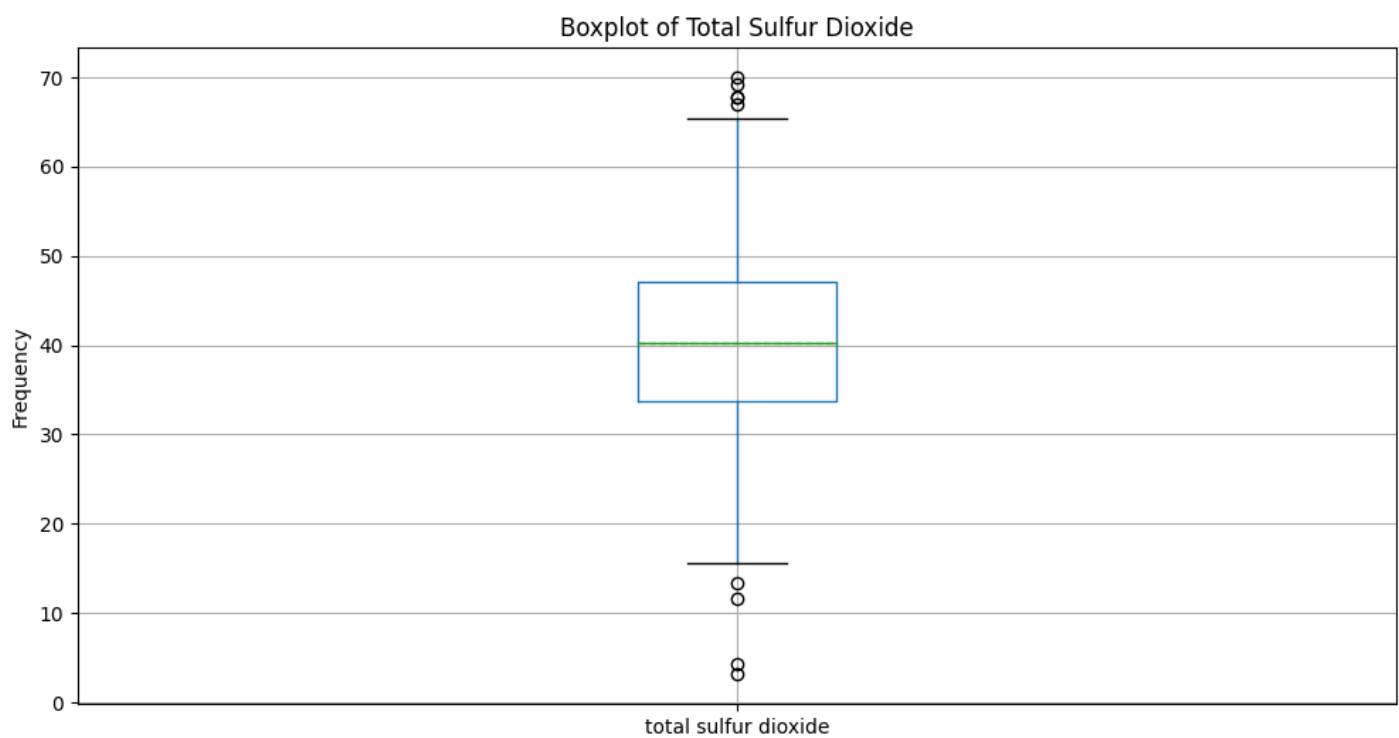
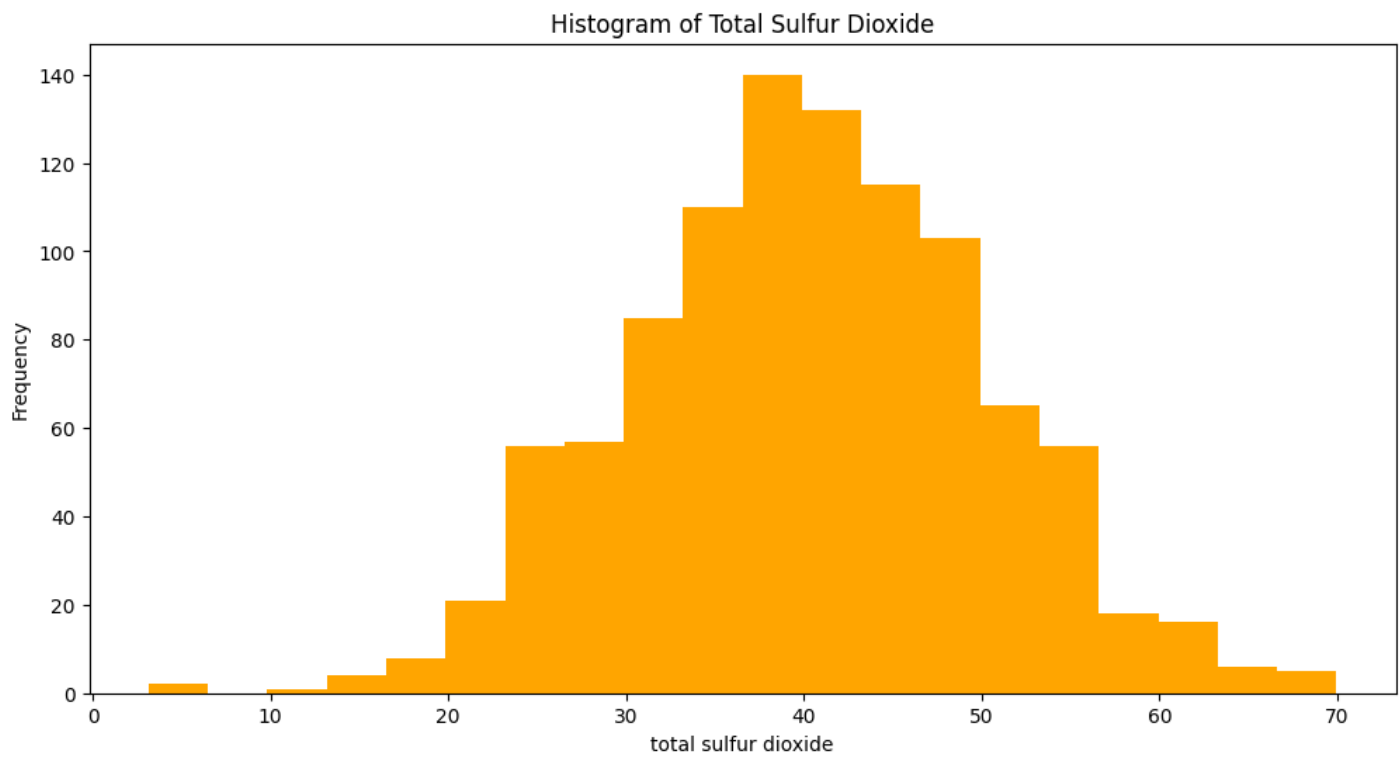


Berdasarkan kedua plot di atas, dapat disimpulkan bahwa:

1. Distribusi free sulfur dioxide memiliki *skewness* normal.
2. Terdapat beberapa data *outlier* yang berada pada rentang 0 - 5. Terlihat juga median berada di sekitar 15.

Total Sulfur Dioxide

```
In [ ]: generate_hist("total sulfur dioxide")
generate_boxplot("total sulfur dioxide")
```

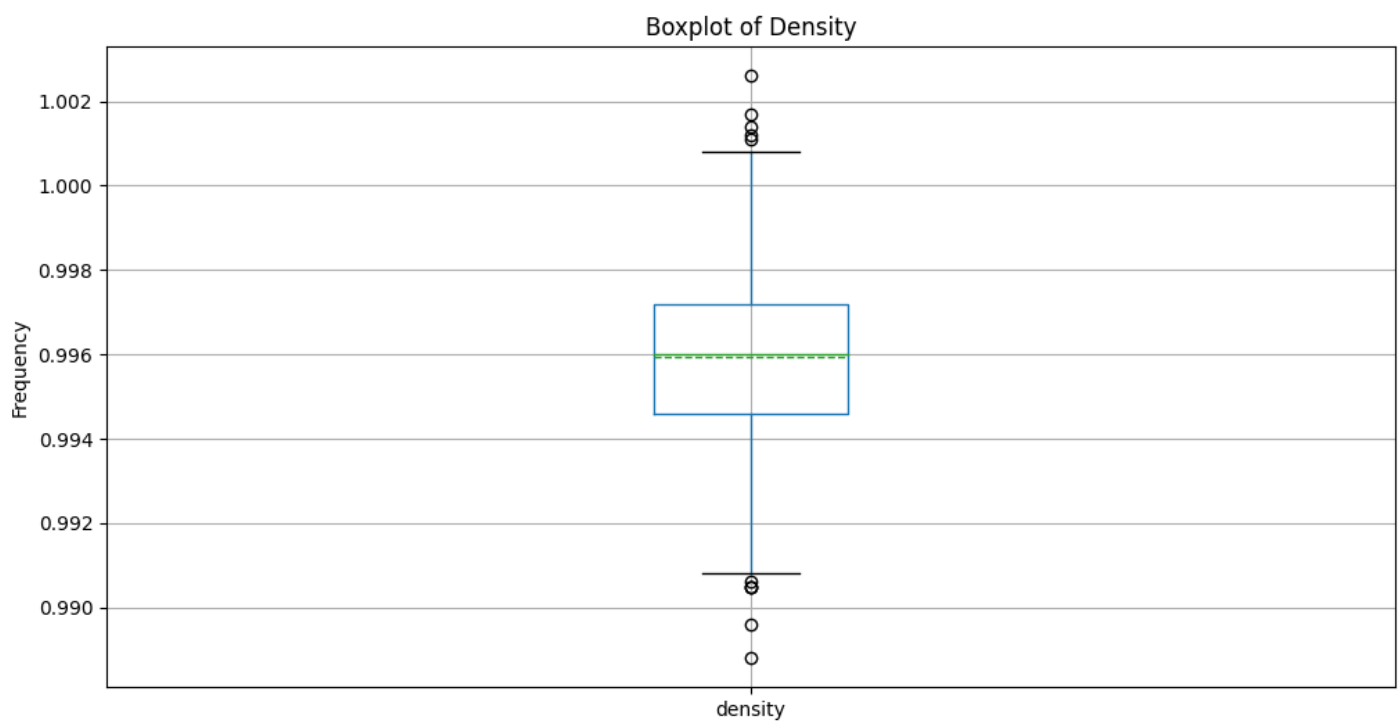
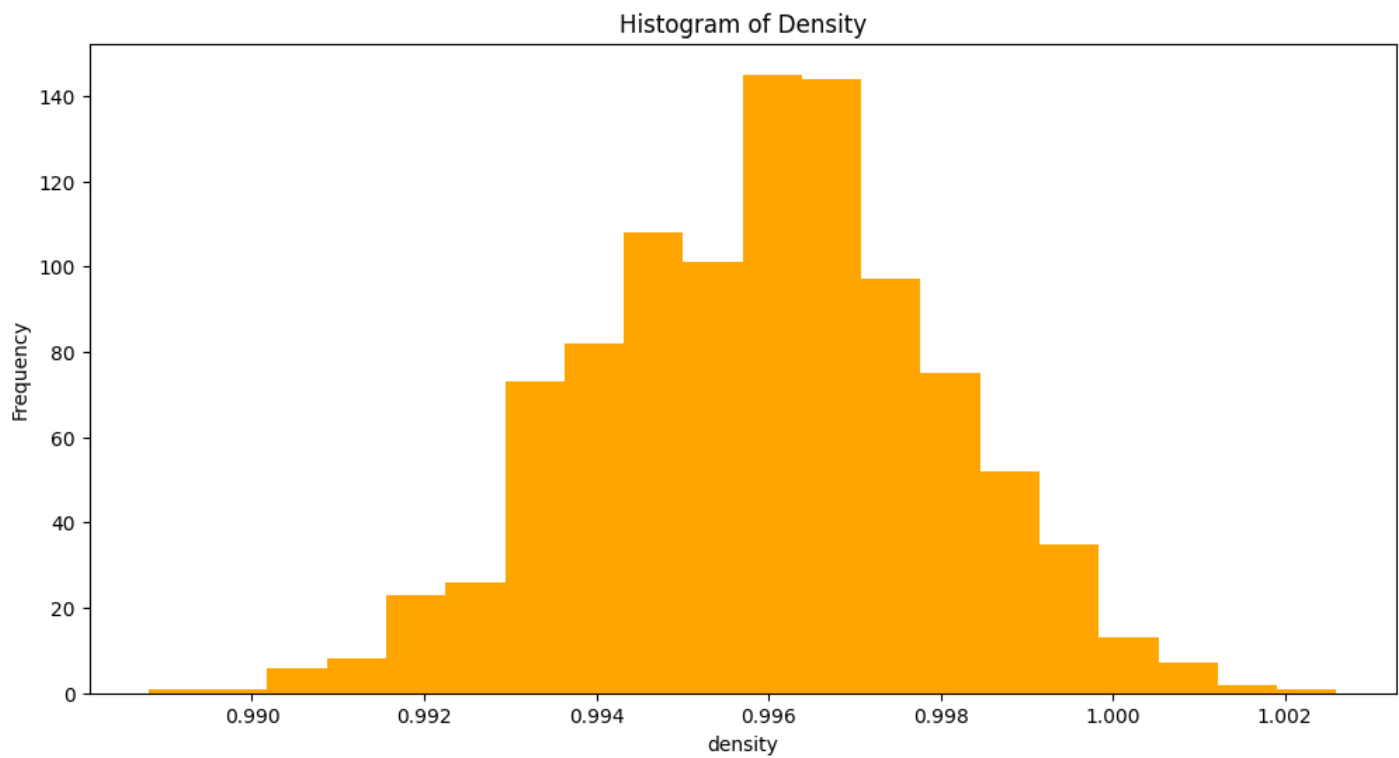


Berdasarkan kedua plot di atas, dapat disimpulkan bahwa:

1. Distribusi total sulfur dioxide memiliki *skewness* normal.
2. Terdapat beberapa data *outlier* yang berada pada rentang 0 - 20 dan 60 - 70. Terlihat juga median berada di sekitar 40.

Density

```
In [ ]: generate_hist("density")
generate_boxplot("density")
```

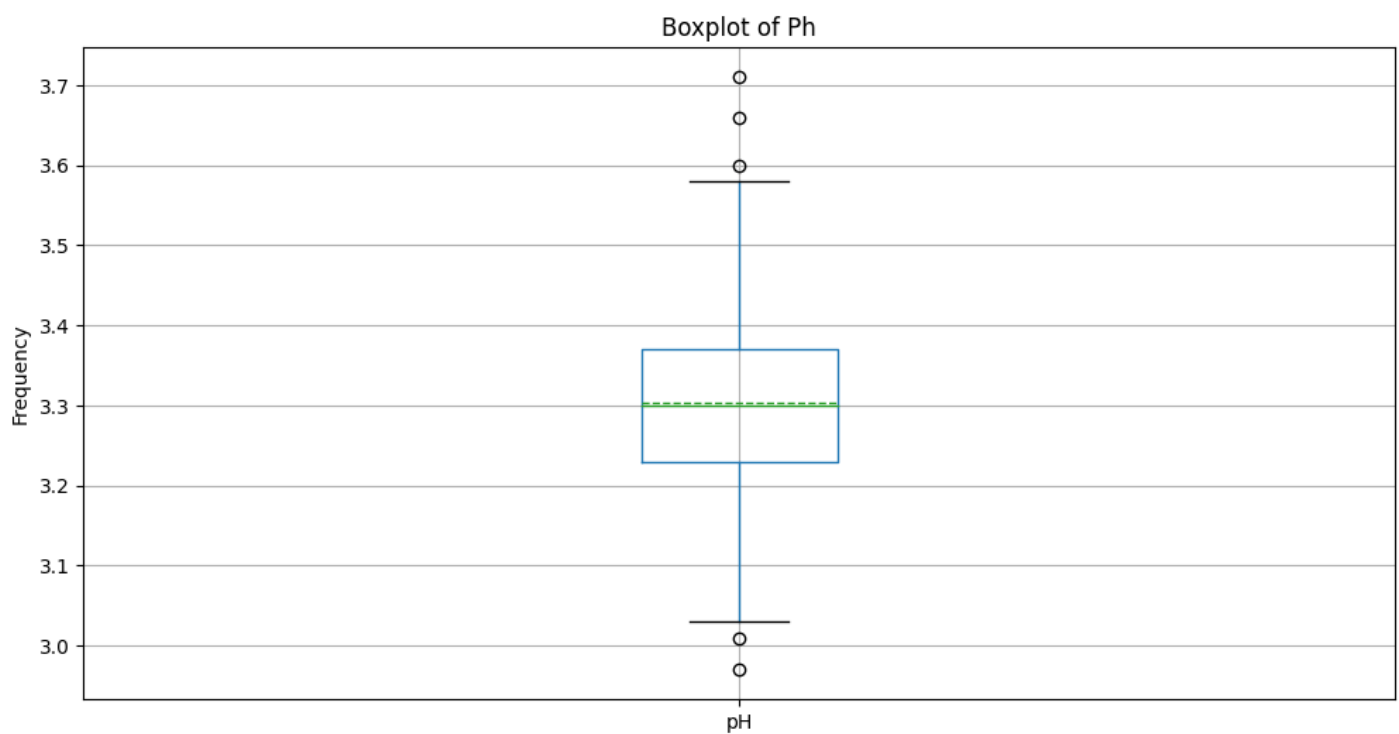
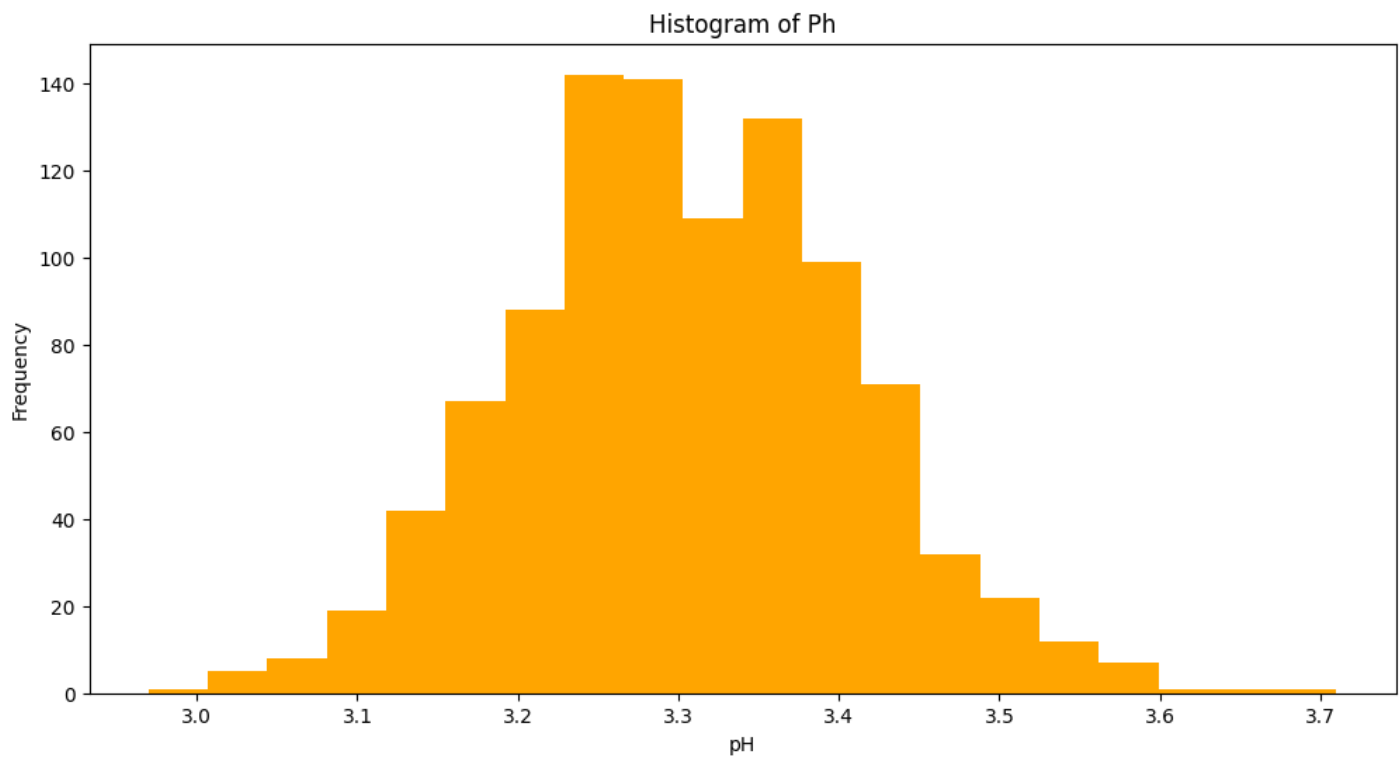


Berdasarkan kedua plot di atas, dapat disimpulkan bahwa:

1. Distribusi density memiliki *skewness* normal.
2. Terdapat beberapa data *outlier* yang berada pada rentang 0.988 - 0.992. Terlihat juga median berada di sekitar 0.996.

pH

```
In [ ]: generate_hist("pH")
        generate_boxplot("pH")
```

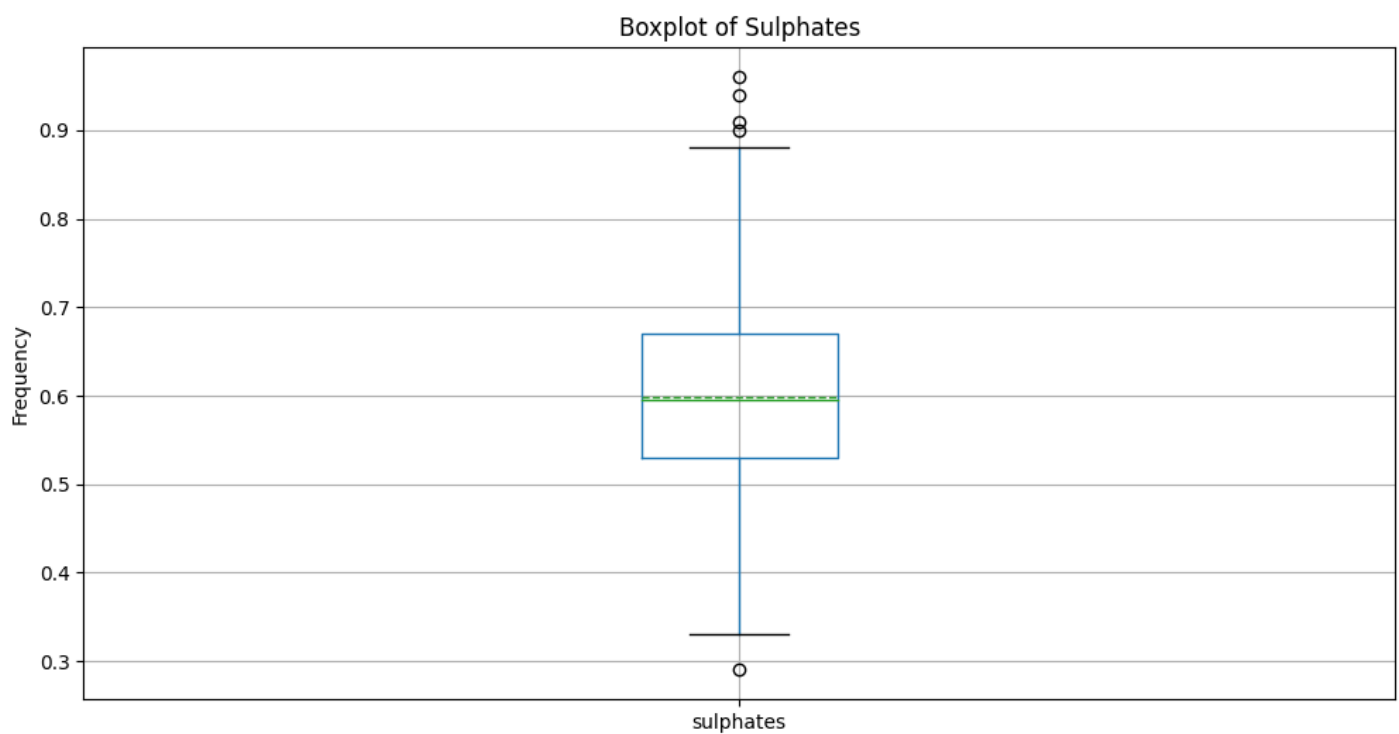
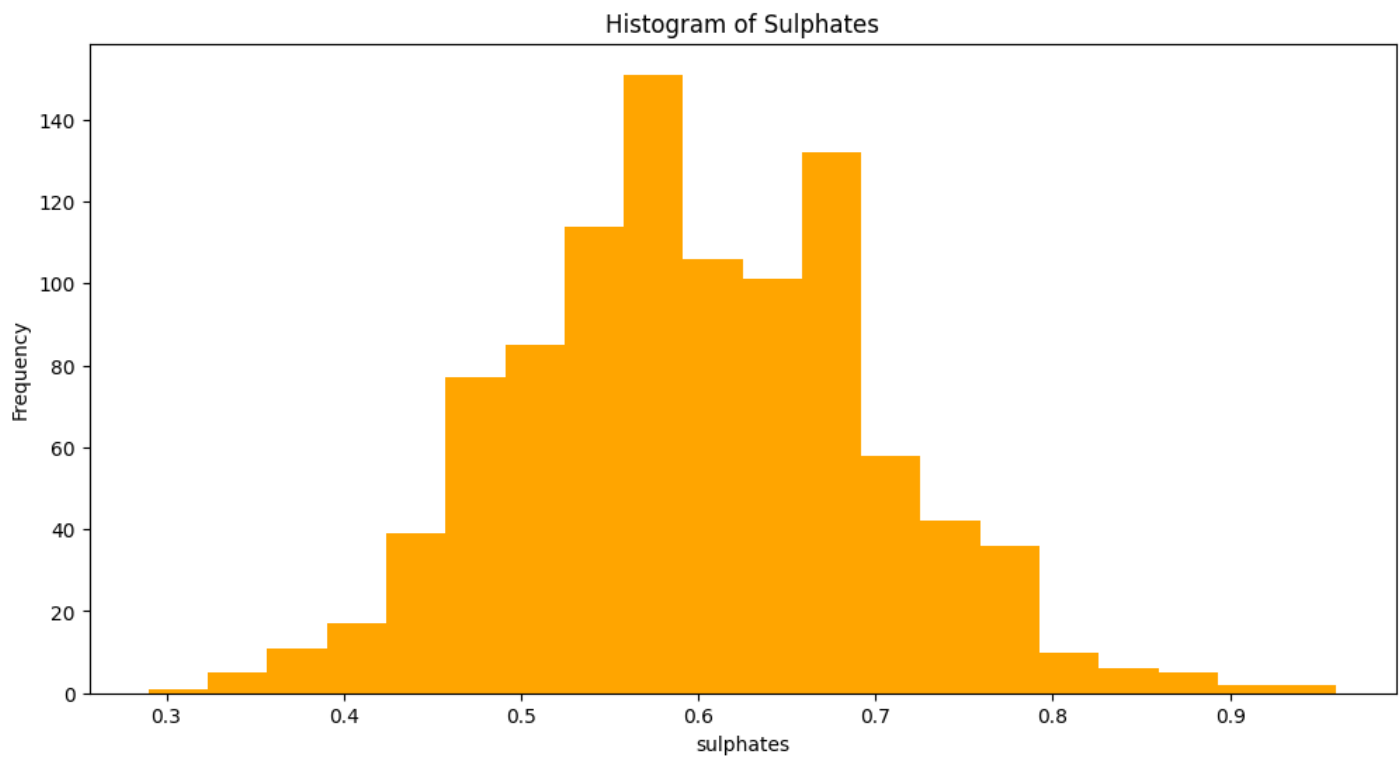


Berdasarkan kedua plot di atas, dapat disimpulkan bahwa:

1. Distribusi pH memiliki *skewness* normal.
2. Terdapat beberapa data *outlier* yang berada pada rentang 2.9 - 3.1 dan 3.5 - 3.8. Terlihat juga median berada di sekitar 3.3

Sulphates

```
In [ ]: generate_hist("sulphates")
generate_boxplot("sulphates")
```

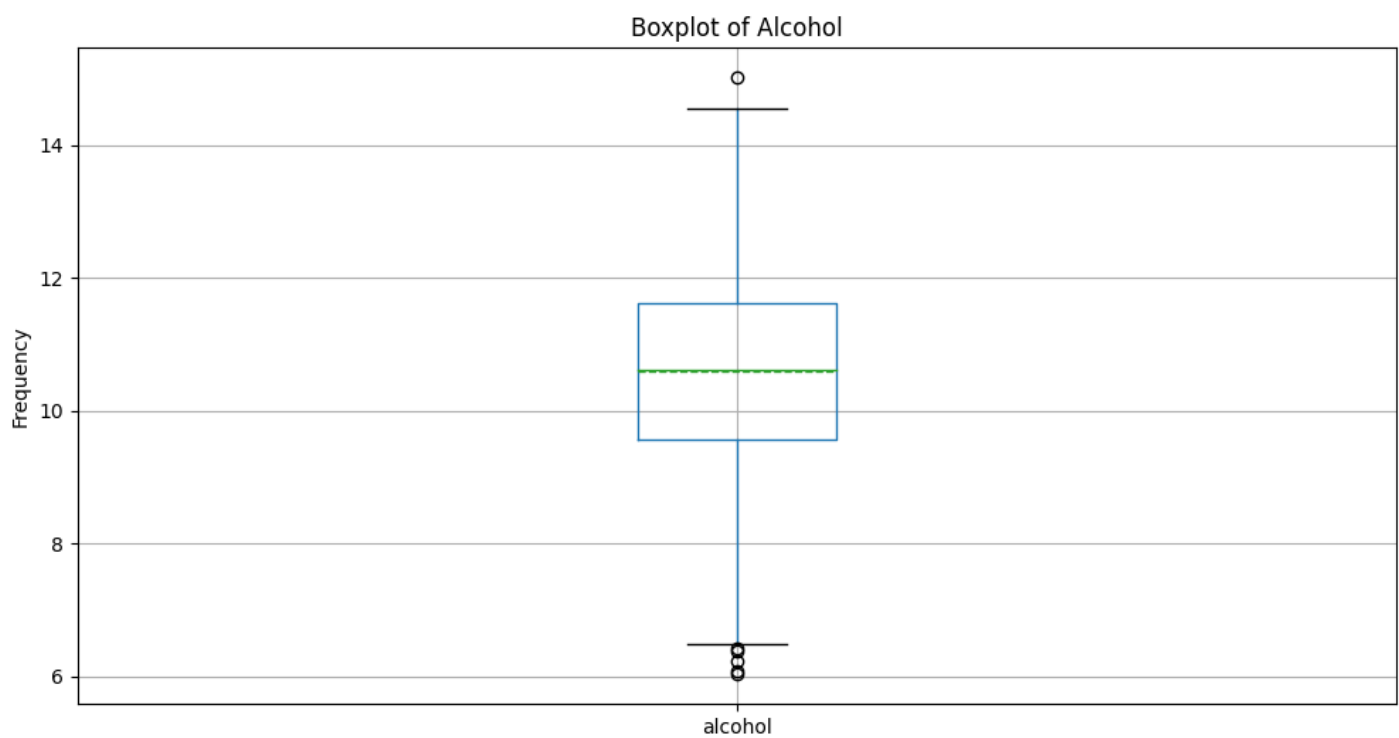
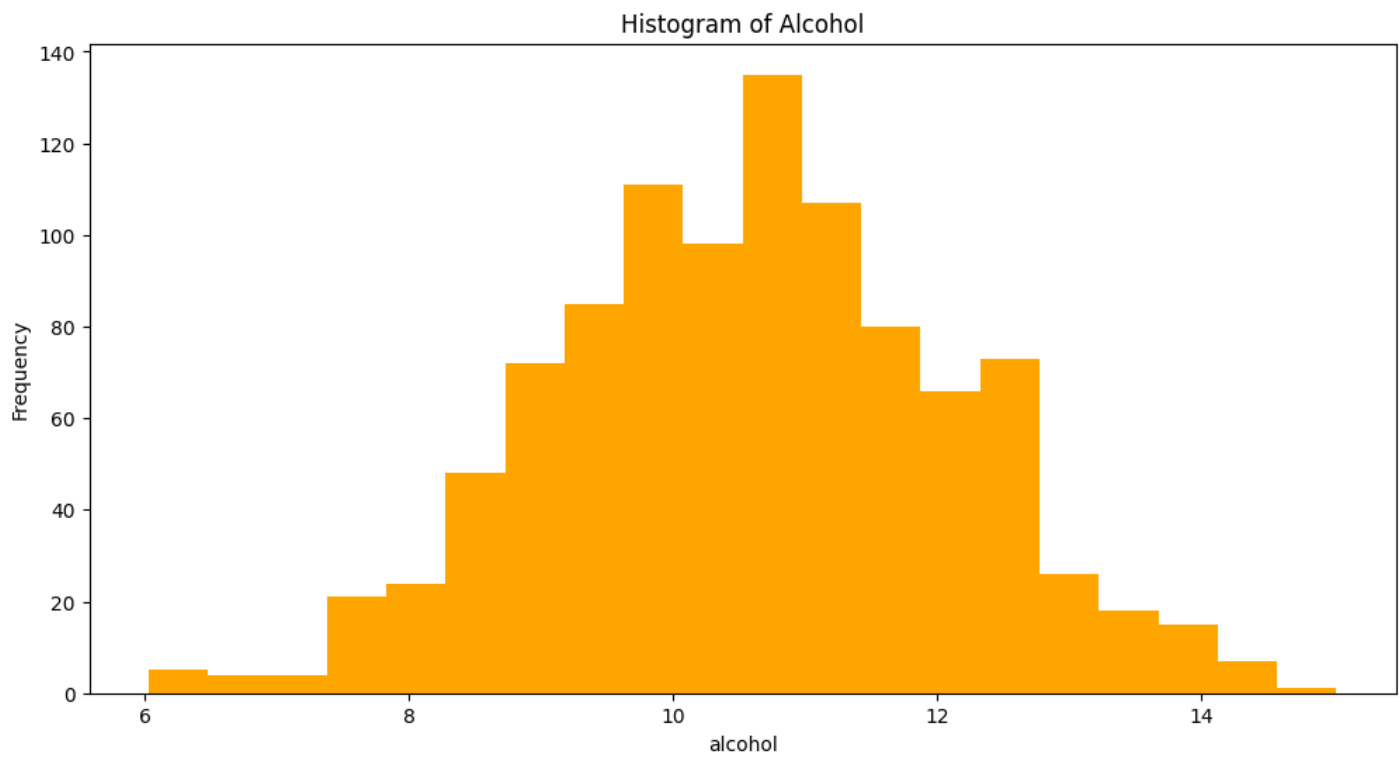



Berdasarkan kedua plot di atas, dapat disimpulkan bahwa:

1. Distribusi sulphates memiliki *skewness* normal.
2. Terdapat beberapa data *outlier* yang berada pada rentang 0.25 - 0.3 dan 0.85 - 1. Terlihat juga median berada di sekitar 0.6.

Alcohol

```
In [ ]: generate_hist("alcohol")
generate_boxplot("alcohol")
```

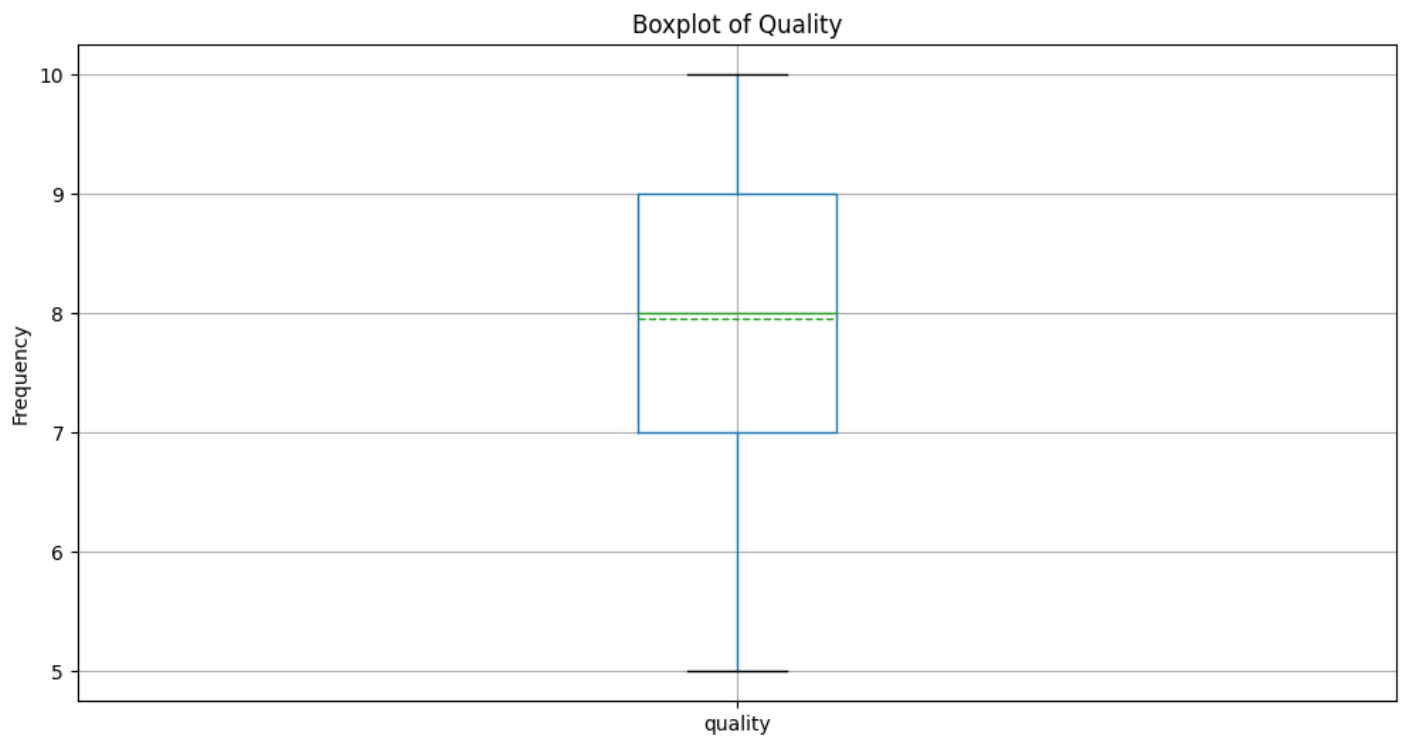
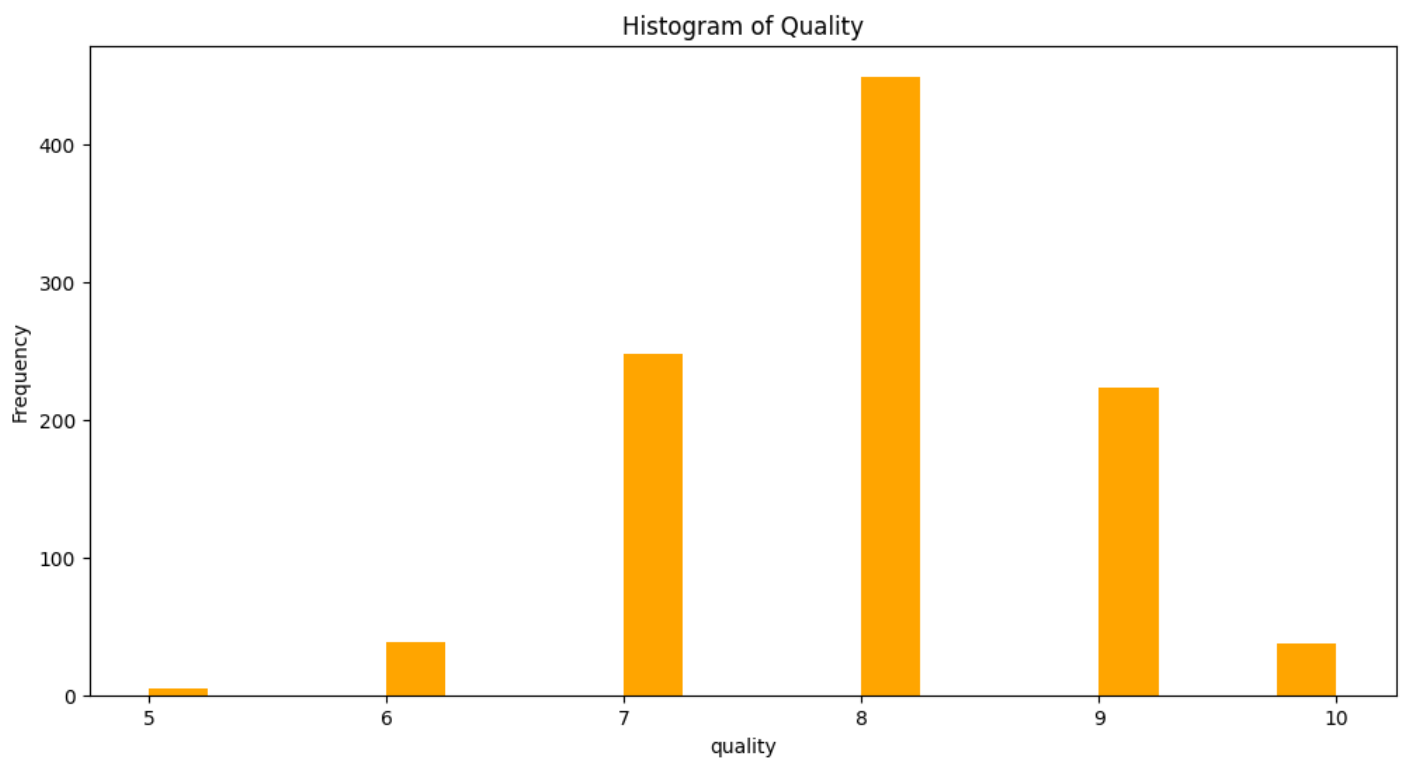


Berdasarkan kedua plot di atas, dapat disimpulkan bahwa:

1. Distribusi alcohol memiliki *skewness* normal.
2. Terdapat beberapa data *outlier* yang berada pada rentang 6 - 7 dan 14 - 16. Terlihat juga median berada di sekitar 11.

Quality

```
In [ ]: generate_hist("quality")
generate_boxplot("quality")
```



Berdasarkan kedua plot di atas, dapat disimpulkan bahwa:

1. Distribusi quality memiliki *left skewed*.
2. Tidak ada data *outlier*. Terlihat juga median berada di sekitar 8.

NOMOR 3

Menentukan setiap kolom numerik berdistribusi normal atau tidak. Gunakan normality test yang dikaitkan dengan histogram plot.

```
In [ ]: import pandas as pd
import scipy.stats as scp
import seaborn as sns
import matplotlib.pyplot as plt

# All columns contain numeric data
df = pd.read_csv('../dataset/anggur.csv')

# Drop target column
df = df.drop(['quality'], axis=1)
```

```
In [ ]: df.head()
```

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol
0	5.90	0.4451	0.1813	2.049401	0.070574	16.593818	42.27	0.9982	3.27	0.71	8.64
1	8.40	0.5768	0.2099	3.109590	0.101681	22.555519	16.01	0.9960	3.35	0.57	10.03
2	7.54	0.5918	0.3248	3.673744	0.072416	9.316866	35.52	0.9990	3.31	0.64	9.23
3	5.39	0.4201	0.3131	3.371815	0.072755	18.212300	41.97	0.9945	3.34	0.55	14.07
4	6.51	0.5675	0.1940	4.404723	0.066379	9.360591	46.27	0.9925	3.27	0.45	11.49

Menentukan setiap kolom numerik berdistribusi normal atau tidak. Gunakan normality test yang dikaitkan dengan histogram plot.

Untuk melakukan normality test untuk setiap kolom numerik, digunakan fungsi `normaltest` dari library `scipy`. Normality test ini didasarkan oleh `D'Agostino` dan `Pearson's` test yang mengombinasikan skewness dan kurtosis untuk mengukur normality suatu data. Hipotesis null (H_0) dari test ini berasal dari data yang terdistribusi normal.

Tingkat Signifikan $\alpha = 0.05$

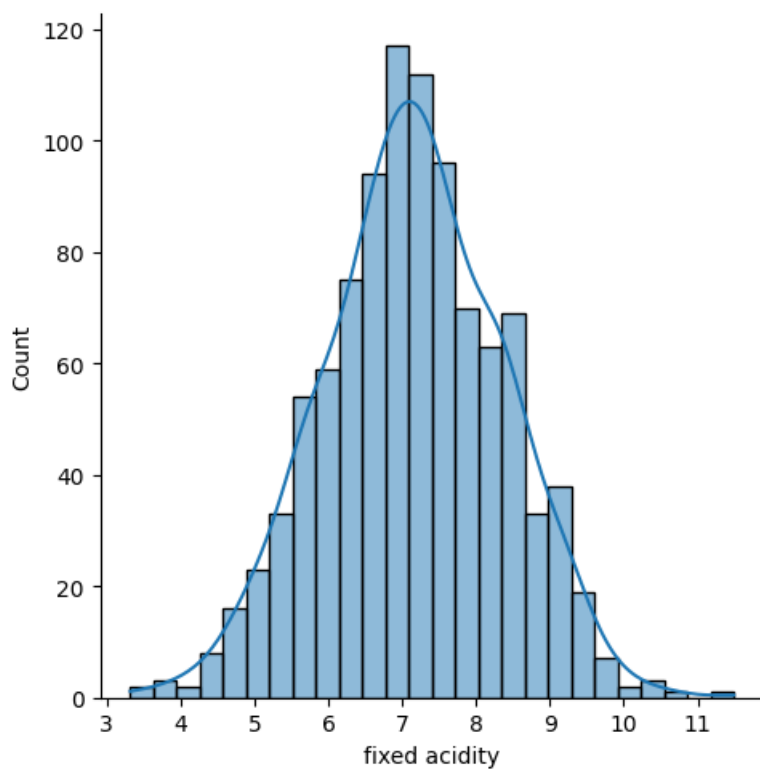
KEPUTUSAN : Tolak H_0 jika $p < \alpha$

```
In [ ]: # Iterate all columns in dataframe
for column in df.columns:
    stat, p = scp.normaltest(df[column])
    alpha = 0.05

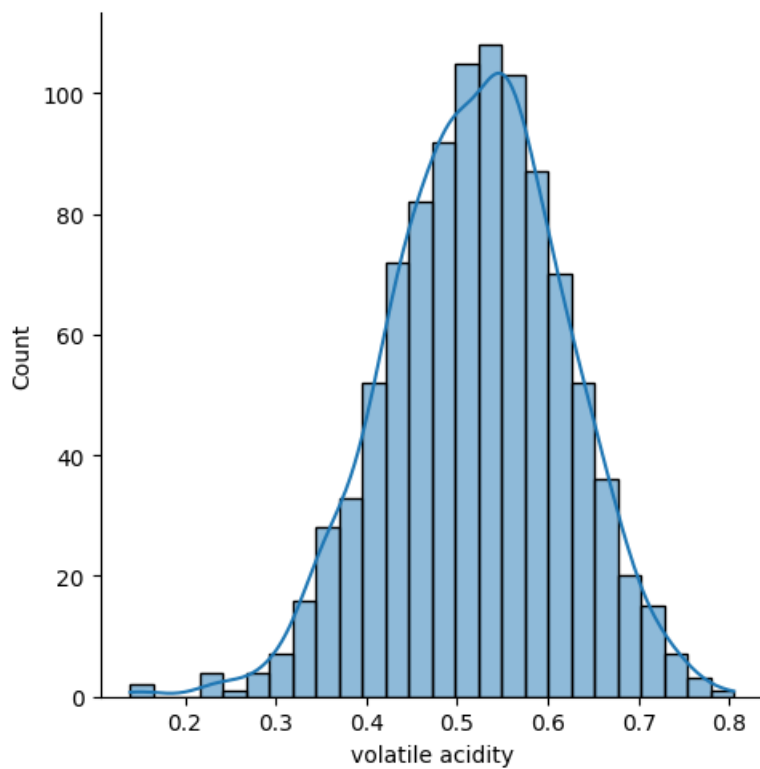
    print('p-value untuk kolom ' + str(column) + ' adalah ' + str(p))
    if p < alpha :
        print('H0 ditolak, kolom ' + str(column) + ' tidak berdistribusi normal')
    else :
        print('H0 diterima, kolom ' + str(column) + ' berdistribusi normal')

sns.displot(data = df[column], kde = True)
plt.show()
print()
```

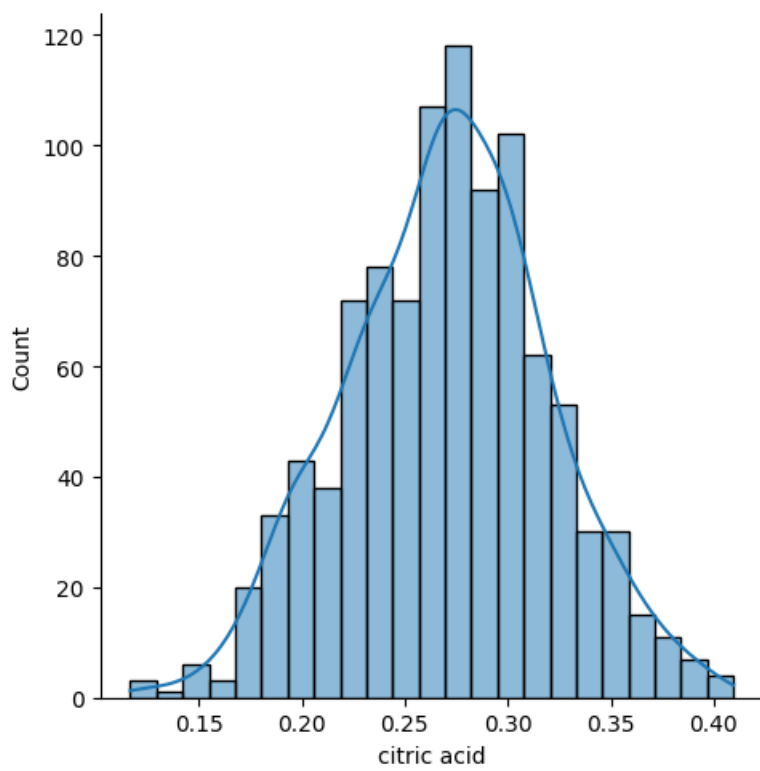
p-value untuk kolom fixed acidity adalah 0.9308584274486692
H0 diterima, kolom fixed acidity berdistribusi normal



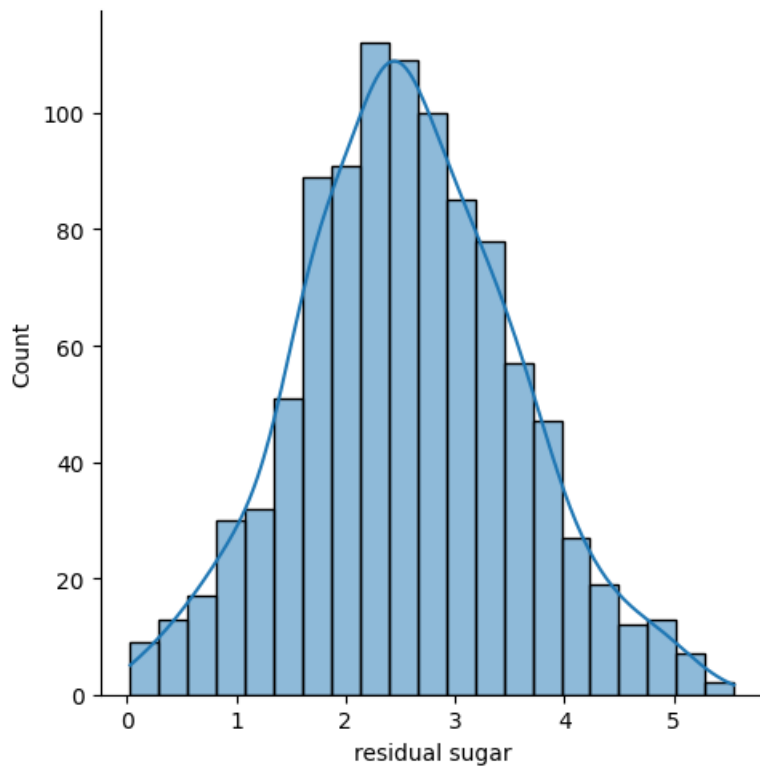
p-value untuk kolom volatile acidity adalah 0.022581461594113835
 H_0 ditolak, kolom volatile acidity tidak berdistribusi normal



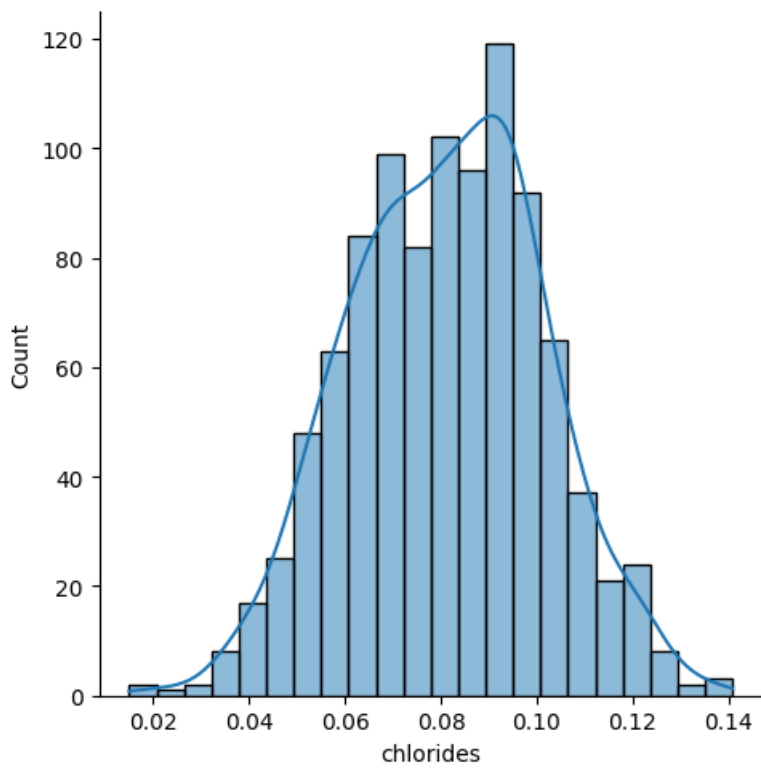
p-value untuk kolom citric acid adalah 0.6816899375976969
 H_0 diterima, kolom citric acid berdistribusi normal



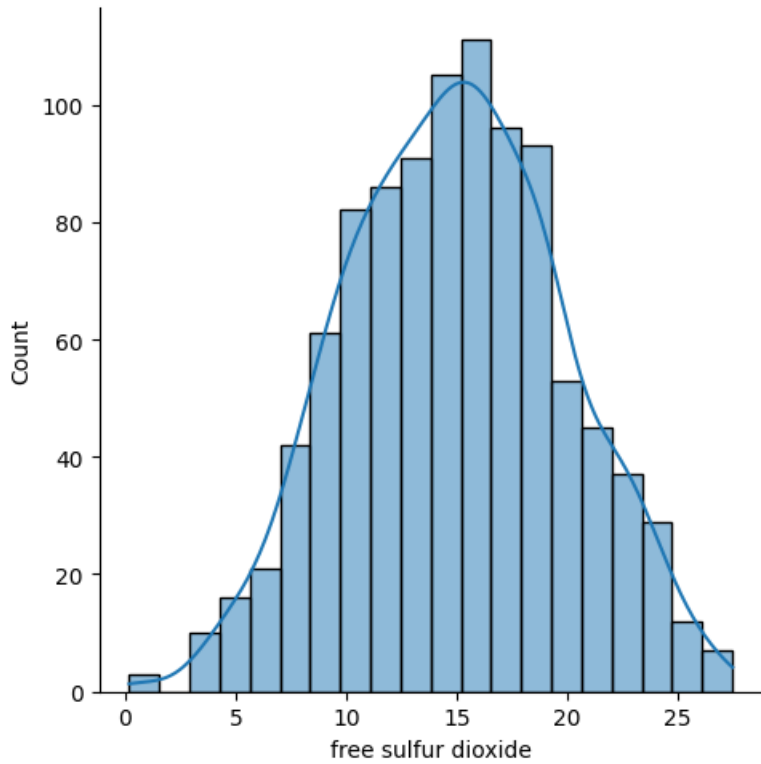
p-value untuk kolom residual sugar adalah 0.2246670332131056
 H_0 diterima, kolom residual sugar berdistribusi normal



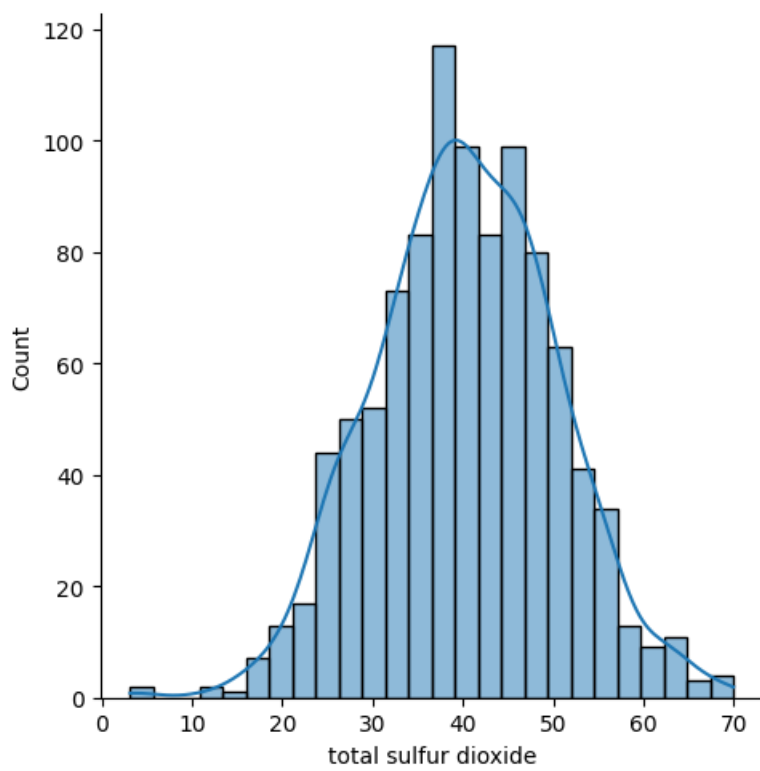
p-value untuk kolom chlorides adalah 0.17048274704296862
 H_0 diterima, kolom chlorides berdistribusi normal



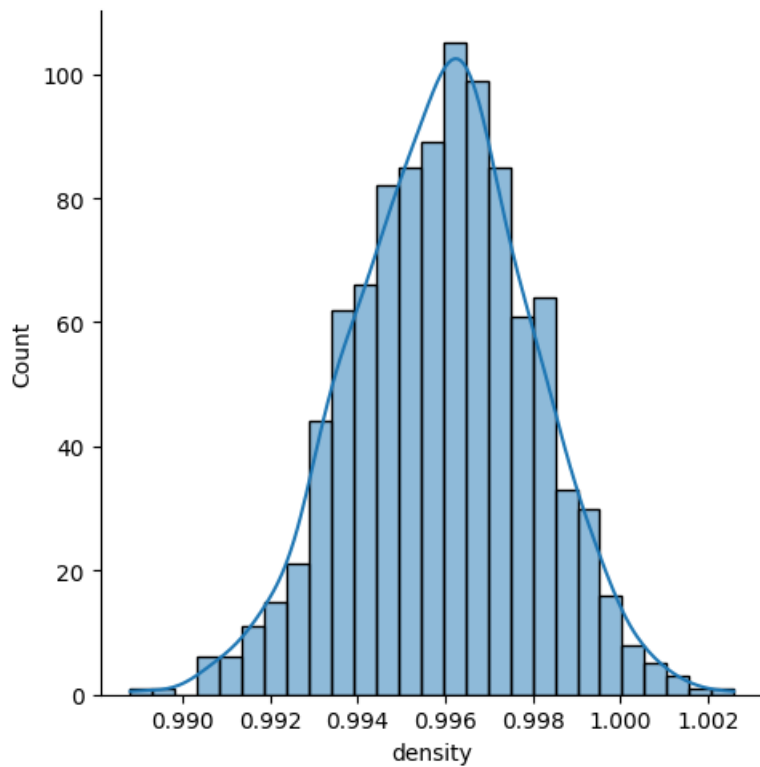
p-value untuk kolom free sulfur dioxide adalah 0.01743043451827735
 H_0 ditolak, kolom free sulfur dioxide tidak berdistribusi normal



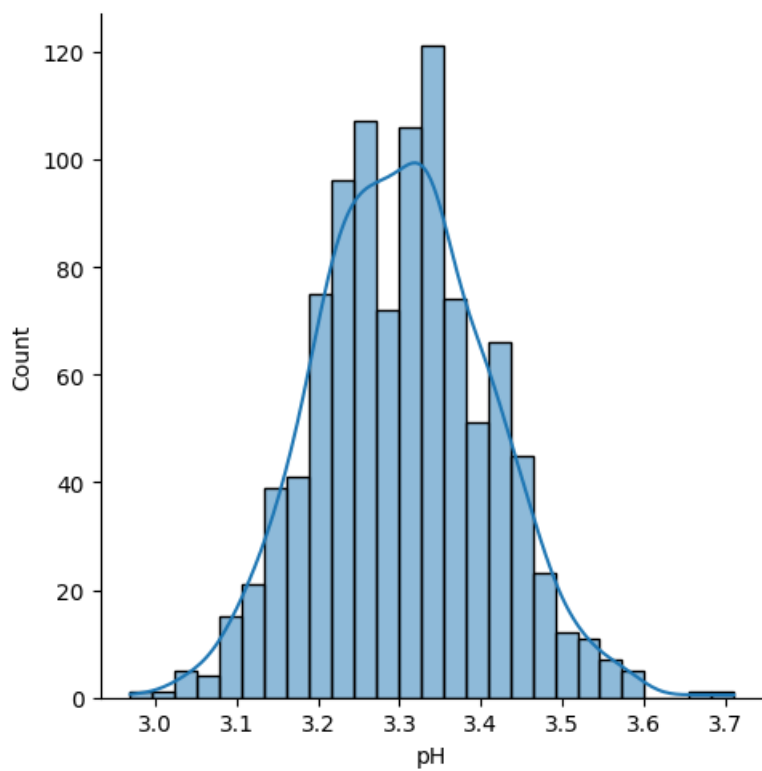
p-value untuk kolom total sulfur dioxide adalah 0.8488846101395726
 H_0 diterima, kolom total sulfur dioxide berdistribusi normal



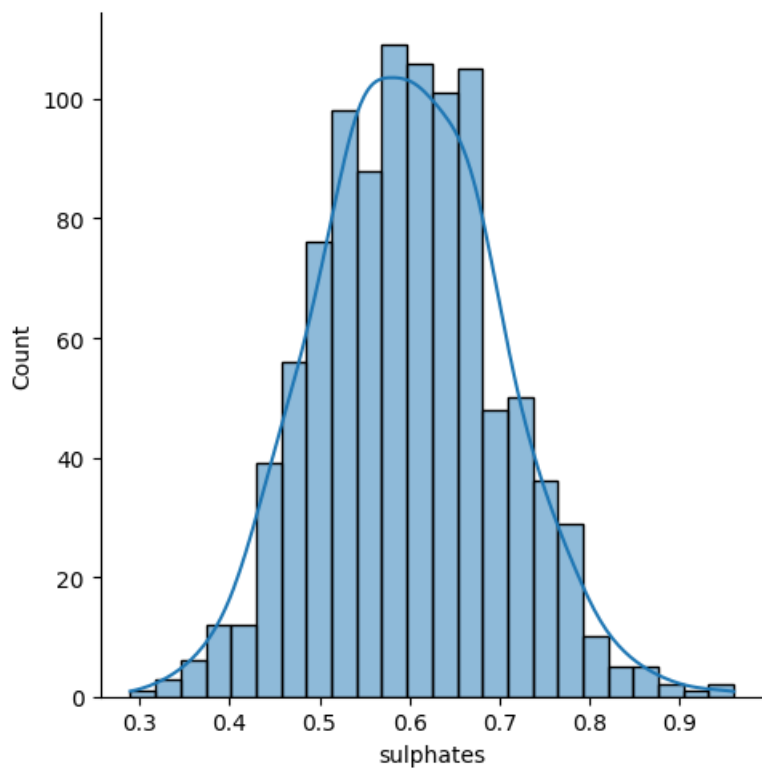
p-value untuk kolom density adalah 0.5985227325531981
 H_0 diterima, kolom density berdistribusi normal



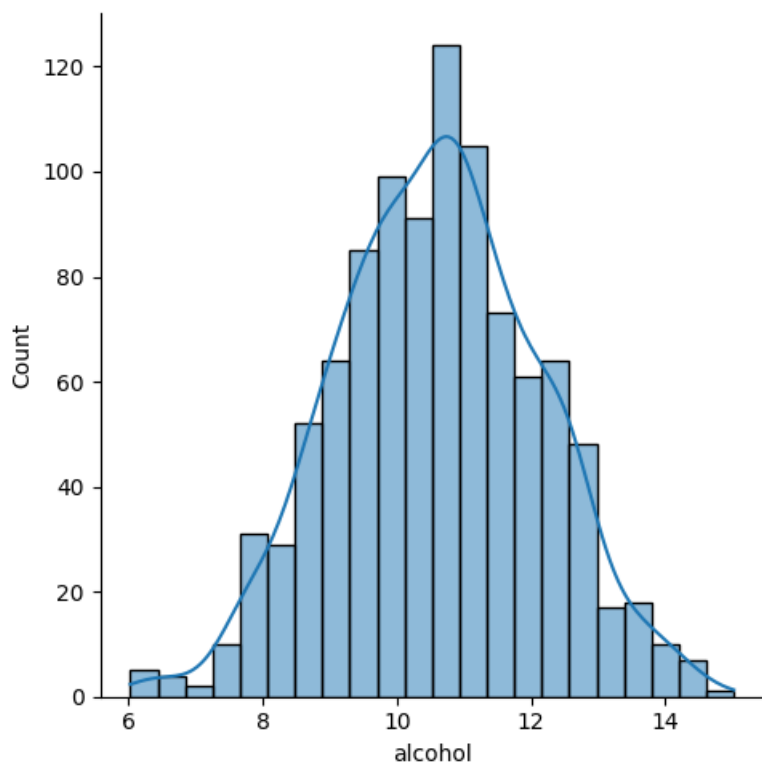
p-value untuk kolom pH adalah 0.13678740824860436
 H_0 diterima, kolom pH berdistribusi normal



p-value untuk kolom sulphates adalah 0.13884318628391681
 H_0 diterima, kolom sulphates berdistribusi normal



p-value untuk kolom alcohol adalah 0.6790884901361043
 H_0 diterima, kolom alcohol berdistribusi normal



NOMOR 4

Melakukan test hipotesis 1 sampel,

- Nilai rata-rata pH di atas 3.29?
- Nilai rata-rata Residual Sugar tidak sama dengan 2.50?
- Nilai rata-rata 150 baris pertama kolom sulphates bukan 0.65?
- Nilai rata-rata total sulfur dioxide di bawah 35?
- Proporsi nilai total Sulfat Dioxide yang lebih dari 40, adalah tidak sama dengan 50% ?

```
In [ ]: import pandas as pd
import scipy.stats as st
import matplotlib.pyplot as plt
from statsmodels.stats.weightstats import ztest
import math

df = pd.read_csv('../dataset/anggur.csv')
df.head()
```

```
Out[ ]:
```

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0	5.90	0.4451	0.1813	2.049401	0.070574	16.593818	42.27	0.9982	3.27	0.71	8.64	7
1	8.40	0.5768	0.2099	3.109590	0.101681	22.555519	16.01	0.9960	3.35	0.57	10.03	8
2	7.54	0.5918	0.3248	3.673744	0.072416	9.316866	35.52	0.9990	3.31	0.64	9.23	8
3	5.39	0.4201	0.3131	3.371815	0.072755	18.212300	41.97	0.9945	3.34	0.55	14.07	9
4	6.51	0.5675	0.1940	4.404723	0.066379	9.360591	46.27	0.9925	3.27	0.45	11.49	8

a) Nilai rata-rata pH di atas 3.29

1. Tentukan nilai hipotesis nol (H_0)

H_0 : Nilai rata rata pH sama dengan 3.29 ($\mu = 3.29$)

2. Tentukan hipotesis alternatif (H_1)

H_1 : Nilai rata rata pH lebih dari 3.29 ($\mu > 3.29$)

3. Tentukan tingkat signifikan

Tingkat Signifikan $\alpha = 0.05$

4. Tentukan uji statistik yang sesuai dan tentukan daerah kritis.

Uji statistik: **One Tailed Mean Test**

Critical section: $z > z_{\alpha} : z > z_{0.05}$

5. Hitung nilai uji statistik

Hitung nilai z dengan rumus

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Hitung nilai P

6. KEPUTUSAN : Tolak H_0 jika $z > z_{\alpha}$ dan $p < \alpha$

```
In [ ]: mu = 3.29
alpha = 0.05

#calculate z
z, p = ztest(df["pH"], value=mu)

#calculate z_alpha
z_alpha = st.norm.ppf(1-alpha)

print(f"Nilai z : {round(z, 4)}")
print(f"Nilai z_alpha : {round(z_alpha, 4)}")
print(f"Nilai P : {p}")

if (z > z_alpha and p < alpha):
    print("H0 ditolak")
```

```

else:
    print("H0 diterima")

df["pH"].plot(kind="box", figsize=(12, 6))
plt.title("pH")
plt.show()

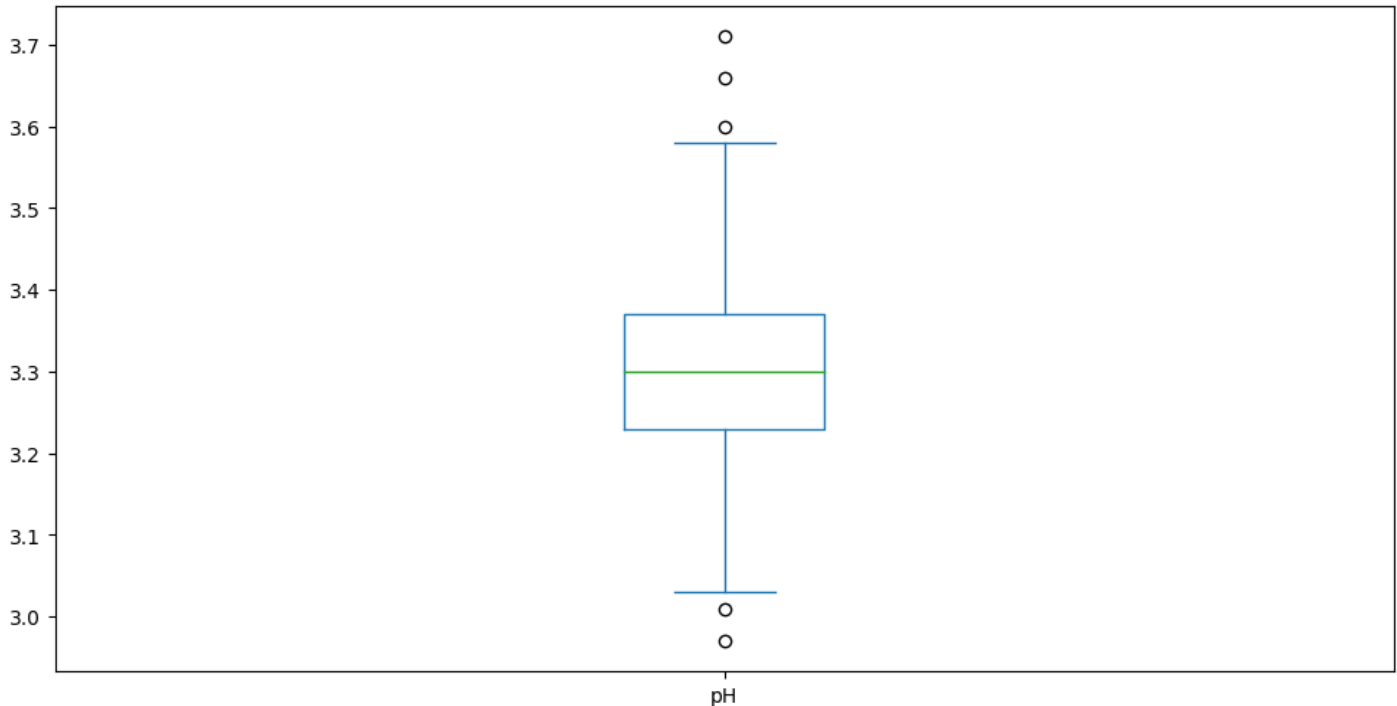
```

Nilai z : 4.1038

Nilai z_alpha : 1.6449

Nilai P : 4.0645260086604666e-05

H0 ditolak



Dari hasil di atas, nampak bahwa H0 ditolak. Jadi, **rata-rata pH > 3.29**

b) Nilai rata-rata Residual Sugar tidak sama dengan 2.50

1. Tentukan nilai hipotesis nol (H_0)

H_0 : Nilai rata rata residual sugar sama dengan 2.50 ($\mu = 2.50$)

2. Tentukan hipotesis alternatif (H_1)

H_1 : Nilai rata rata residual sugar lebih dari 2.50 ($\mu \neq 205$)

3. Tentukan tingkat signifikan

Tingkat Signifikan $\alpha = 0.05$

4. Tentukan uji statistik yang sesuai dan tentukan daerah kritis.

Uji statistik: **Two Tailed Mean Test**

Critical section: $z > z_{\alpha/2}$ atau $z < -z_{\alpha/2}$: $z > z_{0.025}$ atau $z < -z_{0.025}$

5. Hitung nilai uji statistik

Hitung nilai z dengan rumus

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Hitung nilai P

6. KEPUTUSAN : Tolak H_0 jika ($z > z_{\alpha/2}$ atau $z < -z_{\alpha/2}$) dan $p < \alpha$

```

In [ ]: mu = 2.50
alpha = 0.05

#calculate z
z, p = ztest(df["residual sugar"], value=mu)

#calculate z_alpha
z_alpha = st.norm.ppf(1-alpha/2)

print(f"Nilai z : {round(z, 4)}")

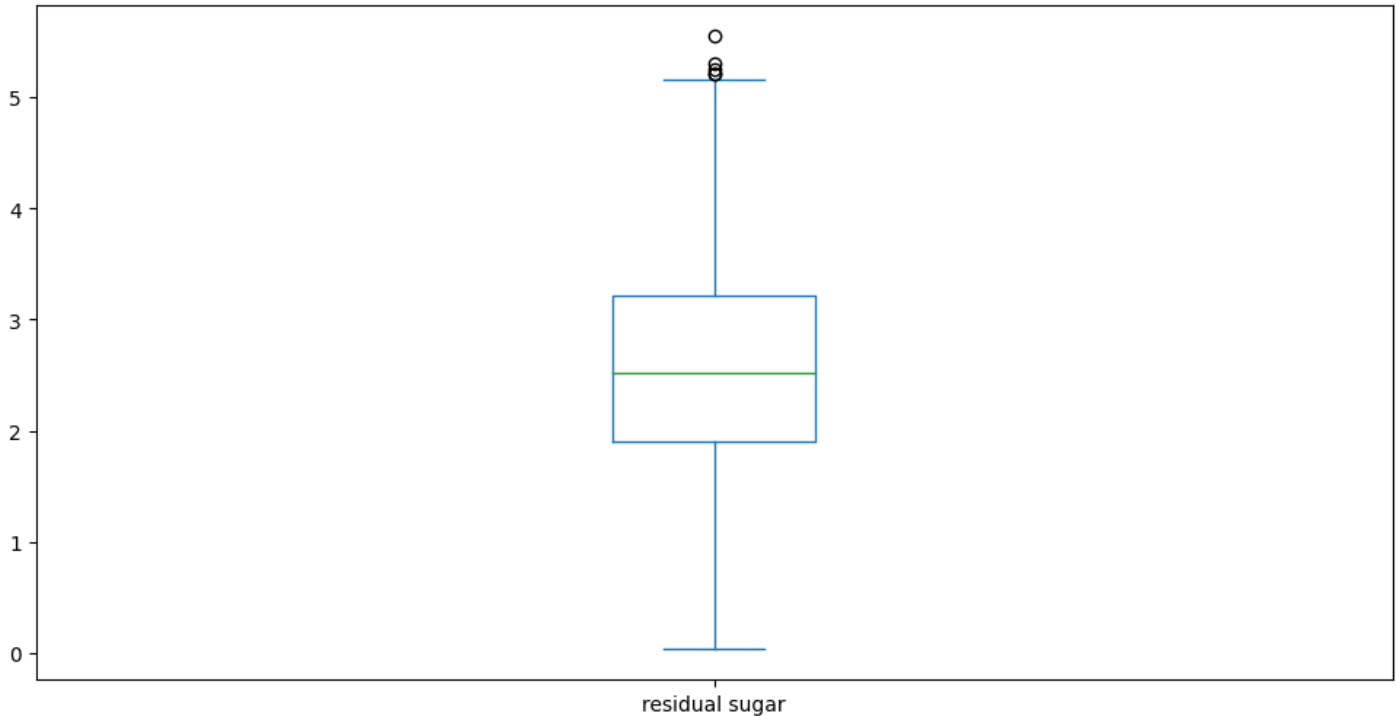
```

```
print(f"Nilai z_alpha/2 : {round(z_alpha, 4)}")
print(f"Nilai P : {p}")

if ((z > z_alpha or z < -z_alpha) and p < alpha):
    print("H0 ditolak")
else:
    print("H0 diterima")

df["residual sugar"].plot(kind="box", figsize=(12, 6))
plt.title("Residual Sugar")
plt.show()
```

Nilai z : 2.148
 Nilai z_alpha/2 : 1.96
 Nilai P : 0.031716778818727434
 H0 ditolak



Dari hasil di atas, nampak bahwa H0 ditolak. Jadi, **rata-rata residual sugar tidak sama dengan 2.50**

c) Nilai rata-rata 150 baris pertama kolom sulphates bukan 0.65

1. Tentukan nilai hipotesis nol (H_0)

H_0 : Nilai rata rata 50 baris pertama kolom sulphates sama dengan 0.65 ($\mu = 0.65$)

2. Tentukan hipotesis alternatif (H_1)

H_1 : Nilai rata rata 50 baris pertama kolom sulphates lebih dari 0.65 ($\mu \neq 0.65$)

3. Tentukan tingkat signifikan

Tingkat Signifikan $\alpha = 0.05$

4. Tentukan uji statistik yang sesuai dan tentukan daerah kritis.

Uji statistik: **Two Tailed Mean Test**

Critical section: $z > z_{\alpha/2}$ atau $z < -z_{\alpha/2}$: $z > z_{0.025}$ atau $z < -z_{0.025}$

5. Hitung nilai uji statistik

Hitung nilai z dengan rumus

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Hitung nilai P

6. KEPUTUSAN : Tolak H_0 jika ($z > z_{\alpha/2}$ atau $z < -z_{\alpha/2}$) dan $p < \alpha$

```
In [ ]: mu = 0.65
alpha = 0.05

#calculate z
z, p = ztest(df["sulphates"].head(150), value = mu)
```

```
#calculate z_alpha
z_alpha = st.norm.ppf(1-alpha/2)

print(f"Nilai z : {round(z, 4)}")
print(f"Nilai z_alpha/2 : {round(z_alpha, 4)}")
print(f"Nilai P : {p}")

if ((z>z_alpha or z < -z_alpha) and p<alpha):
    print("H0 ditolak")
else:
    print("H0 diterima")

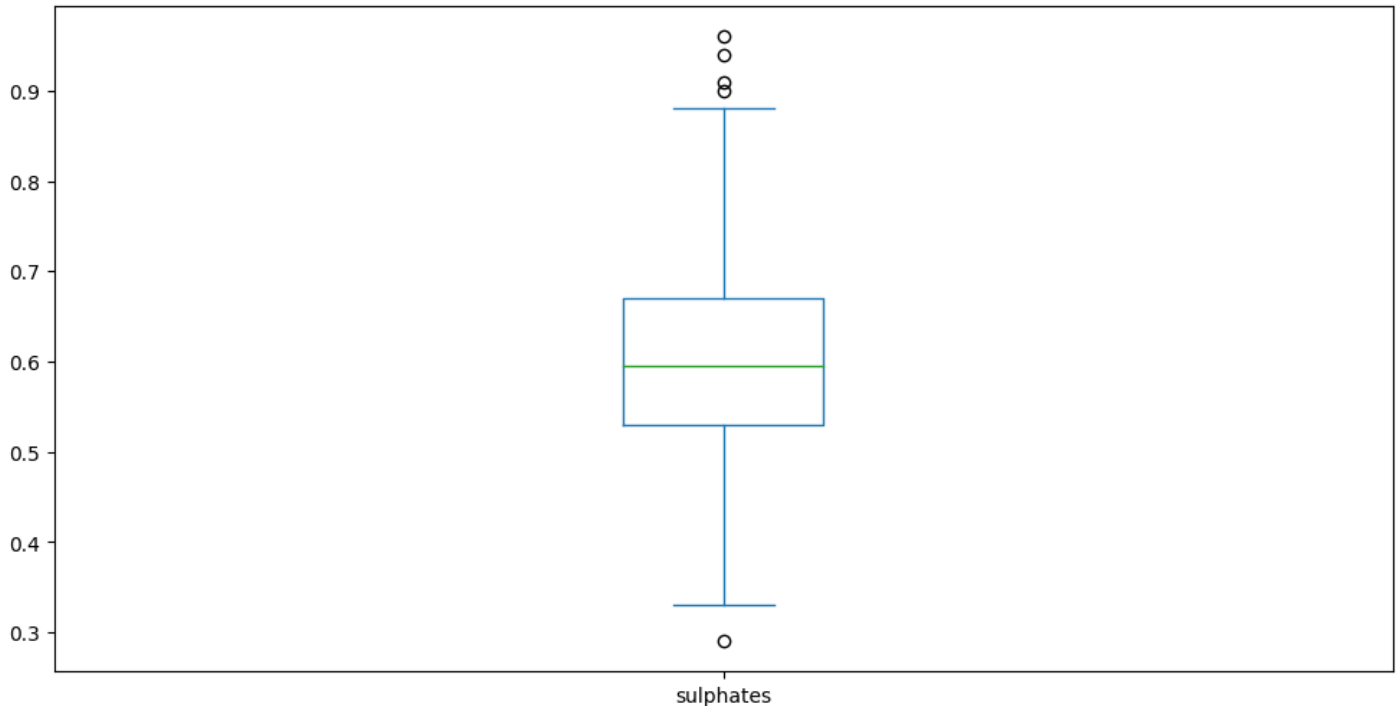
df["sulphates"].plot(kind="box", figsize=(12, 6))
plt.title("Sulphates")
plt.show()
```

Nilai z : -4.9648

Nilai z_alpha/2 : 1.96

Nilai P : 6.875652918327357e-07

H0 ditolak



Dari hasil di atas, nampak bahwa H0 ditolak. Jadi, **rata-rata 150 baris pertama kolom sulphates bukan 0.65**

d) Nilai rata-rata total sulfur dioxide di bawah 35

1. Tentukan nilai hipotesis nol (H_0)

H_0 : Nilai rata rata total sulfur dioxide sama dengan 35 ($\mu = 35$)

2. Tentukan hipotesis alternatif (H_1)

H_1 : Nilai rata rata total sulfur dioxide lebih dari 35 ($\mu < 35$)

3. Tentukan tingkat signifikan

Tingkat Signifikan $\alpha = 0.05$

4. Tentukan uji statistik yang sesuai dan tentukan daerah kritis.

Uji statistik: **One Tailed Mean Test**

Critical section: $z < -z_\alpha : z < -z_{0.05}$

5. Hitung nilai uji statistik

Hitung nilai z dengan rumus

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Hitung nilai P

6. KEPUTUSAN : Tolak H_0 jika $z < -z_\alpha$ dan $p < \alpha$

```
In [ ]: mu = 35
alpha = 0.05

#calculate z
z, p = ztest(df["total sulfur dioxide"], value=mu, alternative="smaller")

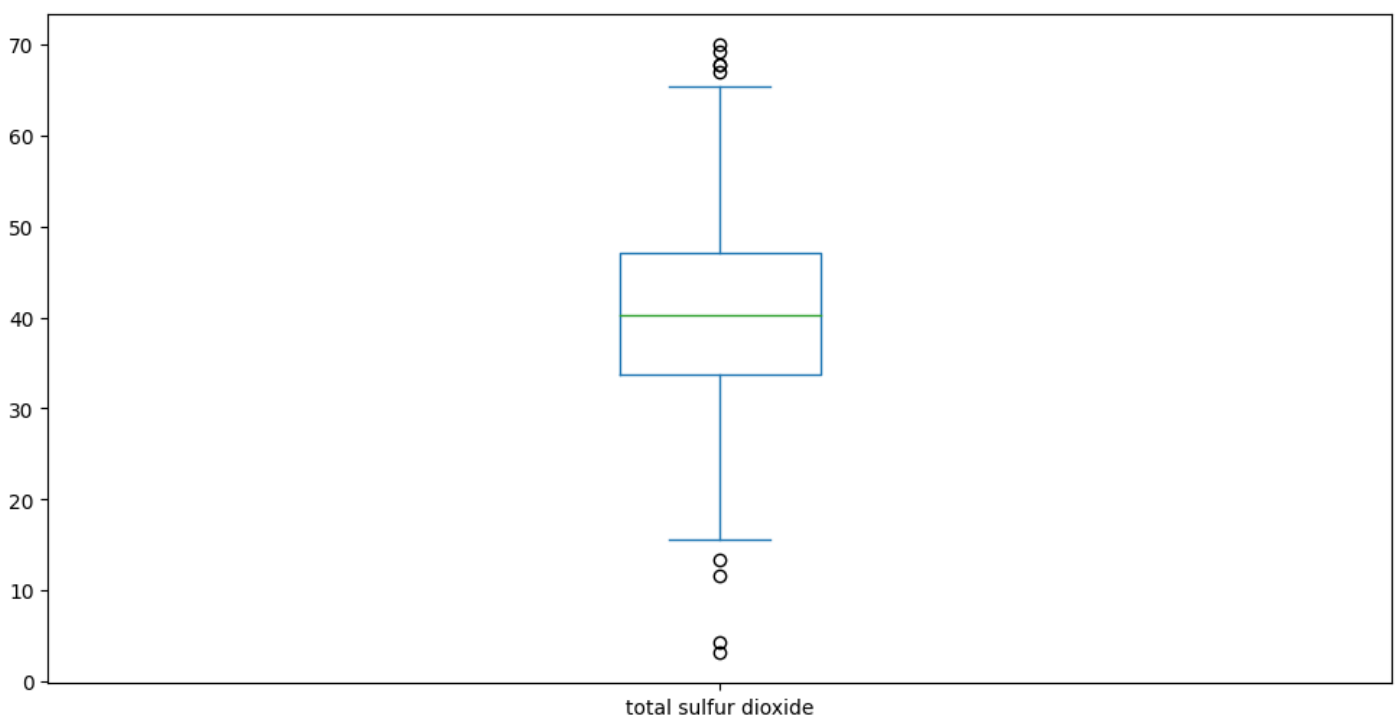
#calculate z_alpha
z_alpha = st.norm.cdf(1-alpha)

print(f"Nilai z : {round(z, 4)}")
print(f"Nilai z_alpha : {round(z_alpha, 4)}")
print(f"Nilai P : {p}")

if (z < -z_alpha and p<alpha):
    print("H0 ditolak")
else:
    print("H0 diterima")

df["total sulfur dioxide"].plot(kind="box", figsize=(12, 6))
plt.title("Total Sulfur Dioxide")
plt.show()
```

Nilai z : 16.7864
 Nilai z_alpha : 0.8289
 Nilai P : 1.0
 H0 diterima



Dari hasil di atas, nampak bahwa H0 diterima. Jadi, **rata-rata total sulfur dioxide tidak berada di bawah 35**

Proporsi nilai total Sulfat Dioxide yang lebih dari 40, adalah tidak sama dengan 50%

1. Tentukan nilai hipotesis nol (H_0)

H_0 : Proporsi nilai total Sulfat Dioxide yang lebih dari 40, adalah sama dengan 50% 0.5 ($p = 0.5$)

2. Tentukan hipotesis alternatif (H_1)

H_1 : Proporsi nilai total Sulfat Dioxide yang lebih dari 40, adalah tidak sama dengan 50% ($p \neq 0.5$)

3. Tentukan tingkat signifikan

Tingkat Signifikan $\alpha = 0.05$

4. Tentukan uji statistik yang sesuai dan tentukan daerah kritis.

Uji statistik: **Uji Variabel Binomial X** dengan $p = p_0$

Critical section: $z > z_{\alpha/2}$ atau $z < -z_{\alpha/2}$: $z > z_{0.025}$ atau $z < -z_{0.025}$

5. Hitung nilai uji statistik

Hitung nilai z dengan rumus

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}}$$

Hitung nilai P dari x (jumlah sukses)

6. KEPUTUSAN : Tolak H_0 jika $(z > z_{\alpha/2}$ atau $z < -z_{\alpha/2})$ dan $p < \alpha$

```
In [ ]: alpha = 0.05
# total sulfur dioxide count
n = len(df["total sulfur dioxide"])
# total sulfur dioxide > 40 count
n_sampel = len(df[df["total sulfur dioxide"] > 40])
p1 = n_sampel / n
p0 = 0.5
q0 = 0.5

#calculate z
z = (p1-p0)/math.sqrt(p0*q0/n)

#calculate z alpha
z_alpha = st.norm.ppf(1-alpha/2)

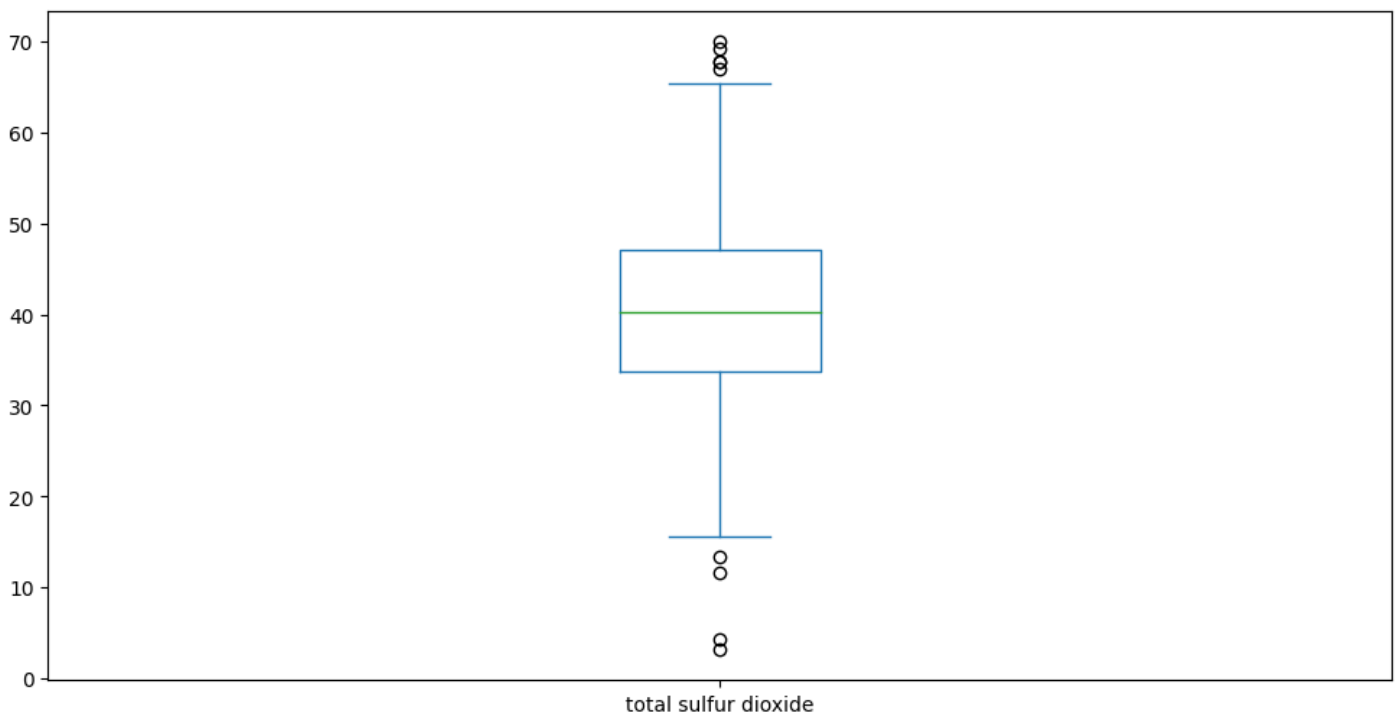
#calculate p
p = st.norm.sf(abs(z)) * 2

print(f"Nilai z : {round(z, 4)}")
print(f"Nilai z_alpha/2 : {round(z_alpha, 4)}")
print(f"Nilai P : {p}")

if ((z>z_alpha or z < -z_alpha) and p<alpha):
    print("H0 ditolak")
else:
    print("H0 diterima")

df["total sulfur dioxide"].plot(kind="box", figsize=(12, 6))
plt.title("Total Sulfur Dioxide")
plt.show()
```

Nilai z : 0.7589
Nilai z_alpha/2 : 1.96
Nilai P : 0.4478844782641115
H0 diterima



Dari hasil di atas, nampak bahwa H0 diterima. Jadi, **proporsi nilai total Sulfat Dioxide yang lebih dari 40, adalah sama dengan 50%**

NOMOR 5

Melakukan test hipotesis 2 sampel,

- a. Data kolom fixed acidity dibagi 2 sama rata: bagian awal dan bagian akhir kolom. Benarkah rata-rata kedua bagian tersebut sama?
- b. Data kolom chlorides dibagi 2 sama rata: bagian awal dan bagian akhir kolom. Benarkah rata-rata bagian awal lebih besar daripada bagian akhir sebesar 0.001?
- c. Benarkah rata-rata sampel 25 baris pertama kolom Volatile Acidity sama dengan rata-rata 25 baris pertama kolom Sulphates ?
- d. Bagian awal kolom residual sugar memiliki variansi yang sama dengan bagian akhirnya?
- e. Proporsi nilai setengah bagian awal alcohol yang lebih dari 7, adalah lebih besar daripada, proporsi nilai yang sama di setengah bagian akhir alcohol?

```
In [ ]: import pandas as pd
import scipy.stats as scp
import matplotlib.pyplot as plt
from statsmodels.stats.weightstats import ztest
from statsmodels.stats.proportion import proportions_ztest
df = pd.read_csv('../dataset/anggur.csv')
df.head()
```

Out []:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0	5.90	0.4451	0.1813	2.049401	0.070574	16.593818	42.27	0.9982	3.27	0.71	8.64	7
1	8.40	0.5768	0.2099	3.109590	0.101681	22.555519	16.01	0.9960	3.35	0.57	10.03	8
2	7.54	0.5918	0.3248	3.673744	0.072416	9.316866	35.52	0.9990	3.31	0.64	9.23	8
3	5.39	0.4201	0.3131	3.371815	0.072755	18.212300	41.97	0.9945	3.34	0.55	14.07	9
4	6.51	0.5675	0.1940	4.404723	0.066379	9.360591	46.27	0.9925	3.27	0.45	11.49	8

a. Data kolom fixed acidity dibagi 2 sama rata: bagian awal dan bagian akhir kolom. Benarkah rata-rata kedua bagian tersebut sama?

- 1. Tentukan nilai hipotesis nol (H_0)
 H_0 : Nilai rata-rata kolom awal fixed acidity sama dengan nilai rata-rata kolom akhir fixed acidity ($\mu_1 - \mu_2 = 0$)
- 2. Tentukan hipotesis alternatif (H_1)
 H_1 : Nilai rata-rata kolom awal fixed acidity tidak sama dengan nilai rata-rata kolom akhir fixed acidity ($\mu_1 - \mu_2 \neq 0$)
- 3. Tentukan tingkat signifikan
Tingkat Signifikan $\alpha = 0.05$
- 4. Tentukan uji statistik yang sesuai dan tentukan daerah kritis.
Uji statistik: **Two Tailed Mean Test**
Critical section: $z > z_{\alpha/2}$ atau $z < -z_{\alpha/2} : z > z_{0.025}$ atau $z < -z_{0.025}$
- 5. Hitung nilai uji statistik
Hitung nilai z dengan rumus

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - \text{delta}}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

Hitung nilai P

6. KEPUTUSAN : Tolak H_0 jika ($z > z_{\alpha/2}$ atau $z < -z_{\alpha/2}$) dan $p < \alpha$

```
In [ ]: # Initialize constant variables
df_fixed_acidity = df["fixed acidity"]
delta = 0
alpha = 0.05

# Divide fixed acidity into two
df_fixed_acidity_head = df_fixed_acidity[:len(df)//2]
df_fixed_acidity_tail = df_fixed_acidity[len(df)//2:]

# find z and p value
z, p = ztest(df_fixed_acidity_head, df_fixed_acidity_tail, value=delta)

# Find z(alpha/2) value
z_alpha_over_2 = scp.norm.ppf(1-alpha/2)

# Print results
print('Nilai z          : ' + str(z))
print('Nilai z(alpha/2) : ' + str(z_alpha_over_2))
print('Nilai p          : ' + str(p))

if ((z > z_alpha_over_2 or z < -z_alpha_over_2) and p < alpha):
    print('Hipotesis H0 ditolak, artinya rata-rata kolom awal fixed acidity tidak sama dengan nilai rata-rata kolom akhir fixed acidity')
else :
    print('Hipotesis H0 diterima, artinya rata-rata kolom awal fixed acidity sama dengan nilai rata-rata kolom akhir fixed acidity')
```

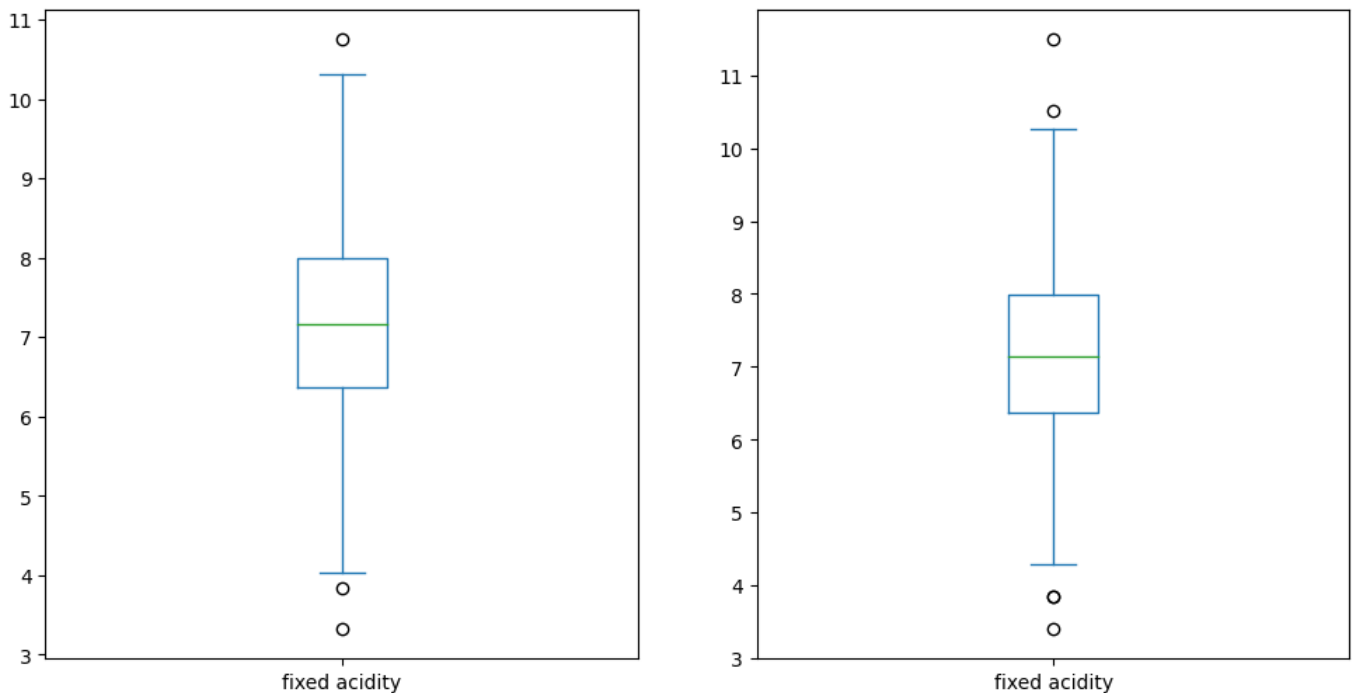
```
# Plot each side
plt.subplot(1, 2, 1)
df_fixed_acidity_head.plot(kind="box", figsize=(12, 6))
plt.subplot(1, 2, 2)
df_fixed_acidity_tail.plot(kind="box", figsize=(12, 6))
plt.show()
```

Nilai z : 0.02604106999906379

Nilai z(alpha/2) : 1.959963984540054

Nilai p : 0.9792245804254097

Hipotesis H_0 diterima, artinya rata-rata kolom awal fixed acidity sama dengan nilai rata-rata kolom akhir fixed acidity



b. Data kolom chlorides dibagi 2 sama rata: bagian awal dan bagian akhir kolom. Benarkah rata-rata bagian awal lebih besar daripada bagian akhir sebesar 0.001?

1. Tentukan nilai hipotesis nol (H_0)

H_0 : Nilai rata-rata kolom chlorides bagian awal lebih besar daripada bagian akhir sebesar 0.001 ($\mu_1 - \mu_2 = 0.001$)

2. Tentukan hipotesis alternatif (H_1)

H_1 : Nilai rata-rata kolom chlorides bagian awal tidak lebih besar daripada bagian akhir sebesar 0.001 ($\mu_1 - \mu_2 \neq 0.001$)

3. Tentukan tingkat signifikan

Tingkat Signifikan $\alpha = 0.05$

4. Tentukan uji statistik yang sesuai dan tentukan daerah kritis.

Uji statistik: **Two Tailed Mean Test**

Critical section: $z > z_{\alpha/2}$ atau $z < -z_{\alpha/2}$: $z > z_{0.025}$ atau $z < -z_{0.025}$

5. Hitung nilai uji statistik

Hitung nilai z dengan rumus

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - \text{delta}}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

Hitung nilai P

6. KEPUTUSAN : Tolak H_0 jika ($z > z_{\alpha/2}$ atau $z < -z_{\alpha/2}$) dan $p < \alpha$

```
In [ ]: # Initialize constant variables
df_chlorides = df["chlorides"]
delta = 0.001
alpha = 0.05

# Divide chlorides column into two
df_chlorides_head = df_chlorides[:len(df)//2]
df_chlorides_tail = df_chlorides[len(df)//2:]

# find z and p value
z, p = ztest(df_chlorides_head, df_chlorides_tail, value=delta)

# Find z(alpha/2) value
z_alpha_over_2 = scp.norm.ppf(1-alpha/2)

# Print results
print('Nilai z : ' + str(z))
print('Nilai z(alpha/2) : ' + str(z_alpha_over_2))
print('Nilai p : ' + str(p))
```

```

if ((z > z_alpha_over_2 or z < -z_alpha_over_2) and p < alpha):
    print('Hipotesis H0 ditolak, artinya rata-rata kolom chlorides bagian awal tidak lebih besar daripada bagian akhir sebesar 0.001')
else :
    print('Hipotesis H0 diterima, artinya rata-rata kolom chlorides bagian awal lebih besar daripada bagian akhir sebesar 0.001')

# Plot each side
plt.subplot(1, 2, 1)
df_chlorides_head.plot(kind="box", figsize=(12, 6))
plt.subplot(1, 2, 2)
df_chlorides_tail.plot(kind="box", figsize=(12, 6))
plt.show()

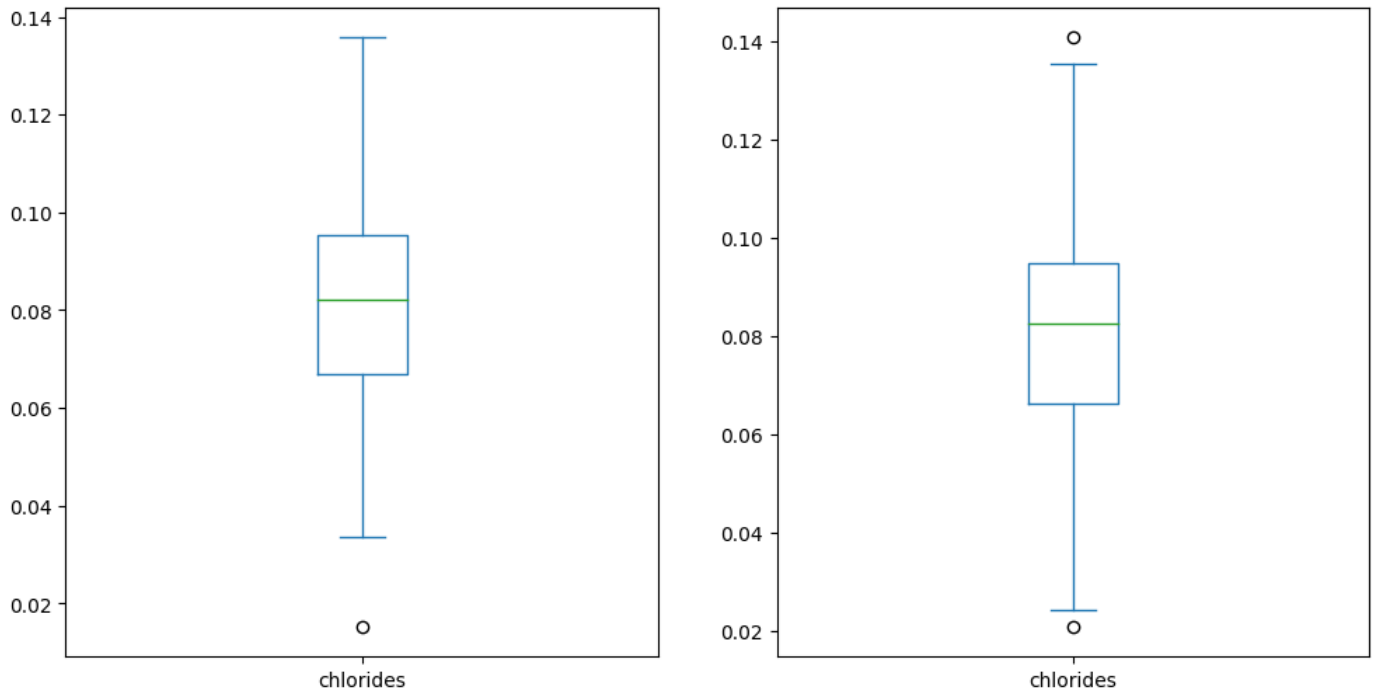
```

Nilai z : -0.467317122852132

Nilai z(alpha/2) : 1.959963984540054

Nilai p : 0.640273007581107

Hipotesis H0 diterima, artinya rata-rata kolom chlorides bagian awal lebih besar daripada bagian akhir sebesar 0.001



c. Benarkah rata-rata sampel 25 baris pertama kolom Volatile Acidity sama dengan rata-rata 25 baris pertama kolom Sulphates ?

1. Tentukan nilai hipotesis nol (H_0)

H_0 : Nilai rata-rata sampel 25 baris pertama kolom Volatile Acidity sama dengan rata-rata 25 baris pertama kolom Sulphates ($\mu_1 - \mu_2 = 0$)

2. Tentukan hipotesis alternatif (H_1)

H_1 : Nilai rata-rata sampel 25 baris pertama kolom Volatile Acidity tidak sama dengan rata-rata 25 baris pertama kolom Sulphates ($\mu_1 - \mu_2 \neq 0$)

3. Tentukan tingkat signifikan

Tingkat Signifikan $\alpha = 0.05$

4. Tentukan uji statistik yang sesuai dan tentukan daerah kritis.

Uji statistik: **Two Tailed Mean Test**

Critical section: $z > z_{\alpha/2}$ atau $z < -z_{\alpha/2}$: $z > z_{0.025}$ atau $z < -z_{0.025}$

5. Hitung nilai uji statistik

Hitung nilai z dengan rumus

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - \text{delta}}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

Hitung nilai P

6. KEPUTUSAN : Tolak H_0 jika ($z > z_{\alpha/2}$ atau $z < -z_{\alpha/2}$) dan $p < \alpha$

```

In [ ]: # Initialize constant variables
df_first_25_volatile_acidity = df.loc[0:24, 'volatile acidity']
df_first_25_sulphates = df.loc[0:24, 'sulphates']

delta = 0
alpha = 0.05

# Find z and p value
z, p = ztest(df_first_25_volatile_acidity, df_first_25_sulphates, value=delta)

# Find z(alpha/2) value
z_alpha_over_2 = scp.norm.ppf(1-alpha/2)

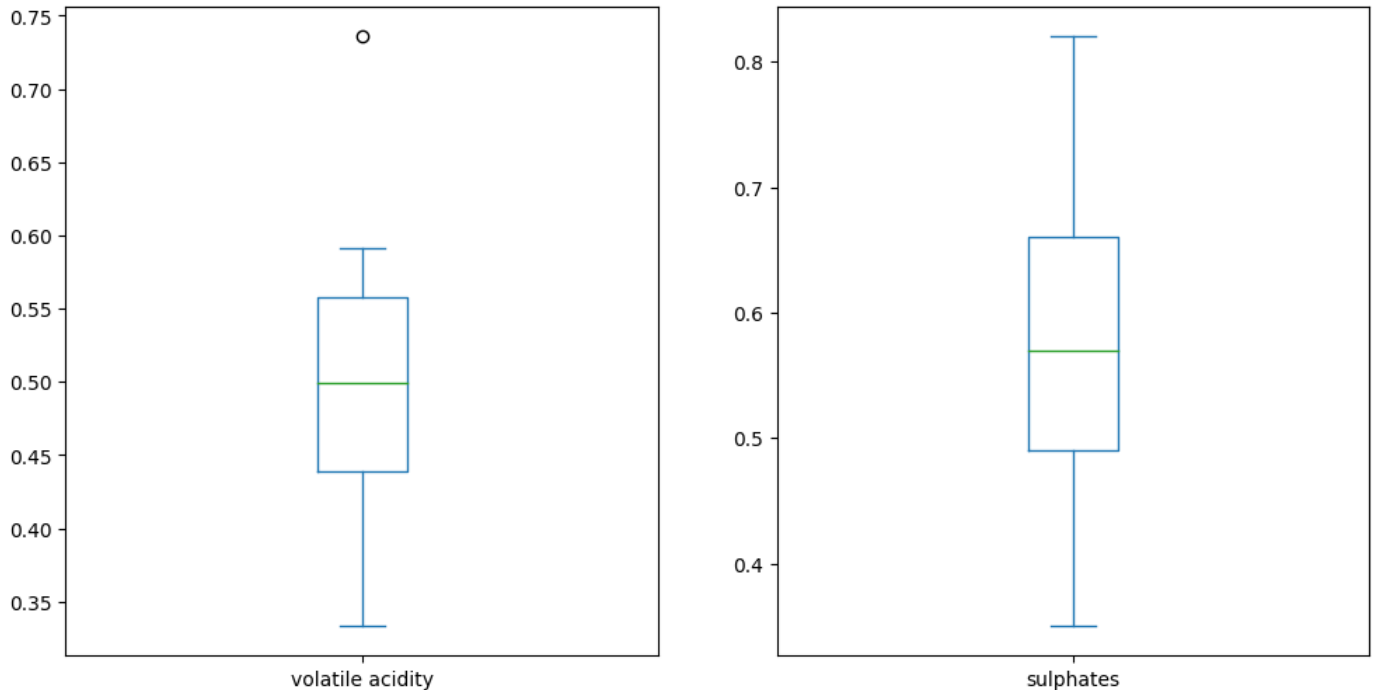
```

```
# Print results
print('Nilai z          : ' + str(z))
print('Nilai z(alpha/2) : ' + str(z_alpha_over_2))
print('Nilai p          : ' + str(p))

if ((z > z_alpha_over_2 or z < -z_alpha_over_2) and p < alpha):
    print('Hipotesis H0 ditolak, artinya rata-rata sampel 25 baris pertama kolom Volatile Acidity
          tidak sama dengan rata-rata 25 baris pertama kolom Sulphates')
else :
    print('Hipotesis H0 diterima, artinya rata-rata sampel 25 baris pertama kolom Volatile Acidity
          sama dengan rata-rata 25 baris pertama kolom Sulphates')

# Plot each side
plt.subplot(1, 2, 1)
df_first_25_volatile_acidity.plot(kind="box", figsize=(12, 6))
plt.subplot(1, 2, 2)
df_first_25_sulphates.plot(kind="box", figsize=(12, 6))
plt.show()
```

Nilai z : -2.6374821676748703
 Nilai z(alpha/2) : 1.959963984540054
 Nilai p : 0.008352401685453743
 Hipotesis H0 ditolak, artinya rata-rata sampel 25 baris pertama kolom Volatile Acidity
 tidak sama dengan rata-rata 25 baris pertama kolom Sulphates



d. Bagian awal kolom residual sugar memiliki variansi yang sama dengan bagian akhirnya?

1. Tentukan nilai hipotesis nol (H_0)

H_0 : Bagian awal kolom residual sugar memiliki variansi yang sama dengan bagian akhirnya ($\sigma_1^2 - \sigma_2^2 = 0$)

2. Tentukan hipotesis alternatif (H_1)

H_1 : Bagian awal kolom residual sugar tidak memiliki variansi yang sama dengan bagian akhirnya ($\sigma_1^2 - \sigma_2^2 \neq 0$)

3. Tentukan tingkat signifikan

Tingkat Signifikan $\alpha = 0.05$

4. Tentukan uji statistik yang sesuai dan tentukan daerah kritis.

Uji statistik: **Two Tailed Variance Test**

Critical section: $f < f_{1-\alpha/2}(v_1, v_2)$ atau $f > f_{\alpha/2}(v_1, v_2)$

5. Hitung nilai uji statistik

Hitung nilai z dengan rumus

$$f = \frac{s_1^2}{s_2^2}$$

dengan

$$v_1 = n_1 - 1, v_2 = n_2 - 2$$

Hitung nilai P

6. KEPUTUSAN : Tolak H_0 jika ($f < f_{1-\alpha/2}(v_1, v_2)$ atau $f > f_{\alpha/2}(v_1, v_2)$) dan $p < \alpha$

```
In [ ]: # Initialize constant variables
df_residual_sugar = df["residual sugar"]
alpha = 0.05
```

```
# Divide residual sugar column into two
df_residual_sugar_head = df_residual_sugar[:len(df)//2]
df_residual_sugar_tail = df_residual_sugar[len(df)//2:]

# Find degrees of freedom for head and tail
v_head = len(df_residual_sugar_head) - 1
v_tail = len(df_residual_sugar_tail) - 1

# Calculate f value
f = df_residual_sugar_head.var() / df_residual_sugar_tail.var()
f_upper = scp.f.ppf(1 - alpha/2, v_head, v_tail)
f_lower = scp.f.ppf(alpha/2, v_head, v_tail)

# Find z and p value
p = scp.f.cdf(f, v_head, v_tail)

# Print results
print('Nilai f          : ' + str(f))
print('Nilai f upper    : ' + str(f_upper))
print('Nilai f lower     : ' + str(f_lower))
print('Nilai p           : ' + str(p))

if ((f > f_lower or f < f_upper) and p < alpha):
    print('Hipotesis H0 ditolak, artinya bagian awal kolom residual sugar tidak memiliki variansi yang sama dengan bagian akhirnya')
else :
    print('Hipotesis H0 diterima, artinya bagian awal kolom residual sugar memiliki variansi yang sama dengan bagian akhirnya')

# Plot each side
plt.subplot(1, 2, 1)
df_residual_sugar_head.plot(kind="box", figsize=(12, 6))
plt.subplot(1, 2, 2)
df_residual_sugar_tail.plot(kind="box", figsize=(12, 6))
plt.show()
```

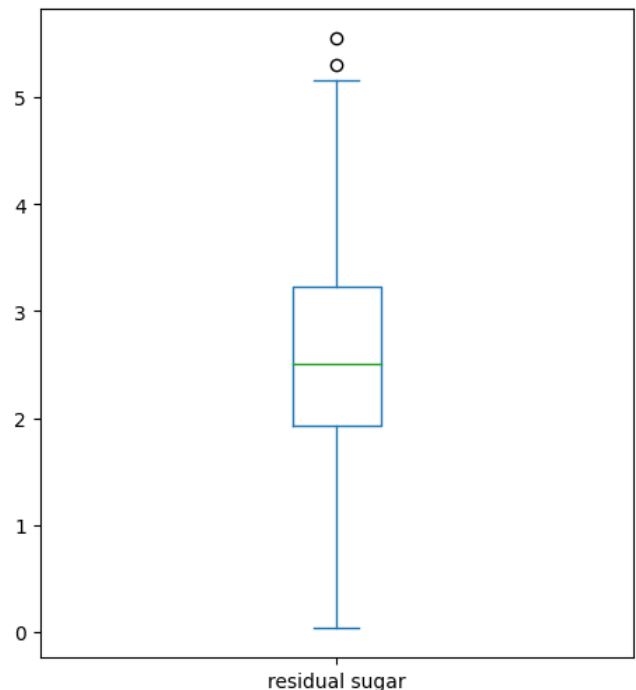
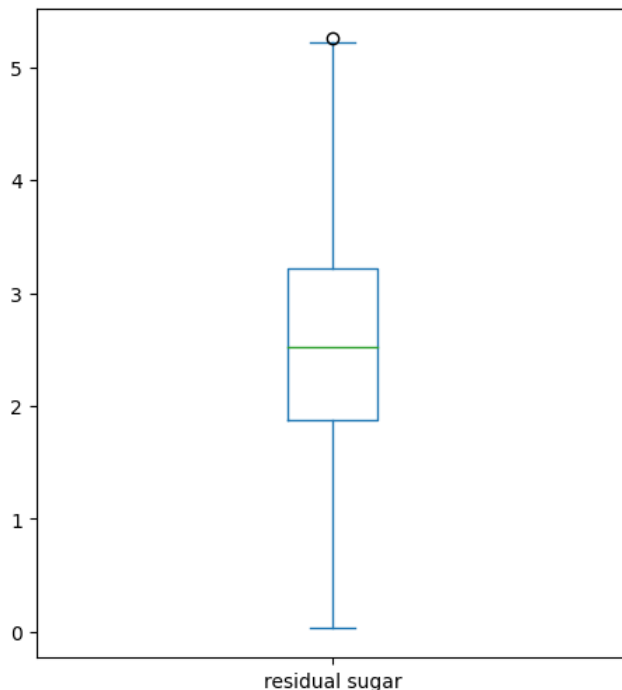
Nilai f : 0.9420041066941615

Nilai f upper : 1.1920574017201653

Nilai f lower : 0.8388857772763105

Nilai p : 0.2524101797623089

Hipotesis H0 diterima, artinya bagian awal kolom residual sugar memiliki variansi yang sama dengan bagian akhirnya



e. Proporsi nilai setengah bagian awal alcohol yang lebih dari 7, adalah lebih besar daripada, proporsi nilai yang sama di setengah bagian akhir alcohol?

1. Tentukan nilai hipotesis nol (H_0)

H_0 : Proporsi nilai setengah bagian awal alcohol yang lebih dari 7 sama dengan proporsi nilai yang sama di setengah bagian akhir alcohol

$(P(X_{head} > 7) = P(X_{tail} > 7))$

2. Tentukan hipotesis alternatif (H_1)

H_1 : Proporsi nilai setengah bagian awal alcohol yang lebih dari 7, adalah lebih besar daripada, proporsi nilai yang sama di setengah bagian akhir alcohol

$(P(X_{head} > 7) > P(X_{tail} > 7))$

3. Tentukan tingkat signifikan

Tingkat Signifikan $\alpha = 0.05$

4. Tentukan uji statistik yang sesuai dan tentukan daerah kritis.

Uji statistik: **One Tailed Proportion Test**

Critical section: $z > z_{\alpha}$

5. Hitung nilai uji statistik

Hitung nilai z dengan rumus

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}(1/n_1 + 1/n_2)}}$$

dengan

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}, \hat{q} = 1 - \hat{p}$$

Hitung nilai P

6. KEPUTUSAN : Tolak H_0 jika $z > z_{\alpha}$ dan $p < \alpha$

```
In [ ]: # Initialize constant variables
df_alcohol = df["alcohol"]
alpha = 0.05

# Divide alcohol column into two
df_alcohol_head = df_alcohol[:len(df)//2]
df_alcohol_tail = df_alcohol[len(df)//2:]

# Calculate the proportion of alcohol values greater than 7 in each half
sum_alcohol_greater_than_7 = [sum(df_alcohol_head > 7), sum(df_alcohol_tail > 7)]
sum_alcohol = [len(df_alcohol_head), len(df_alcohol_tail)]
z, p = proportions_ztest(sum_alcohol_greater_than_7, sum_alcohol, alternative='larger')

# Find z alpha value
z_alpha = scp.norm.ppf(1-alpha)

# Print results
print('Nilai z          : ' + str(z))
print('Nilai p          : ' + str(p))

if (z > z_alpha and p < alpha):
    print('Hipotesis H0 ditolak, artinya proporsi nilai setengah bagian awal alcohol yang lebih dari 7,
          adalah lebih besar daripada, proporsi nilai yang sama di setengah bagian akhir alcohol')
else :
    print('Hipotesis H0 diterima, artinya proporsi nilai setengah bagian awal alcohol yang lebih dari 7
          sama dengan proporsi nilai yang sama di setengah bagian akhir alcohol')

# Plot each side
plt.subplot(1, 2, 1)
df_alcohol_head.plot(kind="box", figsize=(12, 6))
plt.subplot(1, 2, 2)
df_alcohol_tail.plot(kind="box", figsize=(12, 6))
plt.show()
```

Nilai z : 0.0

Nilai p : 0.5

Hipotesis H0 diterima, artinya proporsi nilai setengah bagian awal alcohol yang lebih dari 7
sama dengan proporsi nilai yang sama di setengah bagian akhir alcohol

