

Tugas Besar IF2220 Probabilitas dan Statistika

Penarikan Kesimpulan dan Pengujian Hipotesis

Kelompok 30 K2

- Yanuar Sano Nur Rasyid (13521110)
- Febryan Arota Hia (13521120)

```
In [ ]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from IPython.display import display, Markdown
```

1. Menuliskan 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 (interquartile range), skewness, dan kurtosis. Boleh juga ditambahkan deskripsi lain.

```
In [ ]: df = pd.read_csv('dataset/anggur.csv')
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
---  -
0   fixed acidity          1000 non-null   float64
1   volatile acidity       1000 non-null   float64
2   citric acid            1000 non-null   float64
3   residual sugar         1000 non-null   float64
4   chlorides              1000 non-null   float64
5   free sulfur dioxide    1000 non-null   float64
6   total sulfur dioxide   1000 non-null   float64
7   density                1000 non-null   float64
8   pH                    1000 non-null   float64
9   sulphates              1000 non-null   float64
10  alcohol                1000 non-null   float64
11  quality                1000 non-null   int64
dtypes: float64(11), int64(1)
memory usage: 93.9 KB
```

```
In [ ]: desc = pd.DataFrame()
desc["Mean"] = df.mean()
desc["Median"] = df.median()
desc["Modus"] = df.mode().iloc[0]
desc["Std"] = df.std()
desc["Min"] = df.min()
desc["Max"] = df.max()
desc["Range"] = df.max() - df.min()
desc["Q1"] = df.quantile(0.25)
desc["Q2"] = df.quantile(0.5)
desc["Q3"] = df.quantile(0.75)
desc["IQR"] = desc["Q3"] - desc["Q1"]
desc["Skewness"] = df.skew()
desc["Kurtosis"] = df.kurtosis()
desc["Missing"] = df.isnull().sum()
desc["Unique"] = df.nunique()
desc
```

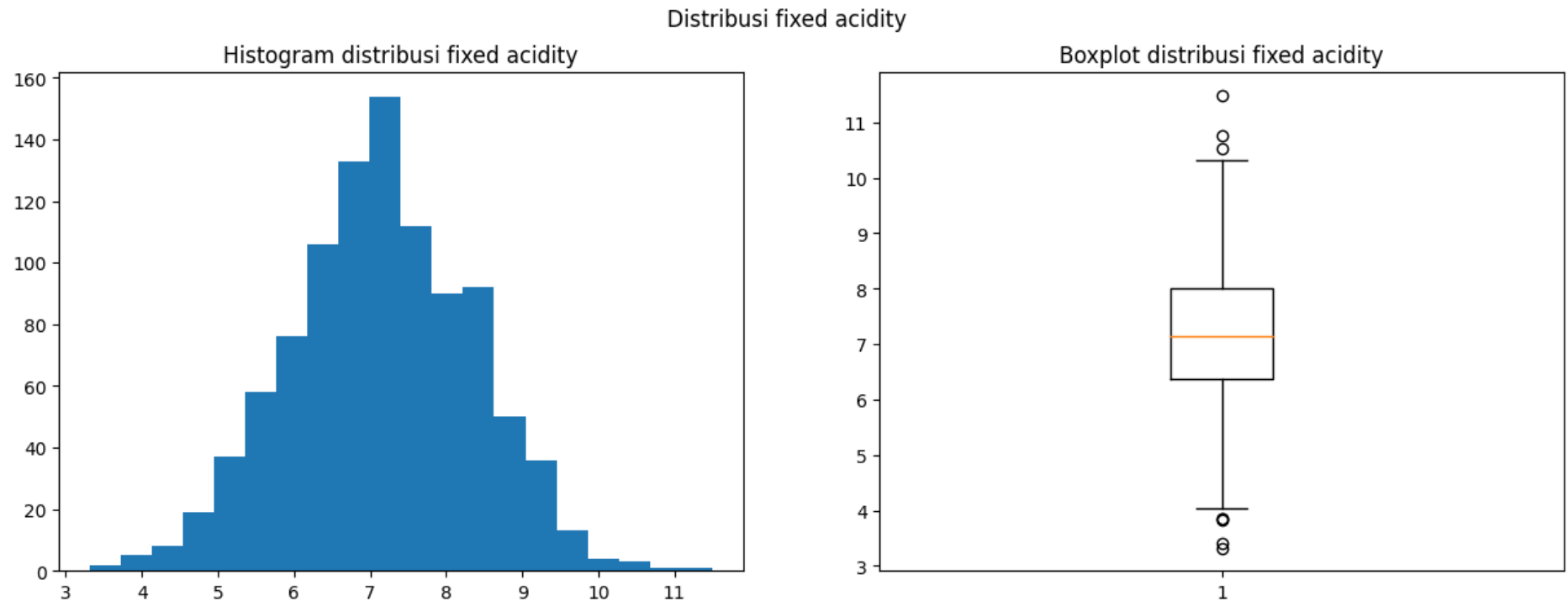
	Mean	Median	Modus	Std	Min	Max	Range	Q1	Q2	Q3	IQR	Skewness	Kurtosis	Missing	Unique
fixed acidity	7.152530	7.150000	6.540000	1.201598	3.320000	11.490000	8.170000	6.377500	7.150000	8.000000	1.622500	-0.028879	-0.019292	0	417
volatile acidity	0.520839	0.524850	0.554600	0.095848	0.139900	0.805100	0.665200	0.456100	0.524850	0.585375	0.129275	-0.197699	0.161853	0	879
citric acid	0.270517	0.272200	0.301900	0.049098	0.116700	0.409600	0.292900	0.237800	0.272200	0.302325	0.064525	-0.045576	-0.104679	0	769
residual sugar	2.567104	2.519430	0.032555	0.987915	0.032555	5.550755	5.518200	1.896330	2.519430	3.220873	1.324544	0.132638	-0.042980	0	1000
chlorides	0.081195	0.082167	0.015122	0.020111	0.015122	0.140758	0.125635	0.066574	0.082167	0.095312	0.028738	-0.051319	-0.246508	0	1000
free sulfur dioxide	14.907679	14.860346	0.194679	4.888100	0.194679	27.462525	27.267847	11.426717	14.860346	18.313098	6.886381	0.007130	-0.364964	0	1000
total sulfur dioxide	40.290150	40.190000	35.200000	9.965767	3.150000	69.960000	66.810000	33.785000	40.190000	47.022500	13.237500	-0.024060	0.063950	0	881
density	0.995925	0.996000	0.995900	0.002020	0.988800	1.002600	0.013800	0.994600	0.996000	0.997200	0.002600	-0.076883	0.016366	0	107
pH	3.303610	3.300000	3.340000	0.104875	2.970000	3.710000	0.740000	3.230000	3.300000	3.370000	0.140000	0.147673	0.080910	0	61
sulphates	0.598390	0.595000	0.590000	0.100819	0.290000	0.960000	0.670000	0.530000	0.595000	0.670000	0.140000	0.149199	0.064819	0	60
alcohol	10.592280	10.610000	9.860000	1.510706	6.030000	15.020000	8.990000	9.560000	10.610000	11.622500	2.062500	-0.018991	-0.131732	0	490
quality	7.958000	8.000000	8.000000	0.902802	5.000000	10.000000	5.000000	7.000000	8.000000	9.000000	2.000000	-0.089054	0.108291	0	6

2. Membuat Visualisasi plot distribusi, dalam bentuk histogram dan boxplot untuk setiap

kolom numerik. Berikan uraian penjelasan kondisi setiap kolom berdasarkan kedua plot tersebut.

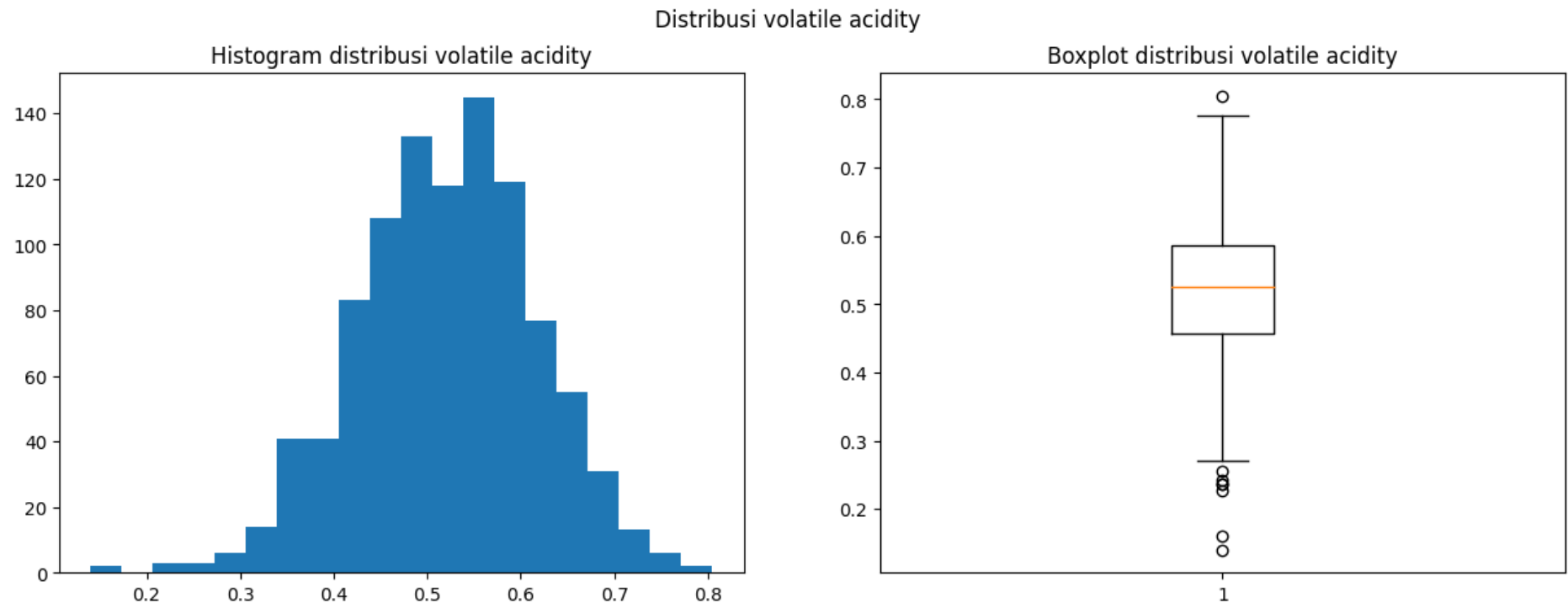
```
In [ ]: # code no 2
def plot_distribusi(data, title):
    fig, ax = plt.subplots(1, 2, figsize=(15, 5))
    fig.suptitle(title)
    ax[0].hist(data, bins=20)
    ax[1].boxplot(data)
    ax[0].set_title("Histogram " + title.lower())
    ax[1].set_title("Boxplot " + title.lower())
    plt.show()
```

```
In [ ]: plot_distribusi(df["fixed acidity"], "Distribusi fixed acidity")
```



Visualisasi data pada kolom fixed acidity dengan histogram dan boxplot menunjukkan bahwa data berbentuk simetris dan terdapat beberapa outlier yang nilainya melebihi maksimum dan kurang dari minimum

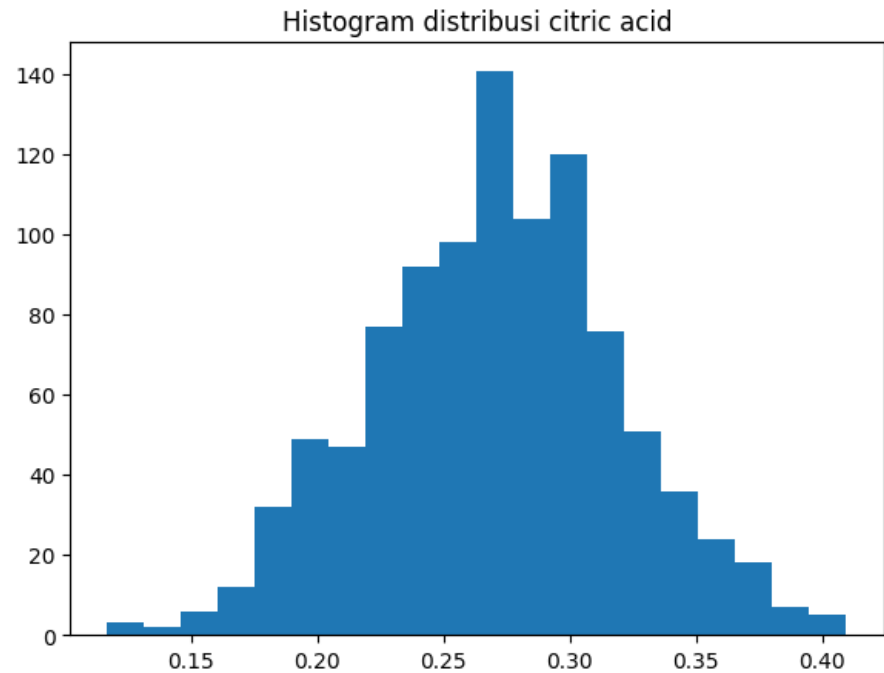
```
In [ ]: plot_distribusi(df["volatile acidity"], "Distribusi volatile acidity")
```



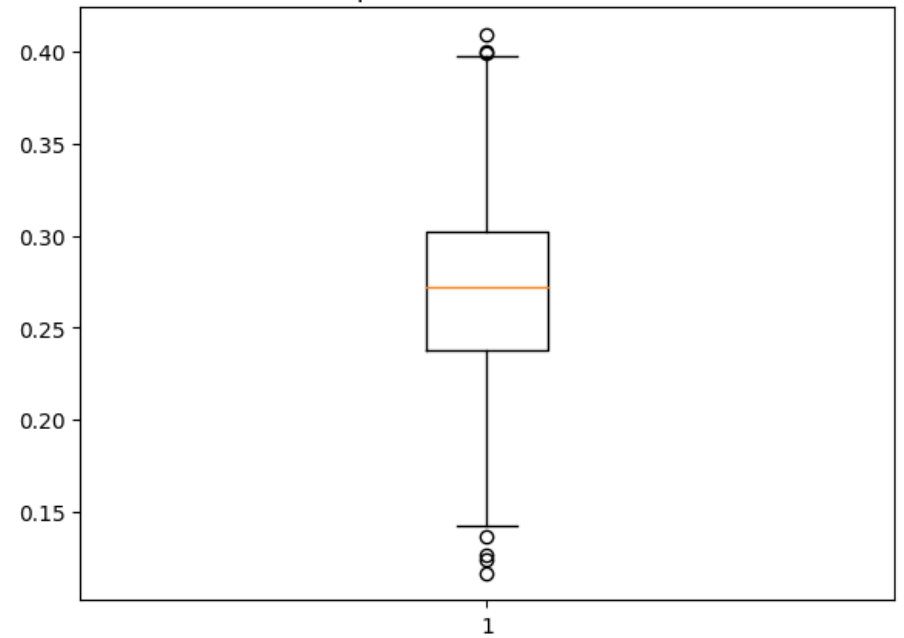
Visualisasi data pada kolom volatile acidity dengan histogram dan boxplot menunjukkan bahwa data sedikit berbentuk left-skewed (negative skew) dan terdapat beberapa outlier pada maksimum dan minimum dengan jumlah yang lebih banyak kurang dari minimum.

```
In [ ]: plot_distribusi(df["citric acid"], "Distribusi citric acid")
```

Distribusi citric acid



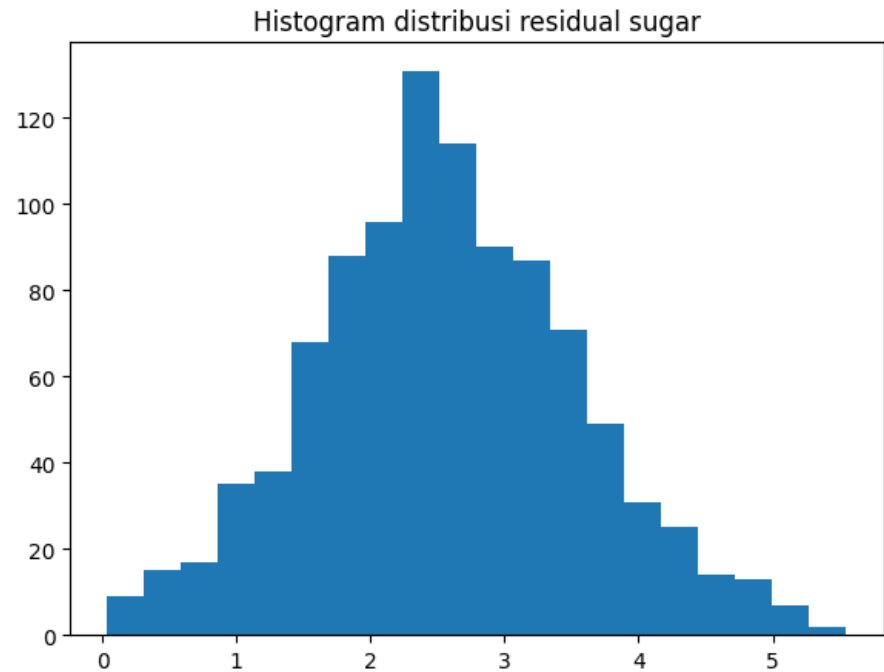
Boxplot distribusi citric acid



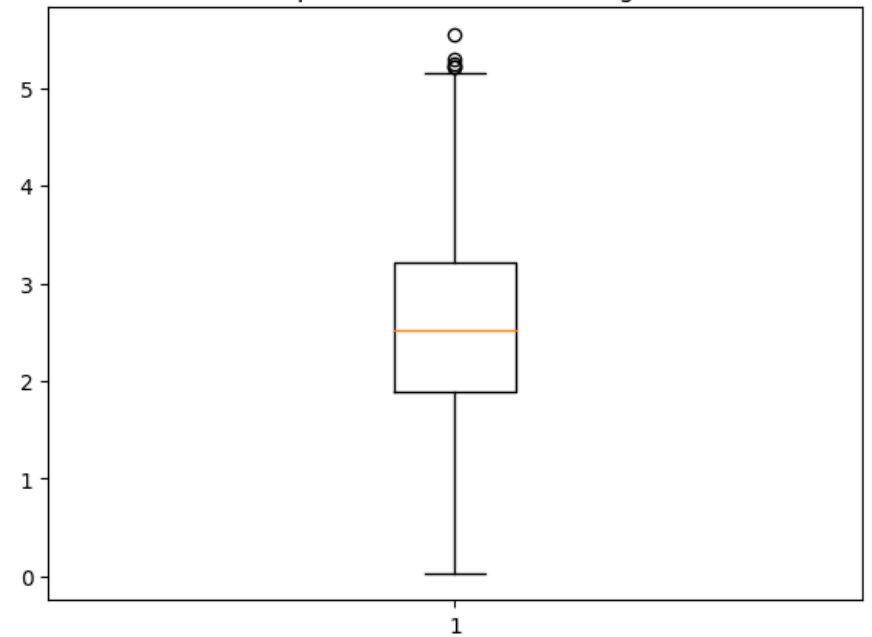
Visualisasi data pada kolom citric acid dengan histogram dan boxplot menunjukkan bahwa data berbentuk simetris dan terdapat beberapa outlier pada maksimum dan minimum dengan jumlah yang lebih banyak kurang dari minimum.

```
In [ ]: plot_distribusi(df["residual sugar"], "Distribusi residual sugar")
```

Distribusi residual sugar



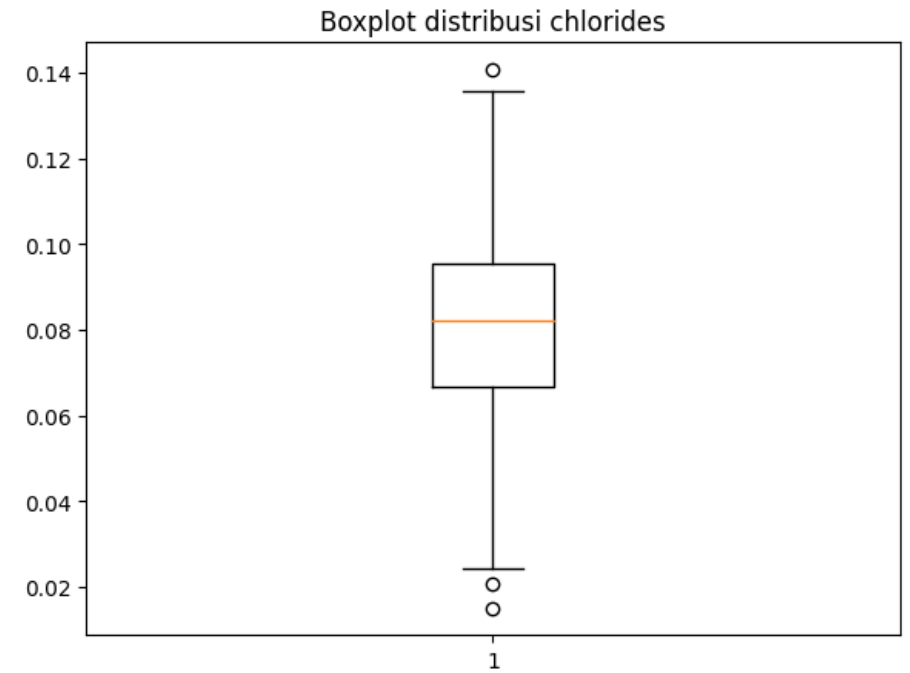
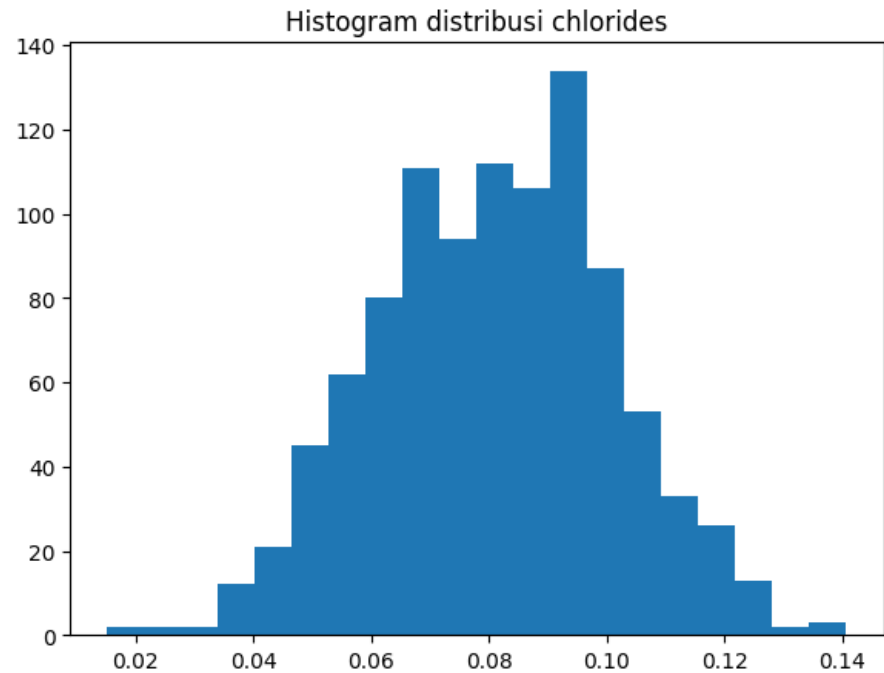
Boxplot distribusi residual sugar



Visualisasi data pada kolom residual sugar dengan histogram dan boxplot menunjukkan bahwa data sedikit berbentuk right-skewed (positive skew) dan terdapat beberapa outlier yang melebihi nilai maksimum.

```
In [ ]: plot_distribusi(df["chlorides"], "Distribusi chlorides")
```

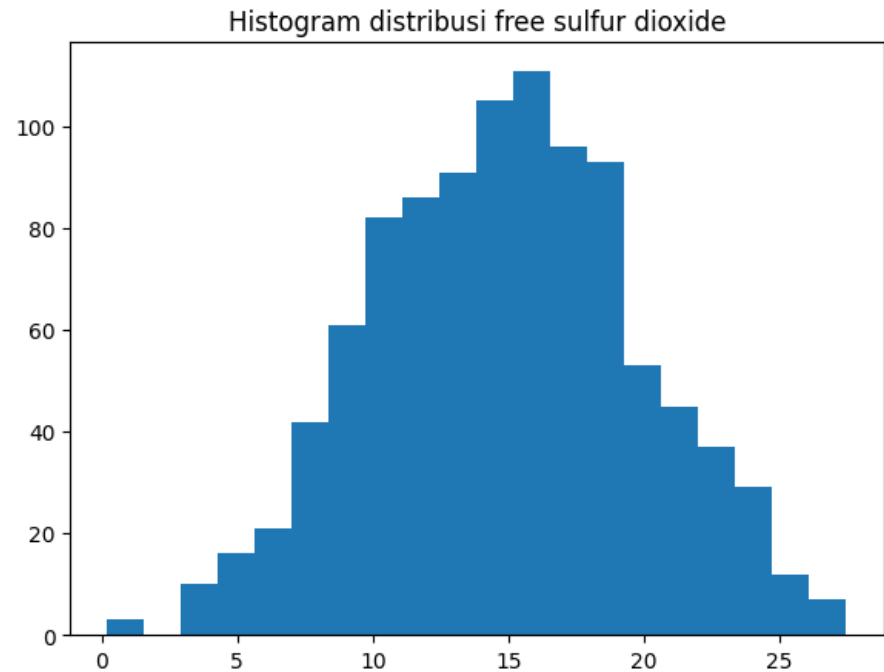
Distribusi chlorides



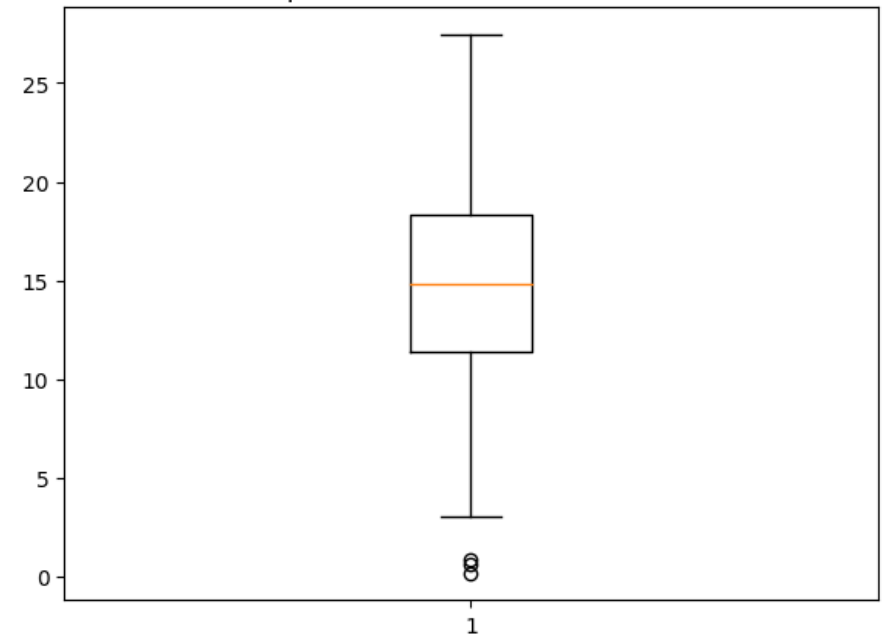
Visualisasi data pada kolom chlorides dengan histogram dan boxplot menunjukkan bahwa data sedikit berbentuk left-skewed (negative skew) dan terdapat beberapa outlier pada maksimum dan minimum dengan jumlah yang lebih banyak kurang dari minimum.

```
In [ ]: plot_distribusi(df["free sulfur dioxide"], "Distribusi free sulfur dioxide")
```

Distribusi free sulfur dioxide



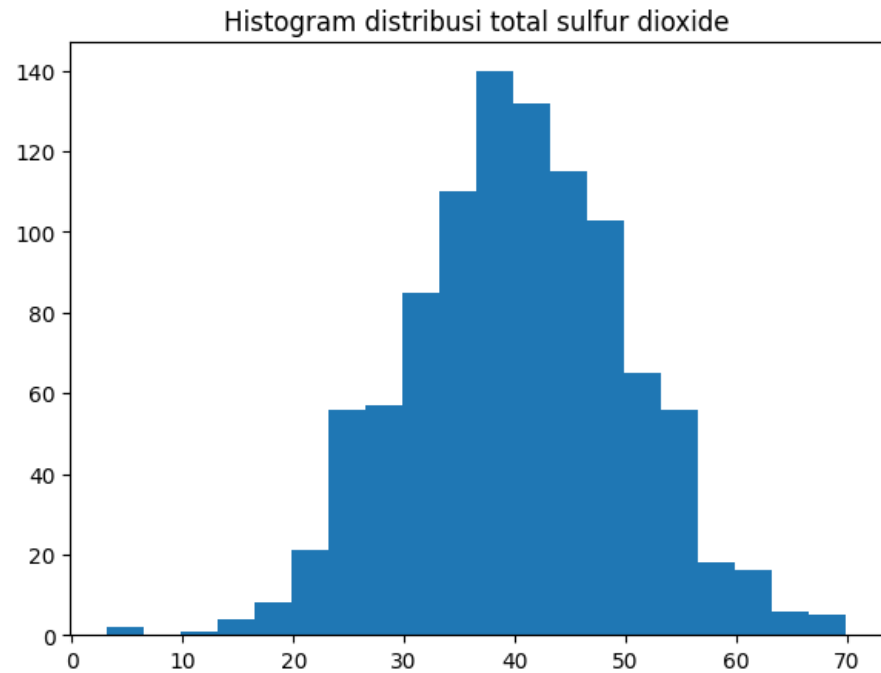
Boxplot distribusi free sulfur dioxide



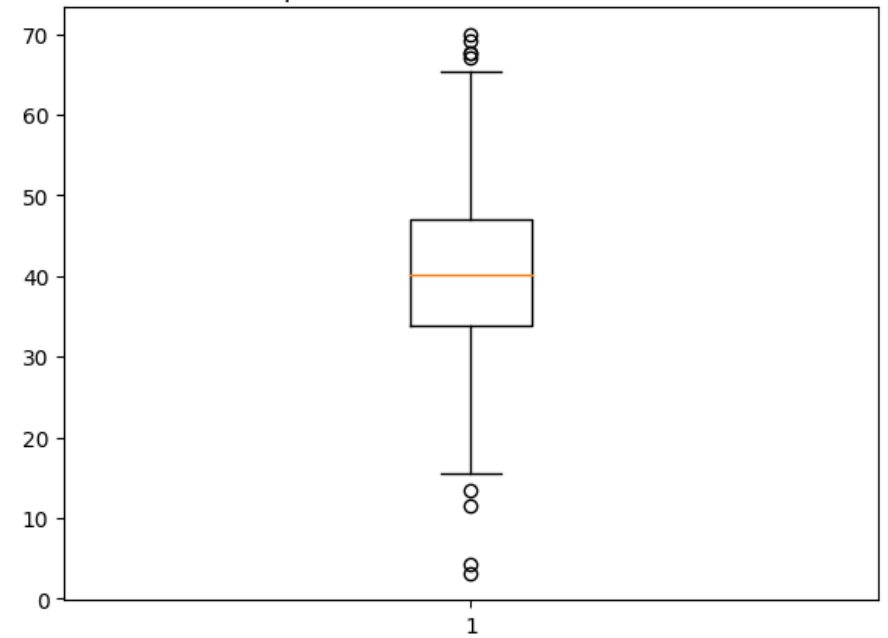
Visualisasi data pada kolom free sulfur dioxide dengan histogram dan boxplot menunjukkan bahwa data sedikit berbentuk left-skewed (negative skew) dan terdapat beberapa outlier yang nilainya melebihi maksimum dan kurang dari minimum.

```
In [ ]: plot_distribusi(df["total sulfur dioxide"], "Distribusi total sulfur dioxide")
```


Distribusi total sulfur dioxide



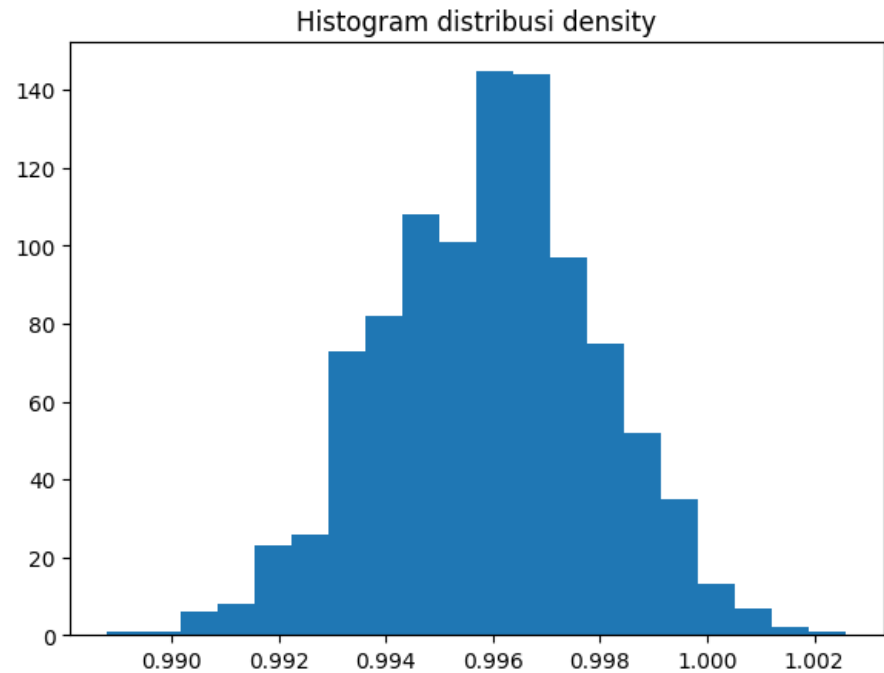
Boxplot distribusi total sulfur dioxide



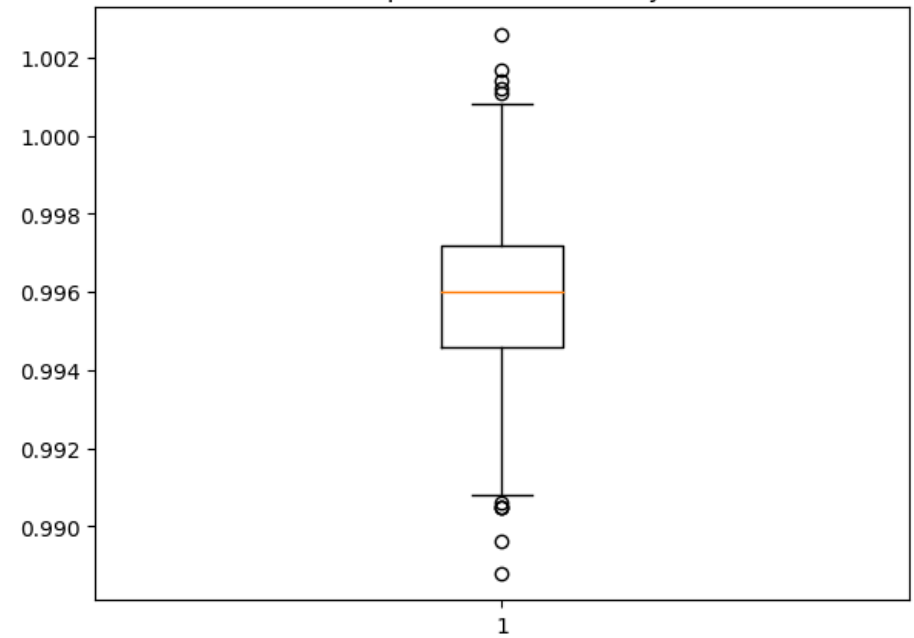
Visualisasi data pada kolom total sulfur dioxide dengan histogram dan boxplot menunjukkan bahwa data berbentuk simetris dan terdapat beberapa outlier pada maksimum dan minimum dengan jumlah yang lebih banyak lebih besar dari maksimum.

```
In [ ]: plot_distribusi(df["density"], "Distribusi density")
```

Distribusi density



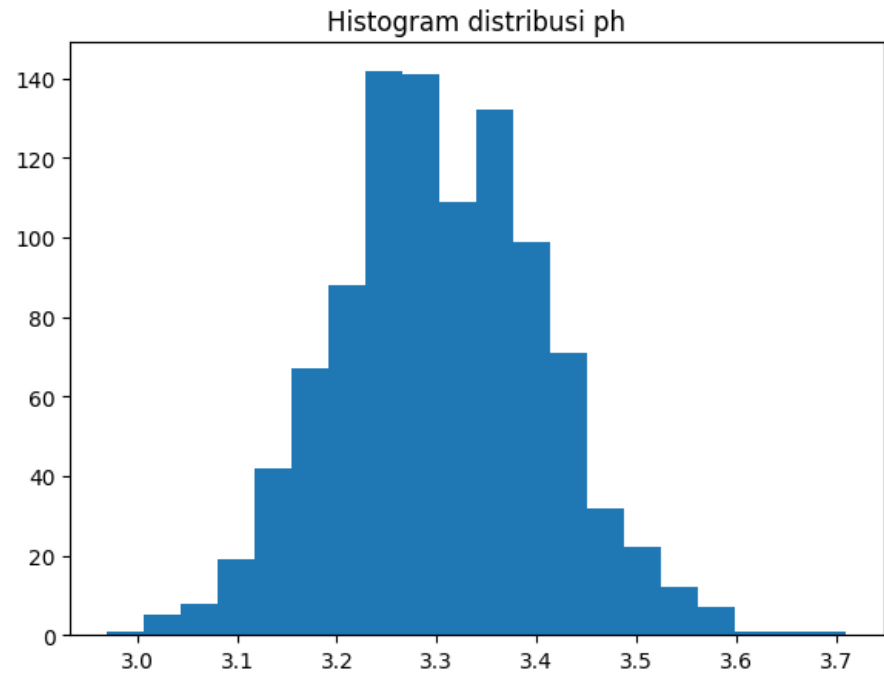
Boxplot distribusi density



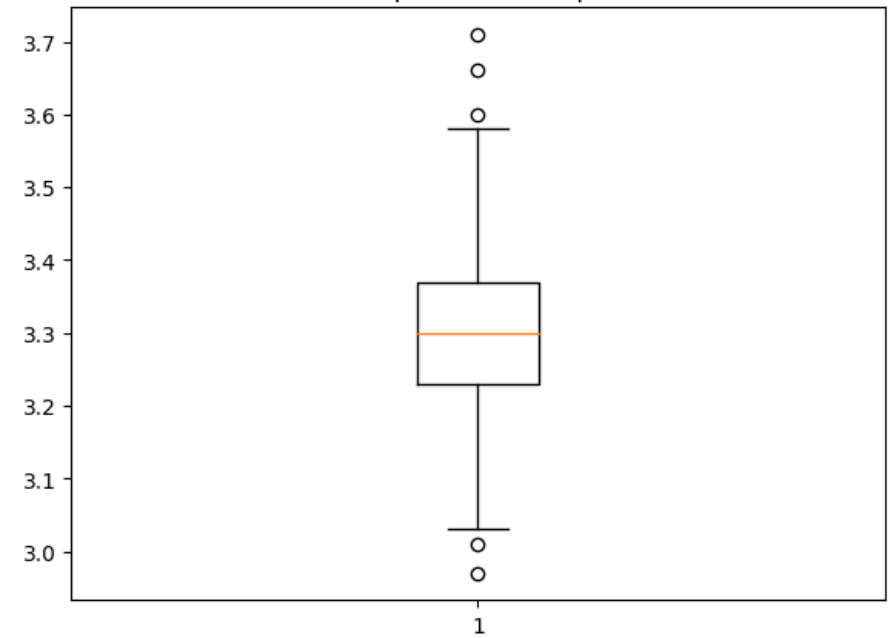
Visualisasi data pada kolom density dengan histogram dan boxplot menunjukkan bahwa data berbentuk simetris dan terdapat beberapa outlier yang melebihi maksimum dan kurang dari minimum dengan outlier maksimum lebih banyak.

```
In [ ]: plot_distribusi(df["pH"], "Distribusi pH")
```

Distribusi pH



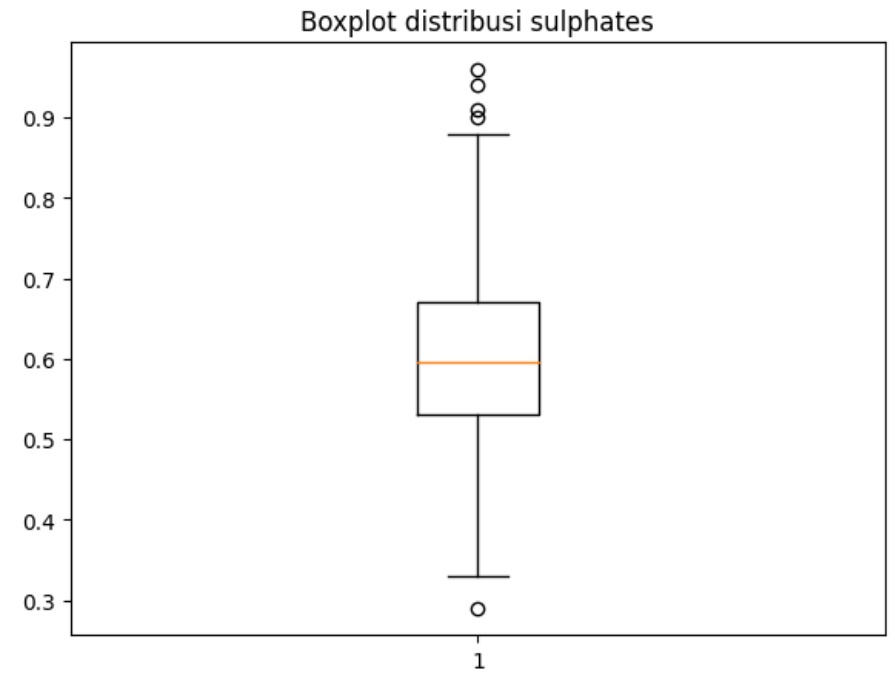
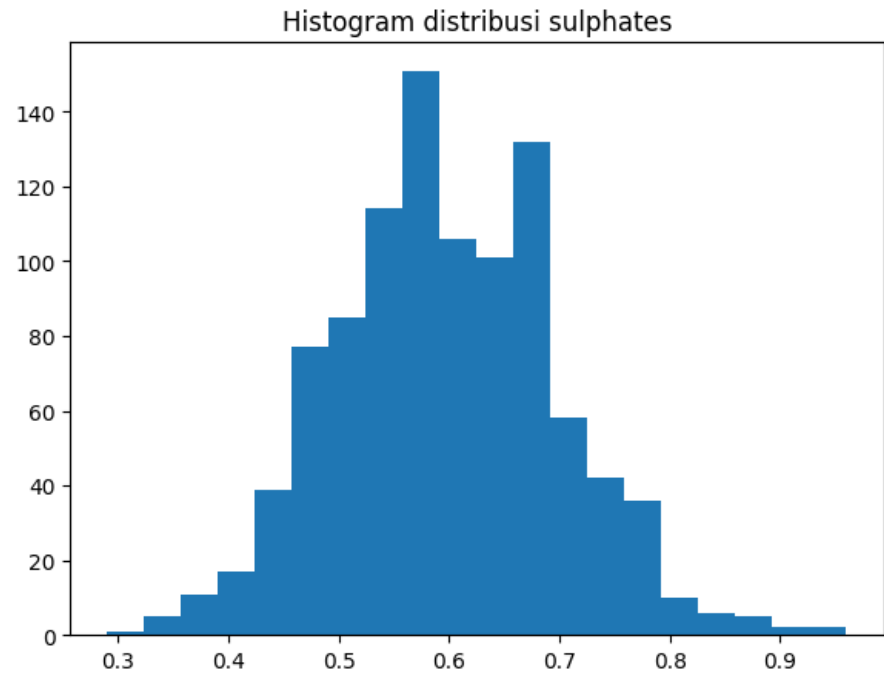
Boxplot distribusi ph



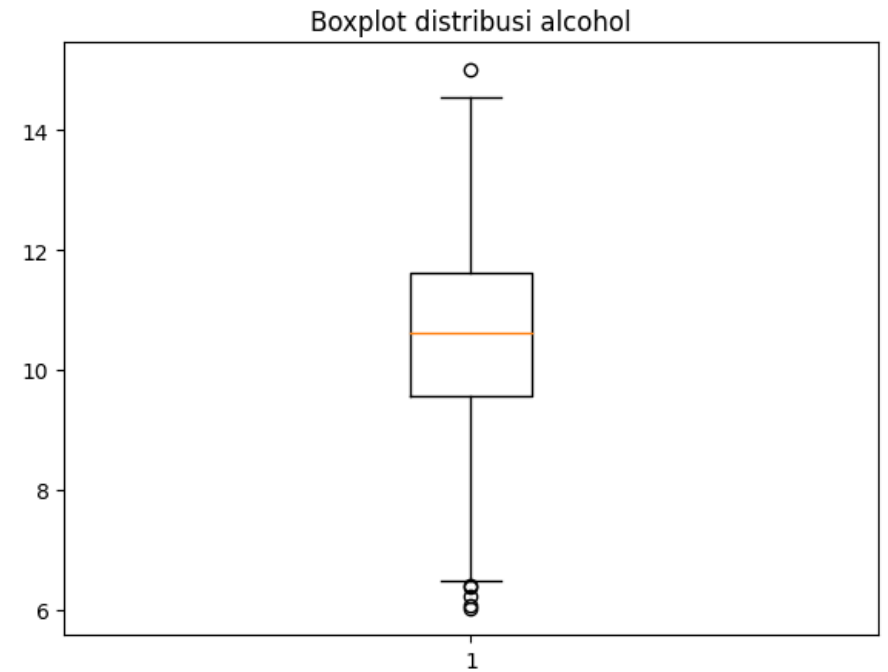
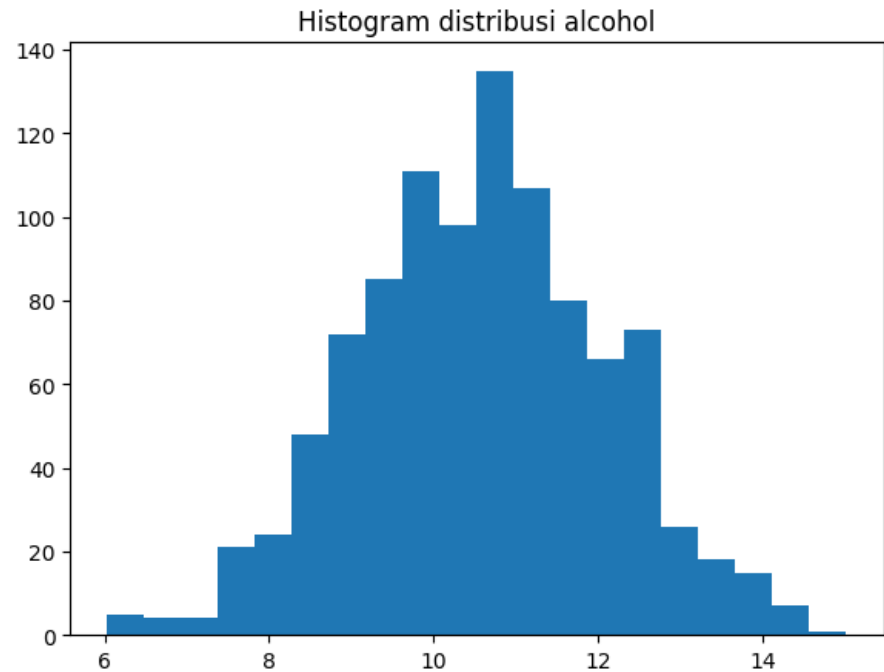
Visualisasi data pada kolom pH dengan histogram dan boxplot menunjukkan bahwa data sedikit berbentuk right-skewed (positive skew) dan terdapat beberapa outlier pada maksimum dan minimum dengan jumlah yang lebih banyak lebih besar dari maksimum.

```
In [ ]: plot_distribusi(df["sulphates"], "Distribusi sulphates")
```

Distribusi sulphates



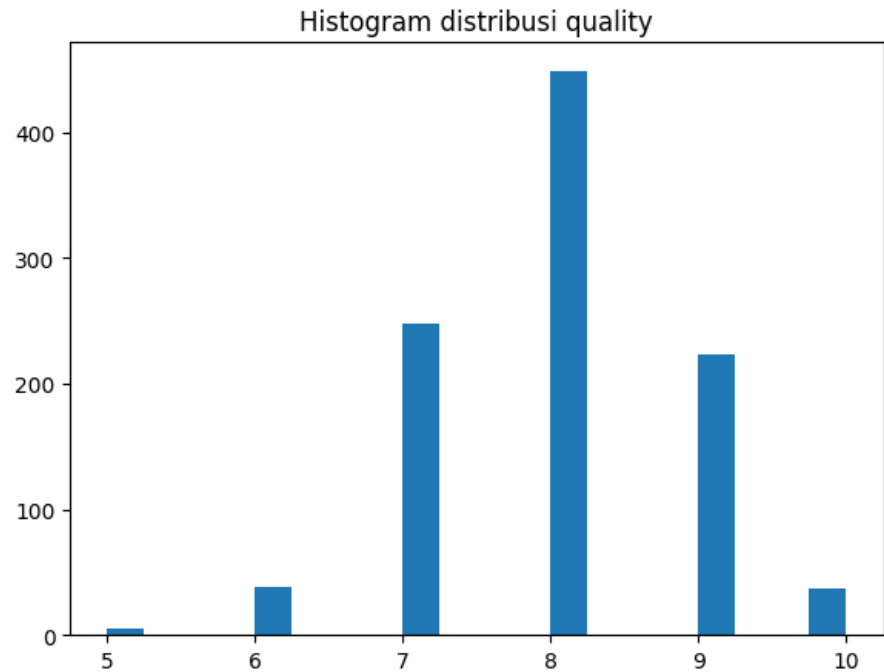
Distribusi alcohol



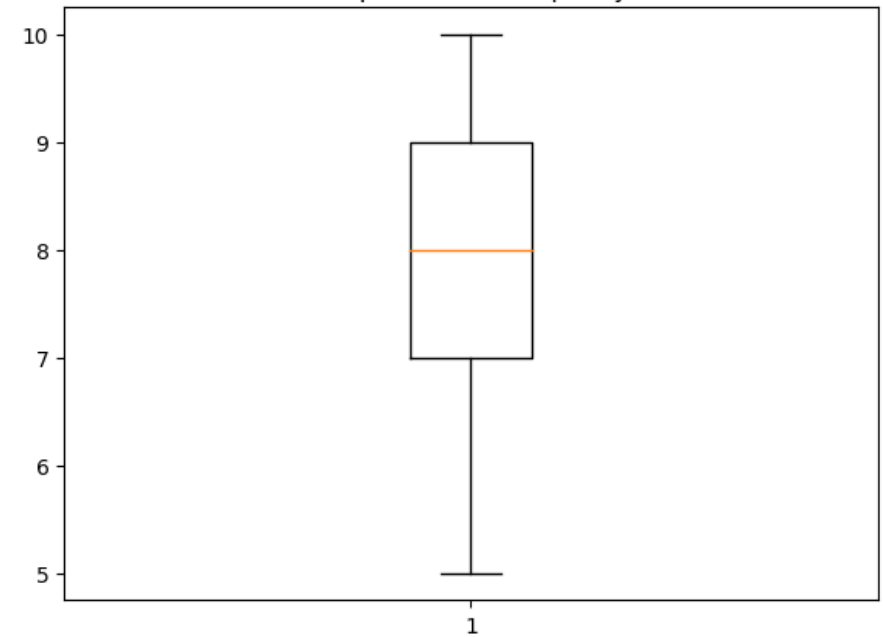
Visualisasi data pada kolom alcohol dengan histogram dan boxplot menunjukkan bahwa data berbentuk simetris dan terdapat beberapa outlier pada maksimum dan minimum dengan jumlah yang lebih banyak kurang dari minimum.

```
In [ ]: plot_distribusi(df["quality"], "Distribusi quality")
```

Distribusi quality



Boxplot distribusi quality



Visualisasi data pada kolom quality dengan histogram dan boxplot menunjukkan bahwa data berbentuk simetris dan tidak terdapat outlier yang ditunjukkan oleh boxplot.

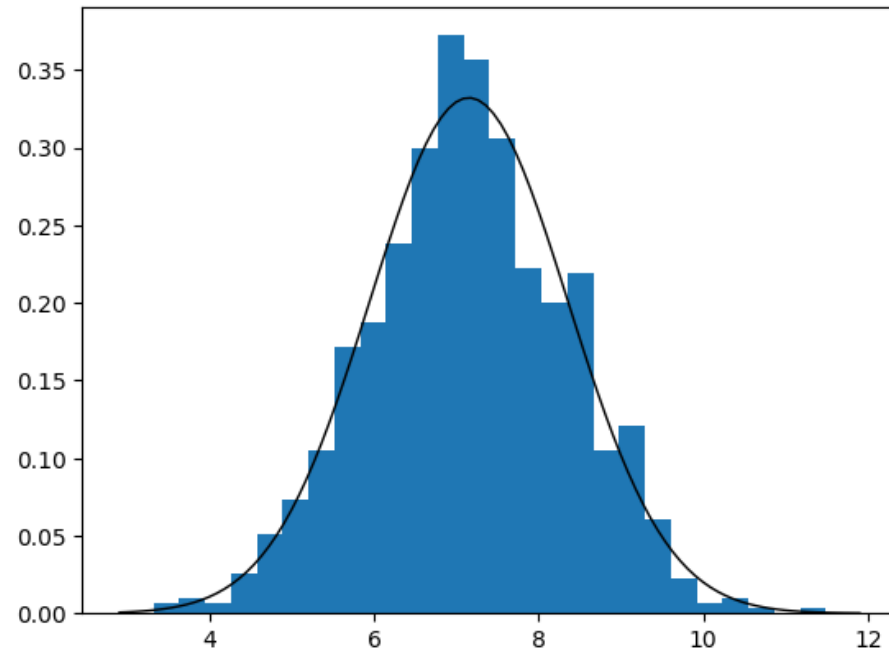
3. Menentukan setiap kolom numerik berdistribusi normal atau tidak. Gunakan normality test yang dikaitkan dengan histogram plot.

```
In [ ]: import scipy.stats as st
```

```
In [ ]: def normalityTest(data):
    k, p = st.normaltest(data)
    alpha = 0.05
    if p < alpha:
        print(f"p = {round(p,4)}, sehingga data tidak terdistribusi dengan normal")
    else:
        print(f"p = {round(p,4)}, sehingga data terdistribusi dengan normal")
    mean, std = st.norm.fit(data)
    plt.hist(data, bins='auto', density=True)
    xmin, xmax = plt.xlim()
    x = np.linspace(xmin, xmax, 100)
    y = st.norm.pdf(x, mean, std)
    plt.plot(x, y, 'k', lw=1)
```

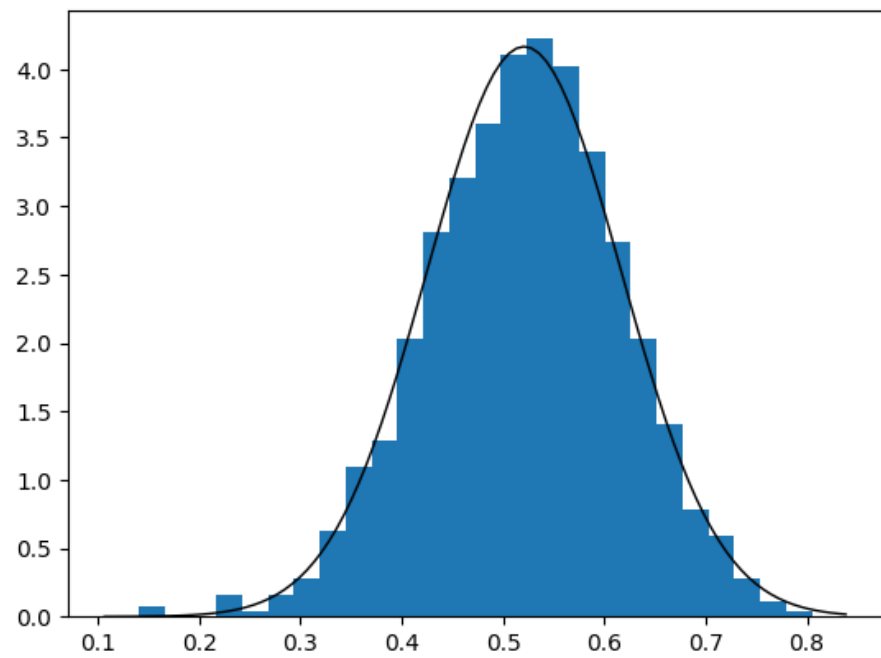
```
In [ ]: normalityTest(df["fixed acidity"])
```

p = 0.9309, sehingga data terdistribusi dengan normal



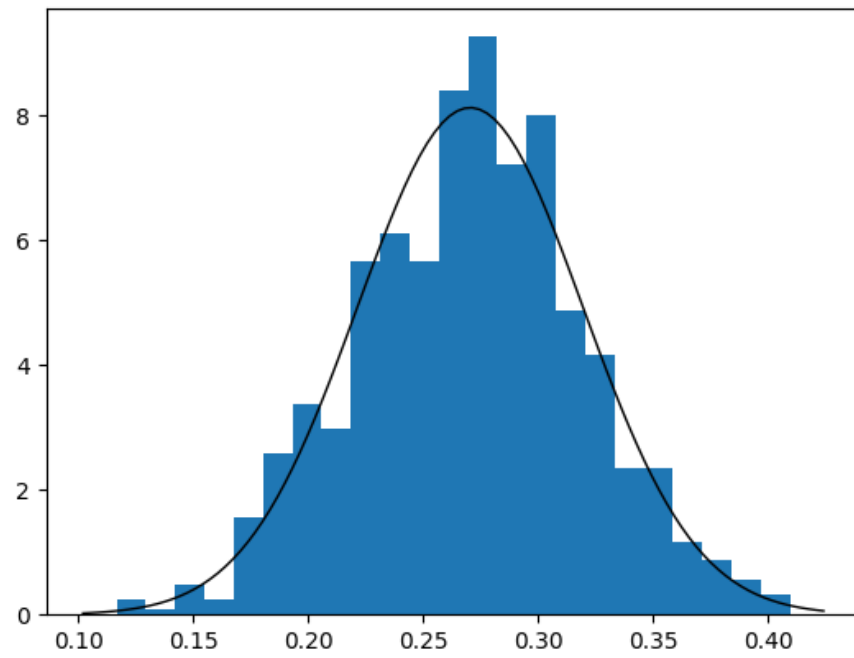
```
In [ ]: normalityTest(df["volatile acidity"])
```

p = 0.0226, sehingga data tidak terdistribusi dengan normal



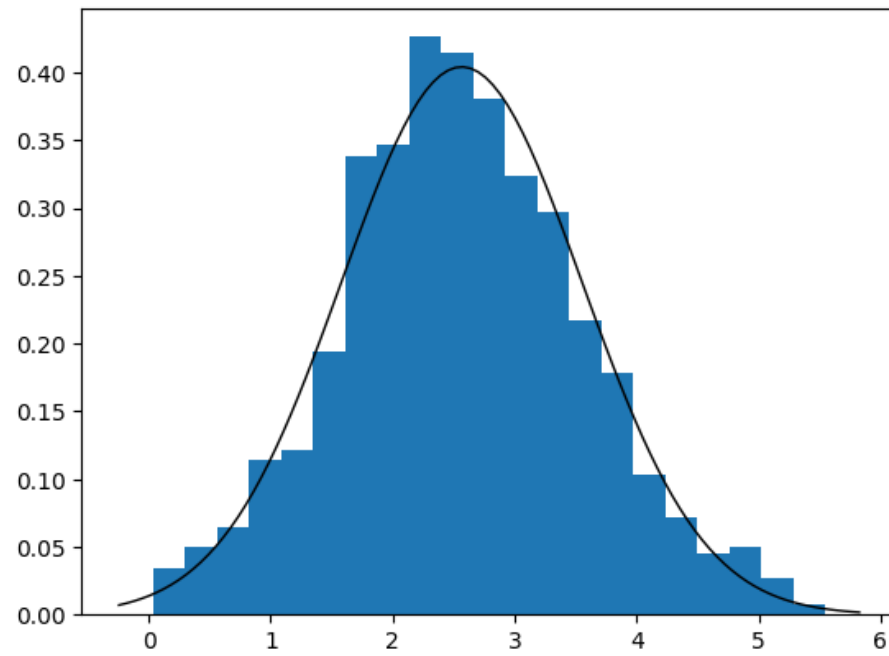
```
In [ ]: normalityTest(df["citric acid"])
```

p = 0.6817, sehingga data terdistribusi dengan normal



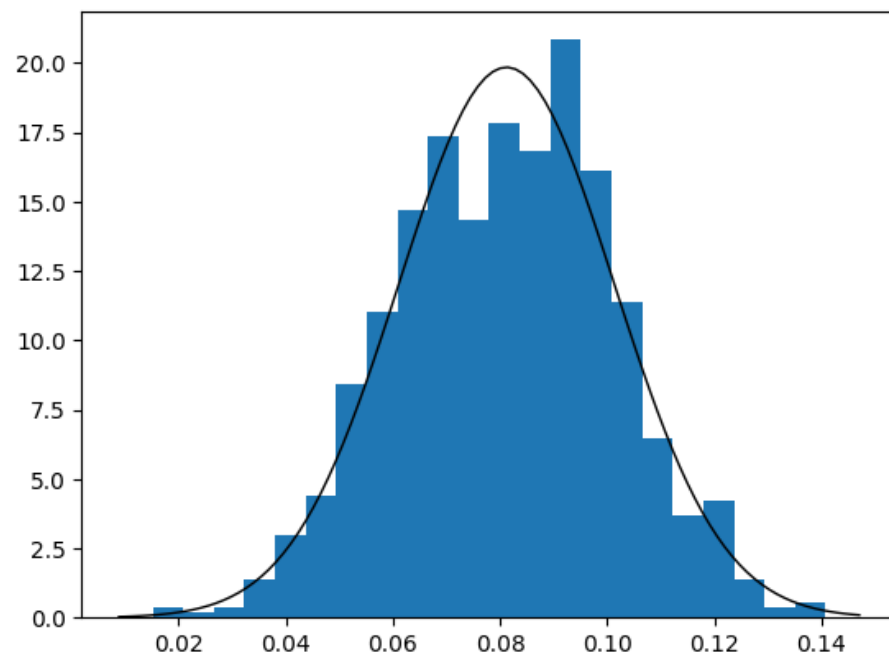
```
In [ ]: normalityTest(df["residual sugar"])
```

p = 0.2247, sehingga data terdistribusi dengan normal



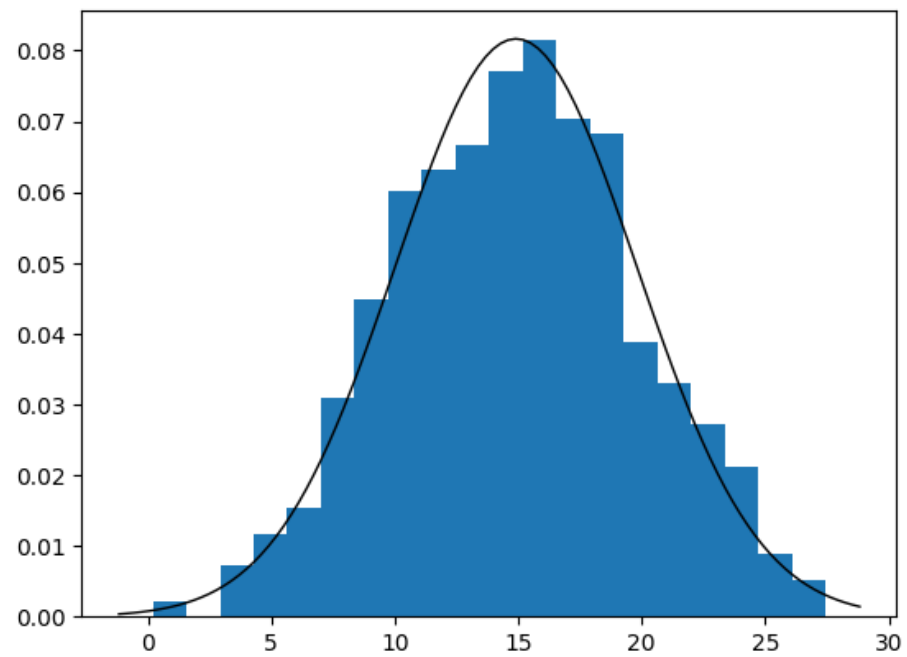
```
In [ ]: normalityTest(df["chlorides"])
```

p = 0.1705, sehingga data terdistribusi dengan normal



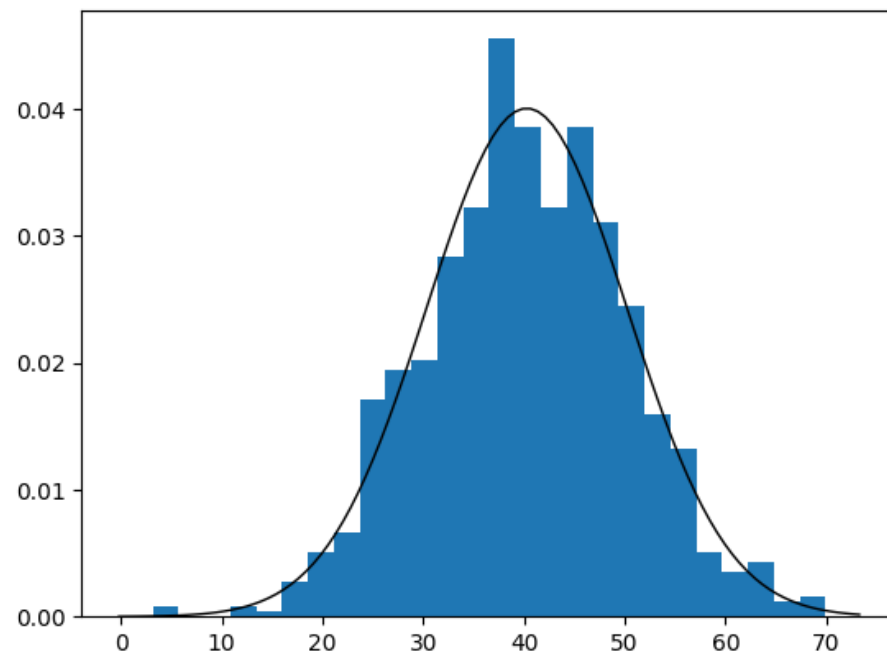
```
In [ ]: normalityTest(df["free sulfur dioxide"])
```

p = 0.0174, sehingga data tidak terdistribusi dengan normal



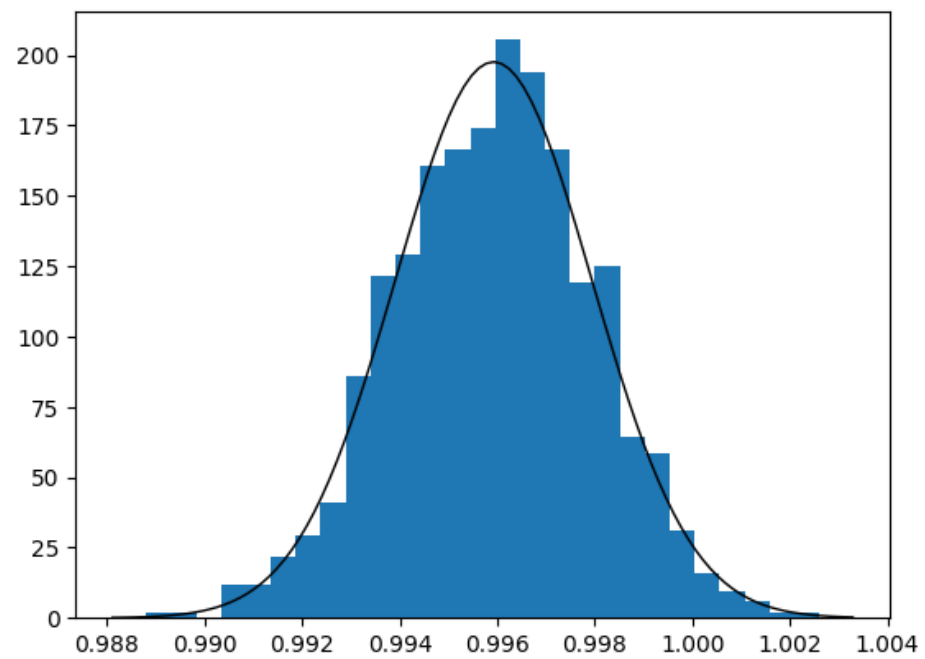
```
In [ ]: normalityTest(df["total sulfur dioxide"])
```

p = 0.8489, sehingga data terdistribusi dengan normal



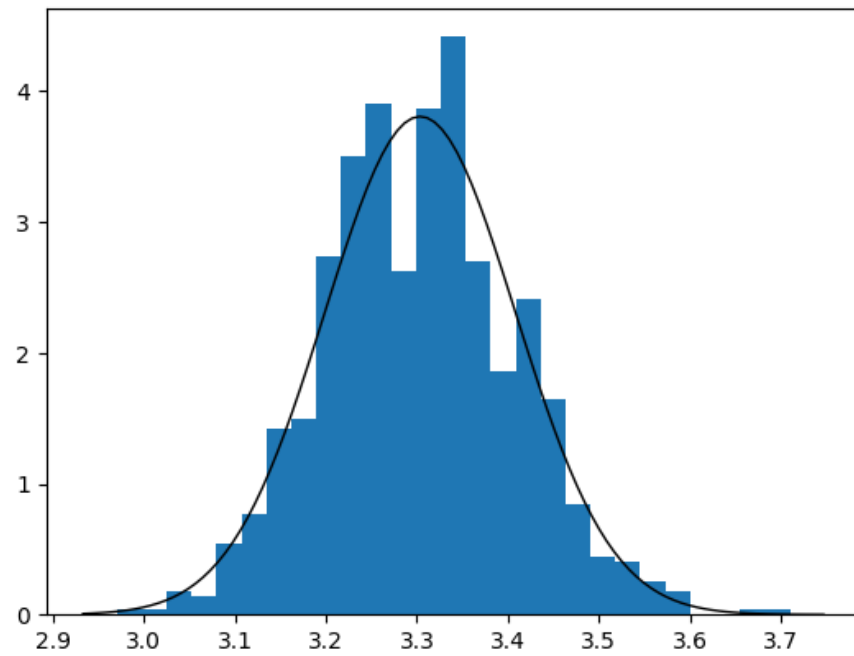
```
In [ ]: normalityTest(df["density"])
```

p = 0.5985, sehingga data terdistribusi dengan normal



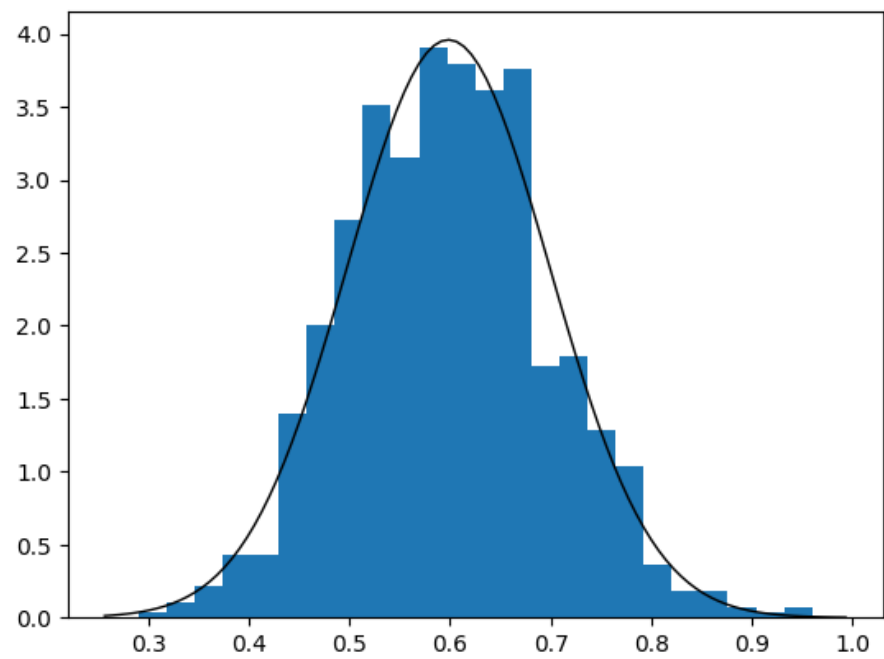
```
In [ ]: normalityTest(df["pH"])
```

p = 0.1368, sehingga data terdistribusi dengan normal



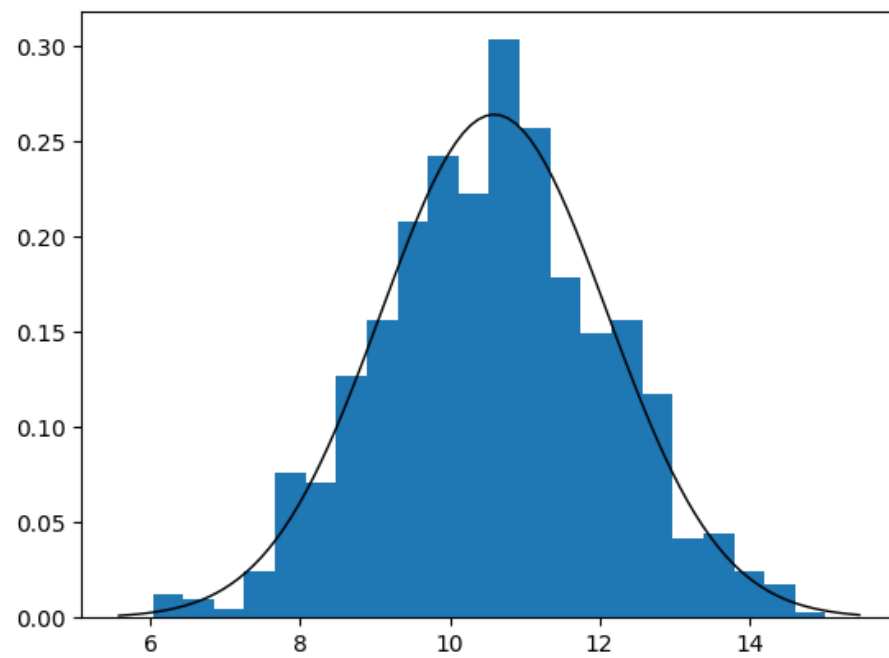
```
In [ ]: normalityTest(df["sulphates"])
```

p = 0.1388, sehingga data terdistribusi dengan normal



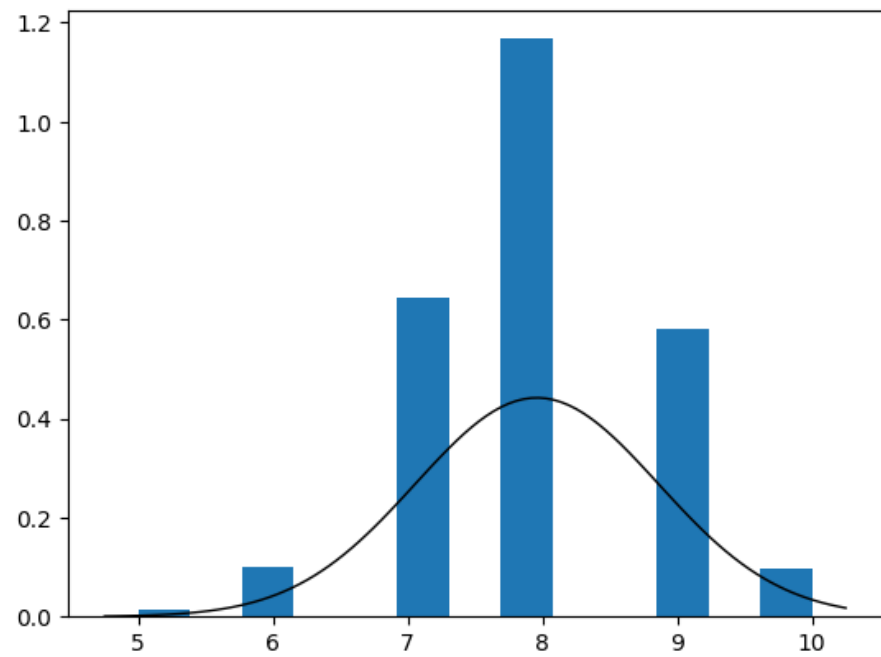
```
In [ ]: normalityTest(df["alcohol"])
```

p = 0.6791, sehingga data terdistribusi dengan normal



```
In [ ]: normalityTest(df["quality"])
```

p = 0.3888, sehingga data terdistribusi dengan normal



4. Melakukan test hipotesis 1 sampel

```
In [ ]: from statsmodels.stats.weightstats import ztest
        from statsmodels.stats.proportion import proportions_ztest
```

4.a Nilai rata-rata pH di atas 3.29

```

In [ ]: alpha = 0.05
pH = df["pH"]

# Uji statistik single sample right tailed Z-test
x_bar = pH.mean()
miu_0 = 3.29
std = pH.std()
n = len(pH)
root_n = np.sqrt(len(pH))
z_alpha = st.norm.ppf(1 - alpha)
z = (x_bar - miu_0) / (std / root_n)
p_value = 1-st.norm.cdf(z)

display(Markdown(f"1. $H_0$: $\mu = \{miu_0\}$"))
display(Markdown(f"2. $H_1$: $\mu > \{miu_0\}$"))
display(Markdown(f"3. $\alpha = 0.05$"))
display(Markdown(f"4. Uji Statistik: $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$"))
display(Markdown(f"Daerah kritis: $z > z_{\alpha}$"))
display(Markdown(f"Jika $z$ berada pada *critical section* ($z > z_{\alpha}$), maka nilai $p < \alpha$ dan $H_0$ ditolak"))
display(Markdown(f"Jika $z$ tidak berada pada *critical section* ($z \leq z_{\alpha}$), maka nilai $p \geq \alpha$ dan $H_0$ diterima"))

display(Markdown(f"5. Komputasi<br> $n = \{n\}$ <br> $\bar{x} = \{\text{round}(x\_bar, 4)\}$ <br> $\sigma = \{\text{round}(std, 4)\}$ <br> $z = \{\text{round}(z, 4)\}$ <br> $z_{\alpha} = \{\text{round}(z\_alpha, 4)\}$"))
display(Markdown(f"6. Test daerah kritis: "))
if z > z_alpha:
    display(Markdown(f"Keputusan: Tolak $H_0$ karena $z > z_{\alpha}$"))
    display(Markdown(f"Tes Signifikan: Tolak $H_0$ karena $p$ lebih kecil dari tingkat signifikan ($p < \alpha$)"))
    display(Markdown(f"Maka, nilai rata-rata pH lebih dari $\{miu_0\}$"))
else:
    display(Markdown(f"Keputusan: Tidak menolak $H_0$ karena $z \leq z_{\alpha}$"))
    display(Markdown(f"Tes Signifikan: Terima $H_0$ dikarenakan $p$ lebih besar atau sama dengan tingkat signifikan ($p \geq \alpha$)"))
    display(Markdown(f"Maka, nilai rata-rata pH sama dengan $\{miu_0\}$"))

df["pH"].plot(kind="box")
plt.title("Boxplot pH")
plt.show()

```

1. $H_0: \mu = 3.29$

2. $H_1: \mu > 3.29$

3. $\alpha = 0.05$

4. Uji Statistik: $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

Daerah kritis: $z > z_{\alpha}$

Jika z berada pada *critical section* ($z > z_{\alpha}$), maka nilai $p < \alpha$ dan H_0 ditolak

Jika z tidak berada pada *critical section* ($z \leq z_\alpha$), maka nilai $p \geq \alpha$ dan H_0 diterima

5. Komputasi

$$n = 1000$$

$$\bar{x} = 3.3036$$

$$\sigma = 0.1049$$

$$z = 4.1038$$

$$z_\alpha = 1.6449$$

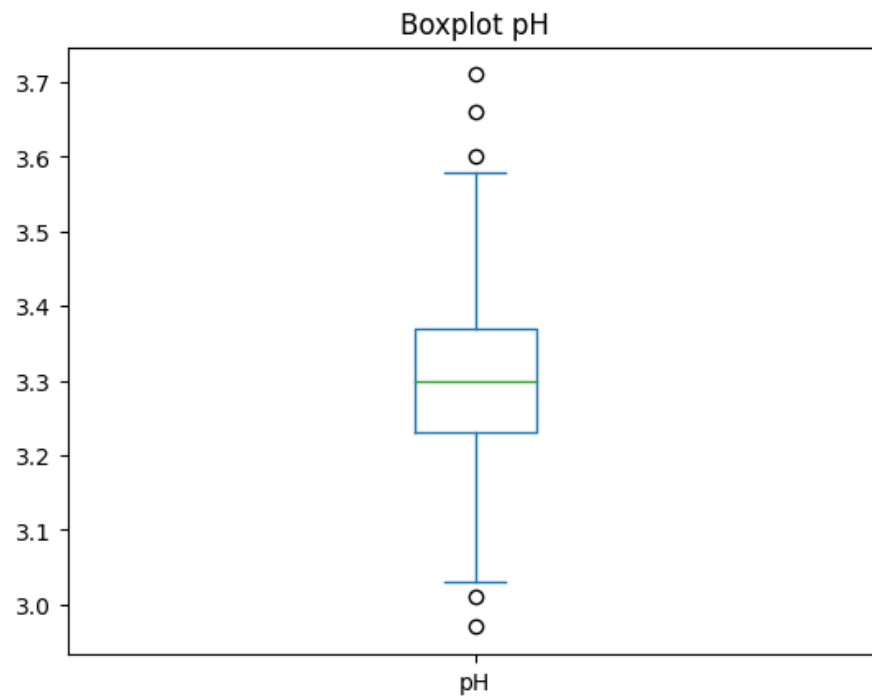
$$p = 2.03e - 05$$

6. Test daerah kritis:

Keputusan: Tolak H_0 karena $z > z_\alpha$

Tes Signifikan: Tolak H_0 karena p lebih kecil dari tingkat signifikan ($p < 0.05$)

Maka, nilai rata-rata pH lebih dari 3.29



4.b Nilai rata-rata Residual Sugar tidak sama dengan 2.50?

```

In [ ]: data = df["residual sugar"]

# Uji statistik single sample two tailed Z-test
miu_0 = 2.50
alpha = 0.05
x_bar = data.mean()
std = data.std()
n = len(data)
root_n = np.sqrt(len(data))
z_alpha = st.norm.ppf(1 - alpha / 2)
z = (x_bar - miu_0) / (std / root_n)
p_value = 1-abs(st.norm.cdf(z)-st.norm.cdf(-1*z))

display(Markdown(f"1. $H_0$: $\mu = \{miu_0\}$"))
display(Markdown(f"2. $H_1$: $\mu \neq \{miu_0\}$"))
display(Markdown(f"3. $\alpha = 0.05$"))
display(Markdown(f"4. Uji Statistik: $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$"))
display(Markdown(f"Daerah kritis: $z_{\alpha/2}$ atau $z < -z_{\alpha/2}$"))
display(Markdown(f"Jika $z$ berada pada *critical section* ( $z_{\alpha/2}$ atau $z < -z_{\alpha/2}$), maka nilai $p < \alpha$ dan $H_0$ ditolak"))
display(Markdown(f"Jika $z$ tidak berada pada *critical section*, maka nilai $p \ge \alpha$ dan $H_0$ diterima"))

display(Markdown(f"5. Komputasi<br> $n = \{n\}$ <br>$\bar{x} = \{\text{round}(x\_bar, 4)\}$ <br> $\sigma = \{\text{round}(std, 4)\}$ <br> $z = \{\text{round}(z, 4)\}$ <br> $z_{\alpha/2} = \{z\_alpha\}$"))
display(Markdown(f"6. Test daerah kritis: "))

if z < -z_alpha or z > z_alpha:
    display(Markdown("Keputusan: Tolak $H_0$ karena $z > z_{\alpha/2}$"))
    display(Markdown(f"Tes Signifikan: Tolak $H_0$ karena $p$ lebih kecil dari tingkat signifikan ($p < \alpha$)"))
    display(Markdown(f"Maka, nilai rata-rata residual sugar tidak sama dengan $\{miu_0\}$"))
else:
    display(Markdown("Keputusan: Tidak menolak $H_0$ karena $-z_{\alpha/2} < z < z_{\alpha/2}$"))
    display(Markdown(f"Tes Signifikan: Tidak menolak $H_0$ dikarenakan $p$ lebih besar atau sama dengan tingkat signifikan ($p \ge \alpha$)"))
    display(Markdown(f"Maka, nilai rata-rata residual sugar sama dengan $\{miu_0\}$"))

data.plot(kind="box")
plt.title("Boxplot Residual Sugar")
plt.show()

```

1. $H_0: \mu = 2.5$

2. $H_1: \mu \neq 2.5$

3. $\alpha = 0.05$

4. Uji Statistik: $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

Daerah kritis: $z_{\alpha/2}$ atau $z < -z_{\alpha/2}$

Jika z berada pada *critical section* ($z_{\alpha/2}$ atau $z < -z_{\alpha/2}$), maka nilai $p < \alpha$ dan H_0 ditolak

Jika z tidak berada pada *critical section*, maka nilai $p \geq \alpha$ dan H_0 diterima

5. Komputasi

$$n = 1000$$

$$\bar{x} = 2.5671$$

$$\sigma = 0.9879$$

$$z = 2.148$$

$$z_{\alpha/2} = 1.96$$

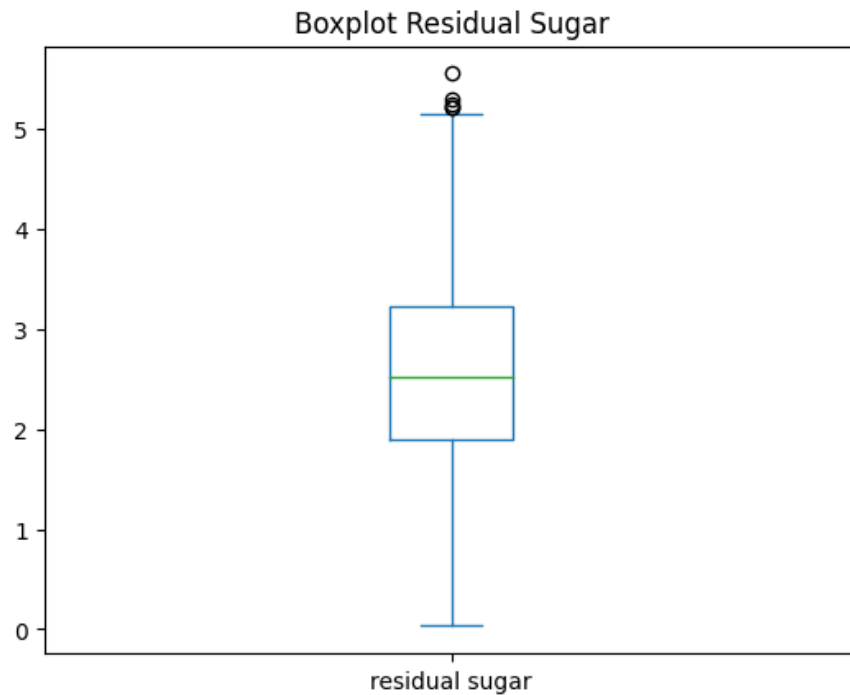
$$p = 0.0317$$

6. Test daerah kritis:

Keputusan: Tolak H_0 karena $z > z_{\alpha/2}$

Tes Signifikan: Tolak H_0 karena p lebih kecil dari tingkat signifikan ($p < 0.05$)

Maka, nilai rata-rata residual sugar tidak sama dengan 2.5



4.c Nilai rata-rata 150 baris pertama kolom sulphates bukan 0.65?

```
In [ ]: data = df["sulphates"].head(150)

# Uji statistik single sample two tailed Z-test
miu_0 = 0.65
alpha = 0.05
x_bar = data.mean()
std = data.std()
n = len(data)
root_n = np.sqrt(len(data))
z_alpha = st.norm.ppf(1 - alpha / 2)
z = (x_bar - miu_0) / (std / root_n)
p_value = 1-abs(st.norm.cdf(z)-st.norm.cdf(-1*z))

display(Markdown(f"1. $H_0$: $\mu = \{miu_0\}$"))
display(Markdown(f"2. $H_1$: $\mu \neq \{miu_0\}$"))
display(Markdown(f"3. $\alpha = 0.05$"))
display(Markdown(f"4. Uji Statistik: $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$"))
display(Markdown(f"Daerah kritis: $z > z_{\alpha/2}$ atau $z < -z_{\alpha/2}$"))
display(Markdown(f"Jika $z$ berada pada *critical section* ( $z_{\alpha/2}$ atau $z < -z_{\alpha/2}$), maka nilai $p < \alpha$ dan $H_0$ ditolak"))
display(Markdown(f"Jika $z$ tidak berada pada *critical section*, maka nilai $p \ge \alpha$ dan $H_0$ diterima"))

display(Markdown(f"5. Komputasi  
 $n = \{n\}$  
 $\bar{x} = \{round(x_bar, 4)\}$  
 $\sigma = \{round(std, 4)\}$  
 $z = \{round(z, 4)\}$  
 $z_{\alpha/2} = \{round(z_alpha, 4)\}$"))
display(Markdown(f"6. Test daerah kritis: "))

if z < -z_alpha or z > z_alpha:
    display(Markdown(f"Keputusan: Tolak $H_0$ karena $z < -z_{\alpha/2}$"))
    display(Markdown(f"Tes Signifikan: Tolak $H_0$ karena $p$ lebih kecil dari tingkat signifikan ($p < \alpha$)"))
    display(Markdown(f"Maka, nilai rata-rata 150 baris pertama kolom sulphates tidak sama dengan $\{miu_0\}$"))
else:
    display(Markdown(f"Keputusan: Tidak menolak $H_0$ karena $-z_{\alpha/2} < z < z_{\alpha/2}$"))
    display(Markdown(f"Tes Signifikan: Tidak menolak $H_0$ dikarenakan $p$ lebih besar atau sama dengan tingkat signifikan ($p \ge \alpha$)"))
    display(Markdown(f"Maka, nilai rata-rata 150 baris pertama kolom sulphates sama dengan $\{miu_0\}$"))

data.head(150).plot(kind="box")
plt.title("Boxplot Sulphates 150 Baris Pertama")
plt.show()
```

1. $H_0: \mu = 0.65$

2. $H_1: \mu \neq 0.65$

3. $\alpha = 0.05$

4. Uji Statistik: $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$

Daerah kritis: $z > z_{\alpha/2}$ atau $z < -z_{\alpha/2}$

Jika z berada pada *critical section* ($z_{\alpha/2}$ atau $z < -z_{\alpha/2}$), maka nilai $p < \alpha$ dan H_0 ditolak

Jika z tidak berada pada *critical section*, maka nilai $p \geq \alpha$ dan H_0 diterima

5. Komputasi

$$n = 150$$

$$\bar{x} = 0.6059$$

$$\sigma = 0.1089$$

$$z = -4.9648$$

$$z_{\alpha/2} = 1.96$$

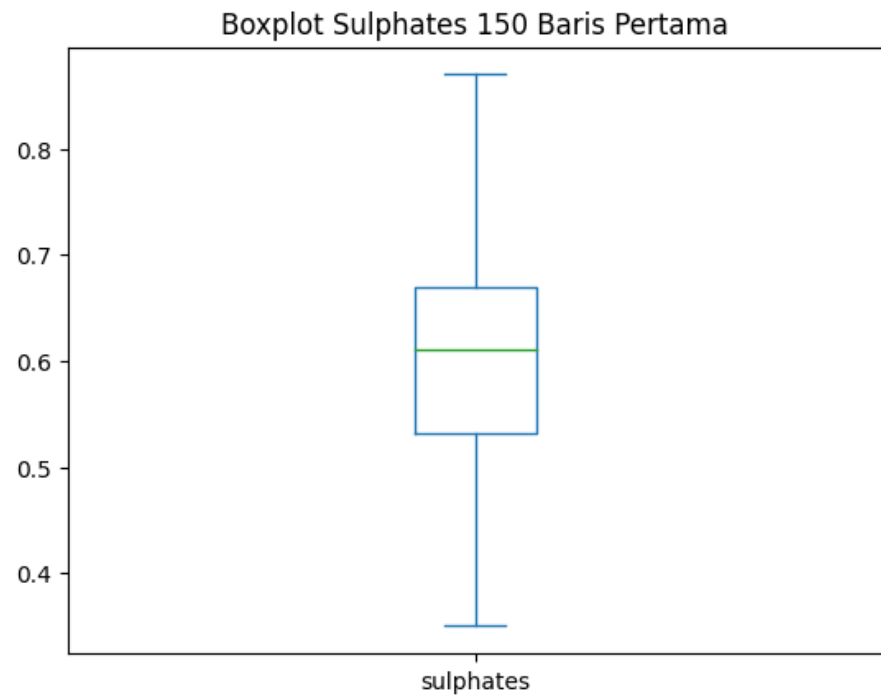
$$p = 6.875652918125752e - 07$$

6. Test daerah kritis:

Keputusan: Tolak H_0 karena $z < -z_{\alpha/2}$

Tes Signifikan: Tolak H_0 karena p lebih kecil dari tingkat signifikan ($p < 0.05$)

Maka, nilai rata-rata 150 baris pertama kolom sulphates tidak sama dengan 0.65



4.d Nilai rata-rata total sulfur dioxide di bawah 35?

```

In [ ]: kolom = "total sulfur dioxide"
data = df[kolom]
miu_0 = 35
alpha = 0.05
x_bar = data.mean()
std = data.std()
n = len(data)
root_n = np.sqrt(len(data))
z_alpha = st.norm.ppf(1 - alpha / 2)
z = (x_bar - miu_0) / (std / root_n)
p_value = 1 - abs(st.norm.cdf(z) - st.norm.cdf(-1 * z))

display(Markdown(f"1.  $H_0$ :  $\mu = \{miu\_0\}$ "))
display(Markdown(f"2.  $H_1$ :  $\mu < \{miu\_0\}$ "))
display(Markdown(f"3.  $\alpha = 0.05$ "))
display(Markdown(f"4. Uji Statistik:  $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$ "))
display(Markdown(f"Daerah kritis:  $z < z_{\alpha}$ "))
display(Markdown(f"Jika  $z$  berada pada *critical section* ( $z < z_{\alpha}$ ), maka nilai  $p > \alpha$  dan  $H_0$  ditolak"))
display(Markdown(f"Jika  $z$  tidak berada pada *critical section* ( $z \geq z_{\alpha}$ ), maka nilai  $p \leq \alpha$  dan  $H_0$  diterima"))

display(Markdown(f""5. Komputasi<br>  $n = \{n\}$  <br>
 $\bar{x} = \{\text{round}(x\_bar, 4)\}$  <br>  $\sigma = \{\text{round}(std, 4)\}$ 
<br>  $z = \{\text{round}(z, 4)\}$  <br>  $z_{\alpha} = \{\text{round}(z\_alpha, 4)\}$ 
<br>  $p = \{\text{round}(p\_value, 7)\}$  <br> ""))
display(Markdown(f"6. Test daerah kritis: "))
if z < z_alpha:
    display(Markdown(f"Keputusan: Tolak  $H_0$  karena  $z < z_{\alpha}$ "))
    display(Markdown(f"Tes Signifikan: Tolak  $H_0$  karena  $p$  lebih kecil dari tingkat signifikan ( $p \geq \alpha$ )"))
    display(Markdown(f"Maka, nilai rata-rata {kolom} kurang dari  $\{miu\_0\}$ "))
else:
    display(Markdown(f"Keputusan: Tidak menolak  $H_0$  karena  $z \geq z_{\alpha}$  "))
    display(Markdown(f"Tes Signifikan: Terima  $H_0$  dikarenakan  $p$  lebih besar atau sama dengan tingkat signifikan ( $p < \alpha$ )"))
    display(Markdown(f"Maka, nilai rata-rata {kolom} sama dengan  $\{miu\_0\}$ "))

df[kolom].plot(kind="box")
plt.title(f"Boxplot {kolom}")
plt.show()

```

1. $H_0: \mu = 35$

2. $H_1: \mu < 35$

3. $\alpha = 0.05$

4. Uji Statistik: $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$

Daerah kritis: $z < z_\alpha$

Jika z berada pada *critical section* ($z < z_\alpha$), maka nilai $p > \alpha$ dan H_0 ditolak

Jika z tidak berada pada *critical section* ($z \geq z_\alpha$), maka nilai $p \leq \alpha$ dan H_0 diterima

5. Komputasi

$$n = 1000$$

$$\bar{x} = 40.2902$$

$$\sigma = 9.9658$$

$$z = 16.7864$$

$$z_\alpha = 1.96$$

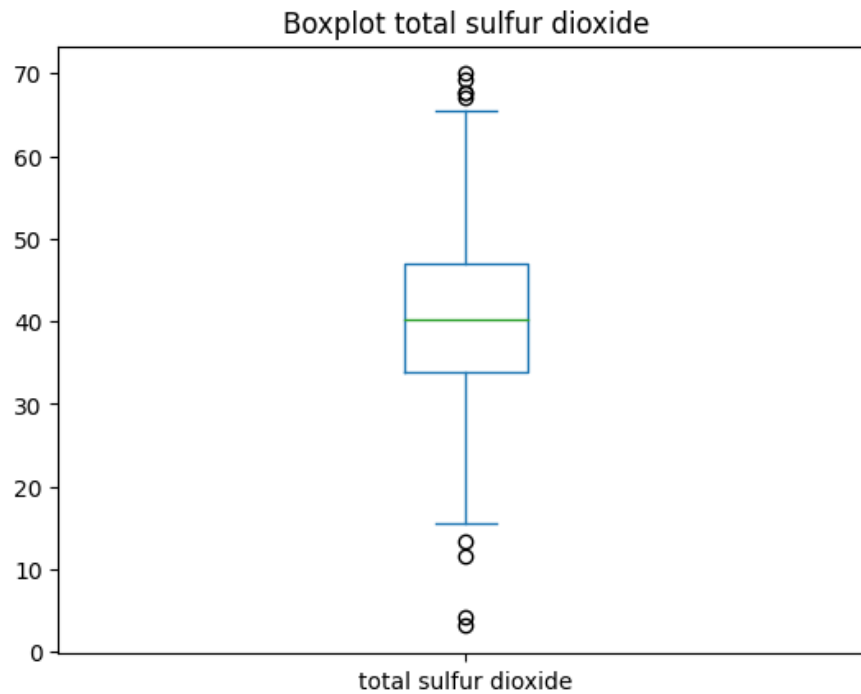
$$p = 0.0$$

6. Test daerah kritis:

Keputusan: Tidak menolak H_0 karena $z \geq z_\alpha$

Tes Signifikan: Terima H_0 dikarenakan p lebih besar atau sama dengan tingkat signifikan ($p < 0.05$)

Maka, nilai rata-rata total sulfur dioxide sama dengan 35



4.e Proporsi nilai total sulfur dioxide yang lebih dari 40, adalah tidak sama dengan 50% ?

```
In [ ]: kolom = "total sulfur dioxide"
data = df[kolom]
p_0 = 0.5
alpha = 0.05
x_bar = data.mean()
std = data.std()
N = len(data)
n = len(data[df[kolom] > 40])
root_n = np.sqrt(len(data))
z_alpha = st.norm.ppf(1 - alpha / 2)

z, p_value = proportions_ztest(n, N, p_0)

display(Markdown(f"1.  $H_0$ :  $p = \{p_0\}$ "))
display(Markdown(f"2.  $H_1$ :  $p \neq \{p_0\}$ "))
display(Markdown(f"3.  $\alpha = 0.05$ "))
display(Markdown(f"4. Uji Statistik:  $z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}}$ "))
display(Markdown(f"Daerah kritis:  $z > z_{\alpha/2}$  atau  $z < -z_{\alpha/2}$ "))
display(Markdown(f"Jika  $z$  berada pada *critical section* ( $z > z_{\alpha/2}$  atau  $z < -z_{\alpha/2}$ ), maka nilai  $p < \alpha$  dan  $H_0$  ditolak"))
display(Markdown(f"Jika  $z$  tidak berada pada *critical section*, maka nilai  $p \geq \alpha$  dan  $H_0$  diterima"))

display(Markdown(f""5. Komputasi  
  $n = \{n\}$    

 $\bar{x} = \{\text{round}(x\_bar, 4)\}$    
  $\sigma = \{\text{round}(std, 4)\}$ 
  
  $z = \{\text{round}(z, 4)\}$    
  $z_{\alpha} = \{\text{round}(z\_alpha, 4)\}$ 
  
  $p = \{\text{round}(p\_value, 7)\}$    
 ""))
display(Markdown(f"6. Test daerah kritis: "))

if z > z_alpha or z < -z_alpha:
    display(Markdown(f"Keputusan: Tolak  $H_0$  karena  $z > z_{\alpha/2}$  atau  $z < -z_{\alpha/2}$  "))
    display(Markdown(f"Tes Signifikan: Tolak  $H_0$  karena  $p$  lebih kecil dari tingkat signifikan ( $p < \alpha$ )"))
    display(Markdown(f"Maka, proporsi nilai total {kolom} yang lebih dari 40 tidak sama dengan  $\{p_0\}$ "))
else:
    display(Markdown(f"Keputusan: Tidak menolak  $H_0$  karena  $z \leq z_{\alpha/2}$  atau  $z \geq -z_{\alpha/2}$  "))
    display(Markdown(f"Tes Signifikan: Terima  $H_0$  dikarenakan  $p$  lebih besar atau sama dengan tingkat signifikan ( $p \geq \alpha$ )"))
    display(Markdown(f"Maka, proporsi nilai total {kolom} yang lebih dari 40 sama dengan  $\{p_0\}$ "))

df[kolom].plot(kind="box")
plt.title(f"Boxplot {kolom}")
plt.show()
```

1. $H_0: p = 0.5$

2. $H_1: p \neq 0.5$

3. $\alpha = 0.05$

4. Uji Statistik: $z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}}$

Daerah kritis: $z > z_{\alpha/2}$ atau $z < -z_{\alpha/2}$

Jika z berada pada *critical section* ($z > z_{\alpha/2}$ atau $z < -z_{\alpha/2}$), maka nilai $p < \alpha$ dan H_0 ditolak

Jika z tidak berada pada *critical section*, maka nilai $p \geq \alpha$ dan H_0 diterima

5. Komputasi

$$n = 512$$

$$\bar{x} = 40.2902$$

$$\sigma = 9.9658$$

$$z = 0.7592$$

$$z_{\alpha} = 1.96$$

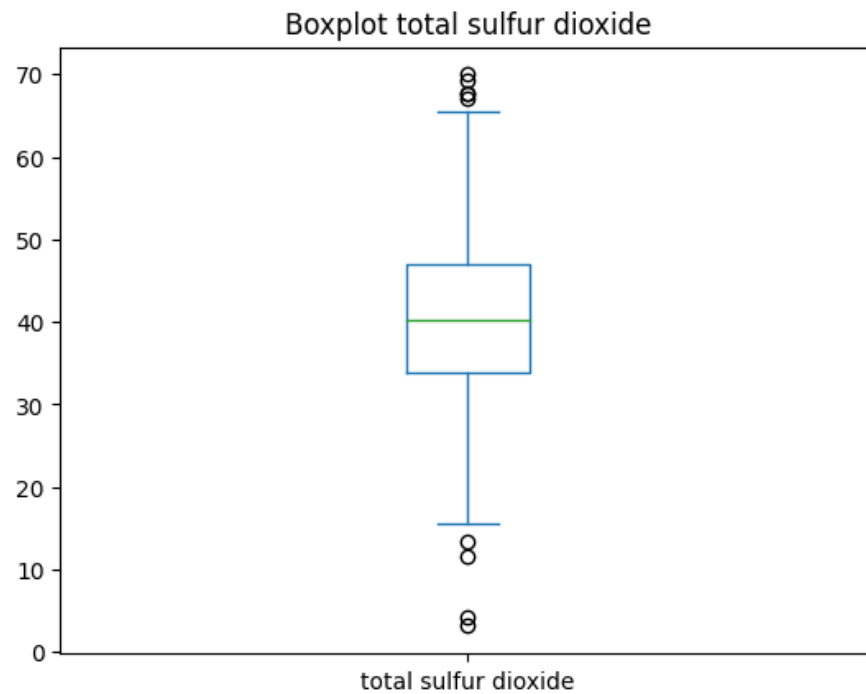
$$p = 0.4477537$$

6. Test daerah kritis:

Keputusan: Tidak menolak H_0 karena $z \leq z_{\alpha/2}$ atau $z \geq -z_{\alpha/2}$

Tes Signifikan: Terima H_0 dikarenakan p lebih besar atau sama dengan tingkat signifikan ($p \geq 0.05$)

Maka, proporsi nilai total total sulfur dioxide yang lebih dari 40 sama dengan 0.5



5. Melakukan test hipotesis 2 sampel,

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

```

In [ ]: kolom = "fixed acidity"
miu_0 = 0
alpha = 0.05
n = len(df)
awal = df[kolom][:n//2]
akhir = df[kolom][n//2:]
x_bar1 = awal.mean()
x_bar2 = akhir.mean()
std1 = awal.std()
std2 = akhir.std()

z, p_value = ztest(awal, akhir, value=miu_0, alternative="two-sided")

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

display(Markdown(f"1. $H_0$: $μ_1 - μ_2 = {miu_0}$"))
display(Markdown(f"2. $H_1$: $μ_1 - μ_2 ≠ {miu_0}$"))
display(Markdown(f"3. $\alpha = 0.05$"))
display(Markdown(f"4. Uji Statistik: $z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$"))
display(Markdown("Daerah kritis: $z > z_{\alpha/2}$ atau $z < -z_{\alpha/2}$"))
display(Markdown("Jika $z$ berada pada *critical section* ( $z > z_{\alpha/2}$ atau $z < -z_{\alpha/2}$), maka nilai $p < \alpha$ dan $H_0$ ditolak"))
display(Markdown(f"Jika $z$ tidak berada pada *critical section*, maka nilai $p \ge \alpha$ dan $H_0$ diterima"))

display(Markdown(f""5. Komputasi<br> n total $ = {n}$ <br>
$\bar{x}_1$ = ${round(x_bar1, 4)}$ <br> $\sigma_1$ = ${round(std1, 4)}$
<br> $\bar{x}_2$ = ${round(x_bar2, 4)}$ <br> $\sigma_2$ = ${round(std2, 4)}$
<br> $z$ = ${round(z, 4)}$ <br> $z_{\alpha}$ = ${round(z_alpha, 4)}$
<br> $p$ = ${round(p_value, 7)}$<br> ""))

display(Markdown(f"6. Test daerah kritis: "))
if z > z_alpha or z < -z_alpha:
    display(Markdown("Keputusan: Tolak $H_0$ karena $z > z_{\alpha/2}$ atau $z < -z_{\alpha/2}$ "))
    display(Markdown(f"Tes Signifikan: Tolak $H_0$ karena $p$ lebih kecil dari tingkat signifikan ($p < \alpha$)"))
    display(Markdown(f"Maka, rata-rata kedua bagian tidak sama"))
else:
    display(Markdown("Keputusan: Tidak menolak $H_0$ karena $z \le z_{\alpha/2}$ atau $z \ge -z_{\alpha/2}$ "))
    display(Markdown(f"Tes Signifikan: Terima $H_0$ dikarenakan $p$ lebih besar atau sama dengan tingkat signifikan ($p \ge \alpha$)"))
    display(Markdown(f"Maka, rata-rata kedua bagian sama"))

awal.plot(kind="box")
plt.title(f"Boxplot {kolom} Awal")
plt.show()

akhir.plot(kind="box")
plt.title(f"Boxplot {kolom} Akhir")
plt.show()

```

1. $H_0: \mu_1 - \mu_2 = 0$

2. $H_1: \mu_1 - \mu_2 \neq 0$

3. $\alpha = 0.05$

4. Uji Statistik: $z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$

Daerah kritis: $z > z_{\alpha/2}$ atau $z < -z_{\alpha/2}$

Jika z berada pada *critical section* ($z > z_{\alpha/2}$ atau $z < -z_{\alpha/2}$), maka nilai $p < \alpha$ dan H_0 ditolak

Jika z tidak berada pada *critical section*, maka nilai $p \geq \alpha$ dan H_0 diterima

5. Komputasi

$n_{\text{total}} = 1000$

$\bar{x}_1 = 7.1535$

$\sigma_1 = 1.2049$

$\bar{x}_2 = 7.1515$

$\sigma_2 = 1.1995$

$z = 0.026$

$z_{\alpha} = 1.96$

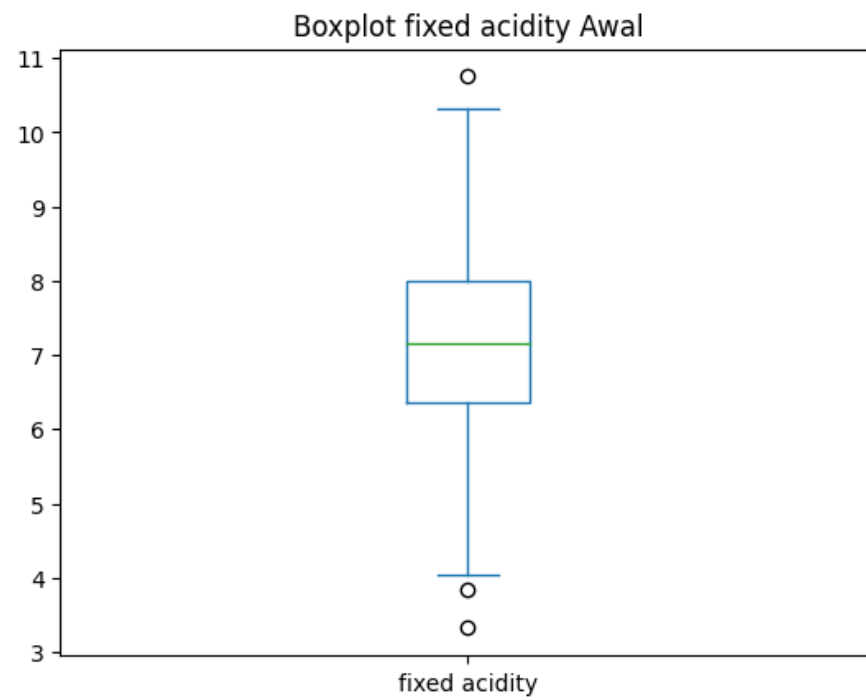
$p = 0.9792246$

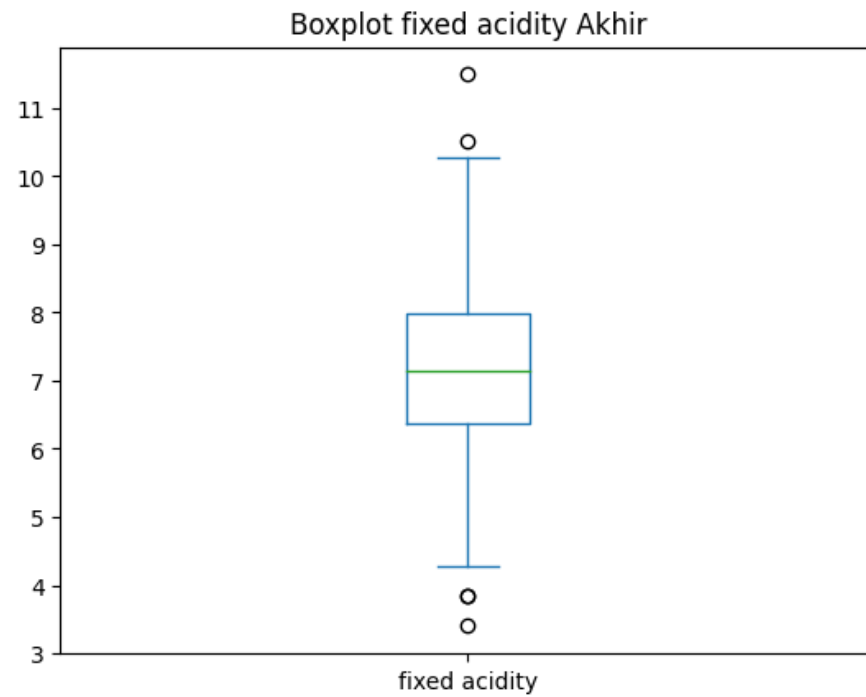
6. Test daerah kritis:

Keputusan: Tidak menolak H_0 karena $z \leq z_{\alpha/2}$ atau $z \geq -z_{\alpha/2}$

Tes Signifikan: Terima H_0 dikarenakan p lebih besar atau sama dengan tingkat signifikan ($p \geq 0.05$)

Maka, rata-rata kedua bagian sama





5.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?

```

In [ ]: kolom = "chlorides"
miu_0 = 0.001
alpha = 0.05
N = len(df)
awal = df[kolom][:N//2]
akhir = df[kolom][N//2:]
x_bar1 = awal.mean()
x_bar2 = akhir.mean()
std1 = awal.std()
std2 = akhir.std()

z, p_value = ztest(awal, akhir, value=miu_0, alternative="two-sided")

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

display(Markdown(f"1. $H_0$:  $\mu_1 - \mu_2 = \{miu\_0\}$ "))
display(Markdown(f"2. $H_1$:  $\mu_1 - \mu_2 \neq \{miu\_0\}$ "))
display(Markdown(f"3.  $\alpha = 0.05$ "))
display(Markdown(f"4. Uji Statistik:  $z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$ "))
display(Markdown("Daerah kritis:  $z > z_{\alpha/2}$  atau  $z < -z_{\alpha/2}$ "))
display(Markdown("Jika  $z$  berada pada *critical section* ( $z > z_{\alpha/2}$  atau  $z < -z_{\alpha/2}$ ), maka nilai  $p < \alpha$  dan  $H_0$  ditolak"))
display(Markdown(f"Jika  $z$  tidak berada pada *critical section*, maka nilai  $p \geq \alpha$  dan  $H_0$  diterima"))

display(Markdown(f""5. Komputasi<br> n total  $= \{n\}$  <br>
 $\bar{x}_1 = \{\text{round}(x\_bar1, 4)\}$  <br>  $\sigma_1 = \{\text{round}(std1, 4)\}$ 
<br>  $\bar{x}_2 = \{\text{round}(x\_bar2, 4)\}$  <br>  $\sigma_2 = \{\text{round}(std2, 4)\}$ 
<br>  $z = \{\text{round}(z, 4)\}$  <br>  $z_{\alpha} = \{\text{round}(z\_alpha, 4)\}$ 
<br>  $p = \{\text{round}(p\_value, 7)\}$ <br> ""))

display(Markdown(f"6. Test daerah kritis: "))
if z > z_alpha or z < -z_alpha:
    display(Markdown("Keputusan: Tolak  $H_0$  karena  $z > z_{\alpha/2}$  atau  $z < -z_{\alpha/2}$  "))
    display(Markdown(f"Tes Signifikan: Tolak  $H_0$  karena  $p$  lebih kecil dari tingkat signifikan ( $p < \alpha$ )"))
    display(Markdown(f"Maka, rata-rata selisih bagian awal dengan bagian akhir tidak lebih besar dari  $\{miu\_0\}$ "))
else:
    display(Markdown("Keputusan: Tidak menolak  $H_0$  karena  $z \leq z_{\alpha/2}$  atau  $z \geq -z_{\alpha/2}$  "))
    display(Markdown(f"Tes Signifikan: Terima  $H_0$  dikarenakan  $p$  lebih besar atau sama dengan tingkat signifikan ( $p \geq \alpha$ )"))
    display(Markdown(f"Maka, rata-rata selisih bagian awal dengan bagian akhir tepat sebesar  $\{miu\_0\}$ "))

awal.plot(kind="box")
plt.title(f"Boxplot {kolom} Awal")
plt.show()

akhir.plot(kind="box")
plt.title(f"Boxplot {kolom} Akhir")
plt.show()

```

1. $H_0: \mu_1 - \mu_2 = 0.001$

2. $H_1: \mu_1 - \mu_2 \neq 0.001$

3. $\alpha = 0.05$

4. Uji Statistik: $z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$

Daerah kritis: $z > z_{\alpha/2}$ atau $z < -z_{\alpha/2}$

Jika z berada pada *critical section* ($z > z_{\alpha/2}$ atau $z < -z_{\alpha/2}$), maka nilai $p < \alpha$ dan H_0 ditolak

Jika z tidak berada pada *critical section*, maka nilai $p \geq \alpha$ dan H_0 diterima

5. Komputasi

$$n \text{ total} = 1000$$

$$\bar{x}_1 = 0.0814$$

$$\sigma_1 = 0.0202$$

$$\bar{x}_2 = 0.081$$

$$\sigma_2 = 0.0201$$

$$z = -0.4673$$

$$z_{\alpha} = 1.96$$

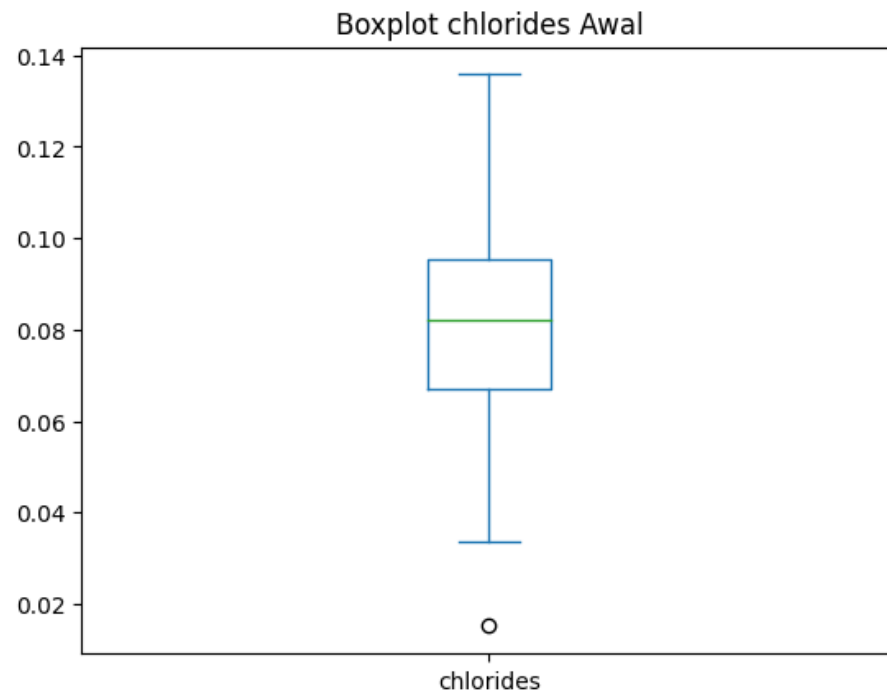
$$p = 0.640273$$

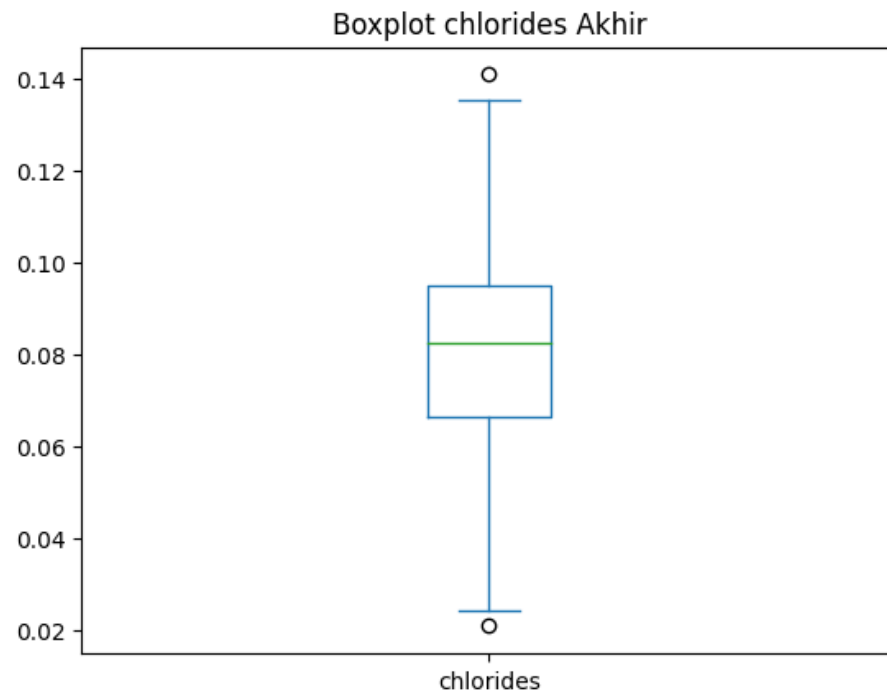
6. Test daerah kritis:

Keputusan: Tidak menolak H_0 karena $z \leq z_{\alpha/2}$ atau $z \geq -z_{\alpha/2}$

Tes Signifikan: Terima H_0 dikarenakan p lebih besar atau sama dengan tingkat signifikan ($p \geq 0.05$)

Maka, rata-rata selisih bagian awal dengan bagian akhir tepat sebesar 0.001





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

```

In [ ]: kolom1 = "volatile acidity"
kolom2 = "sulphates"
miu_0 = 0
alpha = 0.05
awal = df[kolom1].head(25)
akhir = df[kolom2].tail(25)
x_bar1 = awal.mean()
x_bar2 = akhir.mean()
std1 = awal.std()
std2 = akhir.std()

z, p_value = ztest(awal, akhir, value=miu_0, alternative="two-sided")

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

display(Markdown(f"Asumsi data bukan merupakan data sampel"))
display(Markdown(f"1.  $H_0$ :  $\mu_1 - \mu_2 = \mu_0$ "))
display(Markdown(f"2.  $H_1$ :  $\mu_1 - \mu_2 \neq \mu_0$ "))
display(Markdown(f"3.  $\alpha = 0.05$ "))
display(Markdown(f"4. Uji Statistik:  $z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$ "))
display(Markdown(f"Daerah kritis:  $z > z_{\alpha/2}$  atau  $z < -z_{\alpha/2}$ "))
display(Markdown(f"Jika  $z$  berada pada *critical section* ( $z > z_{\alpha/2}$  atau  $z < -z_{\alpha/2}$ ), maka nilai  $p < \alpha$  dan  $H_0$  ditolak"))
display(Markdown(f"Jika  $z$  tidak berada pada *critical section*, maka nilai  $p \geq \alpha$  dan  $H_0$  diterima"))

display(Markdown(f""5. Komputasi<br> n total  $= n$ <br>
 $\bar{x}_1 = \text{round}(\bar{x}_1, 4)$ <br>  $\sigma_1 = \text{round}(\text{std1}, 4)$ 
<br>  $\bar{x}_2 = \text{round}(\bar{x}_2, 4)$ <br>  $\sigma_2 = \text{round}(\text{std2}, 4)$ 
<br>  $z = \text{round}(z, 4)$ <br>  $z_{\alpha} = \text{round}(z_{\alpha}, 4)$ 
<br>  $p = \text{round}(p\_value, 7)$ <br> ""))

display(Markdown(f"6. Test daerah kritis: "))
if z > z_alpha or z < -z_alpha:
    display(Markdown(f"Keputusan: Tolak  $H_0$  karena  $z > z_{\alpha/2}$  atau  $z < -z_{\alpha/2}$  "))
    display(Markdown(f"Tes Signifikan: Tolak  $H_0$  karena  $p$  lebih kecil dari tingkat signifikan ( $p < \alpha$ )"))
    display(Markdown(f"Maka, rata-rata {kolom1} tidak sama dengan {kolom2}"))
else:
    display(Markdown(f"Keputusan: Tidak menolak  $H_0$  karena  $z \leq z_{\alpha/2}$  atau  $z \geq -z_{\alpha/2}$  "))
    display(Markdown(f"Tes Signifikan: Terima  $H_0$  dikarenakan  $p$  lebih besar atau sama dengan tingkat signifikan ( $p \geq \alpha$ )"))
    display(Markdown(f"Maka, rata-rata {kolom1} sama dengan {kolom2}"))

awal.plot(kind="box")
plt.title(f"Boxplot {kolom1} Awal")
plt.show()

akhir.plot(kind="box")
plt.title(f"Boxplot {kolom2} Akhir")
plt.show()

```

Asumsi data bukan merupakan data sampel

1. $H_0: \mu_1 - \mu_2 = 0$

2. $H_1: \mu_1 - \mu_2 \neq 0$

3. $\alpha = 0.05$

4. Uji Statistik: $z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$

Daerah kritis: $z > z_{\alpha/2}$ atau $z < -z_{\alpha/2}$

Jika z berada pada *critical section* ($z > z_{\alpha/2}$ atau $z < -z_{\alpha/2}$), maka nilai $p < \alpha$ dan H_0 ditolak

Jika z tidak berada pada *critical section*, maka nilai $p \geq \alpha$ dan H_0 diterima

5. Komputasi

$$n \text{ total} = 1000$$

$$\bar{x}_1 = 0.5014$$

$$\sigma_1 = 0.0834$$

$$\bar{x}_2 = 0.6164$$

$$\sigma_2 = 0.1171$$

$$z = -3.9978$$

$$z_{\alpha} = 1.96$$

$$p = 6.39e - 05$$

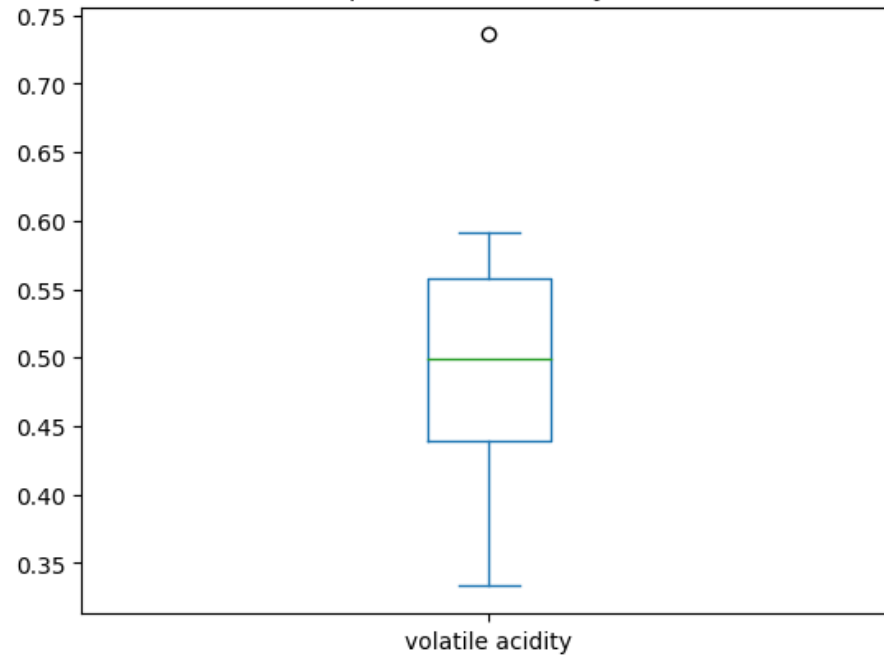
6. Test daerah kritis:

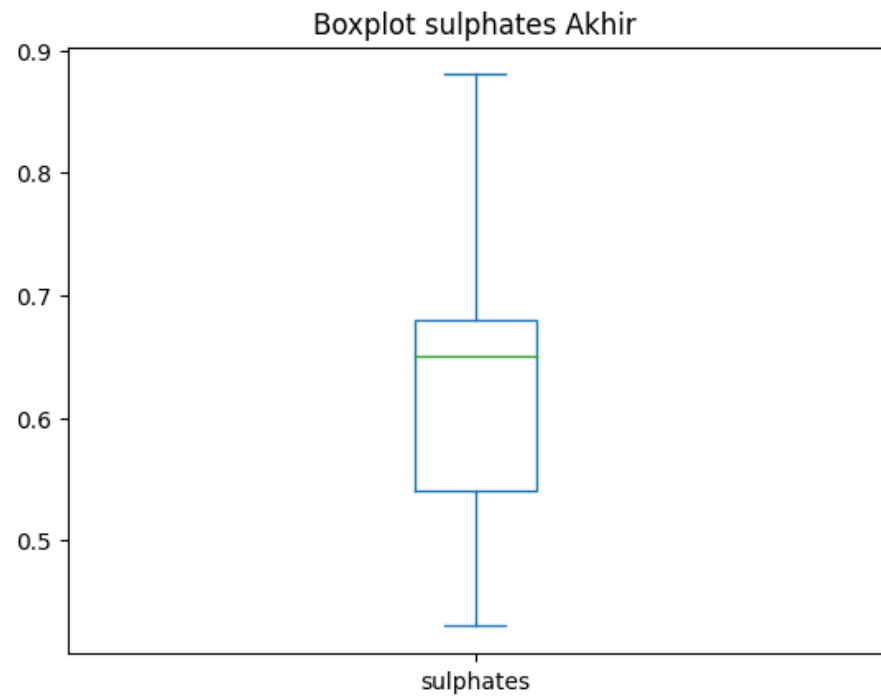
Keputusan: Tolak H_0 karena $z > z_{\alpha/2}$ atau $z < -z_{\alpha/2}$

Tes Signifikan: Tolak H_0 karena p lebih kecil dari tingkat signifikan ($p < 0.05$)

Maka, rata-rata volatile acidity tidak sama dengan sulphates

Boxplot volatile acidity Awal





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

```

In [ ]: kolom = "residual sugar"
alpha = 0.05

# v = n - 1
v_1 = len(df[:len(df) // 2]) - 1
v_2 = len(df[len(df) // 2:]) - 1

# Bagi data menjadi dua bagian
awal = df[kolom][:len(df) // 2]
akhir = df[kolom][len(df) // 2:]

# Uji statistik double sampe two tailed F test
f = awal.var() / akhir.var()
p = 1 - st.f.cdf(f, v_1, v_2)

# Daerah kritis
f_lower = st.f.ppf(alpha / 2, v_1, v_2)
f_upper = st.f.ppf(1 - alpha / 2, v_1, v_2)

display(Markdown("1.  $H_0$ :  $\sigma_1^2 = \sigma_2^2$ "))
display(Markdown("2.  $H_1$ :  $\sigma_1^2 \neq \sigma_2^2$ "))
display(Markdown(f"3.  $\alpha = \{alpha\}$ "))
display(Markdown("4. Uji Statistik:  $f = \frac{S_1^2}{S_2^2}$ "))
display(Markdown(f"Daerah kritis:  $f < f_{\{1 - \alpha / 2\}}(v_1, v_2)$  atau  $f > f_{\{\alpha / 2\}}(v_1, v_2)$ "))
display(Markdown(f"Jika  $f$  berada pada *critical section* ( $f < f_{\{1 - \alpha / 2\}}(v_1, v_2)$  atau  $f > f_{\{\alpha / 2\}}(v_1, v_2)$ ), maka nilai  $p < \alpha$ "))
display(Markdown(f"Jika  $f$  tidak berada pada *critical section* ( $f \geq f_{\{1 - \alpha / 2\}}(v_1, v_2)$  atau  $f \leq f_{\{\alpha / 2\}}(v_1, v_2)$ ), maka nilai  $p \geq \alpha$ "))

display(Markdown(f'''5. Komputasi: <br>
     $f_{\{1 - \alpha / 2\}}(v_1, v_2) = \{round(f\_lower, 4)\}$  <br>
     $f_{\{\alpha / 2\}}(v_1, v_2) = \{round(f\_upper, 4)\}$  <br>
     $f = \{round(f, 4)\}$  <br>
     $p = \{round(p, 4)\}$  <br>
    '''))

if f < f_lower or f > f_upper:
    display(Markdown(f"Keputusan: Tolak  $H_0$  karena  $f$  berada pada *critical section* ( $f < \{round(f\_lower, 4)\}$  atau  $f > \{round(f\_upper, 4)\}$ "))
    display(Markdown(f"Tes Signifikan: Tolak  $H_0$  karena  $p < \{alpha\}$ "))
    display(Markdown(f"Maka, variansi pada bagian awal kolom {kolom} tidak sama dengan variansi pada bagian akhirnya"))
else:
    display(Markdown(f"Keputusan: Tidak menolak  $H_0$  karena  $f$  tidak berada pada *critical section* ( $\{round(f\_lower, 4)\} \leq f \leq \{round(f\_upper, 4)\}$ "))
    display(Markdown(f"Tes Signifikan: Tidak menolak  $H_0$  karena  $p \geq \{alpha\}$ "))
    display(Markdown(f"Maka, variansi pada bagian awal kolom {kolom} sama dengan variansi pada bagian akhirnya"))

awal.plot(kind="box")
plt.title(f"Boxplot {kolom} Awal")
plt.show()

akhir.plot(kind="box")
plt.title(f"Boxplot {kolom} Akhir")
plt.show()

```

pic.show()

1. $H_0: \sigma_1^2 = \sigma_2^2$

2. $H_1: \sigma_1^2 \neq \sigma_2^2$

3. $\alpha = 0.05$

4. Uji Statistik: $f = \frac{s_1^2}{s_2^2}$

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

Jika f berada pada *critical section* ($f < f_{1-\alpha/2}(v_1, v_2)$ atau $f > f_{\alpha/2}(v_1, v_2)$), maka nilai $p < \alpha$ dan H_0 ditolak

Jika f tidak berada pada *critical section* ($f \geq f_{1-\alpha/2}(v_1, v_2)$ atau $f \leq f_{\alpha/2}(v_1, v_2)$), maka nilai $p \geq \alpha$ dan H_0 diterima

5. Komputasi:

$$f_{1-\alpha/2}(v_1, v_2) = 0.8389$$

$$f_{\alpha/2}(v_1, v_2) = 1.1921$$

$$f = 0.942$$

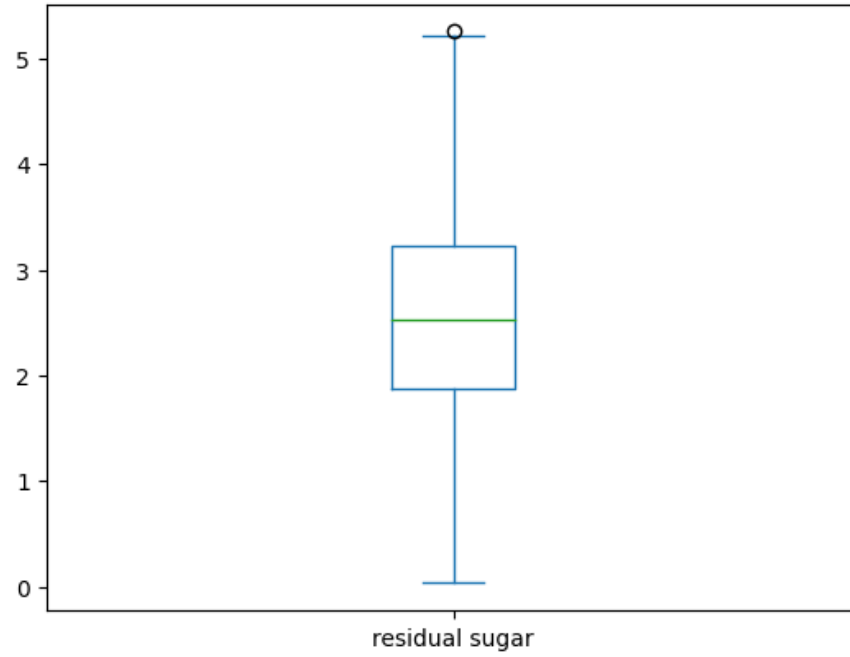
$$p = 0.7476$$

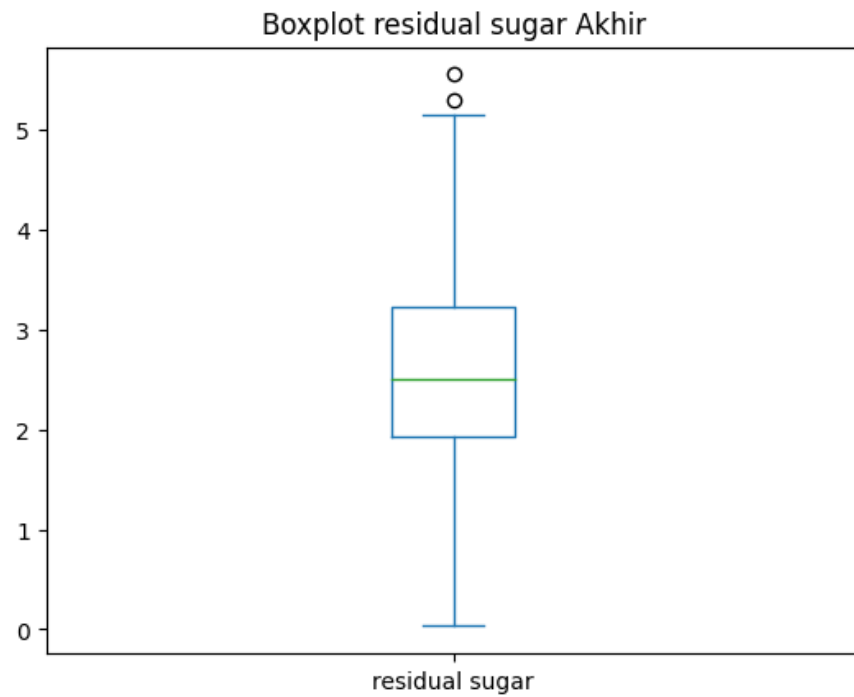
Keputusan: Tidak menolak H_0 karena f tidak berada pada *critical section* ($0.8389 \leq z \leq 1.1921$)

Tes Signifikan: Tidak menolak H_0 karena $p \geq 0.05$

Maka, variansi pada bagian awal kolom residual sugar sama dengan variansi pada bagian akhirnya

Boxplot residual sugar Awal





5.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 [ ]: kolom = "alcohol"
delta_p = 0
alpha = 0.05
z_alpha = st.norm.ppf(1 - alpha)

# Bagi data menjadi dua bagian
awal = df.iloc[:len(df)//2]
akhir = df.iloc[len(df)//2:]

# Uji statistik single sample right tailed one propotion Z test
z_score, p_value = proportions_ztest([len(awal[awal[kolom] > 7]),
                                     len(akhir[akhir[kolom] > 7])],
                                     [len(awal), len(akhir)],
                                     value=delta_p, alternative='larger')

display(Markdown("1. $H_0$: $p_1 - p_2 = 0$"))
display(Markdown("2. $H_1$: $p_1 - p_2 > 0$"))
display(Markdown(f"3. $\alpha = {alpha}$"))
display(Markdown(f"4. Uji Statistik: $z = \frac{p_1 - p_2 - \delta_p}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$"))
display(Markdown(f"Daerah kritis: $z > z_{\alpha}$"))
display(Markdown(f"Jika $z$ berada pada *critical section* ($z > z_{\alpha}$), maka nilai $p < \alpha$ dan $H_0$ ditolak"))
display(Markdown(f"Jika $z$ tidak berada pada *critical section* ($z \leq z_{1-\alpha}$), maka nilai $p \geq \alpha$ dan $H_0$ diterima"))
display(Markdown(f'''5. Komputasi: <br>
    $z_{\alpha} = \text{round}(z_{\alpha}, 4)$ <br>
    $z = \text{round}(z\_score, 4)$ <br> $p = \text{round}(p\_value, 4)$ <br>
    '''))

if z_score > z_alpha:
    display(Markdown(f"Keputusan: Tolak $H_0$ karena $z$ berada pada *critical section* ($z > \text{round}(z_{\alpha}, 4)$)"))
    display(Markdown(f"Tes Signifikan: Tolak $H_0$ karena $p < \alpha$"))
    display(Markdown(f"Maka, proporsi nilai setengah bagian awal alcohol yang lebih dari 7, lebih besar daripada proporsi nilai yang sama di setengah bagian a
else:
    display(Markdown(f"Keputusan: Tidak menolak $H_0$ karena $z$ tidak berada pada *critical section* ($z \leq \text{round}(z_{\alpha}, 4)$)"))
    display(Markdown(f"Tes Signifikan: Tidak menolak $H_0$ karena $p \geq \alpha$"))
    display(Markdown(f"Maka, proporsi nilai setengah bagian awal alcohol yang lebih dari 7, sama besar daripada proporsi nilai yang sama di setengah bagian a

```

1. $H_0: p_1 - p_2 = 0$

2. $H_1: p_1 - p_2 > 0$

3. $\alpha = 0.05$

4. Uji Statistik:
$$z = \frac{p_1 - p_2 - \delta_p}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

Daerah kritis: $z > z_\alpha$

Jika z berada pada *critical section* ($z > z_\alpha$), maka nilai $p < \alpha$ dan H_0 ditolak

Jika z tidak berada pada *critical section* ($z \leq z_{1-\alpha}$), maka nilai $p \geq \alpha$ dan H_0 diterima

5. Komputasi:

$$z_\alpha = 1.6449$$

$$z = 0.0$$

$$p = 0.5$$

Keputusan: Tidak menolak H_0 karena z tidak berada pada *critical section* ($z \leq 1.6449$)

Tes Signifikan: Tidak menolak H_0 karena $p \geq 0.05$

Maka, proporsi nilai setengah bagian awal alcohol yang lebih dari 7, sama besar daripada proporsi nilai yang sama di setengah bagian akhir