

Tugas Besar IF2220 Probabilitas dan Statistika

Penarikan Kesimpulan dan Pengujian Hipotesis

13521109 Rizky Abdillah Rasyid

13521119 Muhammad Rizky Sya'ban

Inisialisasi *library* yang akan digunakan.

```
In [2]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from IPython.display import Markdown, display

df = pd.read_csv("Dataset/anggur.csv")
```

PROBLEM 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.

```
In [3]: display(Markdown("#### Mean, Median, Standar Deviasi, Variansi, Range, Minimum, Maximum, Kuartil, IQR, Skewness, dan Kurtosis"))

stat = pd.DataFrame()

# # Calculate the mode for each column
def get_mode(x):
    try:
        mode = x.mode()[1]
        mode = 'multivalued'
    except:
        mode = x.mode()[0]
    return mode

stat["Mean"] = df.mean()
stat["Median"] = df.median()
stat["Mode"] = df.apply(get_mode)
stat["Deviasi"] = df.std()
stat["Var"] = df.var()
stat["Min"] = df.min()
stat["Max"] = df.max()
stat["Range"] = stat["Max"] - stat["Min"]
stat["Q1"] = df.quantile(0.25)
stat["Q2"] = df.quantile(0.5)
stat["Q3"] = df.quantile(0.75)
stat["IQR"] = stat["Q3"] - stat["Q1"]
stat["Skewness"] = df.skew()
stat["Kurtosis"] = df.kurtosis()
stat
```

Mean, Median, Standar Deviasi, Variansi, Range, Minimum, Maximum, Kuartil, IQR, Skewness, dan Kurtosis

Out[3]:

	Mean	Median	Mode	Deviasi	Var	Min	Max	Range	Q1	Q2	Q3	IQR	Skewness	Kurtosis
fixed acidity	7.152530	7.150000	6.54	1.201598	1.443837	3.320000	11.490000	8.170000	6.377500	7.150000	8.000000	1.622500	-0.028879	-0.019292
volatile acidity	0.520839	0.524850	0.5546	0.095848	0.009187	0.139900	0.805100	0.665200	0.456100	0.524850	0.585375	0.129275	-0.197699	0.161853
citric acid	0.270517	0.272200	0.3019	0.049098	0.002411	0.116700	0.409600	0.292900	0.237800	0.272200	0.302325	0.064525	-0.045576	-0.104679
residual sugar	2.567104	2.519430	multivalued	0.987915	0.975977	0.032555	5.550755	5.518200	1.896330	2.519430	3.220873	1.324544	0.132638	-0.042980
chlorides	0.081195	0.082167	multivalued	0.020111	0.000404	0.015122	0.140758	0.125635	0.066574	0.082167	0.095312	0.028738	-0.051319	-0.246508
free sulfur dioxide	14.907679	14.860346	multivalued	4.888100	23.893519	0.194679	27.462525	27.267847	11.426717	14.860346	18.313098	6.886381	0.007130	-0.364964
total sulfur dioxide	40.290150	40.190000	multivalued	9.965767	99.316519	3.150000	69.960000	66.810000	33.785000	40.190000	47.022500	13.237500	-0.024060	0.063950
density	0.995925	0.996000	multivalued	0.002020	0.000004	0.988800	1.002600	0.013800	0.994600	0.996000	0.997200	0.002600	-0.076883	0.016366
pH	3.303610	3.300000	3.34	0.104875	0.010999	2.970000	3.710000	0.740000	3.230000	3.300000	3.370000	0.140000	0.147673	0.080910
sulphates	0.598390	0.595000	0.59	0.100819	0.010164	0.290000	0.960000	0.670000	0.530000	0.595000	0.670000	0.140000	0.149199	0.064819
alcohol	10.592280	10.610000	multivalued	1.510706	2.282233	6.030000	15.020000	8.990000	9.560000	10.610000	11.622500	2.062500	-0.018991	-0.131732
quality	7.958000	8.000000	8	0.902802	0.815051	5.000000	10.000000	5.000000	7.000000	8.000000	9.000000	2.000000	-0.089054	0.108291

```
In [4]: display(Markdown("#### Modus [*multivalued*]"))
df.mode()
```

Modus [*multivalued*]

Out[4]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0	6.54	0.5546	0.3019	0.032555	0.015122	0.194679	35.20	0.9959	3.34	0.59	9.86	8.0
1	NaN	NaN	NaN	0.033333	0.020794	0.621628	37.25	0.9961	NaN	NaN	10.31	NaN
2	NaN	NaN	NaN	0.051774	0.024259	0.860177	39.64	0.9965	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	0.077156	0.027209	3.032139	40.61	0.9970	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	0.084744	0.032111	3.129885	41.05	NaN	NaN	NaN	NaN	NaN
...
995	NaN	NaN	NaN	5.210260	0.131425	26.630490	NaN	NaN	NaN	NaN	NaN	NaN
996	NaN	NaN	NaN	5.217429	0.133656	26.665773	NaN	NaN	NaN	NaN	NaN	NaN
997	NaN	NaN	NaN	5.252864	0.135368	26.822626	NaN	NaN	NaN	NaN	NaN	NaN
998	NaN	NaN	NaN	5.299524	0.135790	27.006307	NaN	NaN	NaN	NaN	NaN	NaN
999	NaN	NaN	NaN	5.550755	0.140758	27.462525	NaN	NaN	NaN	NaN	NaN	NaN

1000 rows × 12 columns

PROBLEM 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 [5]: dataHeader = df.columns

plt.rcParams['boxplot.boxprops.linewidth'] = 2
plt.rcParams['boxplot.whiskerprops.linewidth'] = 2
plt.rcParams['boxplot.capprops.linewidth'] = 2
plt.rcParams['boxplot.medianprops.linewidth'] = 2
plt.rcParams['boxplot.flierprops.marker'] = 'o'
plt.rcParams['boxplot.flierprops.markersize'] = 8
plt.rcParams['boxplot.flierprops.linewidth'] = 2
plt.rcParams['boxplot.flierprops.markeredgcolor'] = '#ab1ba5'
```

```
In [6]: def problem2(attribute) :

    display(Markdown(f"<h3 style='text-align:center; text-transform: uppercase'> {attribute} </h3>"))

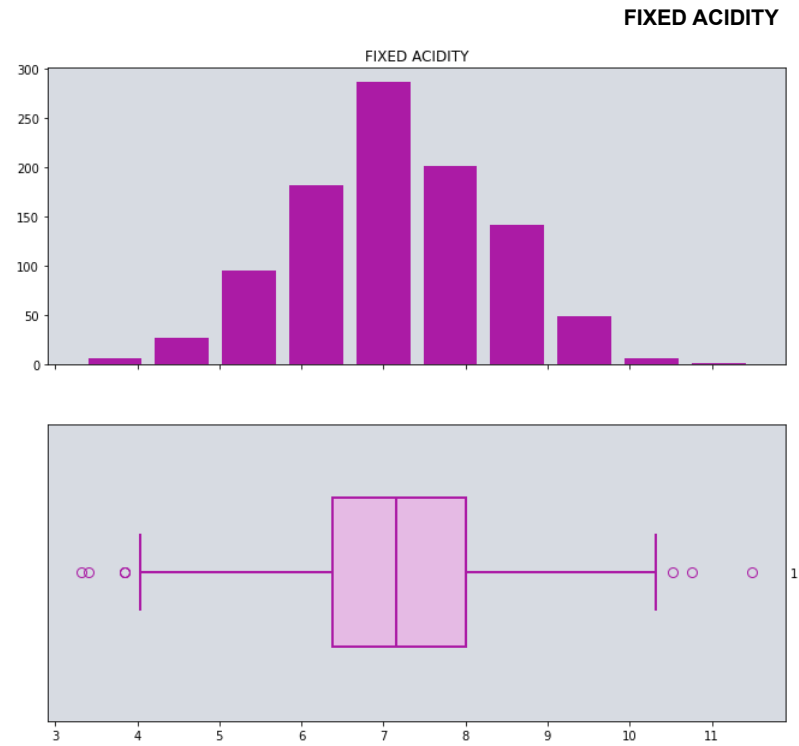
    fig, (ax1, ax2) = plt.subplots(2, 1, figsize=(11, 10), sharex=True)
    plt.subplot(2, 1, 1)

    ax1.set_title(attribute.upper())
    ax1.hist(df[attribute], bins=10, rwidth=0.8, color="#ab1ba5")
    ax1.set_facecolor('#d7dbe2')

    ax2.boxplot(df[attribute], vert=False, widths=0.5, patch_artist=True, boxprops=dict(facecolor='#e5bae4', color='#ab1ba5'), capprops=dict(color='#ab1ba5'), medianprops=dict(color='#ab1ba5'), flierprops=dict(color='#ab1ba5'), whiskerprops=dict(color='#ab1ba5'))
    ax2.set_facecolor('#d7dbe2')
    ax2.tick_params(axis='y', which='both', length=0)
    ax2.yaxis.tick_right()

    plt.show()
```

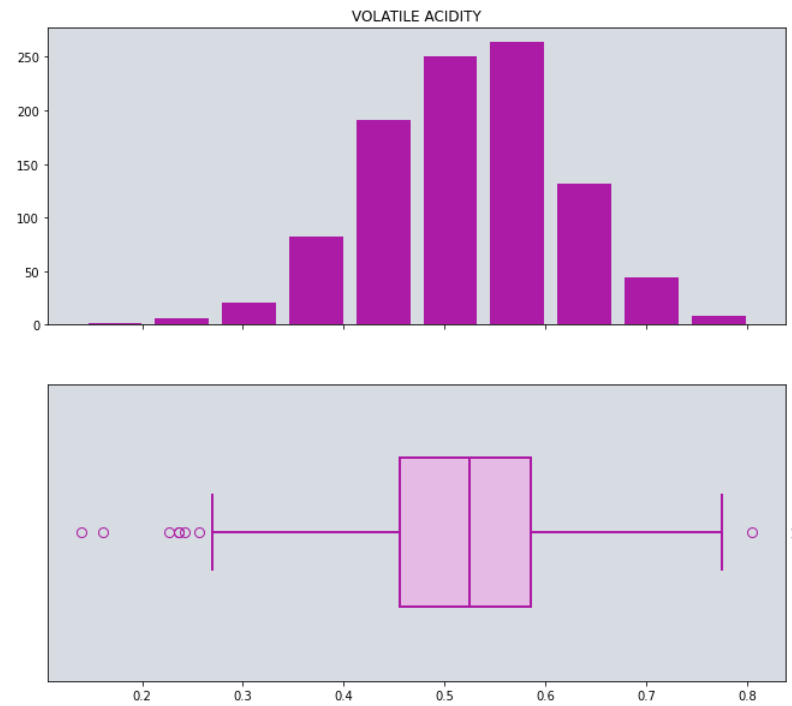
```
In [7]: problem2(dataHeader[0])
```



Berdasarkan histogram dan box plot yang terbentuk dapat dilihat bahwa data *Fixed Acidity* terdistribusi secara **simetris**. Walaupun konsentrasi data cukup besar pada bagian kiri pusat distribusi, namun *skew* tersebut **hampir tidak terlihat**. Dapat dilihat juga cukup sedikit *outliers* yang berada di kanan nilai *maximum whisker* ataupun yang berada di kiri nilai *minimum whisker*. Oleh karena itu, dapat dikatakan bahwa data *Fixed Acidity* cukup konsisten terhadap pusat distribusi.

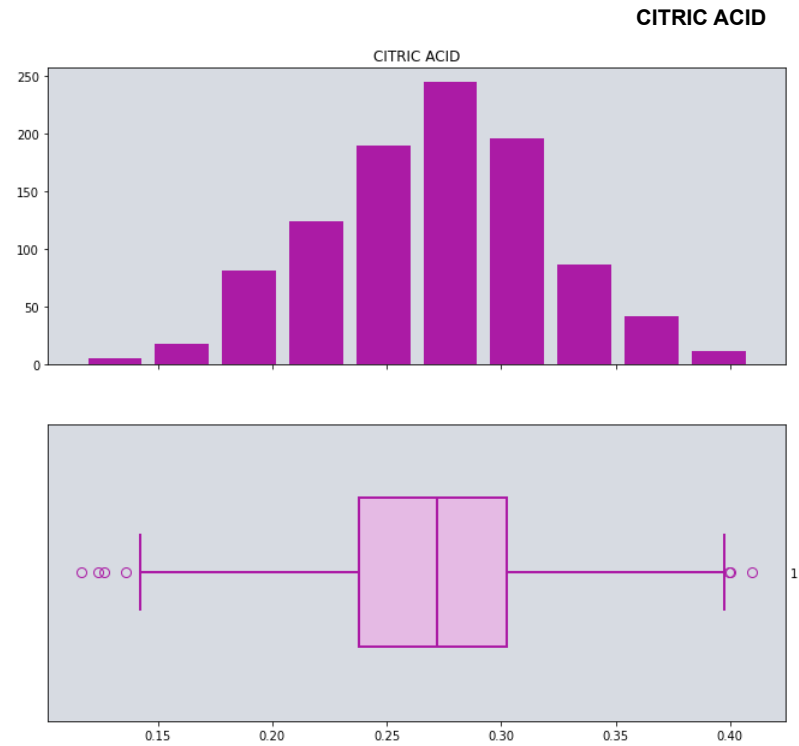
```
In [8]: problem2(dataHeader[1])
```

VOLATILE ACIDITY



Berdasarkan histogram dan box plot yang terbentuk dapat dilihat bahwa data *Volatile Acidity* secara umum terdistribusi secara **simetris** namun data cenderung lebih banyak terkonsentrasi pada sisi kiri pusat distribusi sehingga menyebabkan data tersebut cenderung miring ke kiri (*left-skewed*). Dapat dilihat juga bahwa terdapat cukup banyak *outliers* di kiri nilai *minimum whisker* yang menyebabkan panjangnya ekor distribusi di sisi kiri *mean*. Oleh karena itu, dapat dikatakan bahwa data *Volatile Acidity* **kurang konsisten** terhadap pusat distribusinya.

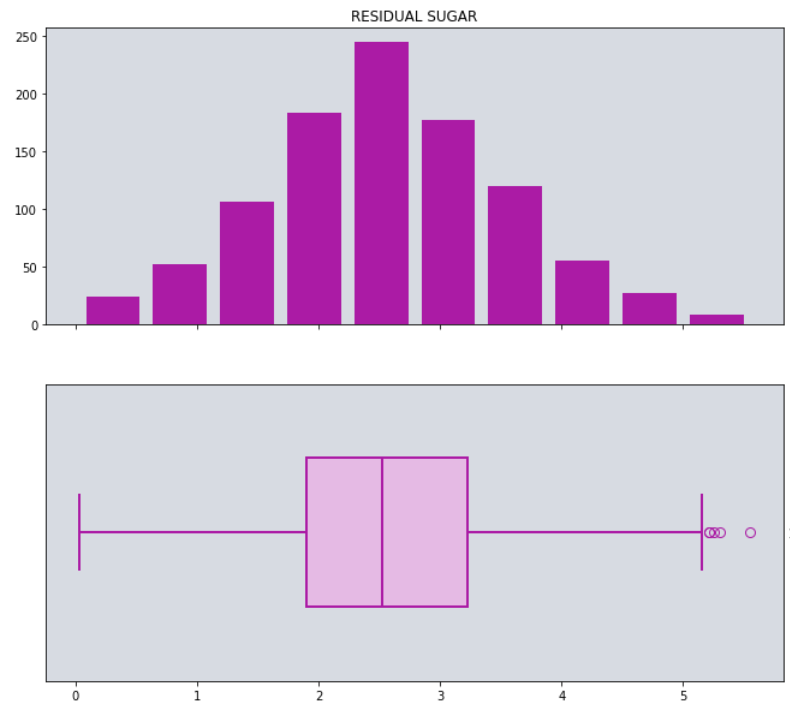
```
In [9]: problem2(dataHeader[2])
```



Berdasarkan histogram dan box plot yang terbentuk dapat dilihat bahwa data *Citric Acid* terdistribusi secara **simetris**. Walaupun konsentrasi data di sisi kiri pusat distribusi agak lebih tinggi dibanding di sisi kanannya, namun *skew* tersebut **hampir tidak terlihat**. Dapat dilihat juga cukup sedikit *outliers* yang berada di kanan nilai *maximum whisker* dan di kiri nilai *minimum whisker*. Oleh karena itu, dapat dikatakan bahwa data *Citric Acid* cukup **konsisten** terhadap pusat distribusi.

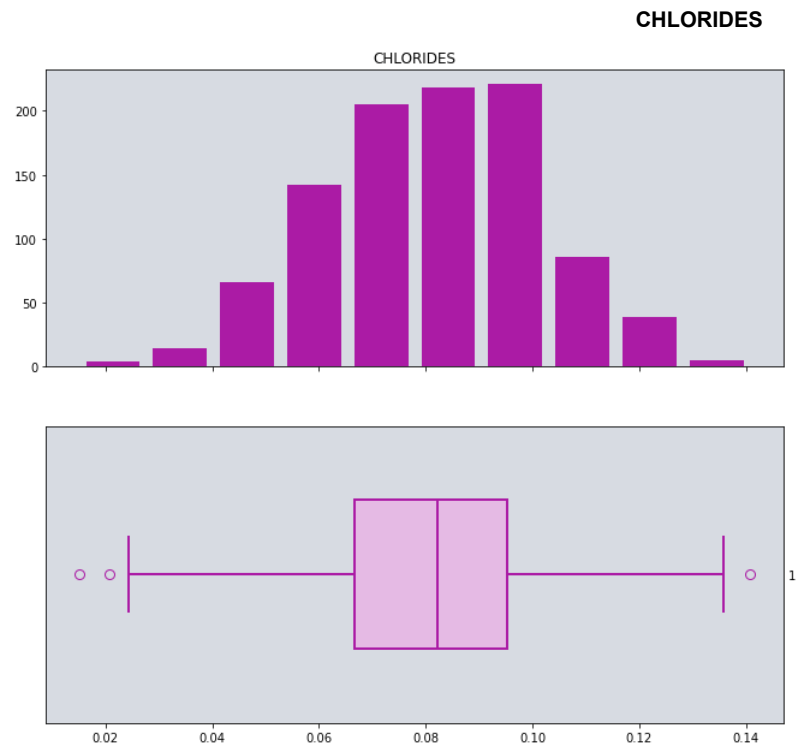
```
In [10]: problem2(dataHeader[3])
```

RESIDUAL SUGAR



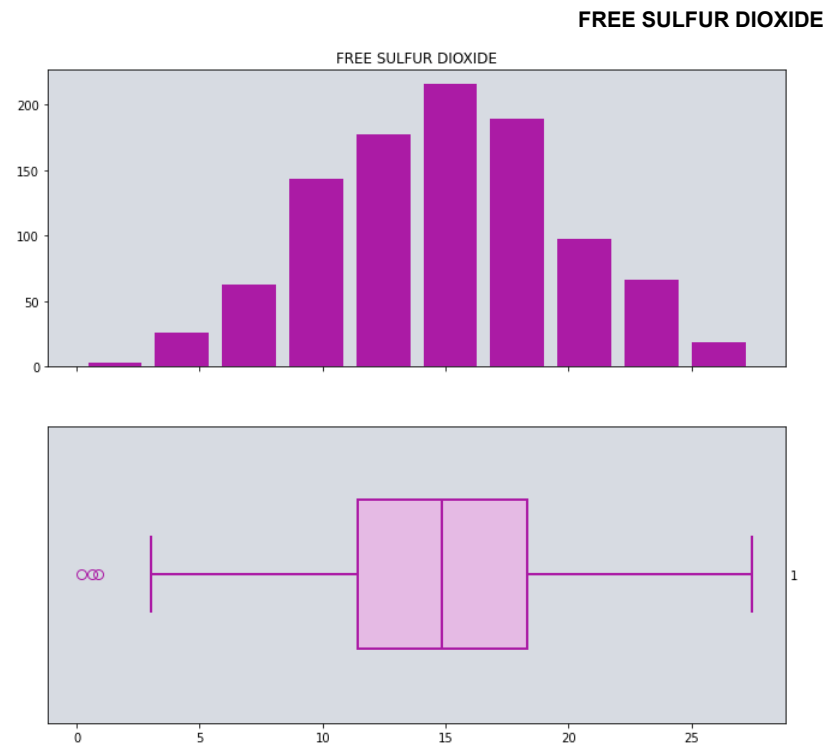
Berdasarkan histogram dan box plot yang terbentuk dapat dilihat bahwa data *Residual Sugar* secara umum terdistribusi secara **simetris**, namun cenderung lebih banyak data yang terkonsentrasi pada sisi kanan pusat distribusi sehingga menyebabkan data tersebut cenderung miring ke kanan (*right-skewed*). Dapat dilihat juga bahwa terdapat cukup banyak *outliers* yang berada di kanan nilai *maximum whisker* yang menyebabkan panjangnya ekor distribusi di sisi kanan pusat distribusi.

```
In [11]: problem2(dataHeader[4])
```



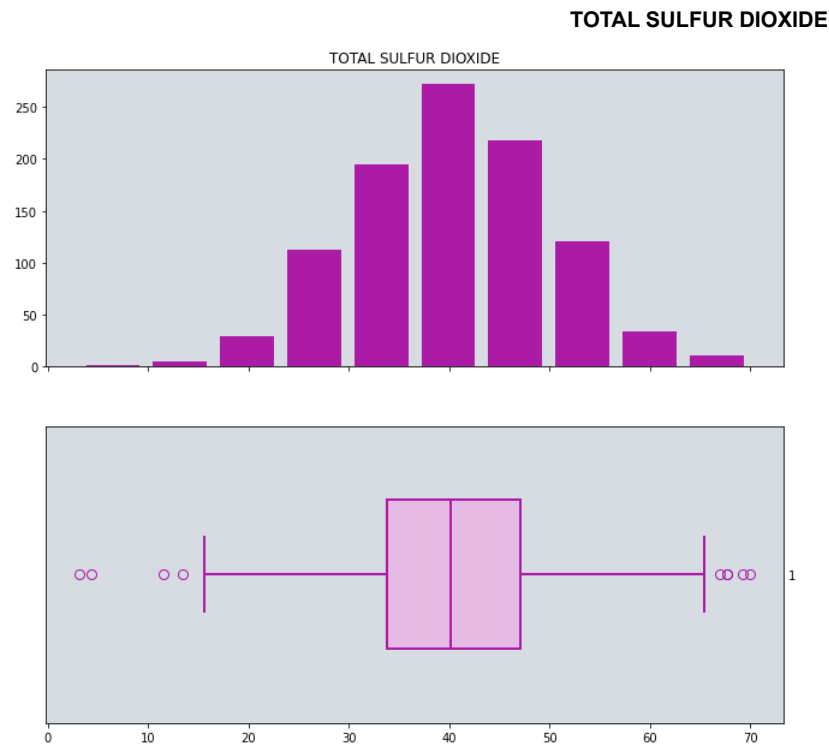
Berdasarkan histogram dan box plot yang terbentuk dapat dilihat bahwa data *Chlorides* secara umum terdistribusi secara **simetris** namun data cenderung lebih banyak terkonsentrasi pada sisi kiri pusat distribusi sehingga menyebabkan data tersebut cenderung miring ke kiri (*left-skewed*). Dapat dilihat juga bahwa hanya terdapat **sedikit outliers** dikiri nilai *minimum whisker* ataupun di kanan nilai *maximum whisker*.

In [12]: problem2(dataHeader[5])



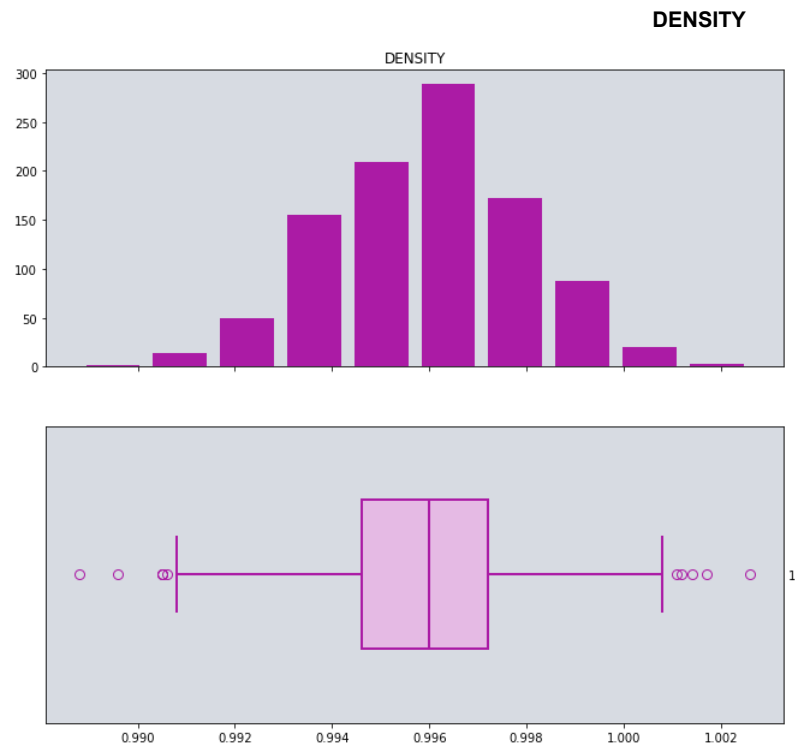
Berdasarkan histogram dan box plot yang terbentuk dapat dilihat bahwa data *Free Sulfur Dioxide* terdistribusi secara **simetris**. Walaupun konsentrasi data sedikit lebih besar pada sisi kanan pusat distribusi, namun *skew* tersebut **hampir tidak terlihat**. Dapat dilihat juga cukup sedikit *outliers* yang berada dikiri nilai *minimum whisker* dan tidak ada *outliers* yang berada dikanan nilai *maximum whisker*. Oleh karena itu, dapat dikatakan bahwa data *Free Sulfur Dioxide* **cukup konsisten** terhadap pusat distribusi.

```
In [13]: problem2(dataHeader[6])
```



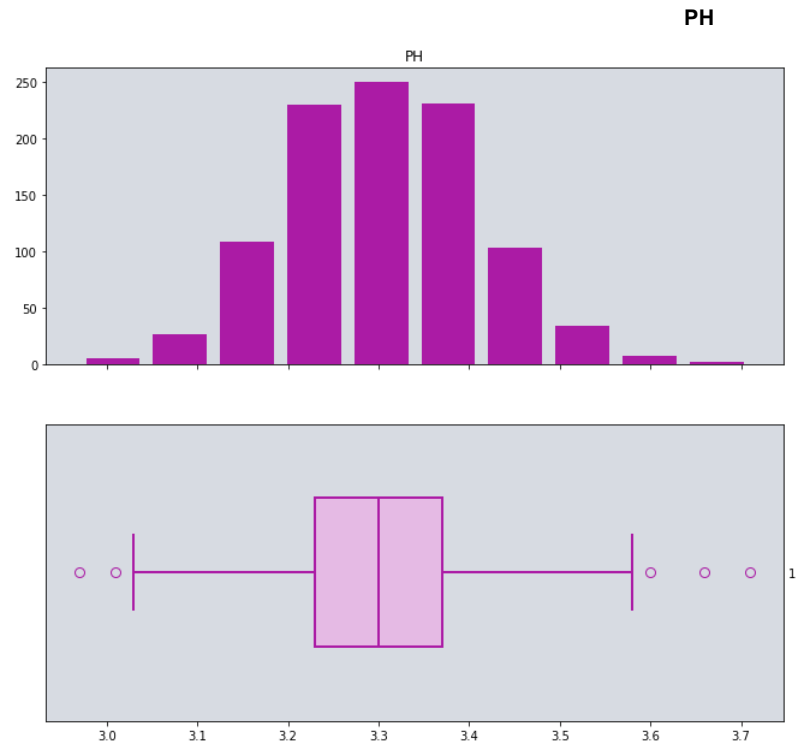
Berdasarkan histogram dan box plot yang terbentuk dapat dilihat bahwa data *Total Sulfur Dioxide* terdistribusi secara **simetris**. Walaupun konsentrasi data sedikit lebih besar pada sisi kanan pusat distribusi, namun *skew* tersebut **hampir tidak terlihat**. Dapat dilihat juga cukup sedikit *outliers* yang berada dikiri nilai *minimum whisker* dan ataupun yang berada dikanan nilai *maximum whisker*, namun *outliers* di sisi kiri berbeda cukup jauh sehingga menyebabkan memanjangnya ekor distribusi data di sisi kiri.

```
In [14]: problem2(dataHeader[7])
```



Berdasarkan histogram dan box plot yang terbentuk dapat dilihat bahwa data *Density* secara umum terdistribusi secara **simetris** namun data cenderung lebih banyak terkonsentrasi pada sisi kiri pusat distribusi, walaupun demikian *skew* tersebut **hampir tidak terlihat**. Dapat dilihat juga bahwa hanya terdapat beberapa *outliers* dikiri nilai *minimum whisker* ataupun dikanan nilai *maximum whisker*. Oleh karena itu, dapat dikatakan bahwa data *Density* **kurang konsisten** terhadap pusat distribusinya.

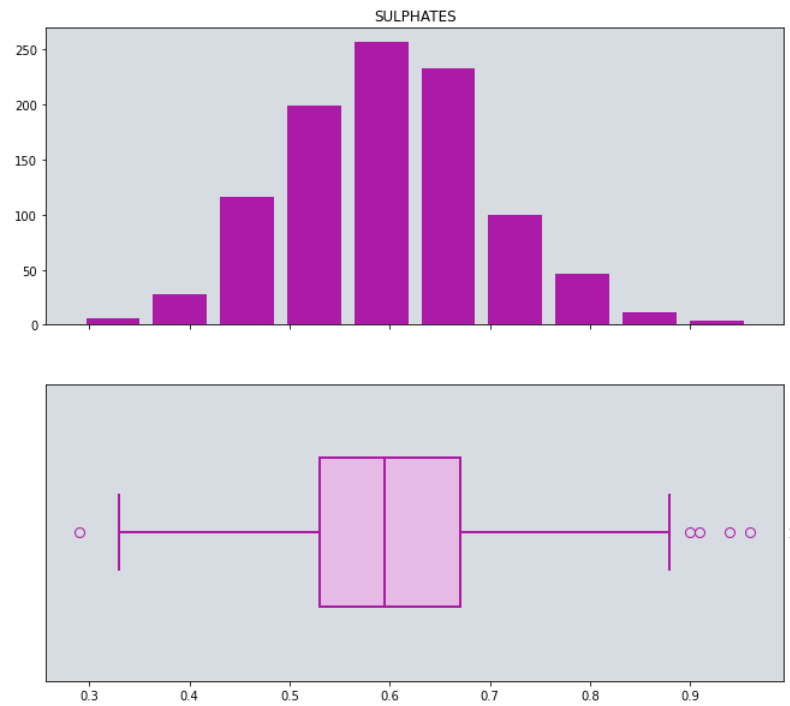
```
In [15]: problem2(dataHeader[8])
```



Berdasarkan histogram dan box plot yang terbentuk dapat dilihat bahwa data *pH* terdistribusi secara **simetris**. Walaupun konsentrasi data cukup besar pada bagian kanan pusat distribusi, namun *skew* tersebut **hampir tidak terlihat**. Dapat dilihat juga cukup sedikit *outliers* yang berada di kanan nilai *maximum whisker* ataupun yang berada di kiri nilai *minimum whisker*. Oleh karena itu, dapat dikatakan bahwa data *pH* cukup konsisten terhadap pusat distribusi walaupun memiliki ekor distribusi yang cukup panjang

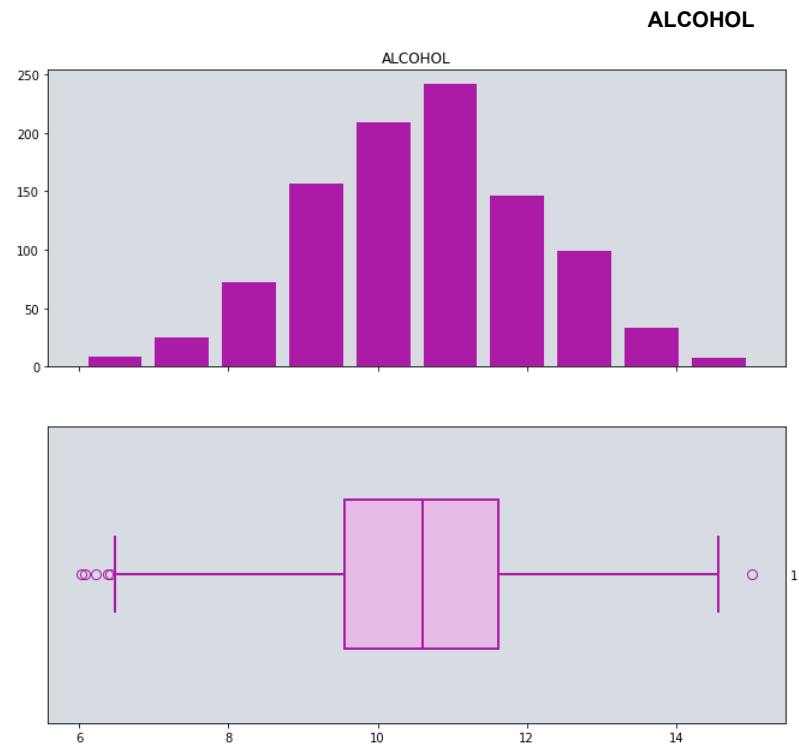
```
In [16]: problem2(dataHeader[9])
```

SULPHATES



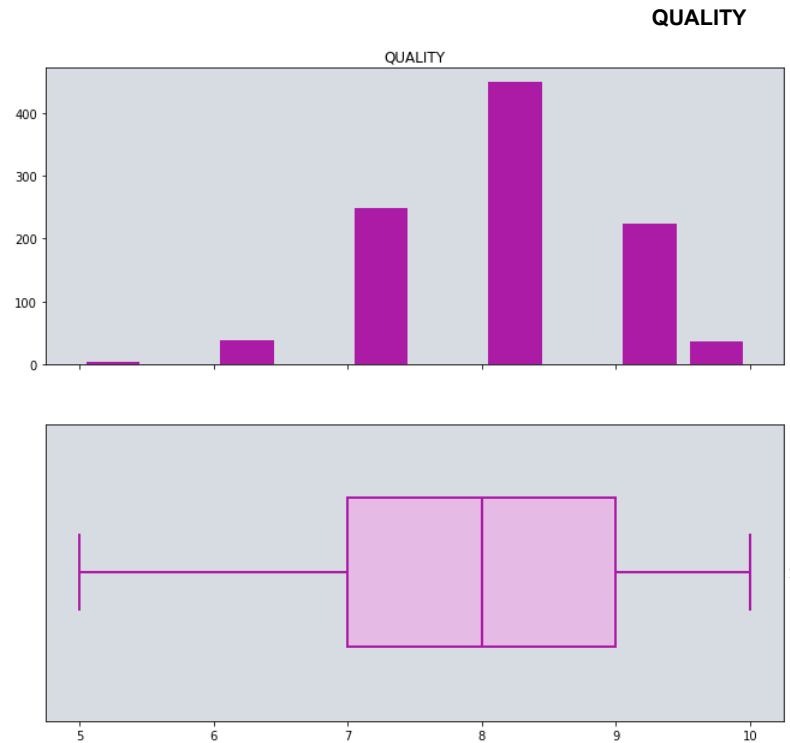
Berdasarkan histogram dan box plot yang terbentuk dapat dilihat bahwa data *Sulphates* terdistribusi secara **simetris**. Walaupun konsentrasi data cukup besar pada bagian kanan pusat distribusi, namun *skew* tersebut **hampir tidak terlihat**. Dapat dilihat juga cukup sedikit *outliers* yang berada di kanan nilai *maximum whisker* ataupun yang berada di kiri nilai *minimum whisker*. Oleh karena itu, dapat dikatakan bahwa data *Sulphates* cukup konsisten terhadap pusat distribusi.

```
In [17]: problem2(dataHeader[10])
```



Berdasarkan histogram dan box plot yang terbentuk dapat dilihat bahwa data *Alcohol* secara umum terdistribusi secara **simetris**, namun cenderung lebih banyak data yang terkonsentrasi pada sisi kiri pusat distribusi sehingga menyebabkan data tersebut cenderung miring ke kiri (*left-skewed*). Dapat dilihat juga bahwa terdapat cukup banyak *outliers* yang berada di kiri nilai *minimum whisker* yang menyebabkan panjangnya ekor distribusi di sisi kiri pusat distribusi.

```
In [18]: problem2(dataHeader[11])
```



Berdasarkan histogram dan box plot yang terbentuk dapat dilihat bahwa data *Quality* terdistribusi secara **simetris**. Walaupun konsentrasi data cukup besar pada bagian kiri pusat distribusi sehingga menyebabkan data tersebut cenderung miring ke kiri (*left-skewed*). Dapat dilihat juga bahwa tidak ada *outliers* yang berada di kanan nilai *maximum whisker* ataupun yang di kiri nilai *minimum whisker* hanya saja ekor distribusi sedikit lebih panjang di sisi kirinya.

PROBLEM 3

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

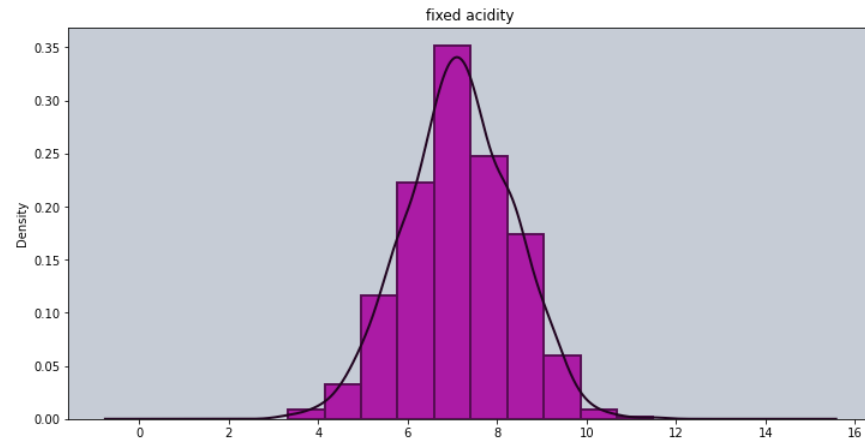
```
In [19]: import scipy.stats as stats
```

```
In [20]: def problem3(attribute):
    stat, p_val = stats.normaltest(df[attribute])
    display(Markdown("### D'Agostino's K-squared Test, $p$ = $%.10f$" % (p_val)))
    if p_val > 0.05 :
        display(Markdown("### Kolom %s Berdistribusi Normal" % (attribute)))
    else :
        display(Markdown("### Kolom %s Berdistribusi Tidak Normal" % (attribute)))
    plt.subplots()
    df[attribute].plot(kind="hist", color="#ab1ba5", linewidth=1.9, edgecolor="#550d52", bins=10, density="True").set_facecolor('#c6ccd6')
    df[attribute].plot(kind="kde", color="#220521", linewidth=2, figsize=(12, 6))
    plt.title(attribute)
    plt.show()
```

```
In [21]: problem3(dataHeader[0])
```

D'Agostino's K-squared Test, $p = 0.9308584274$

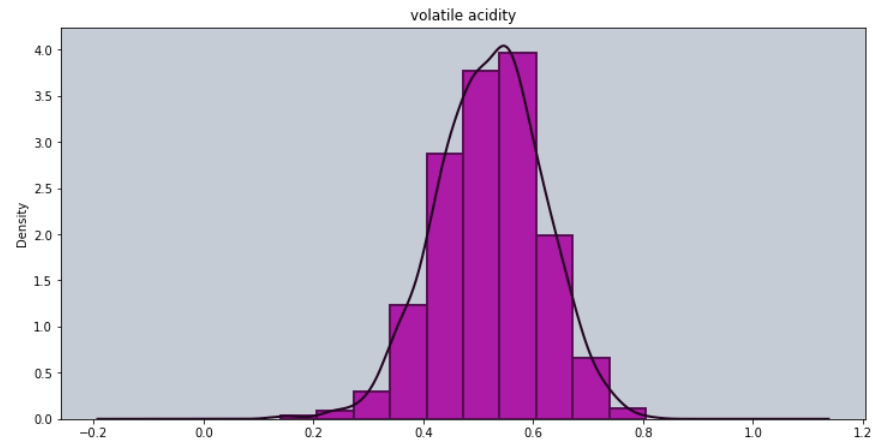
Kolom fixed acidity Berdistribusi Normal



```
In [22]: problem3(dataHeader[1])
```

D'Agostino's K-squared Test, $p = 0.0225814616$

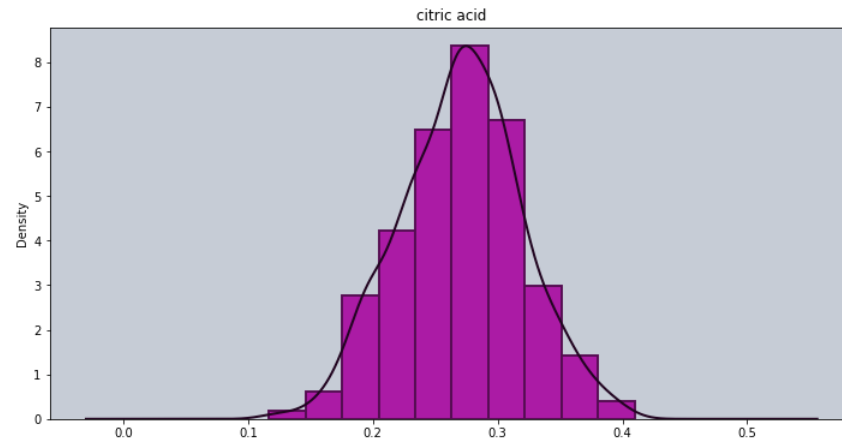
Kolom volatile acidity Berdistribusi Tidak Normal



In [23]: `problem3(dataHeader[2])`

D'Agostino's K-squared Test, $p = 0.6816899376$

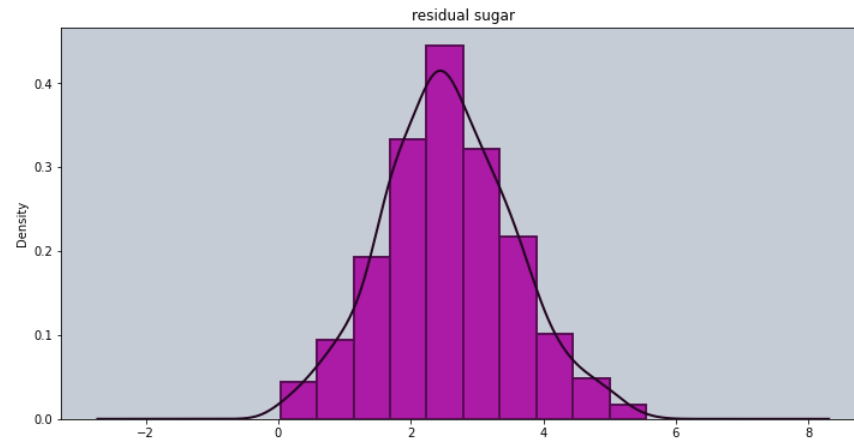
Kolom citric acid Berdistribusi Normal



In [24]: `problem3(dataHeader[3])`

D'Agostino's K-squared Test, $p = 0.2246670332$

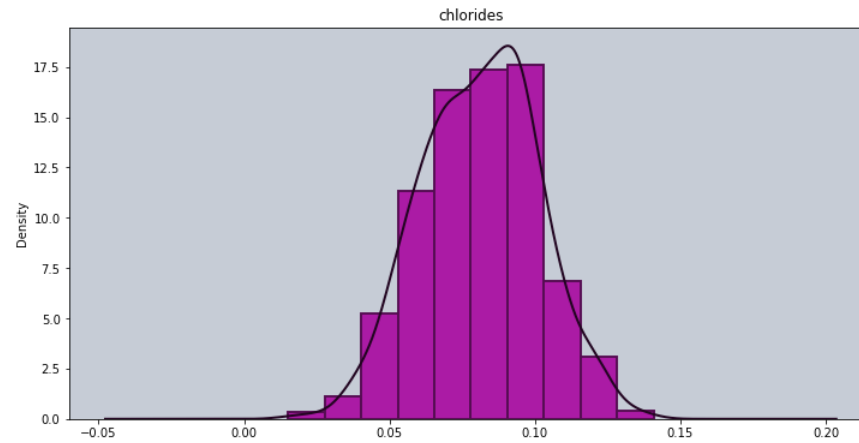
Kolom residual sugar Berdistribusi Normal



```
In [25]: problem3(dataHeader[4])
```

D'Agostino's K-squared Test, $p = 0.1704827470$

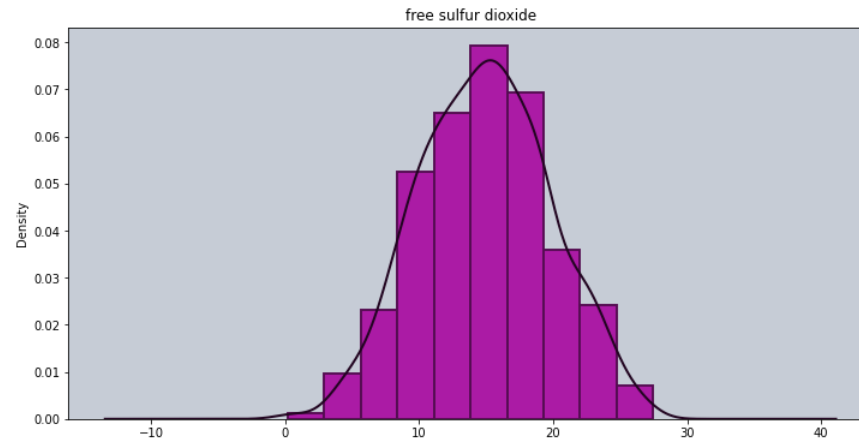
Kolom chlorides Berdistribusi Normal



```
In [26]: problem3(dataHeader[5])
```

D'Agostino's K-squared Test, $p = 0.0174304345$

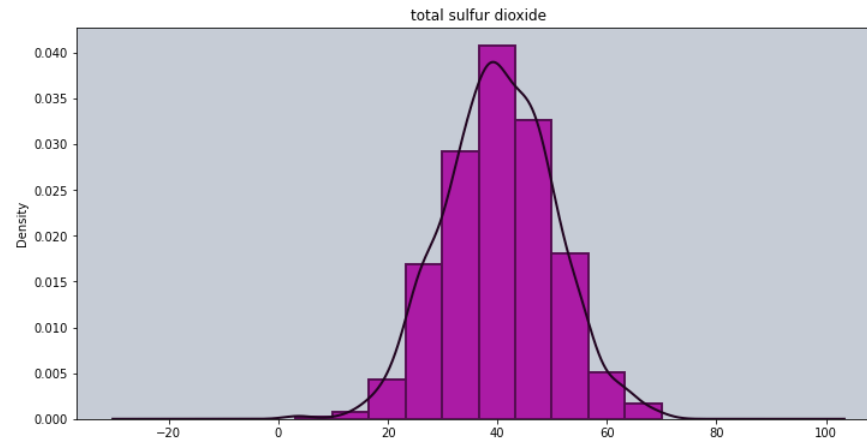
Kolom free sulfur dioxide Berdistribusi Tidak Normal



```
In [27]: problem3(dataHeader[6])
```

D'Agostino's K-squared Test, $p = 0.8488846101$

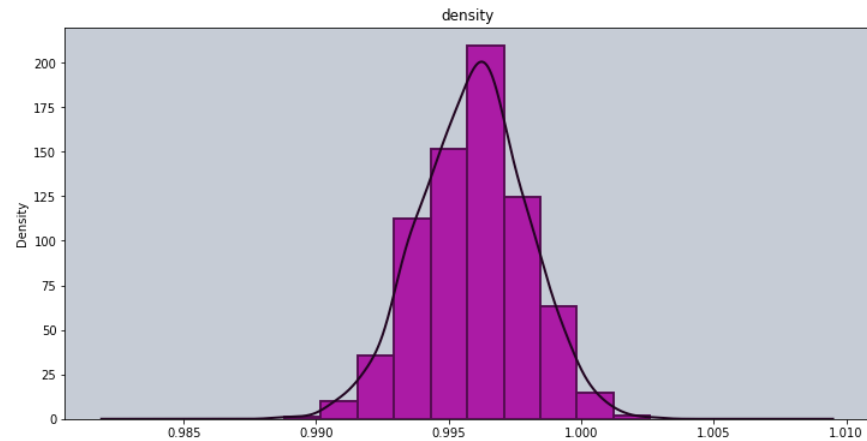
Kolom total sulfur dioxide Berdistribusi Normal



```
In [28]: problem3(dataHeader[7])
```

D'Agostino's K-squared Test, $p = 0.5985227326$

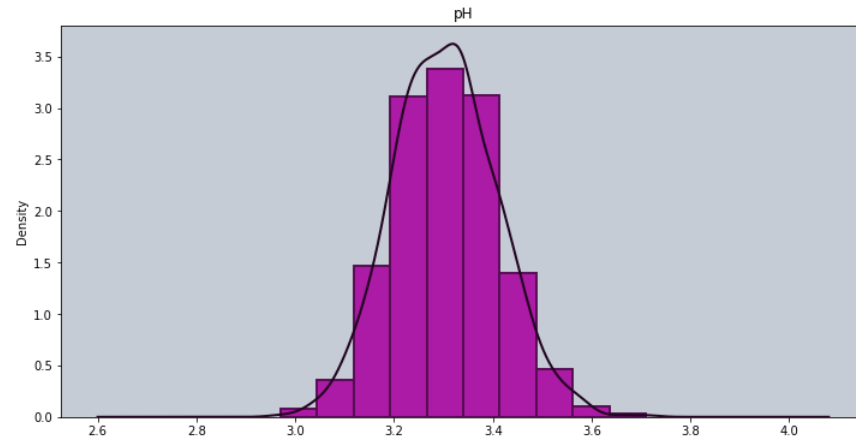
Kolom density Berdistribusi Normal



In [29]: `problem3(dataHeader[8])`

D'Agostino's K-squared Test, $p = 0.1367874082$

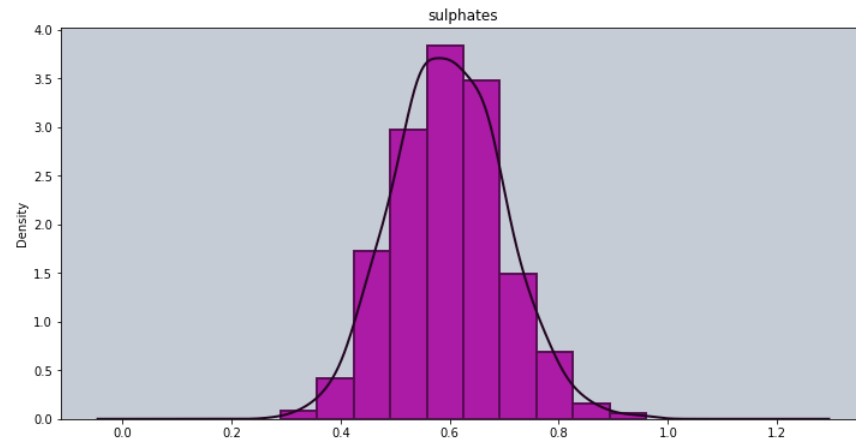
Kolom pH Berdistribusi Normal



In [30]: `problem3(dataHeader[9])`

D'Agostino's K-squared Test, $p = 0.1388431863$

Kolom sulphates Berdistribusi Normal



```
In [31]: problem3(dataHeader[10])
```

D'Agostino's K-squared Test, $p = 0.6790884901$

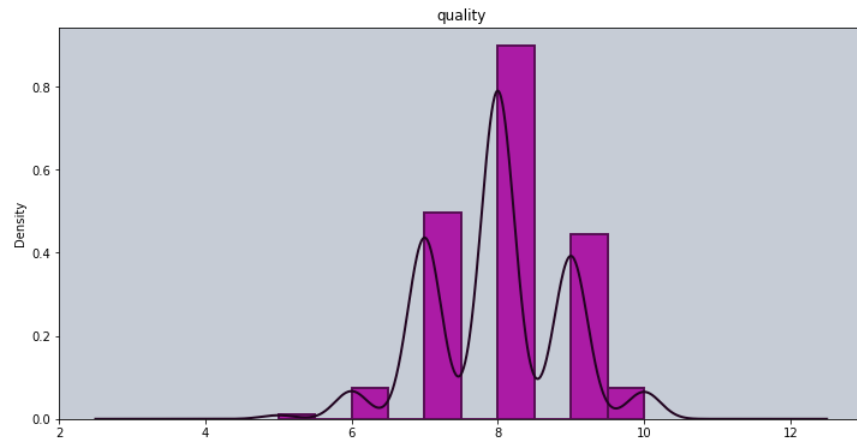
Kolom alcohol Berdistribusi Normal



```
In [32]: problem3(dataHeader[11])
```

D'Agostino's K-squared Test, $p = 0.3888139394$

Kolom quality Berdistribusi Normal



PROBLEM 4

Melakukan Test Hipotesis 1 Sample

```
In [33]: import scipy.stats as stats
from statsmodels.stats.weightstats import ztest
from statsmodels.stats.proportion import proportions_ztest
import numpy as np
```

```

In [34]: def problem4(col, val, sign, kalimat, n, variable, alpha = 0.05):
    if variable == "mean":
        variable = "μ"
    elif variable == "prop":
        variable = "p"
    if sign == "neq":
        sign = "≠"

    # STEP 1
    signStr = ""
    if sign == ">":
        signStr = "lebih dari"
    elif sign == "<":
        signStr = "kurang dari"
    elif sign == "≠":
        signStr = "tidak sama dengan"
    display(Markdown(f"### STEP 1 : Menentukan $H_0$ dan $H_1$ \n$H_0$ : {kalimat} bernilai {val} ${variable} = {val}$<br/>$H_1$ : {kalimat} {signStr} {val} ${variable} {sign} {val}$"))

    # STEP 2
    display(Markdown(f"### STEP 2 : Menentukan Level of Significance ($\alpha$) \n$\alpha$ = {alpha}"))

    # STEP 3
    test = alpha
    if sign == "≠":
        test = alpha / 2

    z_bound = stats.norm.isf(test)
    step3str = ""
    if sign == ">":
        step3str = "$z > " + str(z_bound) + "$"
    elif sign == "<":
        step3str = "$z < -" + str(z_bound) + "$"
    else :
        step3str = f"$z < {(-1 * z_bound)}$ or $z > {z_bound}$"

    display(Markdown(f"### STEP 3 : Menentukan Statistik dan Daerah Kritis\n{step3str}"))

    # STEP 4
    step4str = ""
    z_val = -1
    p_val = -1
    if variable == "μ":
        if sign == ">":
            z_val, p_val = ztest(col, value=val, alternative="larger")
        elif sign == "<":
            z_val, p_val = ztest(col, value=val, alternative="smaller")
        elif sign == "≠":
            z_val, p_val = ztest(col, value=val, alternative="two-sided")
    elif variable == "p":
        # z_val = (Len(col) / n - val) / (np.sqrt(val * (1 - val) / n))
        if sign == ">":
            z_val, p_val = proportions_ztest(len(col), len(df), value=val, alternative="larger", prop_var=val)
        elif sign == "<":
            z_val, p_val = proportions_ztest(len(col), len(df), value=val, alternative="smaller", prop_var=val)
        elif sign == "≠":
            z_val, p_val = proportions_ztest(len(col), len(df), value=val, alternative="two-sided", prop_var=val)

    step4str = f"Didapat Nilai $z$ = ${z_val}$ dan $p$ = ${p_val}$"
    display(Markdown(f"### STEP 4 : Menguji Statistik Sample\n{step4str}"))

    # STEP 5
    h0_z_status = True
    h0_p_status = True

    if sign == ">":
        h0_z_status = not (z_val > z_bound)
    elif sign == "<":
        h0_z_status = not (z_val < (-1 * z_bound))
    elif sign == "≠":
        h0_z_status = not (z_val < (-1 * z_bound) or z_val > z_bound)

    h0_p_status = not (p_val < alpha)

    step5str = f""
    if h0_z_status and h0_p_status:
        step5str = f""Tidak Menolak $H_0$ karena $z$ tidak berada pada *critical section* (*critical section* : {step3str})<br/>
        dan $p$ lebih besar atau sama dengan tingkat signifikansi yang diinginkan ($p \ge \alpha$)<br/>
        **Kesimpulan : {kalimat} bernilai {val}**""

```

```

else :
    signStr = ""
    if sign == ">":
        signStr = "lebih dari"
    elif sign == "<":
        signStr = "kurang dari"
    elif sign == "=":
        signStr = "tidak sama dengan"
    step5str = f""Tolak $H_0$ karena $z$ berada pada *critical section* (*critical section* : {step3str})<br/>
    dan $p$ lebih kecil dari tingkat signifikan yang diinginkan ($p$ < {alpha})<br/><br/>
    **Kesimpulan : {kalimat} {signStr} {val}*****"

display(Markdown(f"### STEP 5 : Pengambilan Keputusan\n{step5str}"))

```

A. Nilai rata-rata pH di atas 3.29?

```

In [35]: problem4(df["pH"], 3.29, ">", "nilai rata-rata pH", len(df), "mean")
fig, ax = plt.subplots(figsize=(7, 3))
ax.set_title("pH")
ax.boxplot(df["pH"], vert=False, widths=0.5, patch_artist=True, boxprops=dict(facecolor='e5bae4', edgecolor='ab1ba5'), capprops=dict(color='ab1ba5'), medianprops=dict(color='ab1ba5'), flierprops=dict(markeredgcolor='ab1ba5'), whiskerprops=dict(color='ab1ba5'));

```

STEP 1 : Menentukan H_0 dan H_1

H_0 : nilai rata-rata pH bernilai 3.29 ($\mu = 3.29$)

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

STEP 2 : Menentukan Level of Significance (α)

$\alpha = 0.05$

STEP 3 : Menentukan Statistik dan Daerah Kritis

$z > 1.6448536269514729$

STEP 4 : Menguji Statistik Sample

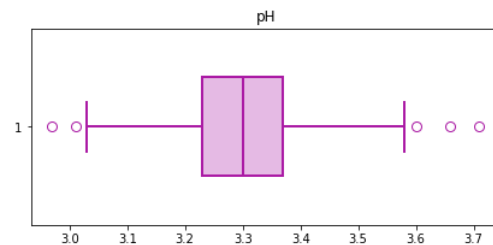
Didapat Nilai $z = 4.1037807933651145$ dan $p = 2.0322630043302333e - 05$

STEP 5 : Pengambilan Keputusan

Tolak H_0 karena z berada pada *critical section* (*critical section* : $z > 1.6448536269514729$)

dan p lebih kecil dari tingkat signifikan yang diinginkan ($p < 0.05$)

Kesimpulan : nilai rata-rata pH lebih dari 3.29



B. Nilai rata-rata Residual Sugar tidak sama dengan 2.50 ?

```
In [36]: problem4(df["residual sugar"], 2.5, "neq", "nilai rata-rata Residual Sugar", len(df), "mean")
fig, ax = plt.subplots(figsize=(7, 3))
ax.set_title("Residual Sugar")
ax.boxplot(df["residual sugar"], vert=False, widths=0.5, patch_artist=True, boxprops=dict(facecolor='#e5bae4', edgecolor='#ab1ba5'), capprops=dict(color='#ab1ba5'), medianprops=dict(color='#ab1ba5'), flierprops=dict(markeredgecolor='#ab1ba5'), whiskerprops=dict(color='#ab1ba5'));
```

STEP 1 : Menentukan H_0 dan H_1

H_0 : nilai rata-rata Residual Sugar bernilai 2.5 ($\mu = 2.5$)

H_1 : nilai rata-rata Residual Sugar tidak sama dengan 2.5 ($\mu \neq 2.5$)

STEP 2 : Menentukan Level of Significance (α)

$\alpha = 0.05$

STEP 3 : Menentukan Statistik dan Daerah Kritis

$z < -1.9599639845400545$ or $z > 1.9599639845400545$

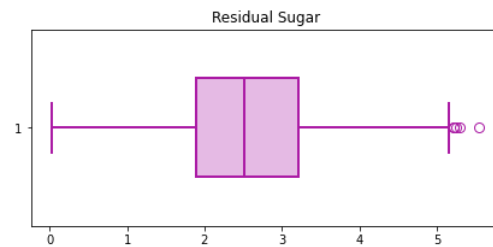
STEP 4 : Menguji Statistik Sample

Didapat Nilai $z = 2.1479619435539523$ dan $p = 0.031716778818727434$

STEP 5 : Pengambilan Keputusan

Tolak H_0 karena z berada pada *critical section* (*critical section* : $z < -1.9599639845400545$ or $z > 1.9599639845400545$) dan p lebih kecil dari tingkat signifikan yang diinginkan ($p < 0.05$)

Kesimpulan : nilai rata-rata Residual Sugar tidak sama dengan 2.5



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

```
In [37]: problem4(df["sulphates"][:150], 0.65, "neq", "nilai rata-rata 150 baris pertama kolom Sulphates", len(df["sulphates"][:150]), "mean")
fig, ax = plt.subplots(figsize=(7, 3))
ax.set_title("150 baris pertama kolom Sulphates")
ax.boxplot(df["sulphates"][:150], vert=False, widths=0.5, patch_artist=True, boxprops=dict(facecolor='#e5bae4', edgecolor='#ab1ba5'), capprops=dict(color='#ab1ba5'), medianprops=dict(color='#ab1ba5'), flierprops=dict(markeredgecolor='#ab1ba5'), whiskerprops=dict(color='#ab1ba5'));
```

STEP 1 : Menentukan H_0 dan H_1

H_0 : nilai rata-rata 150 baris pertama kolom Sulphates bernilai 0.65 ($\mu = 0.65$)

H_1 : nilai rata-rata 150 baris pertama kolom Sulphates tidak sama dengan 0.65 ($\mu \neq 0.65$)

STEP 2 : Menentukan Level of Significance (α)

$\alpha = 0.05$

STEP 3 : Menentukan Statistik dan Daerah Kritis

$z < -1.9599639845400545$ or $z > 1.9599639845400545$

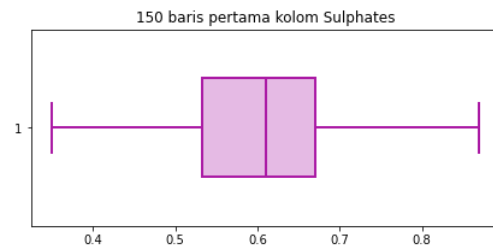
STEP 4 : Menguji Statistik Sample

Didapat Nilai $z = -4.964843393315918$ dan $p = 6.875652918327357e - 07$

STEP 5 : Pengambilan Keputusan

Tolak H_0 karena z berada pada *critical section* (*critical section* : $z < -1.9599639845400545$ or $z > 1.9599639845400545$) dan p lebih kecil dari tingkat signifikan yang diinginkan ($p < 0.05$)

Kesimpulan : nilai rata-rata 150 baris pertama kolom Sulphates tidak sama dengan 0.65



D. Nilai rata-rata total sulfur dioxide di bawah 35?

```
In [38]: problem4(df["total sulfur dioxide"], 35, "<", "nilai rata-rata total sulfur dioxide", len(df["total sulfur dioxide"]), "mean")
fig, ax = plt.subplots(figsize=(7, 3))
ax.set_title("total sulfur dioxide")
ax.boxplot(df["total sulfur dioxide"], vert=False, widths=0.5, patch_artist=True, boxprops=dict(facecolor='#e5bae4', edgecolor='#ab1ba5'), capprops=dict(color='#ab1ba5'), medianprops=dict(color='#ab1ba5'), flierprops=dict(markeredgecolor='#ab1ba5'), whiskerprops=dict(color='#ab1ba5'));
```

STEP 1 : Menentukan H_0 dan H_1

H_0 : nilai rata-rata total sulfur dioxide bernilai 35 ($\mu = 35$)

H_1 : nilai rata-rata total sulfur dioxide kurang dari 35 ($\mu < 35$)

STEP 2 : Menentukan Level of Significance (α)

$\alpha = 0.05$

STEP 3 : Menentukan Statistik dan Daerah Kritis

$z < -1.6448536269514729$

STEP 4 : Menguji Statistik Sample

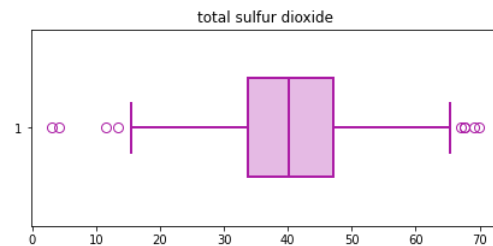
Didapat Nilai $z = 16.786387372296744$ dan $p = 1.0$

STEP 5 : Pengambilan Keputusan

Tidak Menolak H_0 karena z tidak berada pada *critical section* (*critical section* : $z < -1.6448536269514729$)

dan p lebih besar atau sama dengan tingkat signifikansi yang diinginkan ($p \geq 0.05$)

Kesimpulan : nilai rata-rata total sulfur dioxide bernilai 35



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

```
In [39]: problem4(df[df["total sulfur dioxide"] > 40], 0.5, "neq", "Proporsi nilai total sulfur dioxide yang lebih dari 40", len(df), "prop")
fig, ax = plt.subplots(figsize=(7, 3))
ax.set_title("total sulfur dioxide")
ax.boxplot(df["total sulfur dioxide"], vert=False, widths=0.5, patch_artist=True, boxprops=dict(facecolor='#e5bae4', edgecolor='#ab1ba5'), capprops=dict(color='#ab1ba5'), medianprops=dict(color='#ab1ba5'), flierprops=dict(markeredgecolor='#ab1ba5'), whiskerprops=dict(color='#ab1ba5'));
```

STEP 1 : Menentukan H_0 dan H_1

H_0 : Proporsi nilai total sulfur dioxide yang lebih dari 40 bernilai 0.5 ($p = 0.5$)

H_1 : Proporsi nilai total sulfur dioxide yang lebih dari 40 tidak sama dengan 0.5 ($p \neq 0.5$)

STEP 2 : Menentukan Level of Significance (α)

$\alpha = 0.05$

STEP 3 : Menentukan Statistik dan Daerah Kritis

$z < -1.9599639845400545$ or $z > 1.9599639845400545$

STEP 4 : Menguji Statistik Sample

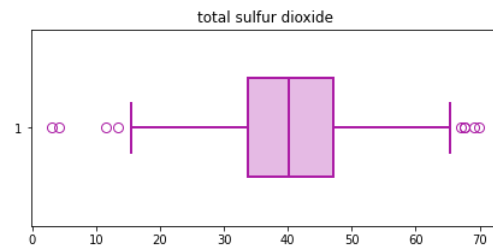
Didapat Nilai $z = 0.7589466384404118$ dan $p = 0.4478844782641115$

STEP 5 : Pengambilan Keputusan

Tidak Menolak H_0 karena z tidak berada pada *critical section* (*critical section* : $z < -1.9599639845400545$ or $z > 1.9599639845400545$)

dan p lebih besar atau sama dengan tingkat signifikan yang diinginkan ($p \geq 0.05$)

Kesimpulan : Proporsi nilai total sulfur dioxide yang lebih dari 40 bernilai 0.5



PROBLEM 5

Test Hipotesis 2 Sampel

```
In [40]: def problem5(col1, col2, val, sign, kalimat1, kalimat2, variable, alpha = 0.05):
    if variable == "mean":
        variable = "μ"
    elif variable == "prop":
        variable = "p"
    elif variable == "var":
        variable = "σ"
    if sign == "neq":
        sign = "≠"

    # STEP 1
    step1str = ""
    if val != 0 :
        step1str = f"$H_0$ : {kalimat1} dikurangi {kalimat2} bernilai {val} (${variable}_1$ - ${variable}_2$ $= {val}$)"
    else :
        step1str = f"$H_0$ : {kalimat1} sama dengan {kalimat2} / {kalimat1} dikurangi {kalimat2} bernilai 0 (${variable}_1$ - ${variable}_2$ $= 0$)"

    if sign == ">":
        step1str += f"<br/> $H_1$ : {kalimat1} dikurangi {kalimat2} lebih besar dari {val} (${variable}_1$ - ${variable}_2$ $> {val}$)"
    elif sign == "<":
```

```

step1str += f"<br/> $H_1$ : {kalimat1} dikurangi {kalimat2} lebih kecil dari {val} ({variable}_1$ - {variable}_2$ < {val})$"
elif sign == "#":
    step1str += f"<br/> $H_1$ : {kalimat1} dikurangi {kalimat2} tidak sama dengan {val} ({variable}_1$ - {variable}_2$ $≠ {val})$"

display(Markdown(f"### STEP 1 : Menentukan $H_0$ dan $H_1$\n{step1str}"))

# STEP 2
display(Markdown(f"### STEP 2 : Menentukan Level of Significance ($α$)\n$α$ = {alpha}"))

# STEP 3
step3str = ""
uji_upper = -1
uji_lower = -1
if variable == "μ":
    uji_upper = stats.norm.isf(alpha/2)
    uji_lower = -1 * stats.norm.isf(alpha/2)
    step3str = f""Uji Statistik menggunakan Z-Test, dengan *critical section* : <br/>
    $z < {uji_lower}$ or $z > {uji_upper}$""
elif variable == "σ":
    uji_upper = stats.f.ppf(1 - alpha, len(col1)-1, len(col2)-1)
    uji_lower = stats.f.ppf(alpha, len(col1)-1, len(col2)-1)
    step3str = f""Uji Statistik menggunakan F-Test, dengan *critical section* : <br/>
    $f < {uji_lower}$ or $f > {uji_upper}$""
elif variable == "p":
    uji_upper = stats.norm.isf(alpha)
    step3str = f""Uji Statistik menggunakan Z-Test, dengan *critical section* : <br/>
    $z > {uji_upper}$""
display(Markdown(f"### STEP 3 : Menentukan Statistik dan Daerah Kritis\n{step3str}"))

# STEP 4
z_val = -1
f_val = -1
p_val = -1
if variable == "μ":
    z_val, p_val = ztest(col1, col2, value=val, alternative="two-sided")
    step4str = f"Didapat Nilai $z$ = ${z_val[0]}$ dan $p$ = ${p_val[0]}$"
elif variable == "p":
    z_val, p_val = proportions_ztest([len(col1[col1.iloc[:,0] > 7]), len(col2[col2.iloc[:,0] > 7])], [len(col1), len(col2)], value=val, alternativ
e="larger")
    # print (z_val, p_val)
    step4str = f"Didapat Nilai $z$ = ${z_val}$ dan $p$ = ${p_val}$"
elif variable == "σ":
    f_val = col1.var()/col2.var()
    p_val = 1 - stats.f.cdf(f_val, len(col1)-1, len(col2)-1)
    step4str = f"Didapat Nilai $f$ = ${f_val[0]}$ dan $p$ = ${p_val[0]}$"

display(Markdown(f"### STEP 4 : Menguji Statistik Sample\n{step4str}"))

# STEP 5
h0_zf_status = True
h0_p_status = True

if variable == "μ":
    if sign == ">":
        h0_zf_status = not (z_val > uji_upper)
    elif sign == "<":
        h0_zf_status = not (z_val < uji_lower)
    elif sign == "#":
        h0_zf_status = not (z_val < uji_lower or z_val > uji_upper)
elif variable == "p":
    if sign == ">":
        h0_zf_status = not (z_val > uji_upper)
elif variable == "σ":
    if sign == ">":
        h0_zf_status = not (f_val[0] > uji_upper)
    elif sign == "<":
        h0_zf_status = not (f_val[0] < uji_lower)
    elif sign == "#":
        h0_zf_status = not (f_val[0] < uji_lower or f_val[0] > uji_upper)

if variable == "p":
    h0_p_status = not (p_val < alpha)
else :
    h0_p_status = not (p_val[0] < alpha)

step5str = f""
if h0_zf_status and h0_p_status:
    symb = ""
    crit = ""
    if variable == "μ":

```

```

        crit = f"$z < {uji_lower}$ or $z > {uji_upper}$"
        symb = "z"
    elif variable == "p":
        crit = f"$z > {uji_upper}$"
        symb = "z"
    else:
        crit = f"$f < {uji_lower}$ or $f > {uji_upper}$"
        symb = "f"
    step5str = f""""Tidak cukup bukti sehingga Tidak Menolak $H_0$, karena ${symb}$ tidak berada pada *critical section* (*critical section* : {crit})<br/>
    dan $p$ lebih besar atau sama dengan tingkat signifikan yang diinginkan ($p \ge \alpha$)<br/>
    **Kesimpulan : {kalimat1} dikurangi {kalimat2} bernilai {val}*****

else :
    signStr = ""
    if sign == ">":
        signStr = "lebih dari"
    elif sign == "<":
        signStr = "kurang dari"
    elif sign == "=":
        signStr = "tidak sama dengan"

    symb = ""
    crit = ""
    if variable == "μ":
        crit = f"$z < {uji_lower}$ or $z > {uji_upper}$"
        symb = "z"
    elif variable == "p":
        crit = f"$z > {uji_upper}$"
        symb = "z"
    else:
        crit = f"$f < {uji_lower}$ or $f > {uji_upper}$"
        symb = "f"

    step5str = f""""Tolak $H_0$ karena $z$ berada pada *critical section* (*critical section* : {crit})<br/>
    dan $p$ lebih kecil dari tingkat signifikan yang diinginkan ($p < \alpha$)<br/><br/>
    **Kesimpulan : {kalimat1} dikurangi {kalimat2} {signStr} {val}*****

display(Markdown(f"### STEP 5 : Pengambilan Keputusan\n{step5str}"))

```

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

```
In [41]: problem5(df[["fixed acidity"]].iloc[:round(len(df)/2)],df[["fixed acidity"]].iloc[round(len(df)/2) :], 0, "neq", "Rata - rata bagian awal kolom fixed acidity", "Rata - rata bagian akhir kolom fixed acidity", "mean")
fig, ax = plt.subplots(figsize=(7, 3))
ax.set_title("fixed acidity bagian awal")
ax.boxplot(df[["fixed acidity"]][:len(df) // 2], vert=False, widths=0.5, patch_artist=True, boxprops=dict(facecolor='e5bae4', edgecolor='ab1ba5'), capprops=dict(color='ab1ba5'), medianprops=dict(color='ab1ba5'), flierprops=dict(markeredgecolor='ab1ba5'), whiskerprops=dict(color='ab1ba5'));
fig, ax = plt.subplots(figsize=(7, 3))
ax.set_title("fixed acidity bagian akhir")
ax.boxplot(df[["fixed acidity"]][len(df) // 2:], vert=False, widths=0.5, patch_artist=True, boxprops=dict(facecolor='e5bae4', edgecolor='ab1ba5'), capprops=dict(color='ab1ba5'), medianprops=dict(color='ab1ba5'), flierprops=dict(markeredgecolor='ab1ba5'), whiskerprops=dict(color='ab1ba5'));
```

STEP 1 : Menentukan H_0 dan H_1

H_0 : Rata - rata bagian awal kolom fixed acidity sama dengan Rata - rata bagian akhir kolom fixed acidity / Rata - rata bagian awal kolom fixed acidity dikurangi Rata - rata bagian akhir kolom fixed acidity bernilai 0 ($\mu_1 - \mu_2 = 0$)

H_1 : Rata - rata bagian awal kolom fixed acidity dikurangi Rata - rata bagian akhir kolom fixed acidity tidak sama dengan 0 ($\mu_1 - \mu_2 \neq 0$)

STEP 2 : Menentukan Level of Significance (α)

$\alpha = 0.05$

STEP 3 : Menentukan Statistik dan Daerah Kritis

Uji Statistik menggunakan Z-Test, dengan *critical section* :

$z < -1.9599639845400545$ or $z > 1.9599639845400545$

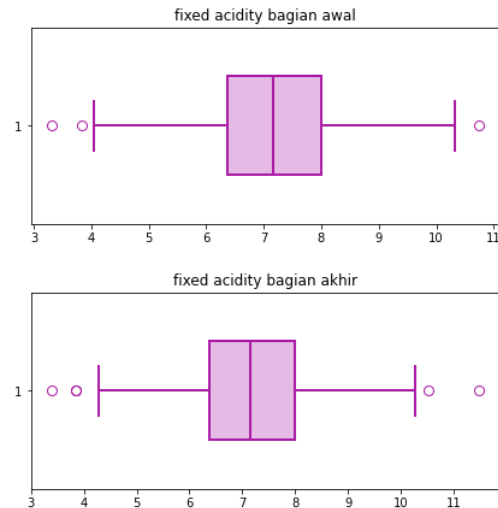
STEP 4 : Menguji Statistik Sample

Didapat Nilai $z = 0.02604106999906379$ dan $p = 0.9792245804254097$

STEP 5 : Pengambilan Keputusan

Tidak cukup bukti sehingga Tidak Menolak H_0 , karena z tidak berada pada *critical section* (*critical section* : $z < -1.9599639845400545$ or $z > 1.9599639845400545$) dan p lebih besar atau sama dengan tingkat signifikan yang diinginkan ($p \geq 0.05$)

Kesimpulan : Rata - rata bagian awal kolom fixed acidity dikurangi Rata - rata bagian akhir kolom fixed acidity bernilai 0



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 [42]: problem5(df[["chlorides"]].iloc[:round(len(df)/2)],df[["chlorides"]].iloc[round(len(df)/2) :], 0.001, ">", "Rata - rata bagian awal kolom chloride  
s","Rata - rata bagian akhir kolom chlorides", "mean")  
fig, ax = plt.subplots(figsize=(7, 3))  
ax.set_title("chlorides bagian awal")  
ax.boxplot(df[["chlorides"]][:len(df) // 2], vert=False, widths=0.5, patch_artist=True, boxprops=dict(facecolor='#e5bae4', edgecolor='#ab1ba5'), capprop  
s=dict(color='#ab1ba5'), medianprops=dict(color='#ab1ba5'), flierprops=dict(markeredgecolor='#ab1ba5'), whiskerprops=dict(color='#ab1ba5'));  
fig, ax = plt.subplots(figsize=(7, 3))  
ax.set_title("chlorides bagian akhir")  
ax.boxplot(df[["chlorides"]][len(df) // 2:], vert=False, widths=0.5, patch_artist=True, boxprops=dict(facecolor='#e5bae4', edgecolor='#ab1ba5'), capprop  
s=dict(color='#ab1ba5'), medianprops=dict(color='#ab1ba5'), flierprops=dict(markeredgecolor='#ab1ba5'), whiskerprops=dict(color='#ab1ba5'));
```

STEP 1 : Menentukan H_0 dan H_1

H_0 : Rata - rata bagian awal kolom chlorides dikurangi Rata - rata bagian akhir kolom chlorides bernilai 0.001 ($\mu_1 - \mu_2 = 0.001$)

H_1 : Rata - rata bagian awal kolom chlorides dikurangi Rata - rata bagian akhir kolom chlorides lebih besar dari 0.001 ($\mu_1 - \mu_2 > 0.001$)

STEP 2 : Menentukan Level of Significance (α)

$\alpha = 0.05$

STEP 3 : Menentukan Statistik dan Daerah Kritis

Uji Statistik menggunakan Z-Test, dengan *critical section* :

$z < -1.9599639845400545$ or $z > 1.9599639845400545$

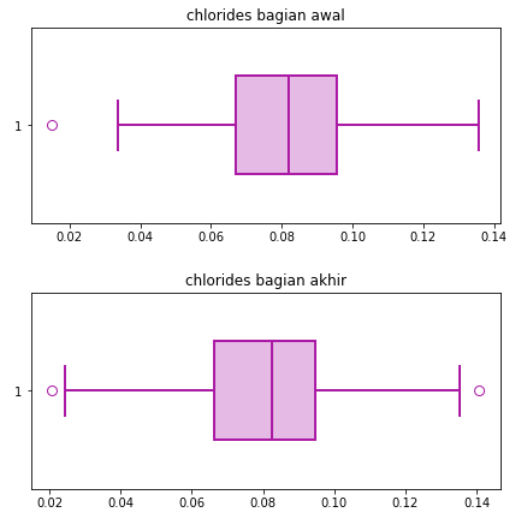
STEP 4 : Menguji Statistik Sample

Didapat Nilai $z = -0.467317122852132$ dan $p = 0.640273007581107$

STEP 5 : Pengambilan Keputusan

Tidak cukup bukti sehingga Tidak Menolak H_0 , karena z tidak berada pada *critical section* (*critical section* : $z < -1.9599639845400545$ or $z > 1.9599639845400545$) dan p lebih besar atau sama dengan tingkat signifikan yang diinginkan ($p \geq 0.05$)

Kesimpulan : Rata - rata bagian awal kolom chlorides dikurangi Rata - rata bagian akhir kolom chlorides bernilai 0.001



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

```
In [43]: problem5(df[["volatile acidity"]].iloc[:25],df[["sulphates"]].iloc[:25], 0, "neq", "Rata - rata 25 baris pertama kolom volatile acidity", "Rata - rata 25 baris pertama kolom sulphates", "mean")
fig, ax = plt.subplots(figsize=(7, 3))
ax.set_title("volatile acidity 25 baris pertama")
ax.boxplot(df[["volatile acidity"]][:25], vert=False, widths=0.5, patch_artist=True, boxprops=dict(facecolor='e5bae4', edgecolor='ab1ba5'), capprops=dict(color='ab1ba5'), medianprops=dict(color='ab1ba5'), flierprops=dict(markeredgecolor='ab1ba5'), whiskerprops=dict(color='ab1ba5'));
fig, ax = plt.subplots(figsize=(7, 3))
ax.set_title("sulphates 25 baris pertama")
ax.boxplot(df[["sulphates"]][:25], vert=False, widths=0.5, patch_artist=True, boxprops=dict(facecolor='e5bae4', edgecolor='ab1ba5'), capprops=dict(color='ab1ba5'), medianprops=dict(color='ab1ba5'), flierprops=dict(markeredgecolor='ab1ba5'), whiskerprops=dict(color='ab1ba5'));
```

STEP 1 : Menentukan H_0 dan H_1

H_0 : Rata - rata 25 baris pertama kolom volatile acidity sama dengan Rata - rata 25 baris pertama kolom sulphates / Rata - rata 25 baris pertama kolom volatile acidity dikurangi Rata - rata 25 baris pertama kolom sulphates bernilai 0 ($\mu_1 - \mu_2 = 0$)

H_1 : Rata - rata 25 baris pertama kolom volatile acidity dikurangi Rata - rata 25 baris pertama kolom sulphates tidak sama dengan 0 ($\mu_1 - \mu_2 \neq 0$)

STEP 2 : Menentukan Level of Significance (α)

$\alpha = 0.05$

STEP 3 : Menentukan Statistik dan Daerah Kritis

Uji Statistik menggunakan Z-Test, dengan *critical section* :

$z < -1.9599639845400545$ or $z > 1.9599639845400545$

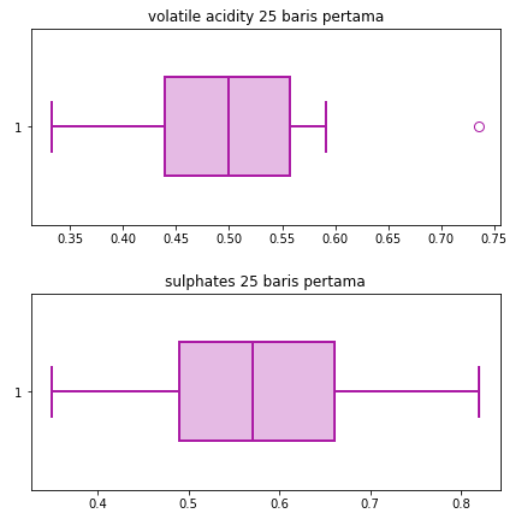
STEP 4 : Menguji Statistik Sample

Didapat Nilai $z = -2.6374821676748703$ dan $p = 0.008352401685453743$

STEP 5 : Pengambilan Keputusan

Tolak H_0 karena z berada pada *critical section* (*critical section* : $z < -1.9599639845400545$ or $z > 1.9599639845400545$) dan p lebih kecil dari tingkat signifikan yang diinginkan ($p < 0.05$)

Kesimpulan : Rata - rata 25 baris pertama kolom volatile acidity dikurangi Rata - rata 25 baris pertama kolom sulphates tidak sama dengan 0



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

```
In [44]: problem5(df[["residual sugar"]].iloc[:round(len(df)/2)],df[["residual sugar"]].iloc[round(len(df)/2) :], 0, "neq", "varians bagian awal kolom residual sugar", "varians bagian akhir kolom residual sugar", "var")
fig, ax = plt.subplots(figsize=(7, 3))
ax.set_title("Bagian awal kolom residual sugar")
ax.boxplot(df[["residual sugar"]][:len(df)//2], vert=False, widths=0.5, patch_artist=True, boxprops=dict(facecolor='e5bae4', edgecolor='ab1ba5'), capprops=dict(color='ab1ba5'), medianprops=dict(color='ab1ba5'), flierprops=dict(markeredgecolor='ab1ba5'), whiskerprops=dict(color='ab1ba5'));
fig, ax = plt.subplots(figsize=(7, 3))
ax.set_title("Bagian akhir kolom residual sugar")
ax.boxplot(df[["residual sugar"]][len(df)//2:], vert=False, widths=0.5, patch_artist=True, boxprops=dict(facecolor='e5bae4', edgecolor='ab1ba5'), capprops=dict(color='ab1ba5'), medianprops=dict(color='ab1ba5'), flierprops=dict(markeredgecolor='ab1ba5'), whiskerprops=dict(color='ab1ba5'));
```

STEP 1 : Menentukan H_0 dan H_1

H_0 : varians bagian awal kolom residual sugar sama dengan varians bagian akhir kolom residual sugar / varians bagian awal kolom residual sugar dikurangi varians bagian akhir kolom residual sugar bernilai 0 ($\sigma_1 - \sigma_2 = 0$)

H_1 : varians bagian awal kolom residual sugar dikurangi varians bagian akhir kolom residual sugar tidak sama dengan 0 ($\sigma_1 - \sigma_2 \neq 0$)

STEP 2 : Menentukan Level of Significance (α)

$\alpha = 0.05$

STEP 3 : Menentukan Statistik dan Daerah Kritis

Uji Statistik menggunakan F-Test, dengan *critical section* :

$f < 0.8629418796107595$ or $f > 1.1588265949626437$

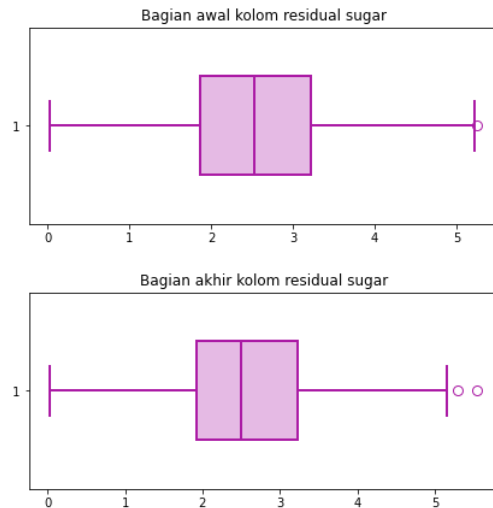
STEP 4 : Menguji Statistik Sample

Didapat Nilai $f = 0.9420041066941615$ dan $p = 0.7475898202376912$

STEP 5 : Pengambilan Keputusan

Tidak cukup bukti sehingga Tidak Menolak H_0 , karena f tidak berada pada *critical section* (*critical section* : $f < 0.8629418796107595$ or $f > 1.1588265949626437$) dan p lebih besar atau sama dengan tingkat signifikan yang diinginkan ($p \geq 0.05$)

Kesimpulan : varians bagian awal kolom residual sugar dikurangi varians bagian akhir kolom residual sugar bernilai 0



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 [45]: problem5(df[["alcohol"]].iloc[:round(len(df)/2)], df[["alcohol"]].iloc[round(len(df)/2):], 0, ">", "Proporsi bagian awal kolom alcohol", "Proporsi bagian akhir kolom alcohol", "prop")
fig, ax = plt.subplots(figsize=(7, 3))
ax.set_title("alcohol bagian awal")
ax.boxplot(df[["alcohol"]][:len(df)//2], vert=False, widths=0.5, patch_artist=True, boxprops=dict(facecolor='e5bae4', edgecolor='ab1ba5'), capprops=dict(color='ab1ba5'), medianprops=dict(color='ab1ba5'), flierprops=dict(markeredgecolor='ab1ba5'), whiskerprops=dict(color='ab1ba5'));
fig, ax = plt.subplots(figsize=(7, 3))
ax.set_title("alcohol bagian akhir")
ax.boxplot(df[["alcohol"]][len(df)//2:], vert=False, widths=0.5, patch_artist=True, boxprops=dict(facecolor='e5bae4', edgecolor='ab1ba5'), capprops=dict(color='ab1ba5'), medianprops=dict(color='ab1ba5'), flierprops=dict(markeredgecolor='ab1ba5'), whiskerprops=dict(color='ab1ba5'));
```

STEP 1 : Menentukan H_0 dan H_1

H_0 : Proporsi bagian awal kolom alcohol sama dengan Proporsi bagian akhir kolom alcohol / Proporsi bagian awal kolom alcohol dikurangi Proporsi bagian akhir kolom alcohol bernilai 0 ($p_1 - p_2 = 0$)

H_1 : Proporsi bagian awal kolom alcohol dikurangi Proporsi bagian akhir kolom alcohol lebih besar dari 0 ($p_1 - p_2 > 0$)

STEP 2 : Menentukan Level of Significance (α)

$\alpha = 0.05$

STEP 3 : Menentukan Statistik dan Daerah Kritis

Uji Statistik menggunakan Z-Test, dengan *critical section* :

$z > 1.6448536269514729$

STEP 4 : Menguji Statistik Sample

Didapat Nilai $z = 0.0$ dan $p = 0.5$

STEP 5 : Pengambilan Keputusan

Tidak cukup bukti sehingga Tidak Menolak H_0 , karena z tidak berada pada *critical section* (*critical section* : $z > 1.6448536269514729$) dan p lebih besar atau sama dengan tingkat signifikan yang diinginkan ($p \geq 0.05$)

Kesimpulan : Proporsi bagian awal kolom alcohol dikurangi Proporsi bagian akhir kolom alcohol bernilai 0

