

P3_Normality_Test

April 6, 2023

```
[ ]: # Import Libraries
import pandas as pd
import matplotlib.pyplot as plt
import scipy.stats as st
import seaborn as sns

significance = 0.05

# Read csv file
df = pd.read_csv("../data/anggur.csv")
```

1 Normality Test

Langkah-langkah: - Menguji normalitas dari setiap kolom A, dengan hipotesis pengujian sebagai berikut. - H_0 = kolom A berdistribusi normal - H_1 = kolom A tidak berdistribusi normal - Tingkat signifikan yang digunakan adalah $\alpha = 0.05$ - Uji statistik yang digunakan adalah normal-test (D'Agostino's K^2 test) - Pengambilan keputusan: - Tolak H_0 jika pvalue < α - H_0 tidak ditolak jika pvalue $\geq \alpha$

```
[ ]: # Print df
display(df)
```

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	\
0	5.90	0.4451	0.1813	2.049401	0.070574	
1	8.40	0.5768	0.2099	3.109590	0.101681	
2	7.54	0.5918	0.3248	3.673744	0.072416	
3	5.39	0.4201	0.3131	3.371815	0.072755	
4	6.51	0.5675	0.1940	4.404723	0.066379	
..	
995	7.96	0.6046	0.2662	1.592048	0.057555	
996	8.48	0.4080	0.2227	0.681955	0.051627	
997	6.11	0.4841	0.3720	2.377267	0.042806	
998	7.76	0.3590	0.3208	4.294486	0.098276	
999	5.87	0.5214	0.1883	2.179490	0.052923	

	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	\
0	16.593818	42.27	0.9982	3.27	0.71	
1	22.555519	16.01	0.9960	3.35	0.57	

2	9.316866	35.52	0.9990	3.31	0.64
3	18.212300	41.97	0.9945	3.34	0.55
4	9.360591	46.27	0.9925	3.27	0.45
..
995	14.892445	44.61	0.9975	3.35	0.54
996	23.548965	25.83	0.9972	3.41	0.46
997	21.624585	48.75	0.9928	3.23	0.55
998	12.746186	44.53	0.9952	3.30	0.66
999	16.203864	24.37	0.9983	3.29	0.70

	alcohol	quality
0	8.64	7
1	10.03	8
2	9.23	8
3	14.07	9
4	11.49	8
..
995	10.41	8
996	9.91	8
997	9.94	7
998	9.76	8
999	10.17	7

[1000 rows x 12 columns]

1.1 Kode Pengujian Hipotesis untuk Setiap Kolom

```
[ ]: for column in df.columns:
    # D'Agostino's K^2 test
    stat, pvalue = st.normaltest(df[column])

    # Plot data and distribution curve
    if (column == "quality"):
        sns.histplot(df[column], discrete=True)
    else:
        sns.displot(df[column], kde=True)
    plt.title(f"Distribusi nilai '{column}'")
    plt.show()

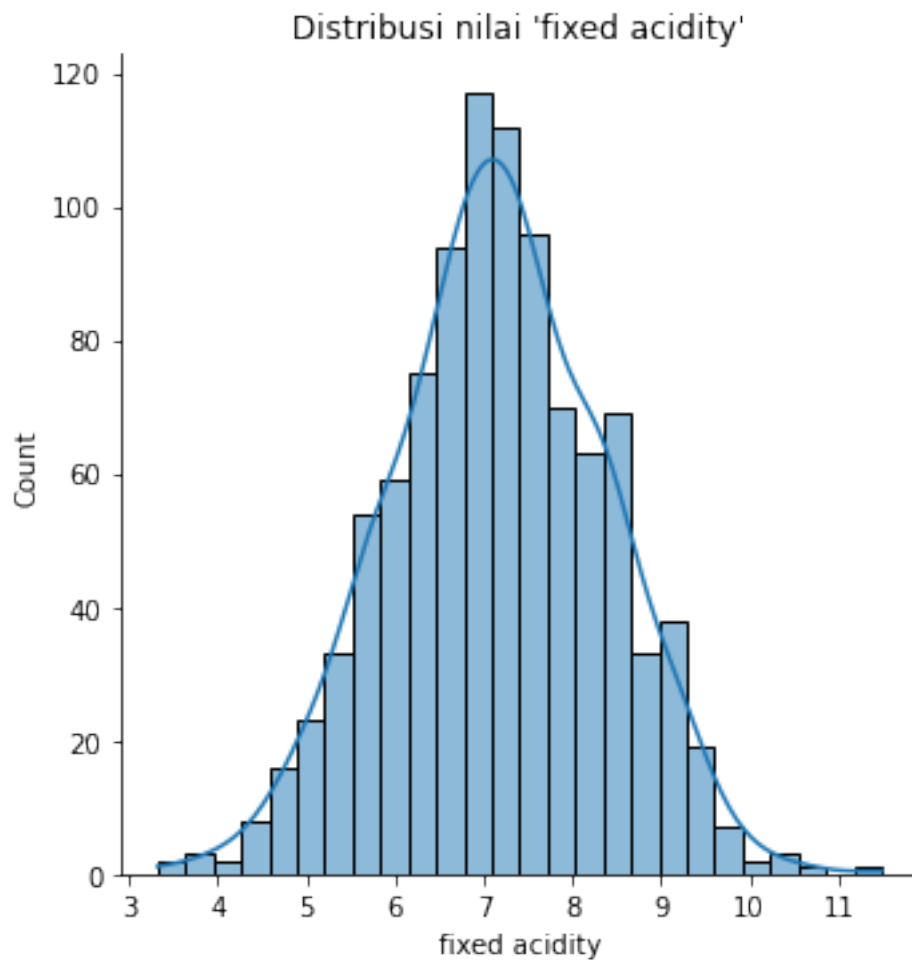
    # K^2 Test Result
    print(f"Statistic:\n K^2 = (Z_skew)^2 + (Z_kurtosis)^2 = {stat}")
    print(f"Two-sided Chi-Squared Probability Test:\n pvalue = {pvalue}")
    print(f"Significance:\n alpha = {significance}")

    # Hypothesis testing
    if pvalue >= significance: # H0 not rejected
        print("\npvalue >= alpha")
```

```

    verdict = f"Kesimpulan: H0 tidak ditolak, '{column}' berdistribusi_
↪normal\n"
    else:
        print("\npvalue < alpha")
        verdict = f"Kesimpulan: H0 ditolak, '{column}' tidak berdistribusi_
↪normal\n"
    print(verdict)

```



Statistic:

$K^2 = (Z_{\text{skew}})^2 + (Z_{\text{kurtosis}})^2 = 0.14329615661430725$

Two-sided Chi-Squared Probability Test:

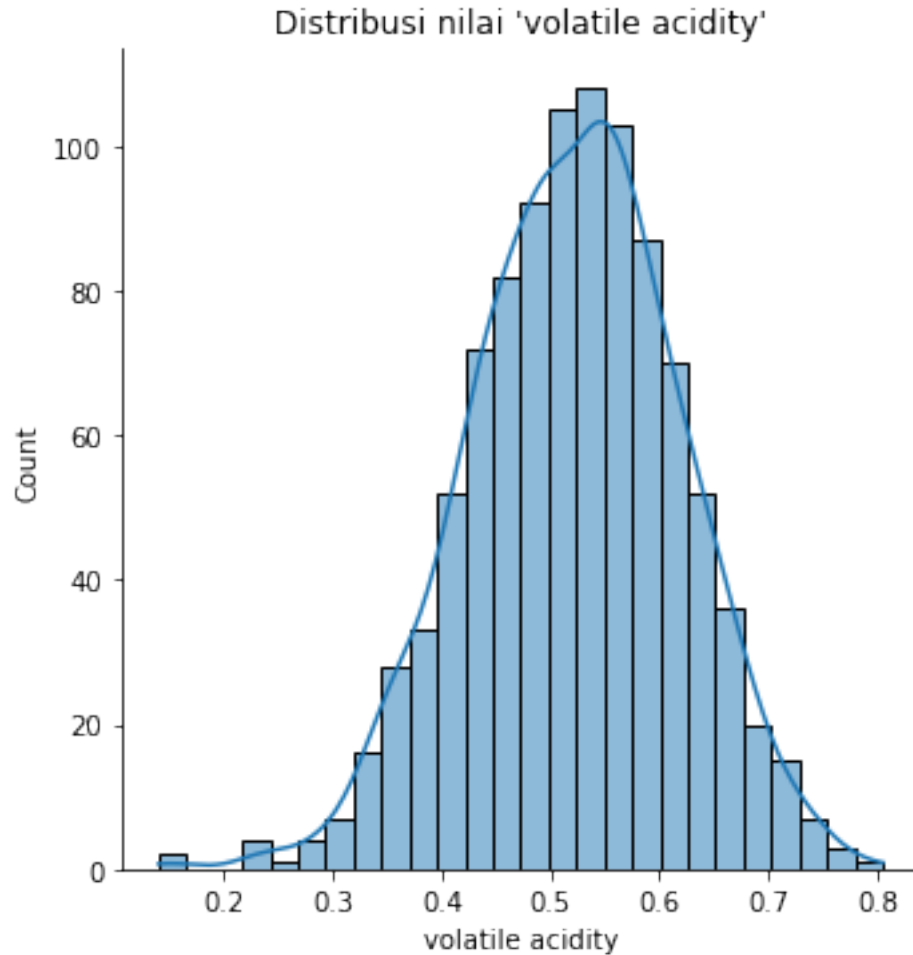
pvalue = 0.9308584274486692

Significance:

alpha = 0.05

pvalue >= alpha

Kesimpulan: H0 tidak ditolak, 'fixed acidity' berdistribusi normal



Statistic:

$$K^2 = (Z_{\text{skew}})^2 + (Z_{\text{kurtosis}})^2 = 7.581251985533493$$

Two-sided Chi-Squared Probability Test:

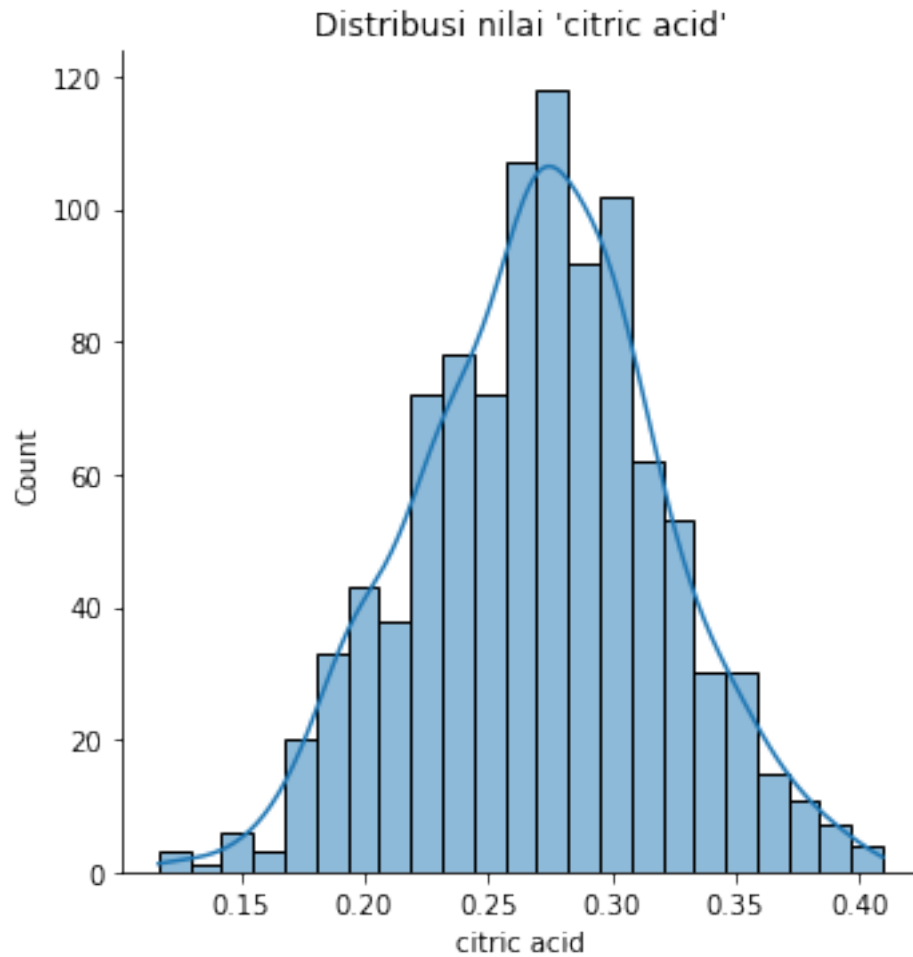
$$p\text{value} = 0.022581461594113835$$

Significance:

$$\alpha = 0.05$$

$$p\text{value} < \alpha$$

Kesimpulan: H_0 ditolak, 'volatile acidity' tidak berdistribusi normal



Statistic:

$$K^2 = (Z_{\text{skew}})^2 + (Z_{\text{kurtosis}})^2 = 0.7663607229418252$$

Two-sided Chi-Squared Probability Test:

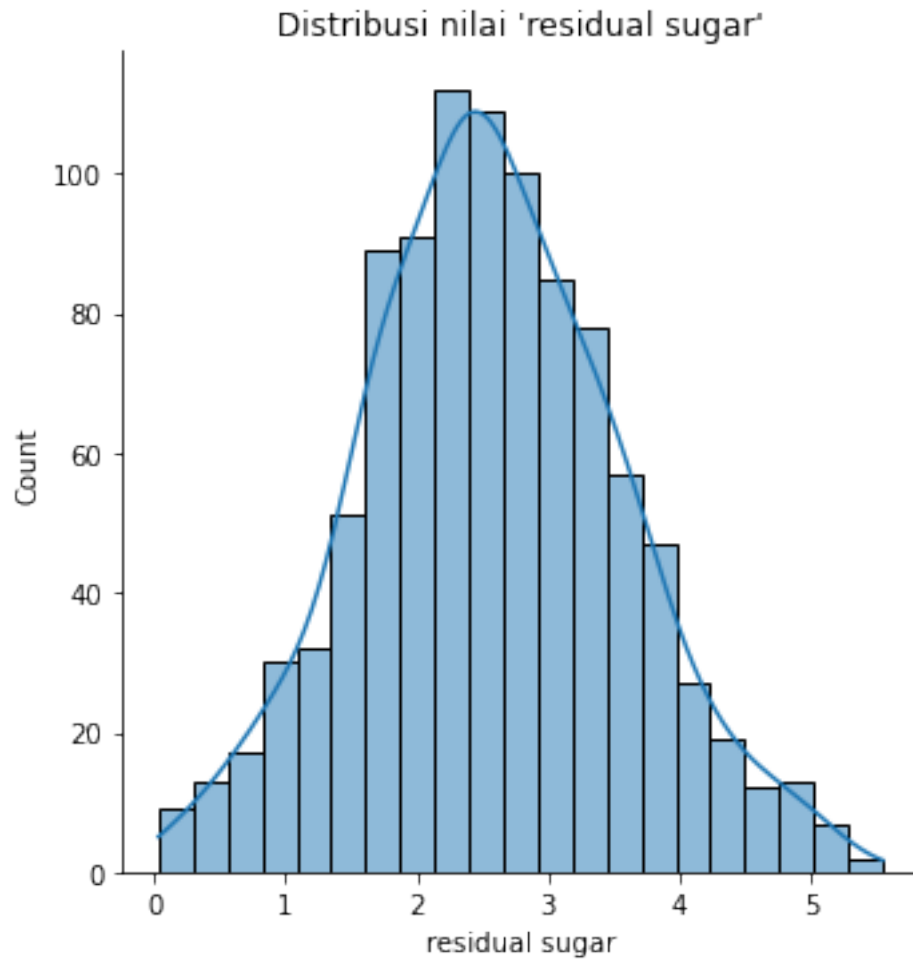
$$p\text{value} = 0.6816899375976969$$

Significance:

$$\alpha = 0.05$$

$$p\text{value} \geq \alpha$$

Kesimpulan: H_0 tidak ditolak, 'citric acid' berdistribusi normal



Statistic:

$$K^2 = (Z_{\text{skew}})^2 + (Z_{\text{kurtosis}})^2 = 2.9862716504538622$$

Two-sided Chi-Squared Probability Test:

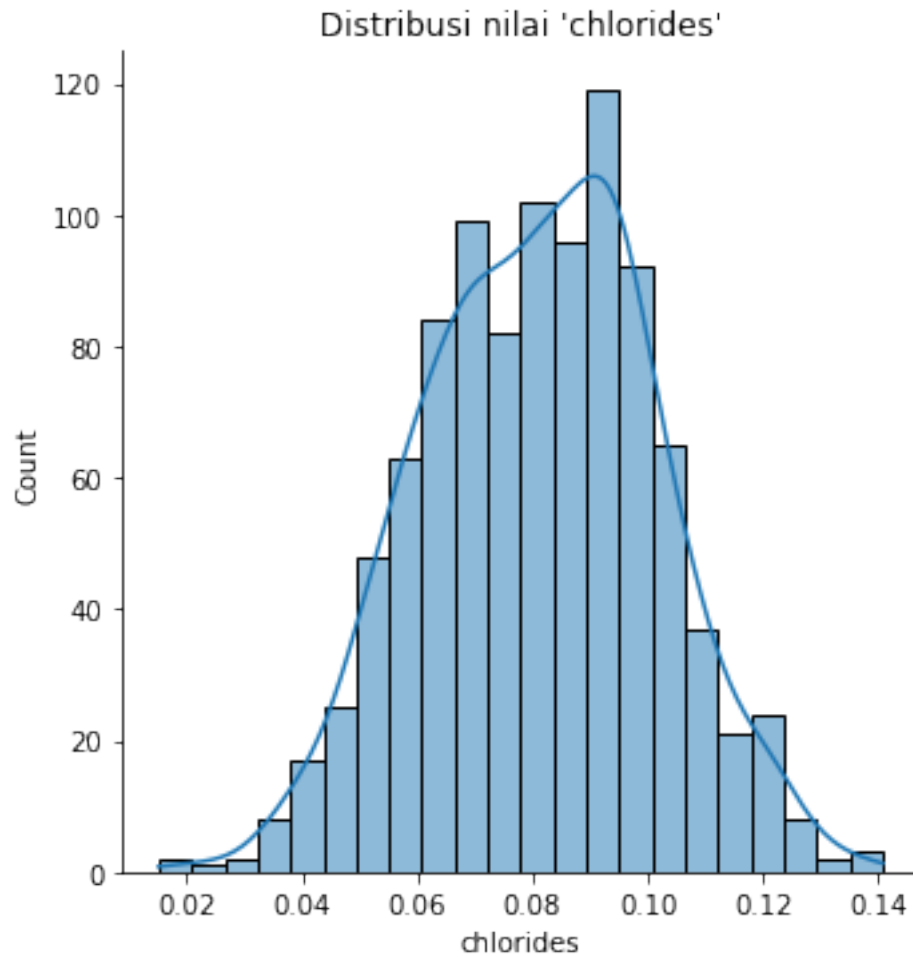
$$p\text{value} = 0.22466703321310558$$

Significance:

$$\alpha = 0.05$$

$$p\text{value} \geq \alpha$$

Kesimpulan: H_0 tidak ditolak, 'residual sugar' berdistribusi normal



Statistic:

$$K^2 = (Z_{\text{skew}})^2 + (Z_{\text{kurtosis}})^2 = 3.538242355484952$$

Two-sided Chi-Squared Probability Test:

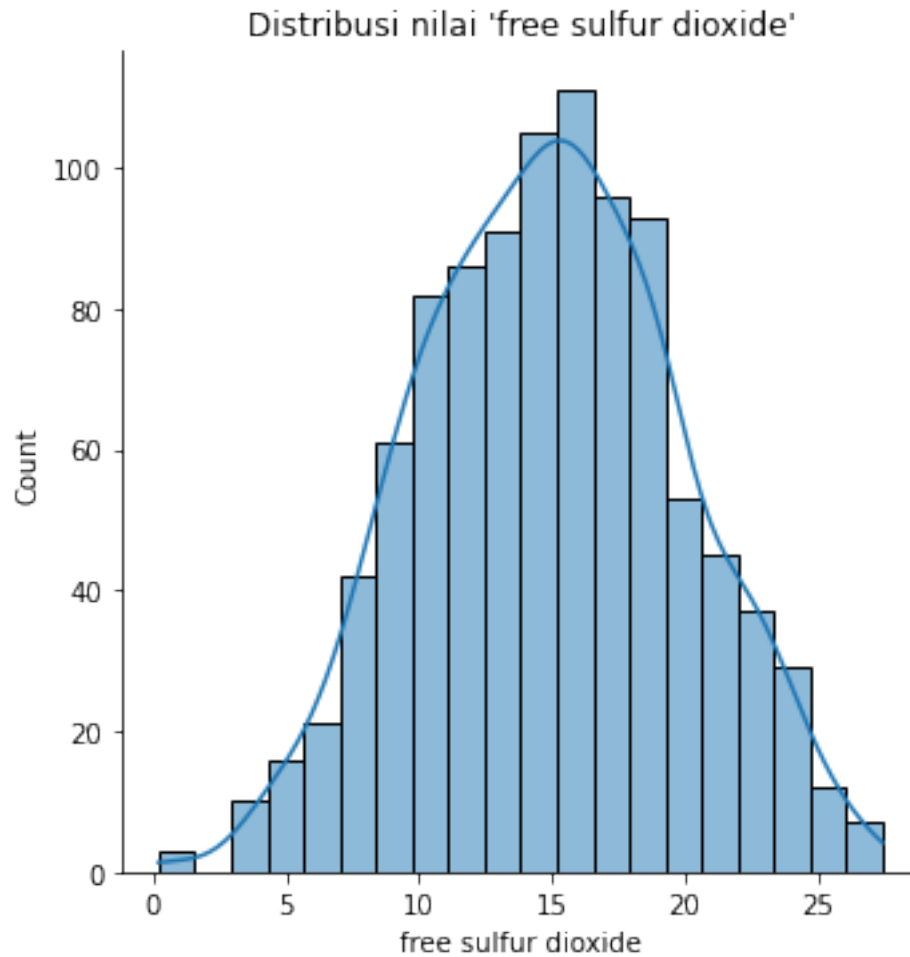
$$pvalue = 0.17048274704296862$$

Significance:

$$\alpha = 0.05$$

$$pvalue \geq \alpha$$

Kesimpulan: H_0 tidak ditolak, 'chlorides' berdistribusi normal



Statistic:

$$K^2 = (Z_{\text{skew}})^2 + (Z_{\text{kurtosis}})^2 = 8.099074980855514$$

Two-sided Chi-Squared Probability Test:

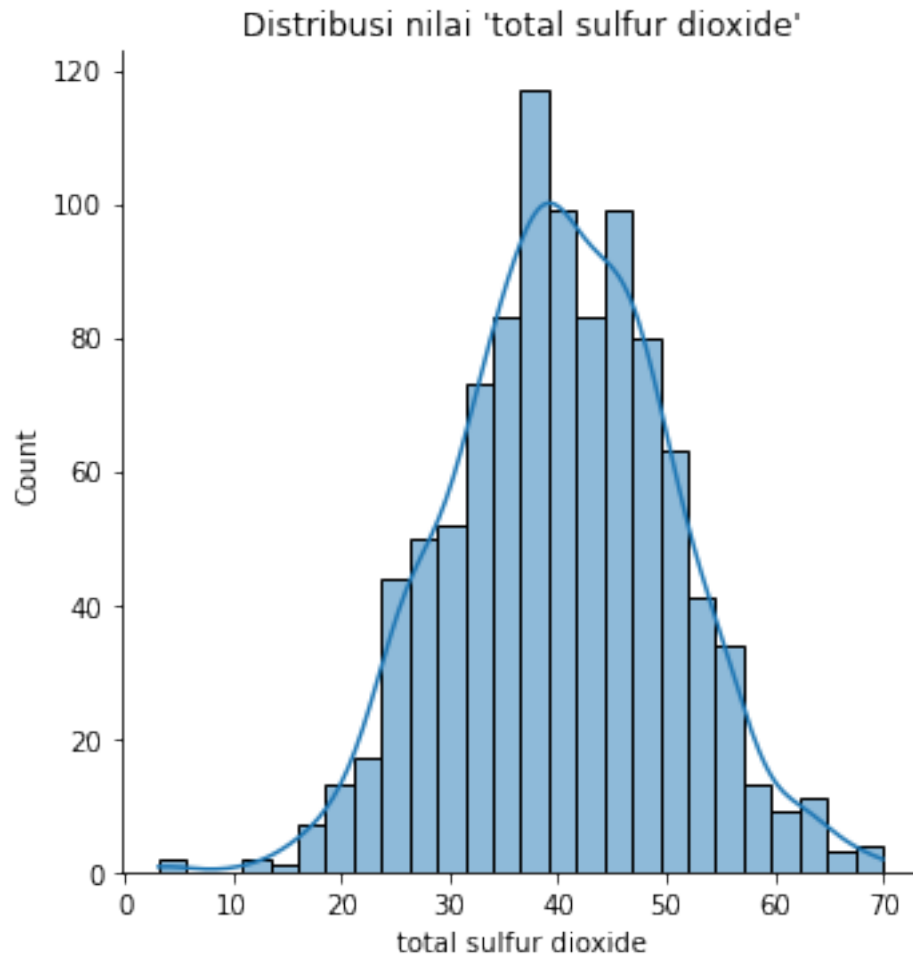
$$p\text{value} = 0.01743043451827735$$

Significance:

$$\alpha = 0.05$$

$$p\text{value} < \alpha$$

Kesimpulan: H_0 ditolak, 'free sulfur dioxide' tidak berdistribusi normal



Statistic:

$$K^2 = (Z_{\text{skew}})^2 + (Z_{\text{kurtosis}})^2 = 0.3276640291639825$$

Two-sided Chi-Squared Probability Test:

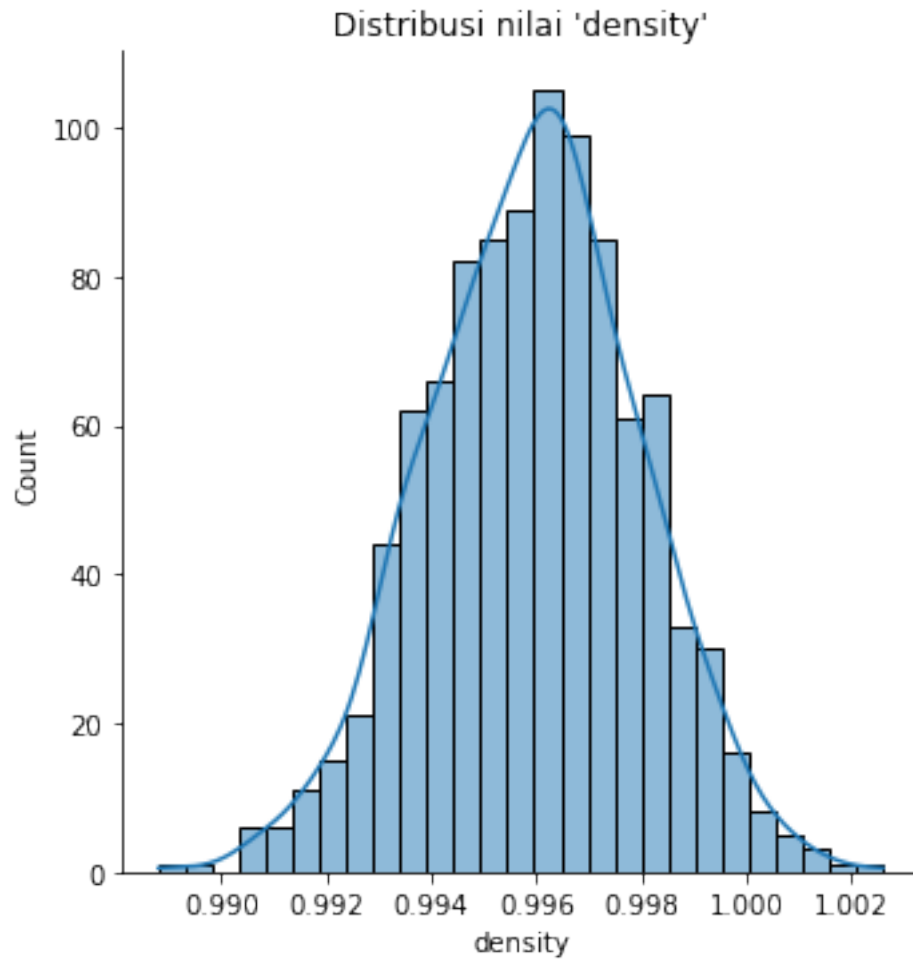
$$p\text{value} = 0.8488846101395726$$

Significance:

$$\alpha = 0.05$$

$$p\text{value} \geq \alpha$$

Kesimpulan: H_0 tidak ditolak, 'total sulfur dioxide' berdistribusi normal



Statistic:

$$K^2 = (Z_{\text{skew}})^2 + (Z_{\text{kurtosis}})^2 = 1.026581544320803$$

Two-sided Chi-Squared Probability Test:

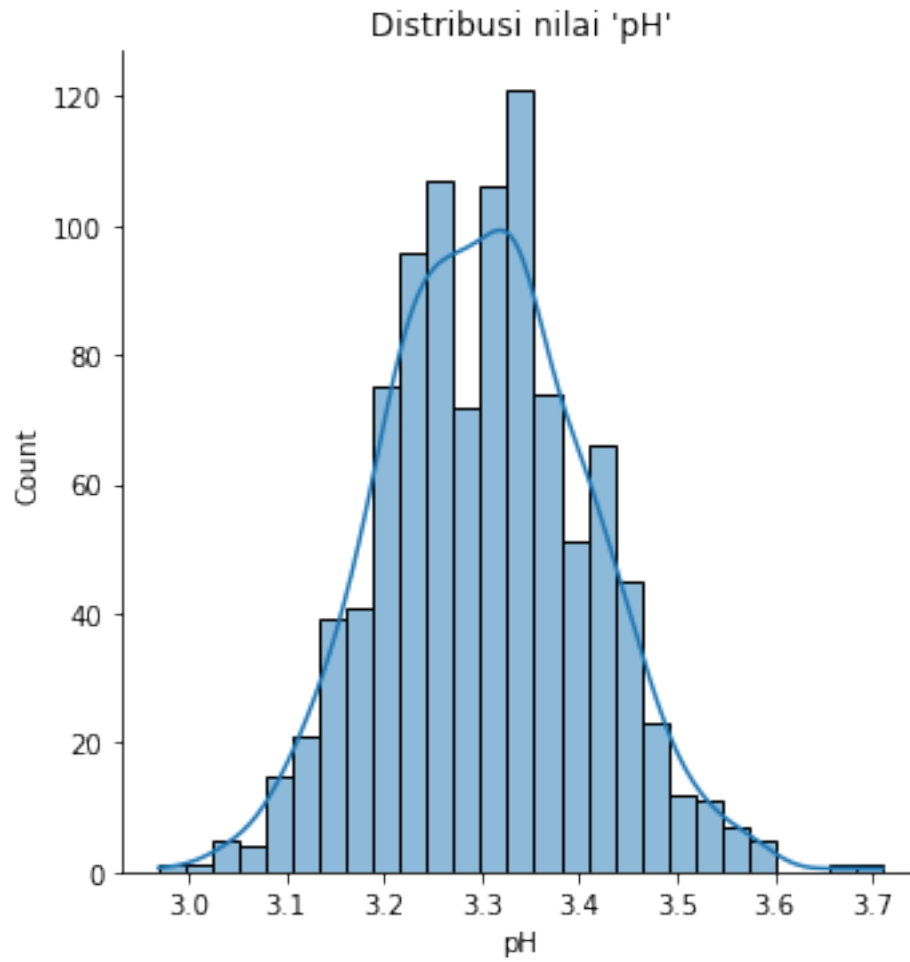
$$p\text{value} = 0.5985227325531981$$

Significance:

$$\alpha = 0.05$$

$$p\text{value} \geq \alpha$$

Kesimpulan: H_0 tidak ditolak, 'density' berdistribusi normal



Statistic:

$$K^2 = (Z_{\text{skew}})^2 + (Z_{\text{kurtosis}})^2 = 3.9786546459928545$$

Two-sided Chi-Squared Probability Test:

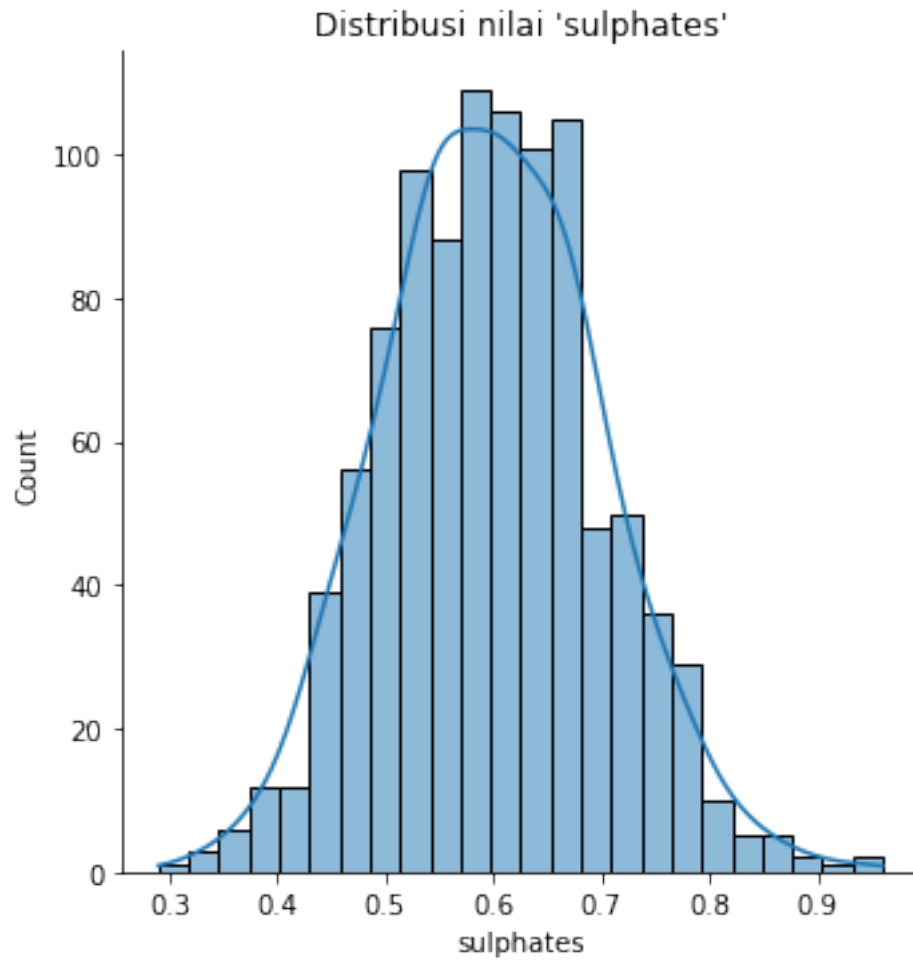
$$p\text{value} = 0.13678740824860436$$

Significance:

$$\alpha = 0.05$$

$$p\text{value} \geq \alpha$$

Kesimpulan: H_0 tidak ditolak, 'pH' berdistribusi normal



Statistic:

$$K^2 = (Z_{\text{skew}})^2 + (Z_{\text{kurtosis}})^2 = 3.948820277859041$$

Two-sided Chi-Squared Probability Test:

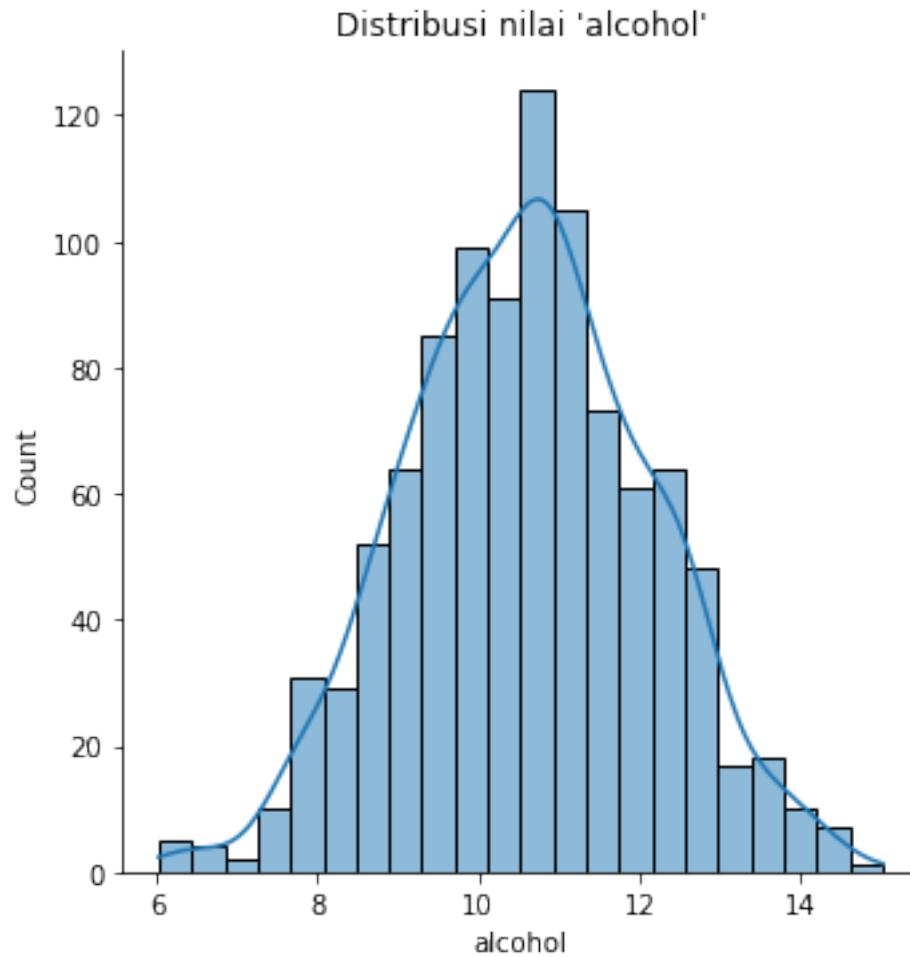
$$p\text{value} = 0.13884318628391681$$

Significance:

$$\alpha = 0.05$$

$$p\text{value} \geq \alpha$$

Kesimpulan: H_0 tidak ditolak, 'sulphates' berdistribusi normal



Statistic:

$$K^2 = (Z_{\text{skew}})^2 + (Z_{\text{kurtosis}})^2 = 0.7740076714171271$$

Two-sided Chi-Squared Probability Test:

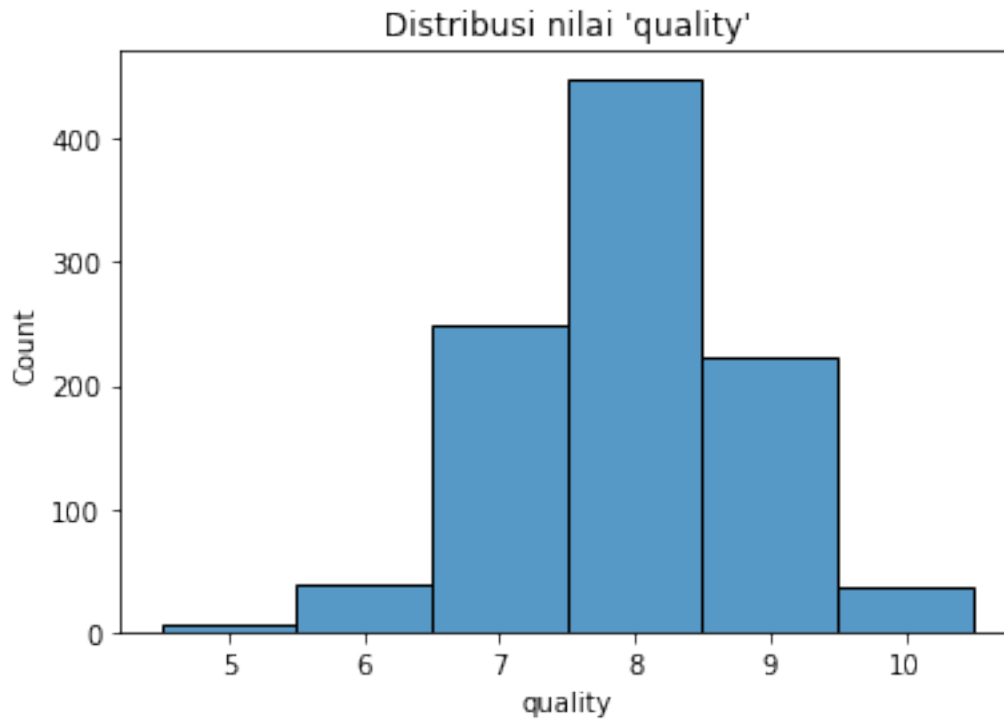
$$p\text{value} = 0.6790884901361043$$

Significance:

$$\alpha = 0.05$$

$$p\text{value} \geq \alpha$$

Kesimpulan: H_0 tidak ditolak, 'alcohol' berdistribusi normal



Statistic:

$$K^2 = (Z_{\text{skew}})^2 + (Z_{\text{kurtosis}})^2 = 1.8893087092494893$$

Two-sided Chi-Squared Probability Test:

$$p\text{value} = 0.3888139394184818$$

Significance:

$$\alpha = 0.05$$

$$p\text{value} \geq \alpha$$

Kesimpulan: H_0 tidak ditolak, 'quality' berdistribusi normal