

Melakukan test hipotesis 1 sampel

menuliskan 6 langkah testing dan menampilkan juga boxplotnya untuk kolom/bagian yang bersesuaian.

```
# import data
```

```
import pandas as pd
import matplotlib.pyplot as plt
```

```
df = pd.read_csv('water_potability.csv', index_col=0)
df.head()
```

	pH	Hardness	Solids	Chloramines	Sulfate \
id					
1	8.316766	214.373394	22018.417441	8.059332	356.886136
2	9.092223	181.101509	17978.986339	6.546600	310.135738
3	5.584087	188.313324	28748.687739	7.544869	326.678363
4	10.223862	248.071735	28749.716544	7.513408	393.663396
5	8.635849	203.361523	13672.091764	4.563009	303.309771

	Conductivity	OrganicCarbon	Trihalomethanes	Turbidity
Potability				
id				
1	363.266516	18.436524	100.341674	4.628771
0				
2	398.410813	11.558279	31.997993	4.075075
0				
3	280.467916	8.399735	54.917862	2.559708
0				
4	283.651634	13.789695	84.603556	2.672989
0				
5	474.607645	12.363817	62.798309	4.401425
0				

Nilai Rata-rata pH di atas 7?

```
import math
import scipy.stats as stats
```

```
df_pH = df['pH']
df.boxplot(['pH'])
print("Menguji apakah rata rata pH diatas 7 ?")
print("Enam langkah testing : ")
```

```
print(" 1. H0 : miu = 7 ")
```

```
print(" 2. H1 : miu > 7 (one-tailed test)")
```

```

print(" 3. alpha = 0.01 ")

print(" 4. Uji statistik yang sesuai dan tentukan daerah kritis ")

print("    Uji yang digunakan pada kasus ini adalah uji hipotesis
terhadap ")
print("    satu sample rataan dengan nilai variance tidak diketahui
karena")
print("    data yang diberikan merupakan sampel bukan populasi,
sehingga variance")
print("    tidak dapat diketahui. Daerah kritis yang digunakan
adalah : ")

alpha = 0.01
print("    alpha = " + str(alpha))
v = df_pH.count()-1
print("    derajat kebebasan = " + str(v))
critical_value = stats.t.ppf(1-alpha, v)
print("    critical value : t > " + str(critical_value))

print(" 5. Hitung nilai uji statistik dari data sample ")
print("    Hitung p-value sesuai dengan uji statistik yang digunakan.
")

n = df_pH.count()
print("    n = " + str(n))
s = df_pH.std()
print("    s = " + str(s));
x_bar = df_pH.mean()
print("    x_bar = " + str(x_bar))
miu = 7
print("    miu = " + str(miu))
t_value = (x_bar - miu)/(s/math.sqrt(n))
print("    t_value = " + str(t_value))

p_value = stats.norm.sf(t_value)
print("    p_value = " + str(p_value))

# 6. Ambil keputusan dengan TOLAK H0 jika nilai uji terletak di daerah
kritis
# atau dengan tes signifikan, TOLAK H0 jika p-value lebih kecil
dibanding tingkat
# signifikansi  $\alpha$  yang diinginkan.
print(" 6. Ambil Keputusan")
if(t_value > critical_value):
    print("    " + str(t_value) + " > " + str(critical_value))
    print("    Null hypothesis ditolak, maka claim bahwa rata-rata pH
> 7 diterima")

```

```

else :
    print("      " + str(t_value) + " < " + str(critical_value))
    print("      Null hypothesis diterima, maka claim bahwa rata-rata pH
> 7 ditolak")

```

Menguji apakah rata rata pH diatas 7 ?

Enam langkah testing :

1. $H_0 : \mu = 7$
2. $H_1 : \mu > 7$ (one-tailed test)
3. $\alpha = 0.01$
4. Uji statistik yang sesuai dan tentukan daerah kritis
 Uji yang digunakan pada kasus ini adalah uji hipotesis terhadap satu sample rata-rata dengan nilai variance tidak diketahui karena data yang diberikan merupakan sampel bukan populasi, sehingga

variance

tidak dapat diketahui. Daerah kritis yang digunakan adalah :

$\alpha = 0.01$

derajat kebebasan = 2009

critical value : $t > 2.3282054763161337$

5. Hitung nilai uji statistik dari data sample

Hitung p-value sesuai dengan uji statistik yang digunakan.

$n = 2010$

$s = 1.5728029470456655$

$\bar{x} = 7.0871927687138285$

$\mu = 7$

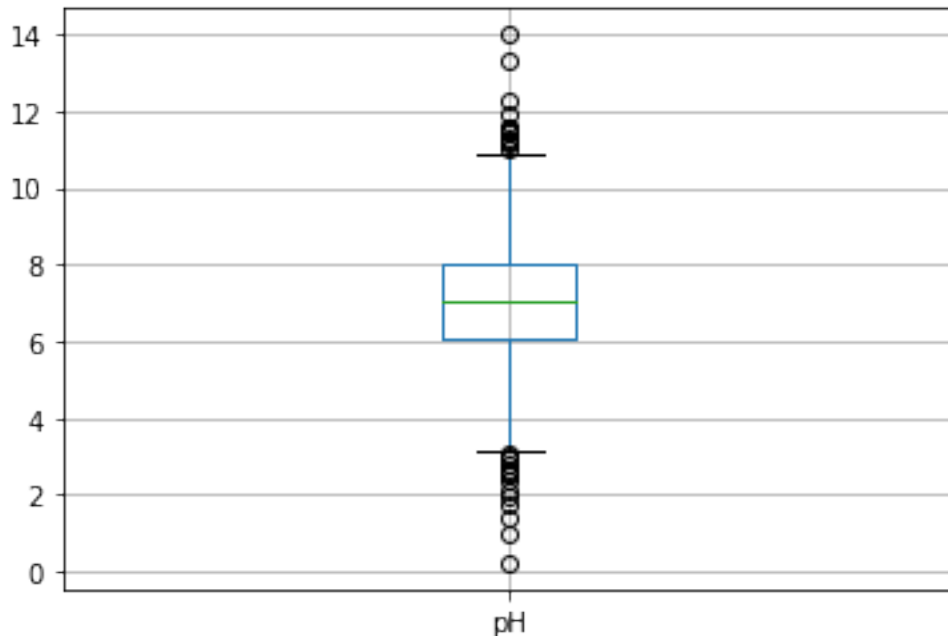
$t_value = 2.485445147379887$

$p_value = 0.006469476288896462$

6. Ambil Keputusan

$2.485445147379887 > 2.3282054763161337$

Null hypothesis ditolak, maka claim bahwa rata-rata pH > 7 diterima



Nilai Rata-rata Hardness tidak sama dengan 205?

```
import math
import scipy.stats as stats

df_Hardness = df['Hardness']
df.boxplot(['Hardness'])
print("Menguji apakah nilai rata rata Hardness tidak sama dengan 205 ?")
print("Enam langkah testing : ")

print(" 1. H0 : miu = 205 ")

print(" 2. H1 : miu != 205 (two-tailed test)")

print(" 3. alpha = 0.01 ")

print(" 4. Uji statistik yang sesuai dan tentukan daerah kritis ")

print("    Uji yang digunakan pada kasus ini adalah uji hipotesis terhadap ")
print("    satu sample rataan dengan nilai variance tidak diketahui karena")
print("    data yang diberikan merupakan sampel bukan populasi, sehingga variance")
print("    tidak dapat diketahui. Daerah kritis yang digunakan adalah : ")

alpha = 0.01
print("    alpha = " + str(alpha))
```

```

alpha = alpha/2
v = df_Hardness.count()-1
print("    derajat kebebasan = " + str(v))
critical_value = stats.t.ppf(1-alpha, v)
critical_value1 = critical_value * -1
print("    critical value : " + " t < " + str(critical_value1) + " or
t > " + str(critical_value))

```

```

print(" 5. Hitung nilai uji statistik dari data sample ")
print("    Hitung p-value sesuai dengan uji statistik yang digunakan.
")

```

```

n = df_Hardness.count()
print("    n = " + str(n))
s = df_Hardness.std()
print("    s = " + str(s));
x_bar = df_Hardness.mean()
print("    x_bar = " + str(x_bar))
miu = 205
print("    miu = " + str(miu))
t_value = (x_bar - miu)/(s/math.sqrt(n))
print("    t_value = " + str(t_value))

```

```

p_value = (1 - stats.norm.sf(t_value)) * 2
print("    p_value = " + str(p_value))

```

```

# 6. Ambil keputusan dengan TOLAK H0 jika nilai uji terletak di daerah
kritis
# atau dengan tes signifikan, TOLAK H0 jika p-value lebih kecil
dibanding tingkat
# signifikansi  $\alpha$  yang diinginkan.
print(" 6. Ambil Keputusan")

```

```

if(t_value > critical_value or t_value < critical_value1):
    print("    " + str(t_value) + " < " + str(critical_value1))
    print("    Null hypothesis ditolak, maka claim bahwa rata-rata
Hardness tidak sama dengan 205 diterima")
elif (t_value < critical_value):
    print("    " + str(t_value) + " < " + str(critical_value))
    print("    Null hypothesis diterima, maka claim bahwa rata-rata
Hardness tidak sama dengan 205 ditolak")
elif (t_value > critical_value1):
    print("    " + str(t_value) + " < " + str(critical_value1))
    print("    Null hypothesis diterima, maka claim bahwa rata-rata
Hardness tidak sama dengan 205 ditolak")

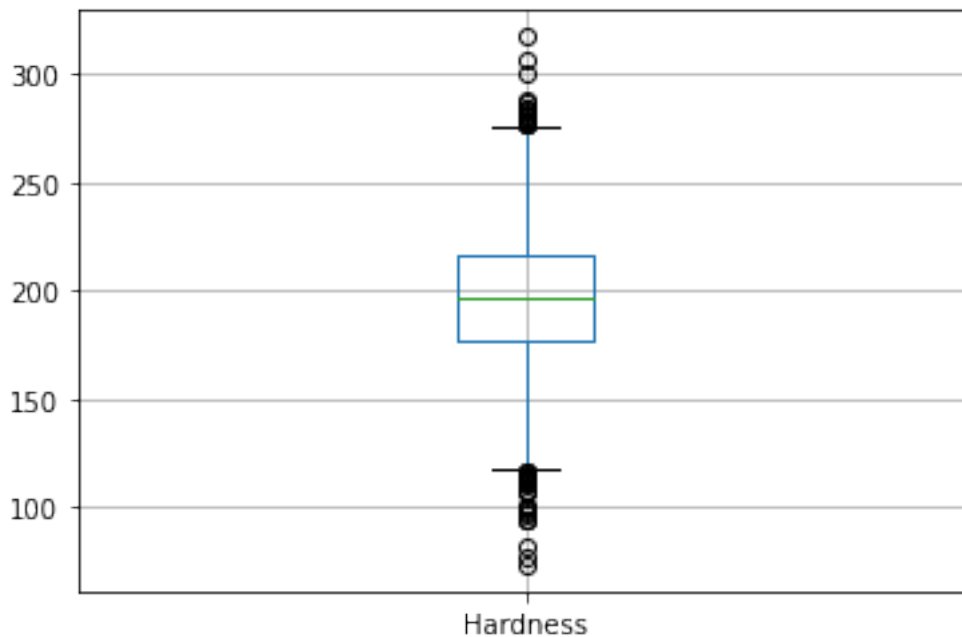
```

Menguji apakah nilai rata rata Hardness tidak sama dengan 205 ?

Enam langkah testing :

1. H_0 : $\mu = 205$

2. $H_1 : \mu \neq 205$ (two-tailed test)
3. $\alpha = 0.01$
4. Uji statistik yang sesuai dan tentukan daerah kritis
 Uji yang digunakan pada kasus ini adalah uji hipotesis terhadap satu sample rata-rata dengan nilai variance tidak diketahui karena data yang diberikan merupakan sampel bukan populasi, sehingga variance tidak dapat diketahui. Daerah kritis yang digunakan adalah :
 $\alpha = 0.01$
 derajat kebebasan = 2009
 critical value : $t < -2.5782787551172976$ or $t > 2.5782787551172976$
5. Hitung nilai uji statistik dari data sample
 Hitung p-value sesuai dengan uji statistik yang digunakan.
 $n = 2010$
 $s = 32.643165859429864$
 $\bar{x} = 195.96920903783524$
 $\mu = 205$
 $t_{\text{value}} = -12.403137170010732$
 $p_{\text{value}} = 0.0$
6. Ambil Keputusan
 $-12.403137170010732 < -2.5782787551172976$
 Null hypothesis ditolak, maka claim bahwa rata-rata Hardness tidak sama dengan 205 diterima



Nilai Rata-rata 100 baris pertama kolom Solids bukan 21900?

```
import math
import scipy.stats as stats
```

```
df_Solids = df['Solids']
```

```

df_solidsSample = df_Solids.head(100)
df.boxplot(['Solids'])
print("Menguji apakah nilai rata rata 100 baris pertama kolom solids
bukan 21900 ?")
print("Enam langkah testing : ")

print(" 1. H0 : miu = 21900 ")

print(" 2. H1 : miu != 21900 (two-tailed test)")

print(" 3. alpha = 0.01 ")

print(" 4. Uji statistik yang sesuai dan tentukan daerah kritis ")

print("    Uji yang digunakan pada kasus ini adalah uji hipotesis
terhadap ")
print("    satu sample rataan dengan nilai variance tidak diketahui
karena")
print("    data yang diberikan merupakan sampel bukan populasi,
sehingga variance")
print("    tidak dapat diketahui. Daerah kritis yang digunakan
adalah : ")

alpha = 0.01
print("    alpha = " + str(alpha))
alpha = alpha/2
v = df_solidsSample.count()-1
print("    derajat kebebasan = " + str(v))
critical_value = stats.t.ppf(1-alpha, v)
critical_value1 = critical_value * -1
print("    critical value : " + " t < " + str(critical_value1) + " or
t > " + str(critical_value))

print(" 5. Hitung nilai uji statistik dari data sample ")
print("    Hitung p-value sesuai dengan uji statistik yang digunakan.
")

n = df_solidsSample.count()
print("    n = " + str(n))
s = df_solidsSample.std()
print("    s = " + str(s));
x_bar = df_solidsSample.mean()
print("    x_bar = " + str(x_bar))
miu = 21900
print("    miu = " + str(miu))
t_value = (x_bar - miu)/(s/math.sqrt(n))
print("    t_value = " + str(t_value))

p_value = (1 - stats.norm.sf(t_value)) * 2

```

```

print("      p_value = " + str(p_value))

# 6. Ambil keputusan dengan TOLAK H0 jika nilai uji terletak di daerah
kritis
# atau dengan tes signifikan, TOLAK H0 jika p-value lebih kecil
dibanding tingkat
# signifikansi  $\alpha$  yang diinginkan.
print(" 6. Ambil Keputusan")

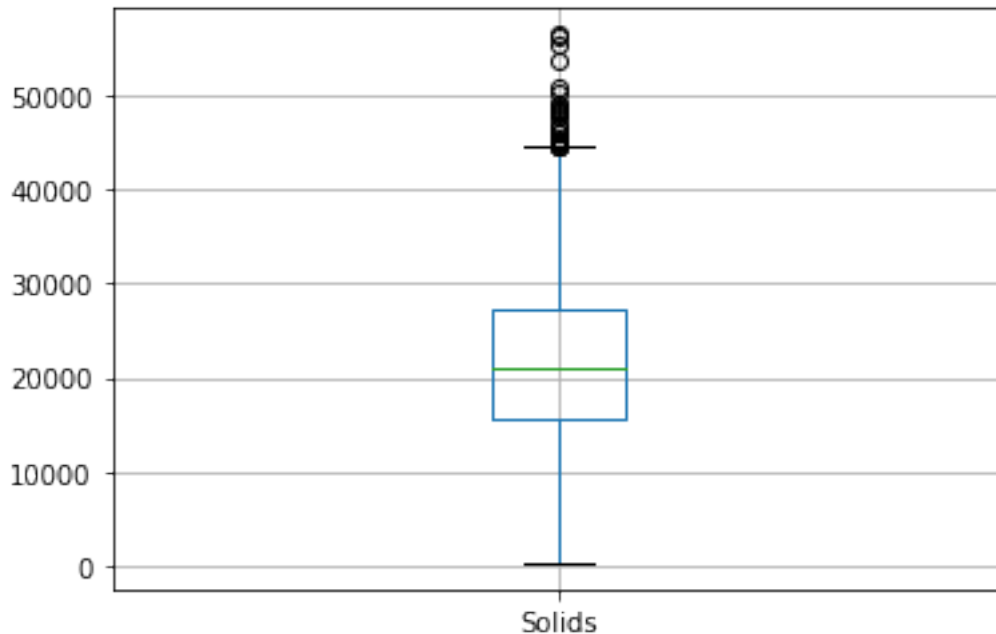
if(t_value > critical_value or t_value < critical_value1):
    print("      " + str(t_value) + " < " + str(critical_value1))
    print("      Null hypothesis ditolak, maka claim bahwa rata-rata 100
baris pertama kolom Solids bukan 21900 diterima")
elif (t_value < critical_value):
    print("      " + str(t_value) + " < " + str(critical_value))
    print("      Null hypothesis diterima, maka claim bahwa rata-rata
100 baris pertama kolom Solids bukan 21900 ditolak")
elif (t_value > critical_value1):
    print("      " + str(t_value) + " < " + str(critical_value1))
    print("      Null hypothesis diterima, maka claim bahwa rata-rata
100 baris pertama kolom Solids bukan 21900 ditolak")

```

Menguji apakah nilai rata rata 100 baris pertama kolom solids bukan 21900 ?

Enam langkah testing :

1. H_0 : $\mu = 21900$
2. H_1 : $\mu \neq 21900$ (two-tailed test)
3. $\alpha = 0.01$
4. Uji statistik yang sesuai dan tentukan daerah kritis
 Uji yang digunakan pada kasus ini adalah uji hipotesis terhadap satu sample rata-rata dengan nilai variance tidak diketahui karena data yang diberikan merupakan sampel bukan populasi, sehingga variance tidak dapat diketahui. Daerah kritis yang digunakan adalah :
 $\alpha = 0.01$
 derajat kebebasan = 99
 critical value : $t < -2.6264054563851857$ or $t > 2.6264054563851857$
5. Hitung nilai uji statistik dari data sample
 Hitung p-value sesuai dengan uji statistik yang digunakan.
 $n = 100$
 $s = 7935.967706199005$
 $\bar{x} = 22347.334446383426$
 $\mu = 21900$
 $t_value = 0.5636797715721551$
 $p_value = 1.4270279135344826$
6. Ambil Keputusan
 $0.5636797715721551 < 2.6264054563851857$
 Null hypothesis diterima, maka claim bahwa rata-rata 100 baris pertama kolom Solids bukan 21900 ditolak



Proporsi nilai Conductivity yang lebih dari 450, adalah tidak sama dengan 10%?

```
import math
import scipy.stats as stats

df_Conductivity = df['Conductivity']
df_conductivitySample = df_Conductivity.head(100)
df.boxplot(['Conductivity'])
print("Menguji apakah proporsi nilai conductivity yang lebih dari 450  
tidak sama dengan 10 %?")
print("Enam langkah testing : ")

print(" 1. H0 : p = 0.1 ")

print(" 2. H1 : p != 0.1 (two-tailed test)")

print(" 3. alpha = 0.01 ")

print(" 4. Uji statistik yang sesuai dan tentukan daerah kritis ")
print("    Uji yang digunakan pada kasus ini adalah uji hipotesis  
terhadap ")
print("    satu sample proporsi. Daerah kritis ")
print("    yang digunakan adalah : ")

alpha = 0.01
print("    alpha = " + str(alpha))
alpha = alpha/2

critical_value = stats.norm.ppf(abs(1-alpha))
critical_value1 = critical_value * -1
```

```
print("    critical value : " + " Z < " + str(critical_value1) + " or  
Z > " + str(critical_value))
```

```
print(" 5. Hitung nilai uji statistik dari data sample ")  
print("    Hitung p-value sesuai dengan uji statistik yang digunakan.  
")
```

```
n = df_conductivitySample.count()  
print("    n = " + str(n))
```

```
p0 = 0.1  
q0 = 1-p0  
print("    p0 = " + str(p0))  
print("    q0 = " + str(q0))
```

```
count = 0  
for data in df_conductivitySample:  
    if (data > 450):  
        count += 1
```

```
print("    Digunakan sebuah sampel data yang berukuran 100, yang  
diketahui proporsi")  
print("    nilai conductivity diatas 450, yaitu : ")  
p_topi = count/n  
print("    p_topi = " + str(p_topi))
```

```
z_score = (p_topi - p0)/math.sqrt((p0*q0)/n)  
print("    Z = " + str(z_score))
```

```
p_value = (1 - stats.norm.sf(z_score)) * 2  
print("    p_value = " + str(p_value))
```

```
# 6. Ambil keputusan dengan TOLAK H0 jika nilai uji terletak di daerah  
kritis  
# atau dengan tes signifikan, TOLAK H0 jika p-value lebih kecil  
dibanding tingkat  
# signifikansi  $\alpha$  yang diinginkan.  
print(" 6. Ambil Keputusan")
```

```
if(z_score > critical_value or z_score < critical_value1):  
    print("    " + str(z_score) + " > " + str(critical_value))  
    print("    Null hypothesis ditolak, maka claim bahwa proporsi  
nilai conductivity")  
    print("    yang lebih dari 450 tidak sama dengan 10% diterima")  
elif (z_score < critical_value):  
    print("    " + str(z_score) + " < " + str(critical_value))  
    print("    Null hypothesis diterima, maka claim bahwa proporsi  
nilai conductivity")  
    print("    yang lebih dari 450 tidak sama dengan 10% ditolak")
```

```

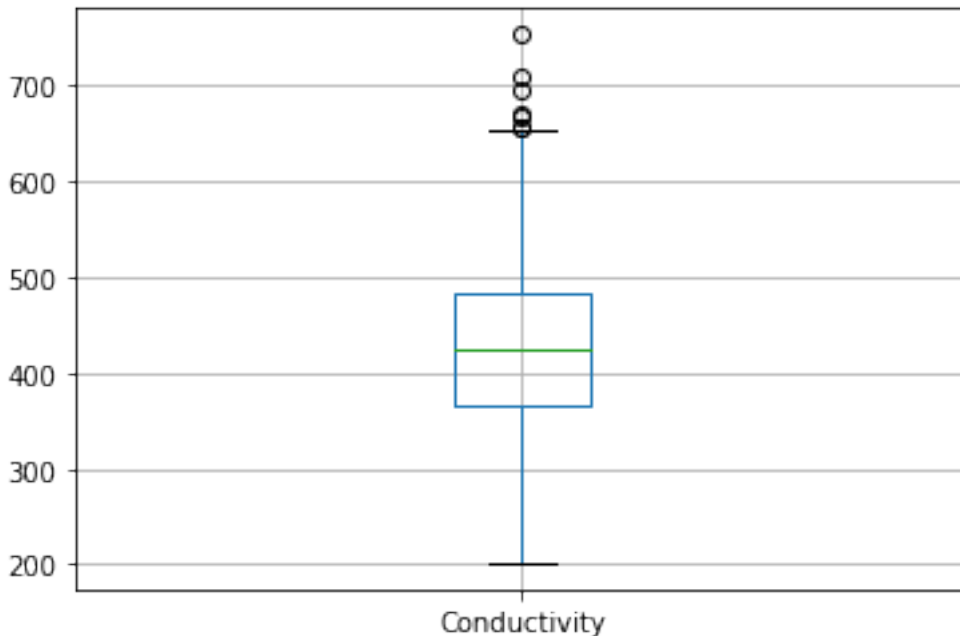
elif (z_score > critical_value1):
    print("      " + str(z_score) + " < " + str(critical_value1))
    print("      Null hypothesis diterima, maka claim bahwa proporsi
nilai conductivity")
    print("      yang lebih dari 450 tidak sama dengan 10% ditolak")

```

Menguji apakah proporsi nilai conductivity yang lebih dari 450 tidak sama dengan 10 %?

Enam langkah testing :

1. $H_0 : p = 0.1$
2. $H_1 : p \neq 0.1$ (two-tailed test)
3. $\alpha = 0.01$
4. Uji statistik yang sesuai dan tentukan daerah kritis
 Uji yang digunakan pada kasus ini adalah uji hipotesis terhadap satu sample proporsi. Daerah kritis yang digunakan adalah :
 $\alpha = 0.01$
 critical value : $Z < -2.5758293035489004$ or $Z > 2.5758293035489004$
5. Hitung nilai uji statistik dari data sample
 Hitung p-value sesuai dengan uji statistik yang digunakan.
 $n = 100$
 $p_0 = 0.1$
 $q_0 = 0.9$
 Digunakan sebuah sampel data yang berukuran 100, yang diketahui proporsi nilai conductivity diatas 450, yaitu :
 $p_{topi} = 0.39$
 $Z = 9.666666666666668$
 $p_value = 2.0$
6. Ambil Keputusan
 $9.666666666666668 > 2.5758293035489004$
 Null hypothesis ditolak, maka claim bahwa proporsi nilai conductivity yang lebih dari 450 tidak sama dengan 10% diterima



Proporsi nilai Trihalomethanes yang kurang dari 40, adalah kurang dari 5%?

```
import math
import scipy.stats as stats

df_Trihalomethanes = df['Trihalomethanes']
df_trihalomethanesSample = df_Trihalomethanes.head(100)
df.boxplot(['Trihalomethanes'])
print("Menguji apakah proporsi nilai trihalomethanes yang kurang dari 40 kurang dari 5% ?")
print("Enam langkah testing : ")

print(" 1. H0 : p = 0.05 ")

print(" 2. H1 : p < 0.05 (one-tailed test)")

print(" 3. alpha = 0.01 ")

print(" 4. Uji statistik yang sesuai dan tentukan daerah kritis ")
print("    Uji yang digunakan pada kasus ini adalah uji hipotesis terhadap ")
print("    satu sample proporsi. Daerah kritis ")
print("    yang digunakan adalah : ")

alpha = 0.01
print("    alpha = " + str(alpha))

critical_value = stats.norm.ppf(abs(1-alpha)) * -1
print("    critical value : " + " Z < " + str(critical_value))
```

```
print(" 5. Hitung nilai uji statistik dari data sample ")
print("      Hitung p-value sesuai dengan uji statistik yang digunakan.
")
```

```
n = df_trihalomethanesSample.count()
print("      n = " + str(n))
```

```
p0 = 0.05
q0 = 1-p0
print("      p0 = " + str(p0))
print("      q0 = " + str(q0))
```

```
count = 0
for data in df_trihalomethanesSample:
    if (data < 40):
        count += 1
```

```
print("      Digunakan sebuah sampel data yang berukuran 100, yang
diketahui proporsi")
print("      nilai trihalomethanes kurang dari 40, yaitu : ")
p_topi = count/n
print("      p_topi = " + str(p_topi))
```

```
z_score = (p_topi - p0)/math.sqrt((p0*q0)/n)
print("      Z = " + str(z_score))
```

```
p_value = stats.norm.cdf(z_score)
print("      p_value = " + str(p_value))
```

```
# 6. Ambil keputusan dengan TOLAK H0 jika nilai uji terletak di daerah
kritis
# atau dengan tes signifikan, TOLAK H0 jika p-value lebih kecil
dibanding tingkat
# signifikansi  $\alpha$  yang diinginkan.
print(" 6. Ambil Keputusan")
```

```
if(z_score < critical_value1):
    print("      " + str(z_score) + " < " + str(critical_value))
    print("      Null hypothesis ditolak, maka claim bahwa proporsi
nilai trihalomethanes")
    print("      yang kurang dari 40, kurang dari 5% diterima")
else :
    print("      " + str(z_score) + " > " + str(critical_value))
    print("      Null hypothesis diterima, maka claim bahwa proporsi
nilai trihalomethanes")
    print("      yang kurang dari 40, kurang dari 5% ditolak")
```

Menguji apakah proporsi nilai trihalomethanes yang kurang dari 40 kurang dari 5% ?
Enam langkah testing :

1. $H_0 : p = 0.05$
2. $H_1 : p < 0.05$ (one-tailed test)
3. $\alpha = 0.01$
4. Uji statistik yang sesuai dan tentukan daerah kritis
 Uji yang digunakan pada kasus ini adalah uji hipotesis terhadap satu sample proporsi. Daerah kritis yang digunakan adalah :
 $\alpha = 0.01$
 critical value : $Z < -2.3263478740408408$
5. Hitung nilai uji statistik dari data sample
 Hitung p-value sesuai dengan uji statistik yang digunakan.
 $n = 100$
 $p_0 = 0.05$
 $q_0 = 0.95$
 Digunakan sebuah sampel data yang berukuran 100, yang diketahui proporsi nilai trihalomethanes kurang dari 40, yaitu :
 $p_{topi} = 0.08$
 $Z = 1.3764944032233706$
 $p_value = 0.9156656905560924$
6. Ambil Keputusan
 $1.3764944032233706 > -2.3263478740408408$
 Null hypothesis diterima, maka claim bahwa proporsi nilai trihalomethanes yang kurang dari 40, kurang dari 5% ditolak

