5

April 18, 2023

[26]:
```
import pandas as pd
import matplotlib.pyplot as plt
import scipy.stats as st
from statsmodels.stats.weightstats import ztest
from statsmodels.stats.proportion import proportions_ztest

dataAnggur = pd.read_csv('../data/anggur.csv')
```

Melakukan test hipotesis 2 sampel,

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

Hipotesis
Misalkan 1 melambangkan bagian awal kolom fixed acidity dan 2 melambangkan bagian akhir kolom fixed acidity.
H_0 : Rata-rata kedua bagian sama ($\mu_1 = \mu_2, \mu_1 - \mu_2 = 0$)
H_1 : Rata-rata kedua bagian berbeda ($\mu_1 \neq \mu_2, \mu_1 - \mu_2 \neq 0$)
Dari kalimat soal, kita dapat menganggap bahwa klaimnya adalah H_0.

Tingkat Signifikansi
$\alpha = 0.05$

Uji Statistik
Pada pengujian hipotesis ini, meskipun variansi populasi tidak diketahui, digunakan z-test, bukan t-test. Hal ini diputuskan karena jumlah sampel yang digunakan jauh lebih banyak dibanding 30.

Digunakan tes statistik z dengan rumus:

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

dengan daerah kritis:
$z < -z_{\alpha/2}$ atau $z > z_{\alpha/2}$ (two-tailed test)

Pengambilan Keputusan

Tes Daerah Kritis

- Reject H_0 jika $z < -z_{\alpha/2}$ atau $z > z_{\alpha/2}$
- Fail to reject H_0 jika $-z_{\alpha/2} \leq z \leq z_{\alpha/2}$

Tes Signifikansi

- Reject H_0 jika $p < \alpha$
- Fail to reject H_0 jika $p \geq \alpha$

[27]:
```
# Diketahui
alpha = 0.05
deltaMean = 0

# Ambil data
nData = len(dataAnggur) // 2
dataAwal = dataAnggur["fixed acidity"][ : nData]
dataAkhir = dataAnggur["fixed acidity"][nData : ]

# Lakukan z-test dengan memanfaatkan library statsmodels untuk mendapatkan␣
↪nilai z dan p
z, p = ztest(dataAwal, dataAkhir, value = deltaMean)

# Hitung z_alpha/2
zAlpha2 = st.norm.ppf(1 - alpha / 2)

# Tampilkan hasil
print(f"Nilai z                 : {round(z, 5)}")
print(f"Nilai z_alpha/2         : {round(zAlpha2, 5)}")
print(f"Nilai p                 : {round(p, 5)}")
```

```
Nilai z                 : 0.02604
Nilai z_alpha/2         : 1.95996
Nilai p                 : 0.97922
```

Hasil Tes

Tes Daerah Kritis
Karena $-z_{\alpha/2} \leq z \leq z_{\alpha/2}$ $(-1.95996 \leq 0.02604 \leq 1.95996)$, fail to reject H_0.

Tes Signifikansi
Karena $p \geq \alpha$ $(0.97922 \geq 0.05)$, fail to reject H_0.

Kesimpulan
Dengan tingkat signifikansi sebesar 0.05, tidak ada bukti yang cukup untuk menolak klaim bahwa rerata bagian awal dan akhir kolom fixed acidity bernilai sama.

0.0.2 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?

Hipotesis
Misalkan 1 melambangkan bagian awal kolom chlorides dan 2 melambangkan bagian akhir kolom chlorides.
H_0 : Rata-rata bagian awal lebih besar daripada bagian akhir sebesar 0.001 ($\mu_1 = \mu_2 + 0.001, \mu_1 - \mu_2 = 0.001$)
H_1 : Rata-rata bagian awal tidak lebih besar daripada bagian akhir sebesar 0.001 ($\mu_1 \neq \mu_2 + 0.001, \mu_1 - \mu_2 \neq 0.001$)
Dari kalimat soal, kita dapat menganggap bahwa klaimnya adalah H_0.

Tingkat Signifikansi
$\alpha = 0.05$

Uji Statistik
Pada pengujian hipotesis ini, meskipun variansi populasi tidak diketahui, digunakan z-test, bukan t-test. Hal ini diputuskan karena jumlah sampel yang digunakan jauh lebih banyak dibanding 30.

Digunakan tes statistik z dengan rumus:

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

dengan daerah kritis:
$z < -z_{\alpha/2}$ atau $z > z_{\alpha/2}$ (two-tailed test)

Pengambilan Keputusan

Tes Daerah Kritis

- Reject H_0 jika $z < -z_{\alpha/2}$ atau $z > z_{\alpha/2}$
- Fail to reject H_0 jika $-z_{\alpha/2} \leq z \leq z_{\alpha/2}$

Tes Signifikansi

- Reject H_0 jika $p < \alpha$
- Fail to reject H_0 jika $p \geq \alpha$

[28]:
```
# Diketahui
alpha = 0.05
deltaMean = 0.001

# Ambil data
nData = len(dataAnggur) // 2
dataAwal = dataAnggur["chlorides"][ : nData]
dataAkhir = dataAnggur["chlorides"][nData : ]
```

```
# Lakukan z-test dengan memanfaatkan library statsmodels untuk mendapatkan␣
↪nilai z dan p
z, p = ztest(dataAwal, dataAkhir, value = deltaMean)

# Hitung z_alpha/2
zAlpha2 = st.norm.ppf(1 - alpha / 2)

# Tampilkan hasil
print(f"Nilai z              : {round(z, 5)}")
print(f"Nilai z_alpha/2      : {round(zAlpha2, 5)}")
print(f"Nilai p              : {round(p, 5)}")
```

```
Nilai z              : -0.46732
Nilai z_alpha/2      : 1.95996
Nilai p              : 0.64027
```

Hasil Tes

Tes Daerah Kritis
Karena $-z_{\alpha/2} \leq z \leq z_{\alpha/2}$ $(-1.95996 \leq -0.46732 \leq 1.95996)$, fail to reject H_0.

Tes Signifikansi
Karena $p \geq \alpha$ $(0.64027 \geq 0.05)$, fail to reject H_0.

Kesimpulan
Dengan tingkat signifikansi sebesar 0.05, tidak ada bukti yang cukup untuk menolak klaim bahwa untuk kolom chlorides, rata-rata bagian awal lebih besar daripada bagian akhir sebesar 0.001.

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

Hipotesis
Misalkan 1 melambangkan 25 baris pertama kolom volatile acidity dan 2 melambangkan 25 baris pertama kolom sulphates.
H_0 : Rata-rata 25 baris pertama kolom volatile acidity sama dengan rata-rata 25 baris pertama kolom sulphates $(\mu_1 = \mu_2, \mu_1 - \mu_2 = 0)$
H_1 : Rata-rata 25 baris pertama kolom volatile acidity tidak sama dengan rata-rata 25 baris pertama kolom sulphates $(\mu_1 \neq \mu_2, \mu_1 - \mu_2 \neq 0)$
Dari kalimat soal, kita dapat menganggap bahwa klaimnya adalah H_0.

Tingkat Signifikansi
$\alpha = 0.05$

Uji Statistik
Pada pengujian hipotesis ini, karena variansi populasi tidak diketahui dan banyak sampel kurang

dari 30, digunakan t-test. Dipilih kasus untuk variansi populasi yang berbeda karena diasumsikan kedua data yang berbeda kolom memiliki variansi populasi yang berbeda.

Digunakan tes statistik t dengan rumus:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

dengan derajat kebebasan:

$$v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$$

dengan daerah kritis:
$t < -t_{\alpha/2}$ atau $t > t_{\alpha/2}$ (two-tailed test)

Pengambilan Keputusan

Tes Daerah Kritis

- Reject H_0 jika $t < -t_{\alpha/2}$ atau $t > t_{\alpha/2}$
- Fail to reject H_0 jika $-t_{\alpha/2} \leq t \leq t_{\alpha/2}$

Tes Signifikansi

- Reject H_0 jika $p < \alpha$
- Fail to reject H_0 jika $p \geq \alpha$

[29]:
```
# Diketahui
alpha = 0.05
deltaMean = 0

# Ambil data
nData = 25
dataVolatileAcidity = dataAnggur["volatile acidity"][ : nData]
dataSulphates = dataAnggur["sulphates"][ : nData]

# Lakukan t-test dengan memanfaatkan library scipy untuk mendapatkan nilai t␣
↪dan p
t, p = st.ttest_ind(a=dataVolatileAcidity, b=dataSulphates, equal_var=False)

# Hitung derajat kebebasan
s1_2 = dataVolatileAcidity.var()
s2_2 = dataSulphates.var()
n1 = len(dataVolatileAcidity)
n2 = len(dataSulphates)
v = (s1_2/n1 + s2_2/n2)**2 / (((s1_2/n1)**2)/(n1-1) + ((s2_2/n2)**2)/(n2-1))

# Hitung t_alpha/2
tAlpha2 = st.t.ppf(q=1-alpha/2,df=v)
```

```python
# Tampilkan hasil
print(f"Nilai t                : {round(t, 5)}")
print(f"Nilai t_alpha/2        : {round(tAlpha2, 5)}")
print(f"Nilai p                : {round(p, 5)}")
```

```
Nilai t               : -2.63748
Nilai t_alpha/2       : 2.01593
Nilai p               : 0.01153
```

Hasil Tes

Tes Daerah Kritis
Karena $t < -t_{\alpha/2}$ $(-2.63748 < -2.01593)$, reject H_0.

Tes Signifikansi
Karena $p < \alpha$ $(0.01153 < 0.05)$, reject H_0.

Kesimpulan
Dengan tingkat signifikansi sebesar 0.05, ada bukti yang cukup untuk menolak klaim bahwa rata-rata 25 baris pertama kolom volatile acidity sama dengan rata-rata 25 baris pertama kolom sulphates.

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

Hipotesis
Misalkan 1 melambangkan bagian awal kolom residual sugar dan 2 melambangkan bagian akhir kolom residual sugar.
H_0 : Variansi bagian awal kolom residual sugar sama dengan bagian akhirnya $(\sigma_1^2 = \sigma_2^2)$
H_1 : Variansi bagian awal kolom residual sugar tidak sama dengan bagian akhirnya $(\sigma_1^2 \neq \sigma_2^2)$
Dari kalimat soal, kita dapat menganggap bahwa klaimnya adalah H_0.

Tingkat Signifikansi
$\alpha = 0.05$

Uji Statistik
Pada uji hipotesis ini, digunakan tes statistik f dengan rumus:

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

dengan daerah kritis:
$f < f_{\alpha/2}(v_1, v_2)$ atau $f > f_{1-\alpha/2}(v_1, v_2)$ (two-tailed test) dengan $v_1 = n_1 - 1$ dan $v_2 = n_2 - 1$

Pengambilan Keputusan

Tes Daerah Kritis

- Reject H_0 jika $f < f_{\alpha/2}(v_1, v_2)$ atau $f > f_{1-\alpha/2}(v_1, v_2)$
- Fail to reject H_0 jika $f_{\alpha/2}(v_1, v_2) \leq f \leq f_{1-\alpha/2}(v_1, v_2)$

Tes Signifikansi

- Reject H_0 jika $p < \alpha$
- Fail to reject H_0 jika $p \geq \alpha$

[30]:
```
# Diketahui
alpha = 0.05

# Ambil data
nData = len(dataAnggur) // 2
dataAwal = dataAnggur["residual sugar"][ : nData]
dataAkhir = dataAnggur["residual sugar"][ nData : ]

# Hitung nilai f
f = dataAwal.var() / dataAkhir.var()

# Tentukan derajat kebebasan
v1 = len(dataAwal) - 1
v2 = len(dataAkhir) - 1

# Hitung f_(1 - alpha/2) dan f_alpha/2 dengan library scipy
f1MinAlpha2 = st.f.ppf(1 - alpha/2, v1, v2)
fAlpha2 = st.f.ppf(alpha/2, v1, v2)

# Hitung nilai p, p untuk two-tailed test adalah 2 kali tail area
p = st.f.cdf(f, v1, v2) * 2

# Tampilkan hasil
print(f"Nilai f                : {round(f, 5)}")
print(f"Nilai f_(1 - alpha/2) : {round(f1MinAlpha2, 5)}")
print(f"Nilai f_alpha/2        : {round(fAlpha2, 5)}")
print(f"Nilai p                : {round(p, 5)}")
```

```
Nilai f                : 0.942
Nilai f_(1 - alpha/2) : 1.19206
Nilai f_alpha/2        : 0.83889
Nilai p                : 0.50482
```

Hasil Tes

Tes Daerah Kritis

Karena $f_{\alpha/2}(v_1, v_2) \leq f \leq f_{1-\alpha/2}(v_1, v_2)$ $(0.83889 \leq 0.942 \leq 1.19206)$, fail to reject H_0.

Tes Signifikansi
Karena $p \geq \alpha$ $(0.74759 \geq 0.05)$, fail to reject H_0.

Kesimpulan
Dengan tingkat signifikansi sebesar 0.05, tidak ada bukti yang cukup untuk menolak klaim bahwa variansi bagian awal dan akhir kolom residual sugar bernilai sama.

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

Hipotesis
Misalkan 1 melambangkan setengah bagian awal kolom alcohol yang lebih dari 7 dan 2 melambangkan setengah bagian akhir kolom kolom alcohol yang lebih dari 7.
H_0 : Proporsi nilai setengah bagian awal alcohol yang lebih dari 7 sama dengan proporsi nilai yang sama di setengah bagian akhir alcohol $(p_1 = p_2, p_1 - p_2 = 0)$
H_1 : Proporsi nilai setengah bagian awal alcohol yang lebih dari 7 lebih besar daripada proporsi nilai yang sama di setengah bagian akhir alcohol $(p_1 > p_2, p_1 - p_2 > 0)$
Dari kalimat soal, kita dapat menganggap bahwa klaimnya adalah H_1.

Tingkat Signifikansi
$\alpha = 0.05$

Uji Statistik
Pada uji hipotesis ini, digunakan tes statistik z dengan rumus:

$$z = \frac{\hat{p_1} - \hat{p_2}}{\sqrt{\hat{p}\hat{q}(1/n_1 + 1/n_2)}}$$

dengan $\hat{p}$:

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

dengan daerah kritis:
$z > z_{\alpha/2}$ (one-tailed test)

Pengambilan Keputusan

Tes Daerah Kritis

- Reject H_0 jika $z > z_\alpha$
- Fail to reject H_0 jika $z \leq z_\alpha$

Tes Signifikansi

- Reject H_0 jika $p < \alpha$
- Fail to reject H_0 jika $p \geq \alpha$

[31]:
```python
# Diketahui
alpha = 0.05
```

```python
deltaProp = 0

# Ambil data
nData = len(dataAnggur) // 2
dataAwal = dataAnggur[ : nData ]
dataAkhir = dataAnggur[ nData : ]

# Lakukan proportions z-test dengan memanfaatkan library statsmodels untuk␣
↪mendapatkan nilai z dan p
xAwal = len(dataAwal[dataAwal["alcohol"] > 7])
nAwal = len(dataAwal)
xAkhir = len(dataAkhir[dataAkhir["alcohol"] > 7])
nAkhir = len(dataAkhir)
z, _ = proportions_ztest([xAwal, xAkhir], [nAwal, nAkhir], value = deltaProp,␣
↪prop_var = deltaProp)
p = 1 - st.norm.cdf(z)

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

# Tampilkan hasil
print(f"Nilai z              : {round(z, 5)}")
print(f"Nilai z_alpha        : {round(zAlpha, 5)}")
print(f"Nilai p              : {round(p, 5)}")
```

```
Nilai z              : 0.0
Nilai z_alpha        : 1.64485
Nilai p              : 0.5
```

Hasil Tes

Tes Daerah Kritis
Karena $z \leq z_\alpha$ $(0.0 \leq 1.64485)$, fail to reject H_0.

Tes Signifikansi
Karena $p \geq \alpha$ $(1.0 \geq 0.05)$, fail to reject H_0.

Kesimpulan
Dengan tingkat signifikansi sebesar 0.05, tidak ada bukti yang cukup untuk mendukung klaim bahwa proporsi nilai setengah bagian awal alcohol yang lebih dari 7 lebih besar daripada proporsi nilai yang sama di setengah bagian akhir alcohol.