### Kelompok 8 - K01

### Anggota:

- Asyifa Nurul Shafira (13521125)
- Ferindya Aulia Berlinty (13521161)

### 5. 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 [33]: import pandas as pd
         import scipy.stats as stats
         print("Berdasarkan data diketahui:")
         data = pd.read_csv('anggur.csv')
         fixed = data['fixed acidity']
         data1 = fixed.head(fixed.size//2)
         data2 = fixed.tail(fixed.size//2)
         mean1 = data1.mean()
         mean2 = data2.mean()
         var1 = data1.var()
         var2 = data2.var()
         size1 = data1.size
         size2 = data2.size
         d0 = 0
         alpha = 0.05
         print("- Mean setengah data awal\t\t= \bar{x}1\t=", "{:.5f}".format(mean1))
         print("- Mean setengah data akhir\t\t= x2\t=", "{:.5f}".format(mean2))
         print("- Selisih rata-rata uji dari kedua bagian\t= d0\t=", d0)
         print("- Variansi setengah data awal\t\t= var1\t=", "{:.5f}".format(var1))
         print("- Variansi setengah data akhir\t\t= var2\t=", "{:.5f}".format(var2))
         print("- Banyak Data setengah bagian awal\t\t= n1\t=", "{:.0f}".format(size1))
         print("- Banyak Data setengah bagian akhir\t\t= n2\t=", "{:.0f}".format(size2))
         print("- Level Signifikan\t\t\t= α\t=", alpha)
         print("\nLangkah-langkah testing :")
         print("- Langkah 1: menentukan H0")
         print("H0 : \mu1 - \mu2 = 0")
         print("\n- Langkah 2: menentukan H1")
         print("H1 : \mu1 - \mu2 \neq 0")
         print("\n- Langkah 3: menentukan tingkat signifikasi, disini kami menggunakan 0.
         print("\alpha = 0.05")
         print("\n- Langkah 4: menentukan uji statistik yang sesuai dan daerah kritis")
         zAlpha = stats.norm.ppf(1 - alpha/2)
         print("Nilai zα :", "{:.5f}".format(zAlpha))
```

```
print("Maka, digunakan daerah kritis z < -z(\alpha/2) atau z > z(\alpha/2) : z <", "{:.5f}
 print("\n- Langkah 5: mencari nilai t-Value dan p-Value")
 z = (mean1 - mean2 - d0) / math.sqrt(var1/size1 + var2/size2)
 p = 2 * (1 - stats.norm.cdf(abs(z)))
 print("Nilai z :", "{:.5f}".format(z))
  print("Nilai p :", "{:.5f}".format(p))
 print("\n- Langkah 6: mengambil kesimpulan")
 print("Berdasarkan uji statistik didapat -za < z < za:", "{:.5f}".format(-zAlpha</pre>
 print("Kesimpulan:")
 if (z > zAlpha or z <-zAlpha):</pre>
      print("menolak H0, Nilai rata-rata kedua bagian tersebut tidak sama")
 else:
      print("gagal menolak H0, Nilai rata-rata kedua bagian tersebut sama")
 print("Berdasarkan uji p-value didapat p > \alpha :", "{:.5f}".format(p), ">", alpha
 print("Kesimpulan:")
 if p < alpha:</pre>
     print("menolak H0, Nilai p lebih kecil dari nilai signifikan sehingga rata-r
 else:
      print("gagal menolak H0, Nilai p lebih besar dari nilai signifikan sehingga
Berdasarkan data diketahui:
- Mean setengah data awal
                                                  = \bar{x}1
                                                         = 7.15352
- Mean setengah data akhir
                                                 = \bar{x}2 = 7.15154
- Selisih rata-rata uji dari kedua bagian
                                                = d0 = 0
                                                 = var1 = 1.45188
- Variansi setengah data awal
- Variansi setengah data akhir
                                                 = var2 = 1.43868
- Banyak Data setengah bagian awal
                                                = n1 = 500
- Banyak Data setengah bagian akhir
                                                = n2 = 500
- Level Signifikan
                                                  = \alpha = 0.05
Langkah-langkah testing :
- Langkah 1: menentukan H0
H0 : \mu 1 - \mu 2 = 0
- Langkah 2: menentukan H1
H1 : \mu1 - \mu2 \neq 0
- Langkah 3: menentukan tingkat signifikasi, disini kami menggunakan 0.05
\alpha = 0.05
- Langkah 4: menentukan uji statistik yang sesuai dan daerah kritis
Nilai z\alpha: 1.95996
Maka, digunakan daerah kritis z < -z(\alpha/2) atau z > z(\alpha/2) : z < -1.95996 atau z >
1.95996
- Langkah 5: mencari nilai t-Value dan p-Value
Nilai z : 0.02604
Nilai p : 0.97922
- Langkah 6: mengambil kesimpulan
Berdasarkan uji statistik didapat -za < z < za: -1.95996 < 0.02604 < 1.95996
Kesimpulan:
gagal menolak H0, Nilai rata-rata kedua bagian tersebut sama
Berdasarkan uji p-value didapat p > \alpha : 0.97922 > 0.05
Kesimpulan:
gagal menolak H0, Nilai p lebih besar dari nilai signifikan sehingga rata-rata ke
dua bagian tersebut 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 [49]: import pandas as pd
         import scipy.stats as stats
         import math
         print("Berdasarkan data diketahui:")
         data = pd.read_csv('anggur.csv')
         chlorides = data['chlorides']
         data1 = chlorides.head(fixed.size//2)
         data2 = chlorides.tail(fixed.size//2)
         mean1 = data1.mean()
         mean2 = data2.mean()
         var1 = data1.var()
         var2 = data2.var()
         size1 = data1.size
         size2 = data2.size
         d0 = 0.001
         alpha = 0.05
         print("- Mean setengah data awal\t\t\t= \bar{x}1\t=", "{:.5f}".format(mean1))
         print("- Mean setengah data akhir\t\t= x1\t=", "{:.5f}".format(mean2))
         print("- Selisih rata-rata uji dari kedua bagian\t= d0\t=", d0)
         print("- Variansi setengah data awal\t\t\= var1\t=", "{:.5f}".format(var1))
         print("- Banyak Data setengah bagian awal\t\t= n1\t=", "{:.0f}".format(size1))
         print("- Banyak Data setengah bagian akhir\t\t= n2\t=", "{:.0f}".format(size2))
         print("- Level Signifikan\t\t\t= \alpha\t=", alpha)
         print("\nTesting :")
         print("- Langkah 1: menentukan H0")
         print("H0 : \mu1 - \mu2 = 0.001")
         print("\n- Langkah 2: menentukan H1")
         print("H1 : \mu1 - \mu2 \neq 0.001")
         print("\n- Langkah 3: menentukan tingkat signifikasi, disini kami menggunakan 0.
         print("\alpha = 0.05")
         print("\n- Langkah 4: menentukan uji statistik yang sesuai dan daerah kritis")
         zAlpha = stats.norm.ppf(1 - alpha/2)
         print("Nilai zα :", "{:.5f}".format(zAlpha))
         print("Maka, digunakan daerah kritis z < -z(\alpha/2) atau z > z(\alpha/2) : z <", "{:.5f}
         print("\n- Langkah 5: mencari nilai t-Value dan p-Value")
         z = (mean1 - mean2 - d0) / math.sqrt(var1/size1 + var2/size2)
         p = 2 * (1 - stats.norm.cdf(abs(z)))
         print("Nilai uji statistik z :", "{:.5f}".format(z))
         print("Nilai p :", "{:.5f}".format(p))
         print("\n- Langkah 6: mengambil kesimpulan")
         print("Berdasarkan uji statistik didapatkan -za < z < za:", "{:.5f}".format(-zAl</pre>
         print("Kesimpulan:")
         if (z > zAlpha or z <-zAlpha):</pre>
             print("menolak H0, Nilai rata-rata bagian awal tidak lebih besar 0.001 darip
             print("gagal menolak H0, Nilai rata-rata bagian awal lebih besar 0.001 darip
         print("Dan")
         print("Berdasarkan uji p-value didapatkan p > \alpha :", "{:.5f}".format(p), ">", al
         print("Kesimpulan:")
```

```
if p < alpha:</pre>
     print("menolak H0, Nilai p lebih kecil dari nilai signifikan sehingga rata-n
 else:
     print("gagal menolak H0, Nilai p lebih besar dari nilai signifikan sehingga
Berdasarkan data diketahui:
- Mean setengah data awal
                                                = \bar{x}1 = 0.08140
- Mean setengah data akhir
                                                = \bar{x}1 = 0.08099
- Selisih rata-rata uji dari kedua bagian = d0 = 0.001
- Variansi setengah data awal
                                                = var1 = 0.00041
- Variansi setengah data akhir
                                               = var2 = 0.00040
- Banyak Data setengah bagian awal
                                                = n1 = 500
- Banyak Data setengah bagian akhir
                                               = n2 = 500
- Level Signifikan
                                                 = \alpha = 0.05
Testing:
- Langkah 1: menentukan H0
H0: \mu 1 - \mu 2 = 0.001
- Langkah 2: menentukan H1
H1 : \mu1 - \mu2 \neq 0.001
- Langkah 3: menentukan tingkat signifikasi, disini kami menggunakan 0.05
\alpha = 0.05
- Langkah 4: menentukan uji statistik yang sesuai dan daerah kritis
Nilai z\alpha : 1.95996
Maka, digunakan daerah kritis z < -z(\alpha/2) atau z > z(\alpha/2) : z < -1.95996 atau z >
1.95996
- Langkah 5: mencari nilai t-Value dan p-Value
Nilai uji statistik z : -0.46732
Nilai p : 0.64027
- Langkah 6: mengambil kesimpulan
Berdasarkan uji statistik didapatkan -za < z < za: -1.95996 < -0.46732 < 1.95996
Kesimpulan:
gagal menolak HO, Nilai rata-rata bagian awal lebih besar 0.001 daripada bagian a
khir
Berdasarkan uji p-value didapatkan p > \alpha : 0.64027 > 0.05
Kesimpulan:
gagal menolak H0, Nilai p lebih besar dari nilai signifikan sehingga rata-rata ba
gian awal lebih besar 0.001 daripada bagian akhir
```

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

```
In [50]: import pandas as pd
import scipy.stats as stats

print("Berdasarkan data diketahui:")
data = pd.read_csv('anggur.csv')
volatile = data['volatile acidity']
data1 = volatile.head(25)
Sulphates = data['sulphates']
data2 = Sulphates.head(25)
```

```
mean1 = data1.mean()
mean2 = data2.mean()
var1 = data1.var()
var2 = data2.var()
size1 = data1.size
size2 = data2.size
d\theta = \theta
alpha = 0.05
print("- Mean 25 data awal volatile acidity\t= \bar{x}1\t=", "{:.5f}".format(mean1))
print("- Mean 25 data awal sulphates\t\t\t= x1\t=", "{:.5f}".format(mean2))
print("- Selisih rata-rata uji kedua data\t\t= d0\t=", d0)
print("- Variansi 25 data awal volatile acidity\t= var1\t=", "{:.5f}".format(var
print("- Variansi 25 data awal sulphates\t\t= var2\t=", "{:.5f}".format(var2))
print("- Banyak Data volatile acidity\t\t= n1\t=", "{:.0f}".format(size1))
print("- Banyak Data sulphates\t\t\t= n2\t=", "{:.0f}".format(size2))
print("- Level Signifikan\t\t\t= \alpha\t=", alpha)
print("\nTesting :")
print("- Langkah 1: menentukan H0")
print("H0 : \mu1 - \mu2 = 0")
print("\n- Langkah 2: menentukan H1")
print("H1 : \mu1 - \mu2 \neq 0")
print("\n- Langkah 3: menentukan tingkat signifikasi, disini kami menggunakan 0.
print("\alpha = 0.05")
print("\n- Langkah 4: menentukan uji statistik yang sesuai dan daerah kritis")
zAlpha = stats.norm.ppf(1 - alpha/2)
print("Nilai zα :", "{:.5f}".format(zAlpha))
print("Maka, digunakan daerah kritis z > z(\alpha/2) atau z < -z(\alpha/2) : z >", "{:.5f}
print("\n- Langkah 5: mencari nilai t-Value dan p-Value")
z = (mean1 - mean2 - d0) / math.sqrt(var1/size1 + var2/size2)
p = 2 * (1 - stats.norm.cdf(abs(z)))
print("Nilai uji statistik z :", "{:.5f}".format(z))
print("Nilai p :", "{:.5f}".format(p))
print("\n- Langkah 6: mengambil kesimpulan")
print("Berdasarkan uji statistik didapatkan z < -za:", "{:.5f}".format(z), "<",</pre>
print("Kesimpulan:")
if (z > zAlpha or z <-zAlpha):</pre>
    print("menolak H0, Nilai rata-rata sampel 25 baris pertama Volatile Acidity
else:
    print("gagal menolak H0, Nilai rata-rata sampel 25 baris pertama Volatile Ac
print("Dan")
print("Berdasarkan uji p-value didapatkan p < \alpha :", "\{:.5f\}".format(p), "<", al
print("Kesimpulan:")
if p < alpha:</pre>
    print("menolak H0, Nilai rata-rata sampel 25 baris pertama Volatile Acidity
else:
    print("gagal menolak H0, Nilai rata-rata sampel 25 baris pertama Volatile Ac
```

```
Berdasarkan data diketahui:
- Mean 25 data awal volatile acidity
                                             = \bar{x}1 = 0.50142
- Selisih rata-rata uji kedua data = 0.57680
- Variansi 25 data awal volatile acidity = var1 = 0.00695
- Variansi 25 data awal sulphates
- Banyak Data volatile acidity
                                                   = n1 = 25
- Banyak Data sulphates
                                                   = n2 = 25
                                                    = \alpha = 0.05
- Level Signifikan
Testing:
- Langkah 1: menentukan H0
H0 : \mu 1 - \mu 2 = 0
- Langkah 2: menentukan H1
H1 : \mu1 - \mu2 \neq 0
- Langkah 3: menentukan tingkat signifikasi, disini kami menggunakan 0.05
\alpha = 0.05
- Langkah 4: menentukan uji statistik yang sesuai dan daerah kritis
Nilai z\alpha : 1.95996
Maka, digunakan daerah kritis z > z(\alpha/2) atau z < -z(\alpha/2) : z > 1.95996 atau z <
-1.95996
- Langkah 5: mencari nilai t-Value dan p-Value
Nilai uji statistik z : -2.63748
Nilai p: 0.00835
- Langkah 6: mengambil kesimpulan
Berdasarkan uji statistik didapatkan z < -za: -2.63748 < -1.95996
Kesimpulan:
menolak H0, Nilai rata-rata sampel 25 baris pertama Volatile Acidity tidak sama d
engan rata-rata 25 baris pertama Sulphates
Berdasarkan uji p-value didapatkan p < \alpha : 0.00835 < 0.05
Kesimpulan:
menolak H0, Nilai rata-rata sampel 25 baris pertama Volatile Acidity tidak sama d
engan rata-rata 25 baris pertama Sulphates
```

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

```
import pandas as pd
import scipy.stats as stats

print("Berdasarkan data diketahui:")
data = pd.read_csv('anggur.csv')
residual = data['residual sugar']
data1 = residual.head(residual.size//2)
data2 = residual.tail(residual.size//2)
var1 = data1.var()
var2 = data2.var()
size1 = data1.size
size2 = data2.size
print("- Variansi bagian awal\t\t\t\t= s1^(2) =", "{:.5f}".format(var1))
print("- Variansi bagian akhir\t\t\t\t= s2^(2) =", "{:.5f}".format(var2))
print("- Banyak Data residual sugar bagian awal\t= n1\t=", "{:.0f}".format(size)
```

```
print("- Banyak Data residual sugar bagian akhir\t= n2\t =", "{:.0f}".format(siz
print("\nTesting :")
print("- Langkah 1: menentukan H0")
print("H0 : var1 - var2 = 0")
print("\n- Langkah 2: menentukan H1")
print("H1 : var1 - var2 ≠ 0\t(Distribusi f)")
print("\n- Langkah 3: menentukan tingkat signifikasi, disini kami menggunakan 0.
print("\alpha = 0.05")
print("\n- Langkah 4: menentukan uji statistik yang sesuai dan daerah kritis")
print("Uji statistik menggunakan distribusi F:")
print("Daerah kritis : f < f1 - \alpha/2(v1,v2) or f > f\alpha/2(v1,v2)")
print("\n- Langkah 5: mencari nilai t-Value dan p-Value")
if var1 > var2:
    f = var1/var2
else:
    f = var2/var1
f1 = stats.f.ppf(q = 1 - alpha/2, dfn = size1 - 1, dfd = size2 - 1)
f2 = 1 / stats.f.ppf(q = 1 - alpha/2, dfn = size2 - 1, dfd = size1 - 1)
p = 2 * (1 - stats.f.cdf(f, size2 - 1, size1 - 1))
print("Nilai f : {:.5f}".format(f))
print("p-value : {:.5f}".format(p))
print("\n- Langkah 6: mengambil kesimpulan")
print("Berdasarkan uji distribusi f didapatkan f < f1 - \alpha/2(v1,v2):", "{:.5f}"
if(f > f1 or f < f2):
    print("menolak H0, Bagian awal residual sugar tidak memiliki variansi yang s
else:
    print("gagal menolak H0, Bagian awal residual sugar memiliki variansi yang s
print("Dan")
print("Berdasarkan uji p-value didapatkan : ")
if(p < alpha):</pre>
    print("menolak H0, Bagian awal residual sugar tidak memiliki variansi yang s
else:
    print("gagal menolak H0, Bagian awal residual sugar memiliki variansi yang s
```

```
Berdasarkan data diketahui:
- Variansi bagian awal
                                                  = s1^{(2)} = 0.94777
- Variansi bagian akhir
                                                 = s2^{(2)} = 1.00612
- Banyak Data residual sugar bagian awal = n1 = 500

- Banyak Data residual sugar bagian akhir = n2 = 500
Testing:
- Langkah 1: menentukan H0
H0 : var1 - var2 = 0
- Langkah 2: menentukan H1
H1 : var1 - var2 ≠ 0 (Distribusi f)
- Langkah 3: menentukan tingkat signifikasi, disini kami menggunakan 0.05
\alpha = 0.05
- Langkah 4: menentukan uji statistik yang sesuai dan daerah kritis
Uji statistik menggunakan distribusi F:
Daerah kritis : f < f1 - \alpha/2(v1,v2) or f > f\alpha/2(v1,v2)
- Langkah 5: mencari nilai t-Value dan p-Value
Nilai f : 1.06157
p-value : 0.50482
- Langkah 6: mengambil kesimpulan
Berdasarkan uji distribusi f didapatkan f < f1 - \alpha/2(v1,v2) : 1.06157 < 1.19206
gagal menolak HO, Bagian awal residual sugar memiliki variansi yang sama dengan b
agian akhirnya
Berdasarkan uji p-value didapatkan :
gagal menolak HO, Bagian awal residual sugar memiliki variansi yang sama dengan b
agian akhirnya
```

## 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 [56]:
         import pandas as pd
         import scipy.stats as stats
         print("Berdasarkan data diketahui:")
         data = pd.read_csv('anggur.csv')
         alcohol = data['alcohol']
         data1 = alcohol.head(alcohol.size//2)
         data2 = alcohol.tail(alcohol.size//2)
         x1 = data1.loc[data1 > 7]
         x2 = data2.loc[data2 > 7]
         size1 = data1.size
         size2 = data2.size
         p = (x1.size + x2.size)/(size1 + size2)
         q = 1 - p
         p1 = x1.size/size1
         p2 = x2.size/size2
         d\theta = \theta
         alpha = 0.05
         zAlpha = stats.norm.ppf(1 - alpha/2)
```

```
print("- Nilai p \t\t= p =", "{:.5f}".format(p))
print("- Nilai p1 \t\t= p1 =", "{:.5f}".format(p1))
print("- Nilai p2 \t\t= p2 =", "{:.5f}".format(p2))
print("- Nilai selisih \t= d0 =", d0)
print("- Level Signifikan\t= \alpha =", alpha)
print("\nTesting :")
print("- Langkah 1: menentukan H0")
print("H0 : p1 - p2 = 0")
print("\n- Langkah 2: menentukan H1")
print("H1 : p1 - p2 \neq 0")
print("\n- Langkah 3: menentukan tingkat signifikasi, disini kami menggunakan 0.
print("\alpha = 0.05")
print("\n- Langkah 4: menentukan uji statistik yang sesuai dan daerah kritis")
print("Uji statistik: Uji statistik satu parameter populasi pengujian proporsi d
print("Daerah kritis : z > zα")
print("\n- Langkah 5: mencari nilai t-Value dan p-Value")
z = (p1 - p2)/(math.sqrt((p*q/size1) + (p*q/size2)))
p = 1 - stats.norm.cdf(abs(z))
print("Nilai z : {:.5f}".format(z))
print("p-value : {:.5f}".format(p))
print("\n- Langkah 6: mengambil kesimpulan")
print("Berdasarkan uji statistik didapatkan -za < z < za:", "{:.5f}".format(-zAl</pre>
if z > zAlpha:
    print("H0 ditolak, Proporsi nilai setengah bagian awal alcohol yang lebih da
else:
    print("H0 gagal ditolak, Proporsi nilai setengah bagian awal alcohol yang le
print("Dan")
print("Berdasarkan uji p-value :")
if p < alpha:</pre>
   print("HO ditolak, Proporsi nilai setengah bagian awal alcohol yang lebih da
    print("H0 gagal ditolak, Proporsi nilai setengah bagian awal alcohol yang le
```

Berdasarkan data diketahui:

#### Testing:

- Langkah 1: menentukan H0

H0: p1 - p2 = 0

- Langkah 2: menentukan H1

 $H1 : p1 - p2 \neq 0$ 

- Langkah 3: menentukan tingkat signifikasi, disini kami menggunakan 0.05  $\alpha$  = 0.05

- Langkah 4: menentukan uji statistik yang sesuai dan daerah kritis Uji statistik: Uji statistik satu parameter populasi pengujian proporsi dua sampe l one-tailed tes normal, z Daerah kritis : z > z $\alpha$ 

- Langkah 5: mencari nilai t-Value dan p-Value

Nilai z : 0.00000 p-value : 0.50000

- Langkah 6: mengambil kesimpulan

Berdasarkan uji statistik didapatkan -za < z < za: -1.95996 < 0.00000 < 1.95996 H0 gagal ditolak, Proporsi nilai setengah bagian awal alcohol yang lebih dari 7 t idak lebih besar daripada proporsi nilai yang sama di setengah bagian akhir alcoh ol

Dan

Berdasarkan uji p-value :

H0 gagal ditolak, Proporsi nilai setengah bagian awal alcohol yang lebih dari 7 t idak lebih besar daripada proporsi nilai yang sama di setengah bagian akhir alcoh ol