

02-T1-IF2220-13519081

April 15, 2021

## Tugas Besar IF2220 Probabilitas dan Statistika Penarikan Kesimpulan dan Pengujian Hipotesis

Oleh:

- Isabella Handayani Sumantri 13519081
- Girvin Junod 13519096

```
[325]: #setup
import pandas as pd
import matplotlib.pyplot as plt
import scipy.stats as s
import seaborn as sns
import math

def zscore(x, mean, std, n):
    return (x-mean)*math.sqrt(n)/std
def ptoz(p):
    return s.norm.ppf(p)
def ztop(z):
    return s.norm.cdf(z)
def ztwotestmean(d0, x1, x2, var1, var2, n1, n2):
    return ((x1 - x2) - d0)/math.sqrt((var1/n1) + (var2/n2))

alpha = 0.05 #nilai alpha
colnames = ['id', 'daerah', 'sumbu utama', 'sumbu kecil', 'keunikan', 'area_
↳bulatan', 'diameter', 'kadar air', 'keliling', 'bulatan', 'ransum', 'kelas']
url = 'https://drive.google.com/file/d/1z0EyNb2c5PD-ECNgsS_6f7KV6bm2ya10/view?
↳usp=sharing' # copy of original dataset, in personal drive
path = 'https://drive.google.com/uc?export=download&id='+url.split('/')[2]
g = pd.read_csv(path, header=None)
g.columns = colnames
gandum = g[g.columns[1:11]] #dataset selain kolom id dan kelas
```

## 1 Descriptive Statistics

### 1.1 Mean

```
[326]: gandum.mean() #mean untuk kolom daerah-ransum
```

```
[326]: daerah          4801.246000  
       sumbu utama     116.045171  
       sumbu kecil     53.715246  
       keunikan        0.878764  
       area bulatan    4937.048000  
       diameter        77.771158  
       kadar air       0.648372  
       keliling        281.479722  
       bulatan         0.761737  
       ransum          2.150915  
       dtype: float64
```

## 1.2 Median

---

```
[327]: gandum.median() #median untuk kolom daerah-ransum
```

```
[327]: daerah          4735.000000  
       sumbu utama     115.405140  
       sumbu kecil     53.731199  
       keunikan        0.890045  
       area bulatan    4857.000000  
       diameter        77.645277  
       kadar air       0.626117  
       keliling        280.045500  
       bulatan         0.761288  
       ransum          2.193599  
       dtype: float64
```

## 1.3 Modus

---

```
[328]: for i in gandum.columns:  
       mod = gandum[i].mode()  
       print("Modus " + i + ": ", end="")  
       if(len(mod) != 500): #jika tidak semua modus  
           for j in range(len(mod)):  
               if j != len(mod) -1:  
                   print(mod[j], end= ", ")  
               else:  
                   print(mod[j])  
       else:  
           print("Semua data muncul sekali") #semua data unik, semua modus  
       print()
```

Modus daerah: 3992, 4881, 5642, 6083

Modus sumbu utama: Semua data muncul sekali

Modus sumbu kecil: Semua data muncul sekali

Modus keunikan: Semua data muncul sekali

Modus area bulatan: 3802, 4913

Modus diameter: 71.29356396, 78.83325579, 84.75622403, 88.00634154

Modus kadar air: 0.735849057, 0.824404762

Modus keliling: Semua data muncul sekali

Modus bulatan: Semua data muncul sekali

Modus ransum: Semua data muncul sekali

## 1.4 Standar Deviasi

---

```
[329]: gandum.std() #standar deviasi untuk kolom daerah-ransum
```

```
[329]: daerah          986.395491
      sumbu utama    18.282626
      sumbu kecil     4.071075
      keunikan        0.036586
      area bulatan   1011.696255
      diameter        8.056867
      kadar air       0.094367
      keliling       37.335402
      bulatan         0.061702
      ransum          0.249767
      dtype: float64
```

## 1.5 Variansi

---

```
[330]: gandum.var() # variansi untuk kolom daerah-ransum
```

```
[330]: daerah          9.729761e+05
      sumbu utama     3.342544e+02
      sumbu kecil     1.657365e+01
```

```
keunikan      1.338528e-03
area bulatan  1.023529e+06
diameter      6.491311e+01
kadar air     8.905149e-03
keliling     1.393932e+03
bulatan       3.807194e-03
ransum        6.238350e-02
dtype: float64
```

## 1.6 Range

---

```
[331]: gandum.max() - gandum.min() #range untuk kolom daerah-ransum
```

```
[331]: daerah      4931.000000
sumbu utama    153.795469
sumbu kecil    29.071182
keunikan       0.194085
area bulatan   5141.000000
diameter      40.747172
kadar air      0.468972
keliling      291.822000
bulatan        0.730158
ransum         1.024013
dtype: float64
```

## 1.7 Min

---

```
[332]: gandum.min() #nilai minimum untuk kolom daerah-ransum
```

```
[332]: daerah      2522.000000
sumbu utama     74.133114
sumbu kecil     39.906517
keunikan        0.719916
area bulatan    2579.000000
diameter       56.666658
kadar air       0.409927
keliling       197.015000
bulatan         0.174590
ransum          1.440796
dtype: float64
```

## 1.8 Max

---

```
[333]: gandum.max() #nilai maksimum untuk kolom daerah-ransum
```

```
[333]: daerah          7453.000000
sumbu utama         227.928583
sumbu kecil         68.977700
keunikan            0.914001
area bulatan        7720.000000
diameter            97.413830
kadar air           0.878899
keliling            488.837000
bulatan             0.904748
ransum              2.464809
dtype: float64
```

## 1.9 Kuartil

```
[334]: gandum1 = pd.DataFrame(data = gandum, columns = ['daerah','sumbu utama', 'sumbu_
↳kecil', 'keunikan', 'area bulatan'])
gandum2 = pd.DataFrame(data = gandum, columns = ['diameter','kadar air',_
↳'keliling', 'bulatan', 'ransum'])
print(gandum1.quantile([0.25,0.5,0.75])) #kuartil (quantil 0.25, 0.5, 0.75)_
↳untuk kolom daerah-areabulatan
print()
print(gandum2.quantile([0.25,0.5,0.75])) #kuartil (quantil 0.25, 0.5, 0.75)_
↳untuk kolom diameter-ransum
```

	daerah	sumbu utama	sumbu kecil	keunikan	area bulatan
0.25	4042.75	104.116098	51.193576	0.863676	4170.25
0.50	4735.00	115.405140	53.731199	0.890045	4857.00
0.75	5495.50	129.046792	56.325158	0.907578	5654.25

	diameter	kadar air	keliling	bulatan	ransum
0.25	71.745308	0.572632	255.8830	0.731991	1.983939
0.50	77.645277	0.626117	280.0455	0.761288	2.193599
0.75	83.648598	0.726633	306.0625	0.796361	2.381612

## 1.10 IQR

```
[335]: gandum.quantile(0.75) - gandum.quantile(0.25) #IQR untuk kolom daerah-ransum
```

```
[335]: daerah          1452.750000
sumbu utama         24.930694
sumbu kecil         5.131582
keunikan            0.043902
```

```
area bulatan      1484.000000
diameter          11.903290
kadar air         0.154001
keliling          50.179500
bulatan           0.064370
ransum            0.397673
dtype: float64
```

### 1.11 Skewness

---

```
[336]: gandum.skew() #Skew untuk kolom daerah-ransum
```

```
[336]: daerah          0.238144
sumbu utama        0.761529
sumbu kecil       -0.010828
keunikan          -1.623472
area bulatan       0.257560
diameter           0.002725
kadar air          0.493661
keliling           0.733627
bulatan           -3.599237
ransum            -0.658188
dtype: float64
```

### 1.12 Kurtosis

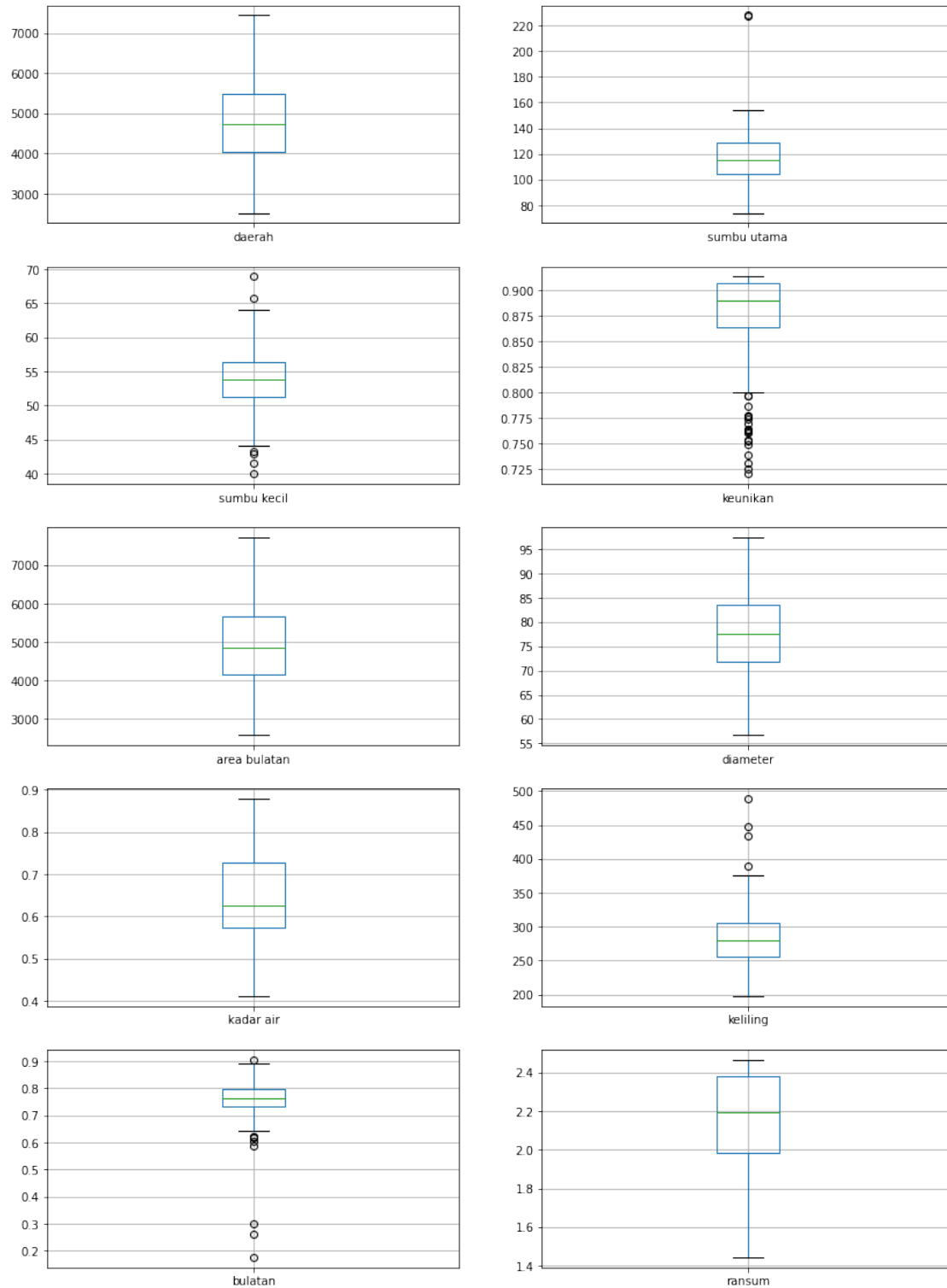
---

```
[337]: gandum.kurtosis() # Kurtosis untuk kolom daerah-ransum
```

```
[337]: daerah          -0.434631
sumbu utama         4.330534
sumbu kecil         0.475568
keunikan            2.917256
area bulatan       -0.409685
diameter           -0.466455
kadar air          -0.740326
keliling           2.272685
bulatan            29.975096
ransum            -0.428656
dtype: float64
```

## 2 Visualisasi

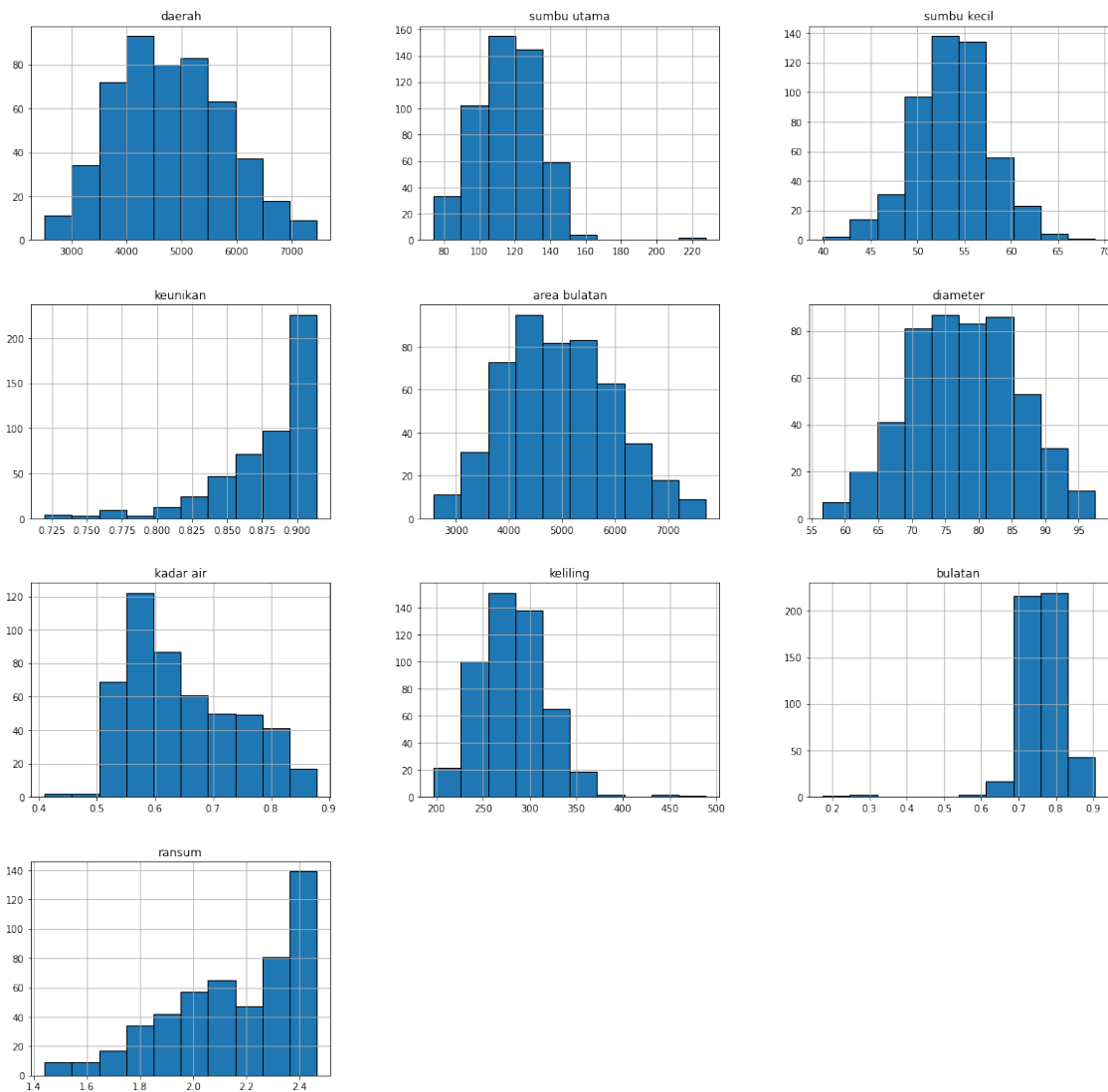
```
[338]: fig, axes = plt.subplots(nrows = 5, ncols=2, figsize=(14,20))
row = 0
for i, col in enumerate(gandum.columns):
    gandum.boxplot(column=col, ax=axes[row, i%2])
    if i % 2 == 1:
        row += 1
plt.show()
#Menunjukkan boxplot untuk kolom daerah - ransom
```



```
[339]: gandum.hist(figsize=(20, 20), edgecolor='black')
        #Menunjukkan histogram untuk kolom daerah - ransum
```

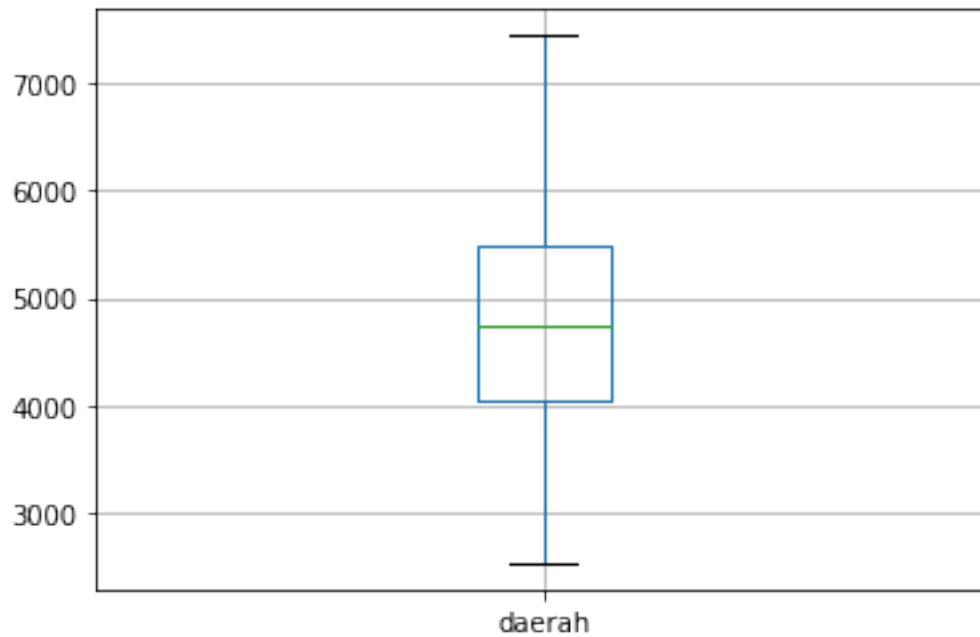


```
[339]: array([[<matplotlib.axes._subplots.AxesSubplot object at 0x7f6cf8509110>,
<matplotlib.axes._subplots.AxesSubplot object at 0x7f6d02473e50>,
<matplotlib.axes._subplots.AxesSubplot object at 0x7f6cfcf8dfd0>],
[<matplotlib.axes._subplots.AxesSubplot object at 0x7f6cfe516150>,
<matplotlib.axes._subplots.AxesSubplot object at 0x7f6d00e72690>,
<matplotlib.axes._subplots.AxesSubplot object at 0x7f6cfe679950>],
[<matplotlib.axes._subplots.AxesSubplot object at 0x7f6d01c7f1d0>,
<matplotlib.axes._subplots.AxesSubplot object at 0x7f6cfe438750>,
<matplotlib.axes._subplots.AxesSubplot object at 0x7f6cfe709c10>],
[<matplotlib.axes._subplots.AxesSubplot object at 0x7f6cfebbe650>,
<matplotlib.axes._subplots.AxesSubplot object at 0x7f6cfd12f350>,
<matplotlib.axes._subplots.AxesSubplot object at 0x7f6cfe8bf490>]],
dtype=object)
```

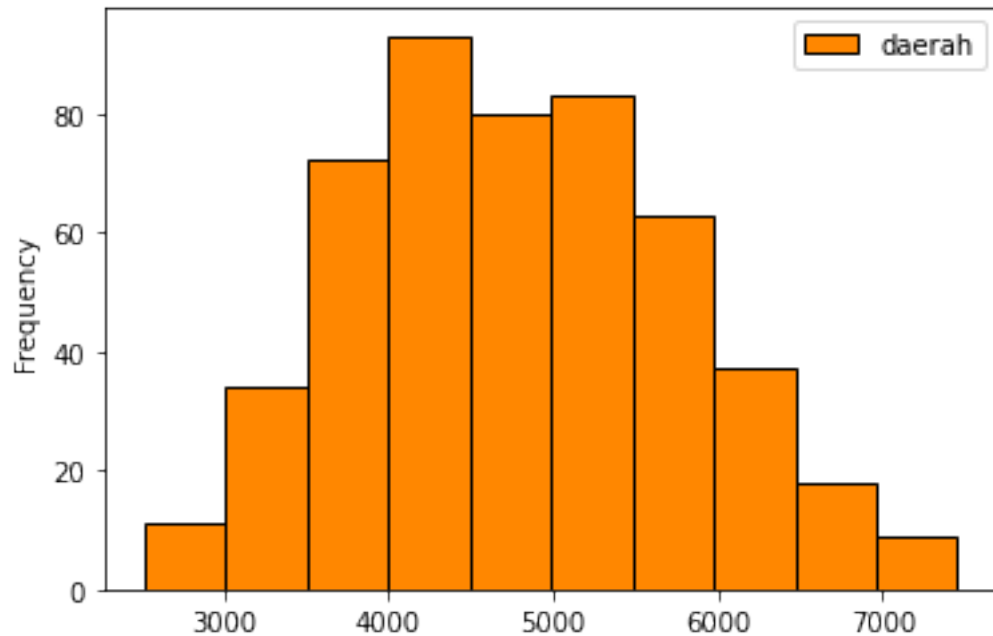


## 2.1 Daerah

```
[340]: df = pd.DataFrame(data = gandum, columns = ['daerah'])  
       boxplot(column = ['daerah'])
```



```
[341]: df = pd.DataFrame(data = gandum, columns = ['daerah'])  
       hist = df.plot.hist(bins = 10, color = '#FF8700', edgecolor='black')
```



Uraian:

Tidak terdapat outlier pada boxplot karena tidak ada nilai yang kurang dari  $Q1 - 1.5IQR$  dan lebih dari  $Q3 + 1.5IQR$  1452.75

Lower tail menandakan nilai  $Q1 - 1.5IQR$  dan Upper tail menandakan nilai  $Q3 + 1.5IQR$ . Posisi kedua tail sesuai karena berdasarkan perhitungan lower tail : 2179.125 dan upper tail 7674.625

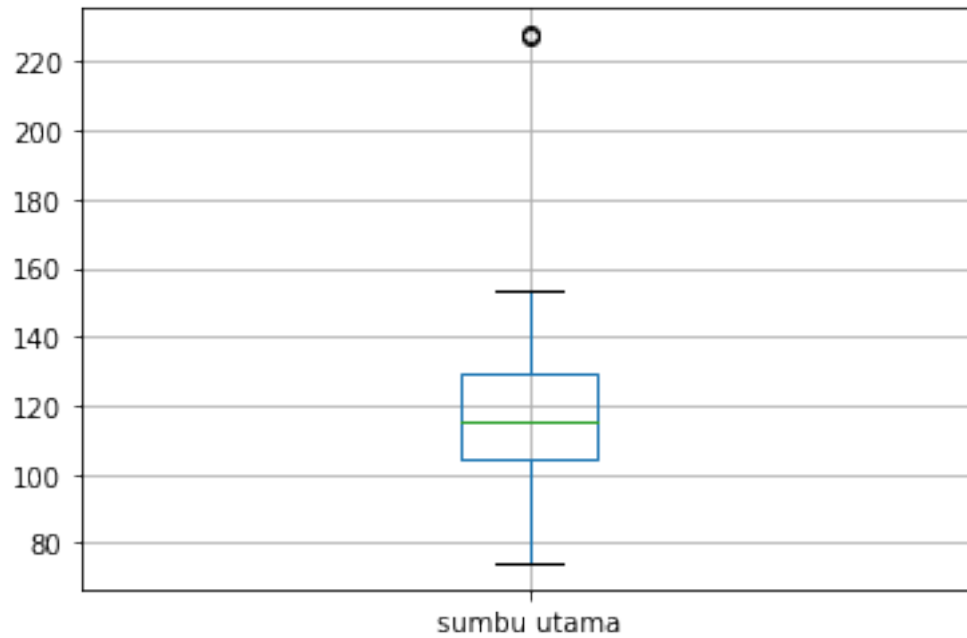
Pada box ditunjukkan  $Q1$ ,  $Q2$ ,  $Q3$ .  $Q1$  adalah bagian bawah dari box,  $Q2$  adalah bagian tengah dan juga median,  $Q3$  adalah bagian atas dari box. Posisi ketiga kuartil sesuai karena  $Q1$ : 4042.75,  $Q2$ : 4735,  $Q3$ : 5495.50. Jarak antara Bagian atas dan bagian bawah adalah nilai IQR.

Histogram memiliki positive skew karena memiliki kecenderungan ke kiri. memiliki tipe kurtosis platykurtic karena bentuknya lebih lebar dari histogram yang berdistribusi normal.

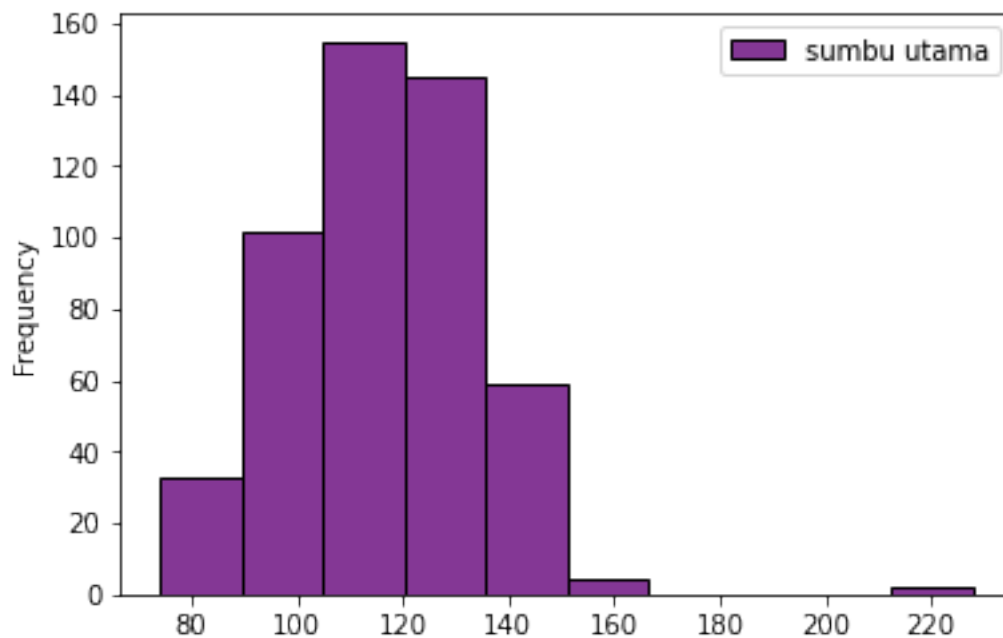
## 2.2 Sumbu Utama

---

```
[342]: df = pd.DataFrame(data = gandum, columns = ['sumbu utama'])
        boxplot = df.boxplot(column = ['sumbu utama'])
```



```
[343]: df = pd.DataFrame(data = gandum, columns = ['sumbu utama'])
hist = df.plot.hist(bins = 10, color = '#843795', edgecolor='black')
```



Uraian:

Terdapat outlier pada boxplot hal ini dapat dibuktikan dengan adanya nilai max 227.928583 yang melebihi Upper tail :  $Q3 + 1.5 \cdot IQR = 166.442833$

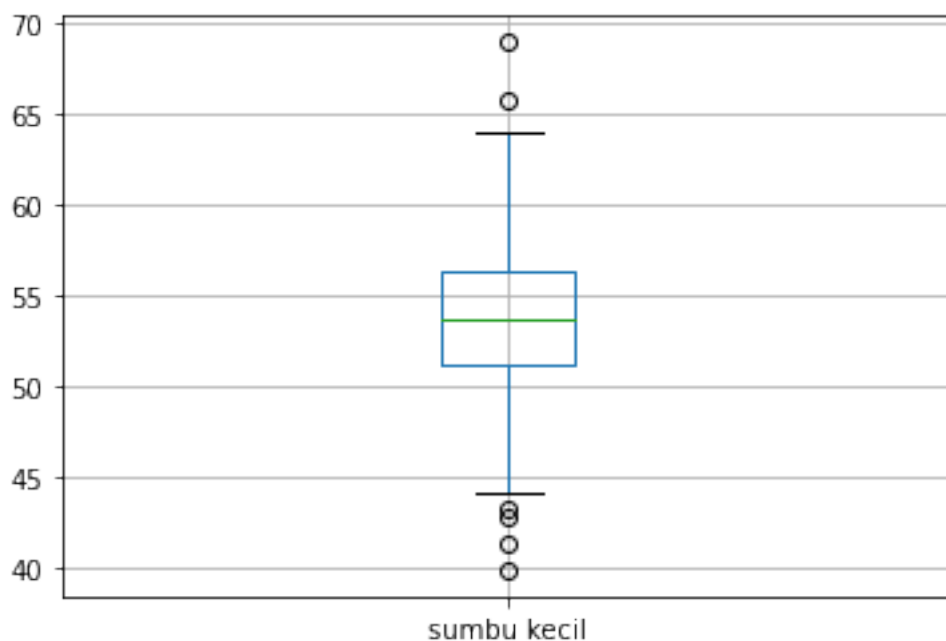
Lower tail menandakan nilai  $Q1 - 1.5 \cdot IQR$  dan Upper tail menandakan nilai  $Q3 + 1.5 \cdot IQR$ . Posisi kedua tail sesuai karena berdasarkan perhitungan lower tail : 66.720058 dan upper tail 166.442833

Pada box ditunjukkan  $Q1$ ,  $Q2$ ,  $Q3$ .  $Q1$  adalah bagian bawah dari box,  $Q2$  adalah bagian tengah dan juga median,  $Q3$  adalah bagian atas dari box. Posisi ketiga kuartil sesuai karena  $Q1$ : 104.116098,  $Q2$ : 115.405140,  $Q3$ : 129.046792. Jarak antara Bagian atas dan bagian bawah adalah nilai IQR.

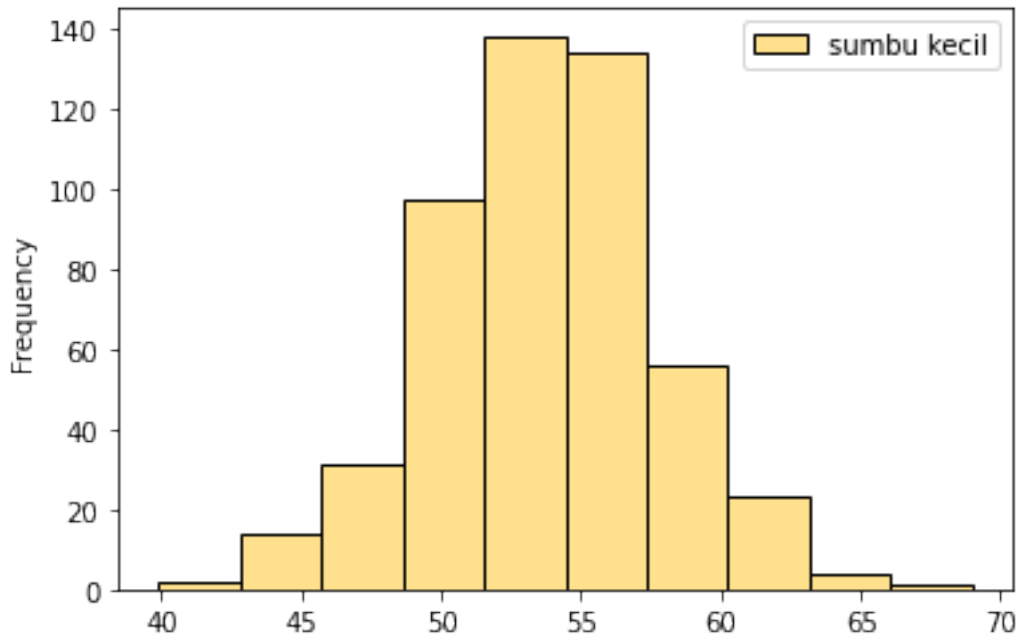
Histogram memiliki positive skew karena memiliki kecendrungan ke kiri. Memiliki tipe kurtosis leptokurtic karena bentuknya lebih ramping dari histogram yang berdistribusi normal. Terdapat outlier di ujung kanan

## 2.3 Sumbu Kecil

```
[344]: df = pd.DataFrame(data = gandum, columns = ['sumbu kecil'])  
boxplot = df.boxplot(column = ['sumbu kecil'])
```



```
[345]: df = pd.DataFrame(data = gandum, columns = ['sumbu kecil'])  
hist = df.plot.hist(bins = 10, color = '#ffe08d', edgecolor='black')
```



Uraian: Terdapat outlier pada boxplot hal ini dapat dibuktikan dengan adanya nilai max dan min yang lebih dari upper dan kurang dari lower tail.

Lower tail menandakan nilai  $Q1 - 1.5IQR$  dan Upper tail menandakan nilai  $Q3 + 1.5IQR$ . Posisi kedua tail sesuai karena berdasarkan perhitungan lower tail : 43.496203 dan upper tail : 64.022531

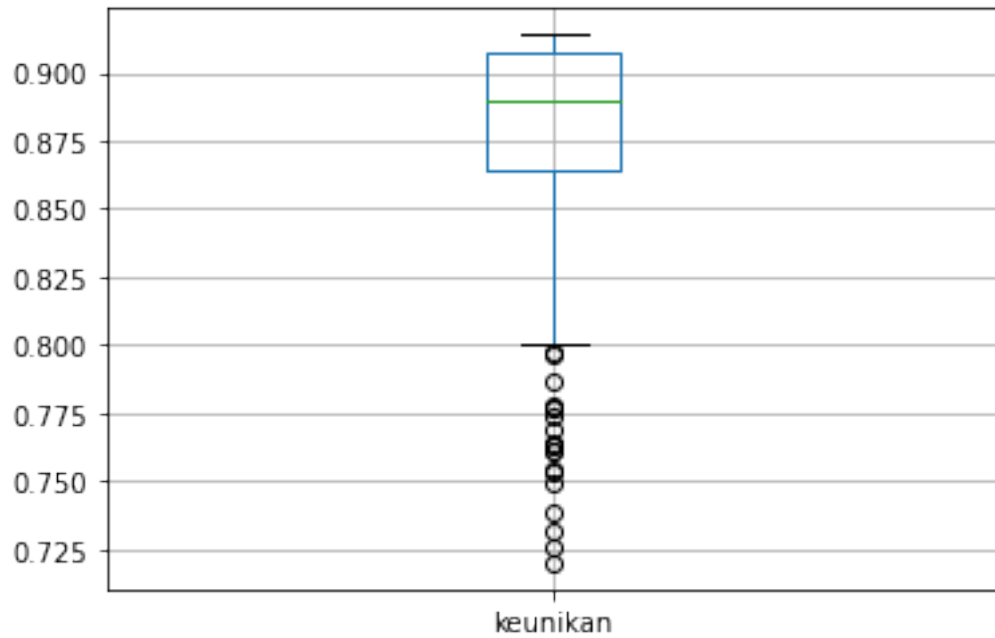
Pada box ditunjukkan Q1, Q2, Q3. Q1 adalah bagian bawah dari box, Q2 adalah bagian tengah dan juga median, Q3 adalah bagian atas dari box. Posisi ketiga kuartil sesuai karena Q1: 51.193576, Q2: 53.731199, Q3: 56.325158 . Jarak antara Bagian atas dan bagian bawah adalah nilai IQR.

Histogram memiliki skew mendekati 0. Memiliki kurtosis mesokurtic karena nilai kurtosis mendekati 0. Histogram menyerupai bell curve dan kolom sumbu kecil dapat dianggap berdistribusi normal karena memiliki nilai skew dan kurtosis yang mendekati 0 (di pandas, kurtosis normal bernilai 0).

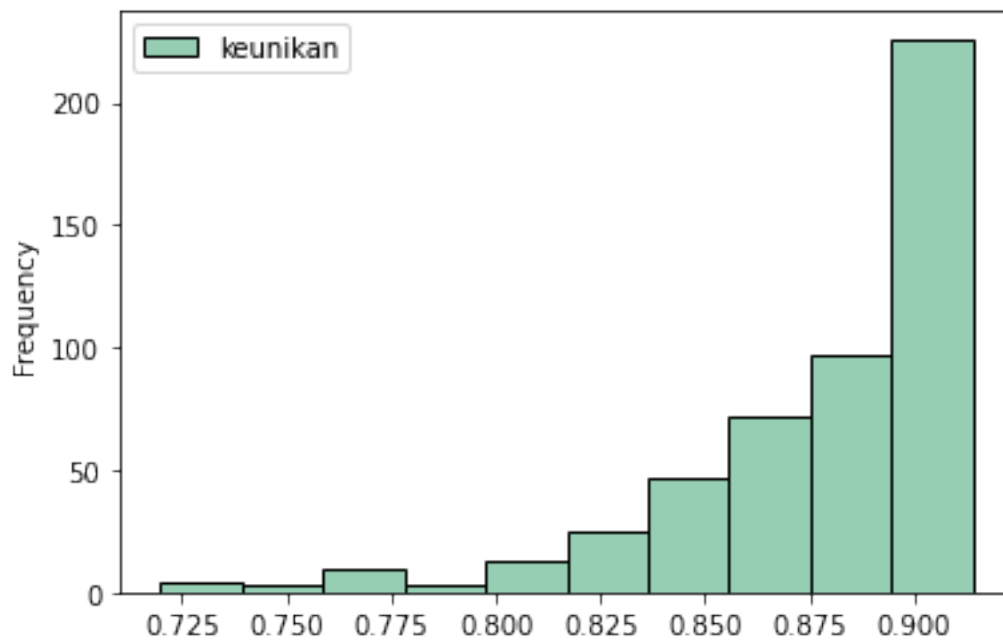
## 2.4 Keunikan

---

```
[346]: df = pd.DataFrame(data = gandum, columns = ['keunikan'])
        boxplot = df.boxplot(column = ['keunikan'])
```



```
[347]: df = pd.DataFrame(data = gandum, columns = ['keunikan'])
hist = df.plot.hist(bins = 10, color = '#96ceb4', edgecolor='black')
```



Uraian:

Terdapat outlier pada boxplot hal ini dapat dibuktikan dengan adanya nilai min yang kurang dari lower tail

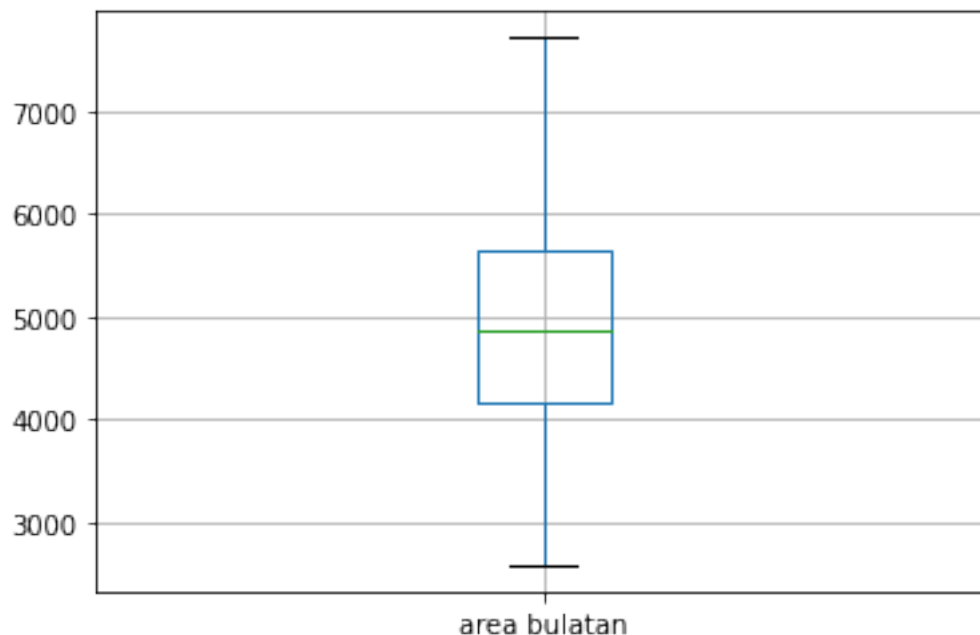
Lower tail menandakan nilai  $Q1 - 1.5IQR$  dan *Upper tail* menandakan nilai  $Q3 + 1.5IQR$ . Posisi kedua tail sesuai karena berdasarkan perhitungan lower tail : 0.797823 dan upper tail : 0.973431

Pada box ditunjukkan Q1, Q2, Q3. Q1 adalah bagian bawah dari box, Q2 adalah bagian tengah dan juga median, Q3 adalah bagian atas dari box. Posisi ketiga kuartil sesuai karena Q1: 0.863676, Q2: 0.890045, Q3: 0.907578. Jarak antara Bagian atas dan bagian bawah adalah nilai IQR.

Histogram memiliki negative skew. Memiliki kurtosis leptokurtic karena bentuknya lebih ramping dari histogram yang berdistribusi normal.

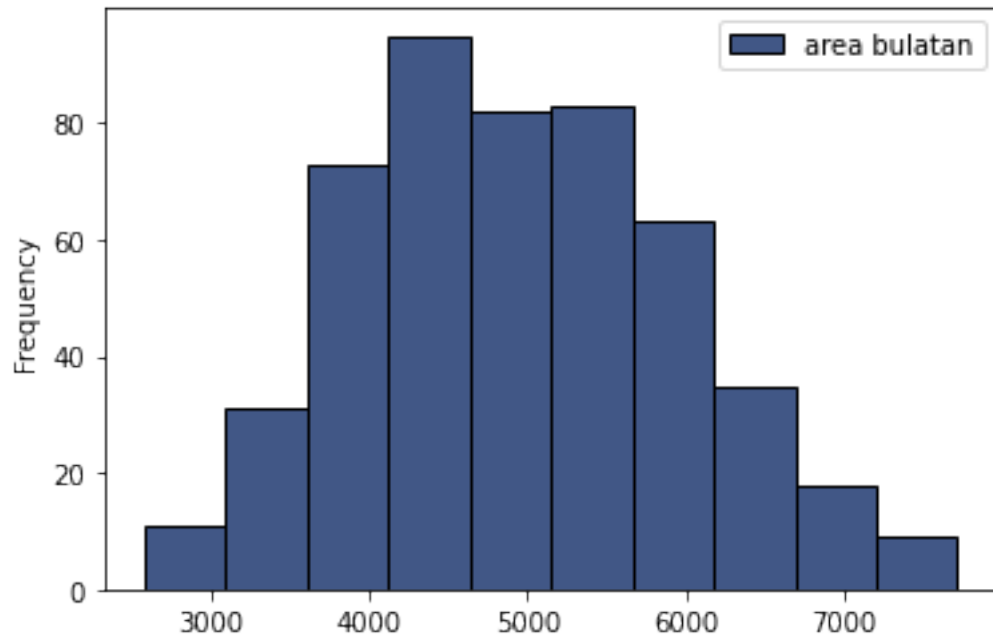
## 2.5 Area Bulatan

```
[348]: df = pd.DataFrame(data = gandum, columns = ['area bulatan'])  
boxplot = df.boxplot(column = ['area bulatan'])
```



```
[349]: df = pd.DataFrame(data = gandum, columns = ['area bulatan'])  
hist = df.plot.hist(bins = 10, color = '#415786', edgecolor='black')
```





Uraian:

Tidak terdapat outlier pada boxplot

Lower tail menandakan nilai  $Q1 - 1.5IQR$  dan *Upper tail* menandakan nilai  $Q3 + 1.5IQR$ . Posisi kedua tail sesuai karena berdasarkan perhitungan lower tail : 2,686.25 dan upper tail : 7,138.25

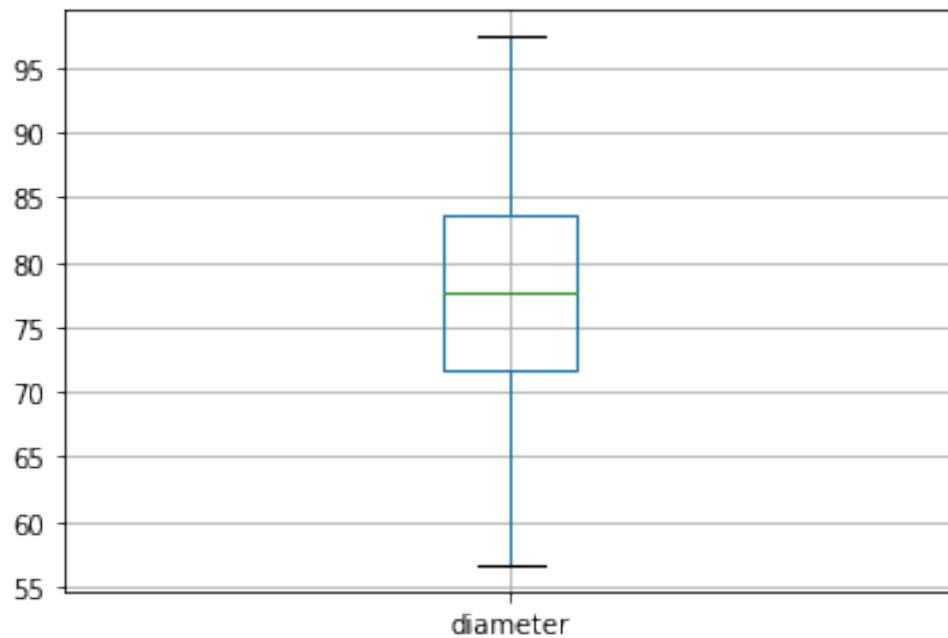
Pada box ditunjukkan Q1, Q2, Q3. Q1 adalah bagian bawah dari box, Q2 adalah bagian tengah dan juga median, Q3 adalah bagian atas dari box. Posisi ketiga kuartil sesuai karena Q1: 4170.25, Q2: 4857.00 , Q3: 5654.25. Jarak antara Bagian atas dan bagian bawah adalah nilai IQR.

Histogram memiliki positive skew. Memiliki kurtosis platykurtic karena bentuknya lebih lebar dari histogram yang berdistribusi normal.

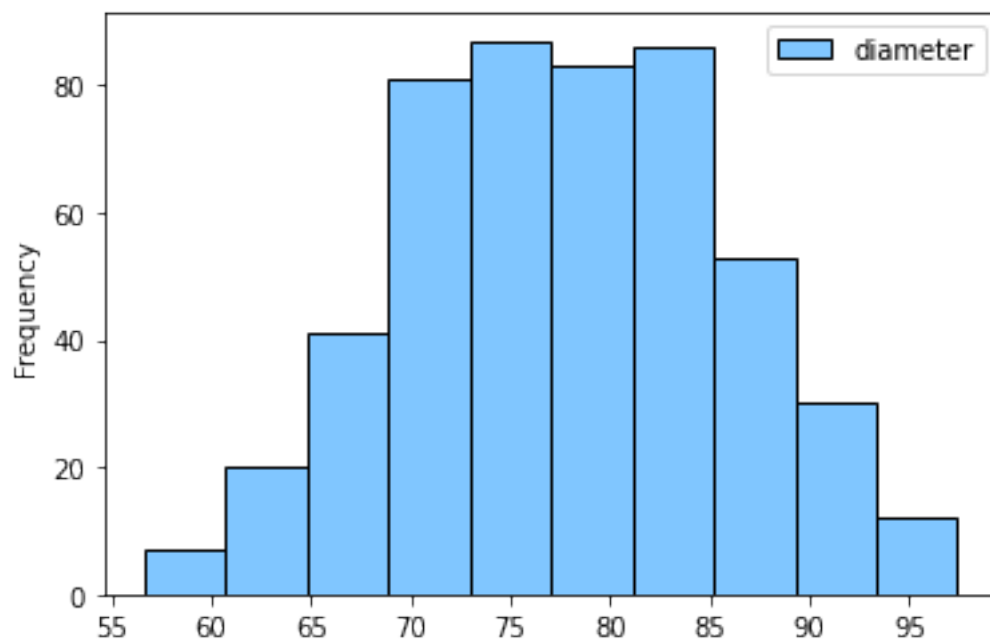
## 2.6 Diameter

---

```
[350]: df = pd.DataFrame(data = gandum, columns = ['diameter'])
       boxplot = df.boxplot(column = ['diameter'])
```



```
[351]: df = pd.DataFrame(data = gandum, columns = ['diameter'])
hist = df.plot.hist(bins = 10, color = '#80C6FF', edgecolor='black')
```



Uraian:

Tidak terdapat outlier

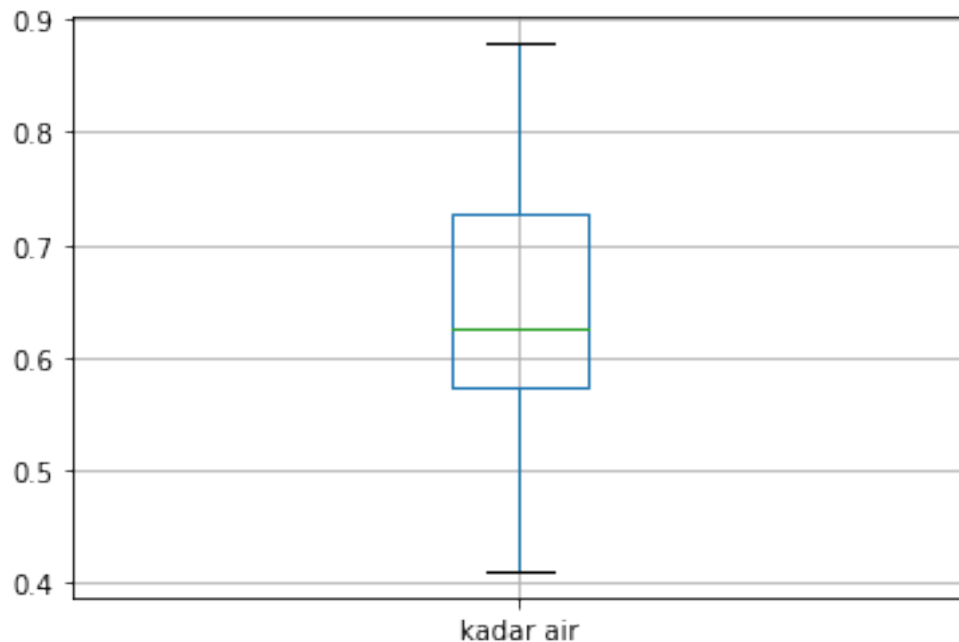
Lower tail menandakan nilai  $Q1 - 1.5IQR$  dan *Upper tail* menandakan nilai  $Q3 + 1.5IQR$ . Posisi kedua tail sesuai karena berdasarkan perhitungan lower tail : 53.890373 dan upper tail : 101.503533

Pada box ditunjukkan Q1, Q2, Q3. Q1 adalah bagian bawah dari box, Q2 adalah bagian tengah dan juga median, Q3 adalah bagian atas dari box. Posisi ketiga kuartil sesuai karena Q1: 71.745308, Q2: 77.645277, Q3: 83.648598. Jarak antara Bagian atas dan bagian bawah adalah nilai IQR.

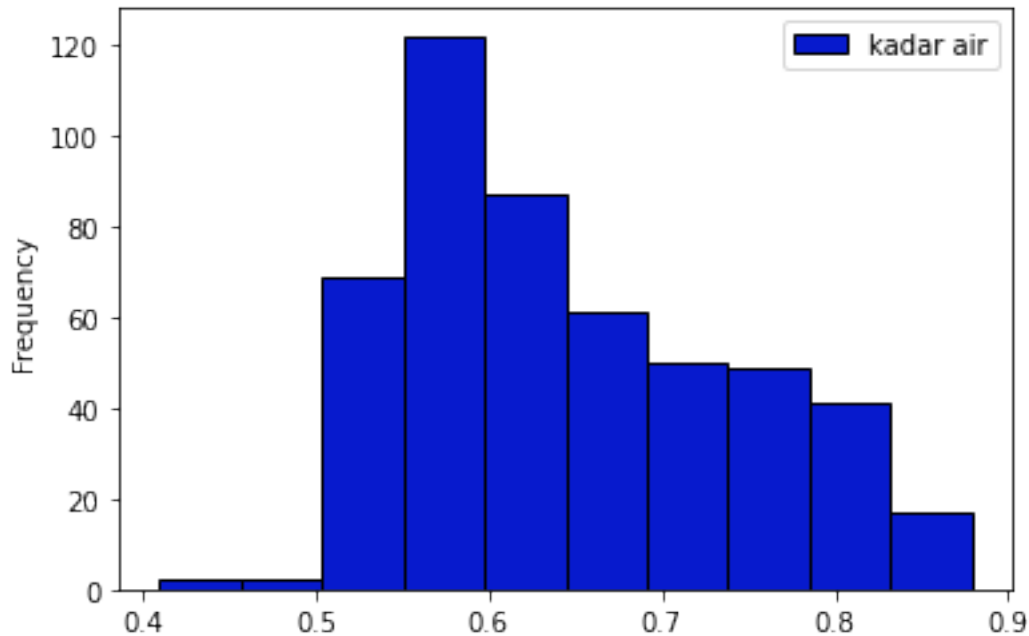
Histogram memiliki skew mendekati 0. Memiliki kurtosis mesokurtic karena nilai kurtosis mendekati 0. Histogram juga menyerupai bell curve. Kolom diameter dapat dianggap berdistribusi normal karena nilai skew dan kurtosis yang mendekati 0 (di pandas, kurtosis normal itu 0).

## 2.7 Kadar Air

```
[352]: df = pd.DataFrame(data = gandum, columns = ['kadar air'])  
       boxplot = df.boxplot(column = ['kadar air'])
```



```
[353]: df = pd.DataFrame(data = gandum, columns = ['kadar air'])  
       hist = df.plot.hist(bins = 10, color = '#071ACD', edgecolor='black')
```



Uraian:

Tidak ada outlier

Lower tail menandakan nilai  $Q1 - 1.5IQR$  dan *Upper tail* menandakan nilai  $Q3 + 1.5IQR$ . Posisi kedua tail sesuai karena berdasarkan perhitungan lower tail :0.3416305 dan upper tail : 0.9576345

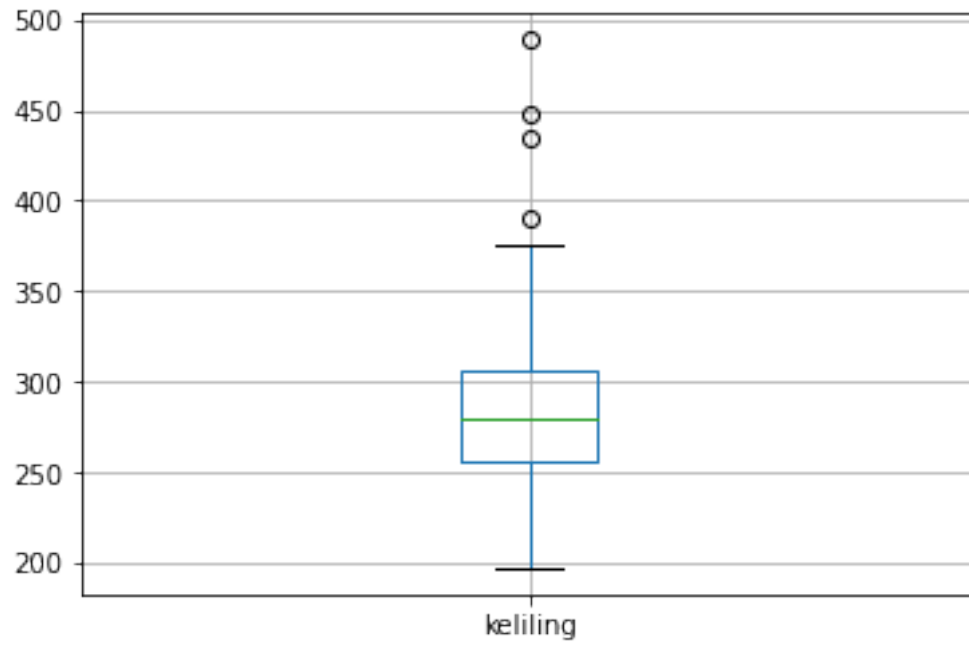
Pada box ditunjukkan Q1, Q2, Q3. Q1 adalah bagian bawah dari box, Q2 adalah bagian tengah dan juga median, Q3 adalah bagian atas dari box. Posisi ketiga kuartil sesuai karena Q1: 0.572632, Q2: 0.626117, Q3: 0.726633. Jarak antara Bagian atas dan bagian bawah adalah nilai IQR.

Histogram memiliki positive skew. Memiliki kurtosis platykurtic karena bentuknya lebih lebar dari histogram yang berdistribusi normal.

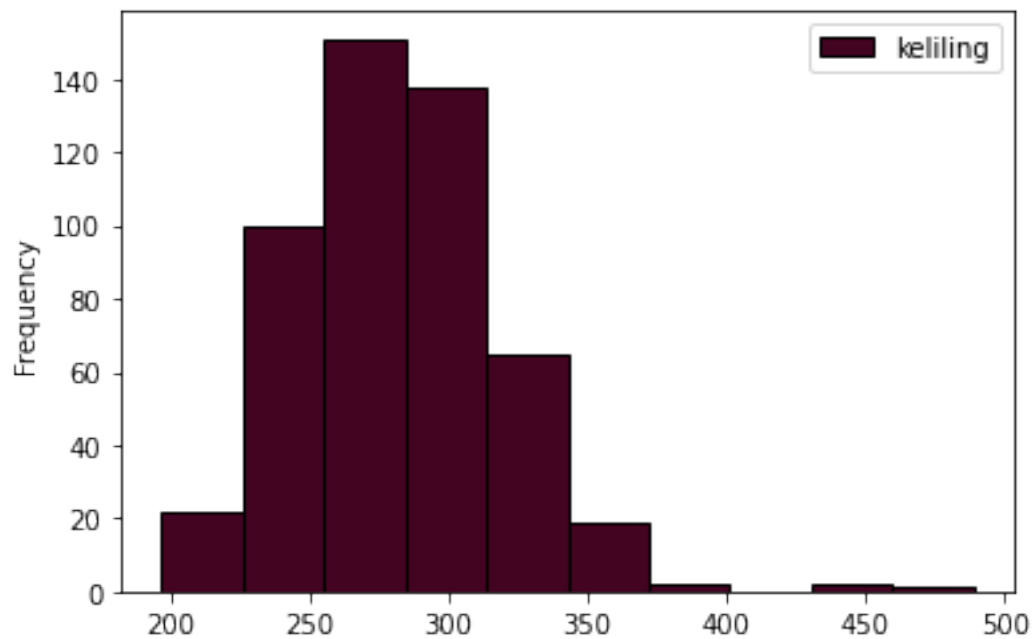
## 2.8 Keliling

---

```
[354]: df = pd.DataFrame(data = gandum, columns = ['keliling'])
      boxplot = df.boxplot(column = ['keliling'])
```



```
[355]: df = pd.DataFrame(data = gandum, columns = ['keliling'])
hist = df.plot.hist(bins = 10, color = '#420420', edgecolor='black')
```



Uraian:

Terdapat outlier pada boxplot hal ini dapat dibuktikan dengan adanya nilai yang lebih besar dari upper tail

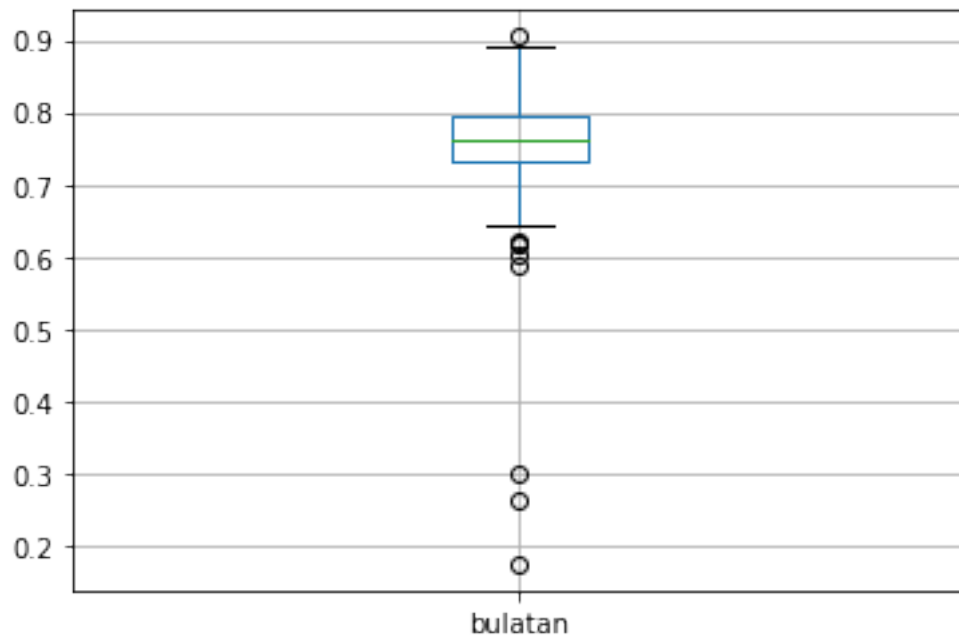
Lower tail menandakan nilai  $Q1 - 1.5IQR$  dan *Upper tail* menandakan nilai  $Q3 + 1.5IQR$ . Posisi kedua tail sesuai karena berdasarkan perhitungan lower tail : 180.61375 dan upper tail : 381.33175

Pada box ditunjukkan Q1, Q2, Q3. Q1 adalah bagian bawah dari box, Q2 adalah bagian tengah dan juga median, Q3 adalah bagian atas dari box. Posisi ketiga kuartil sesuai karena Q1: 255.8830, Q2: 280.0455, Q3: 306.0625. Jarak antara Bagian atas dan bagian bawah adalah nilai IQR.

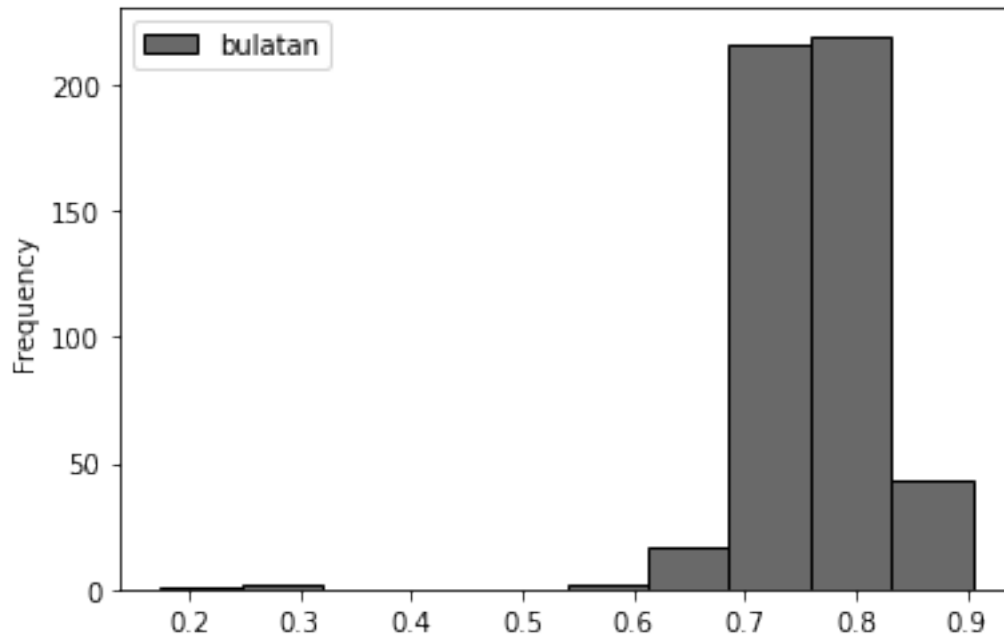
Histogram memiliki negative skew. Memiliki kurtosis leptokurtic karena bentuknya lebih ramping dari histogram yang berdistribusi normal. Terdapat outlier pada histogram.

## 2.9 Bulatan

```
[356]: df = pd.DataFrame(data = gandum, columns = ['bulatan'])  
       boxplot = df.boxplot(column = ['bulatan'])
```



```
[357]: df = pd.DataFrame(data = gandum, columns = ['bulatan'])  
       hist = df.plot.hist(bins = 10, color = '#696969', edgecolor='black')
```



Uraian:

Terdapat outlier pada boxplot. Hal ini dapat dibuktikan dengan adanya nilai max dan min yang lebih dari upper tail dan kurang dari lower tail

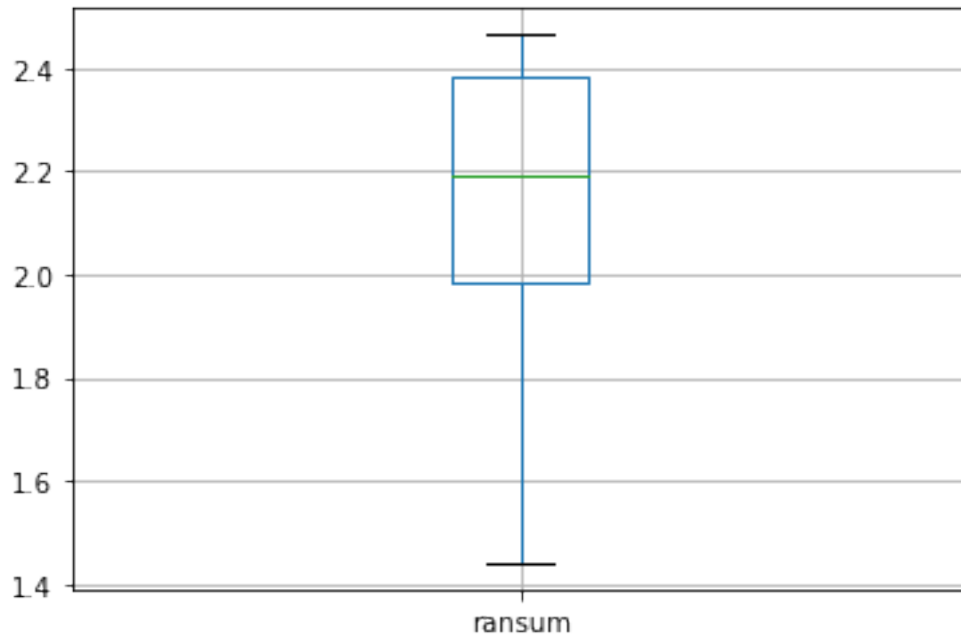
Lower tail menandakan nilai  $Q1 - 1.5IQR$  dan Upper tail menandakan nilai  $Q3 + 1.5IQR$ . Posisi kedua tail sesuai karena berdasarkan perhitungan lower tail : 0.635436 dan upper tail : 0.892916

Pada box ditunjukkan Q1, Q2, Q3. Q1 adalah bagian bawah dari box, Q2 adalah bagian tengah dan juga median, Q3 adalah bagian atas dari box. Posisi ketiga kuartil sesuai karena Q1: 0.731991, Q2: 0.761288, Q3: 0.796361. Jarak antara Bagian atas dan bagian bawah adalah nilai IQR.

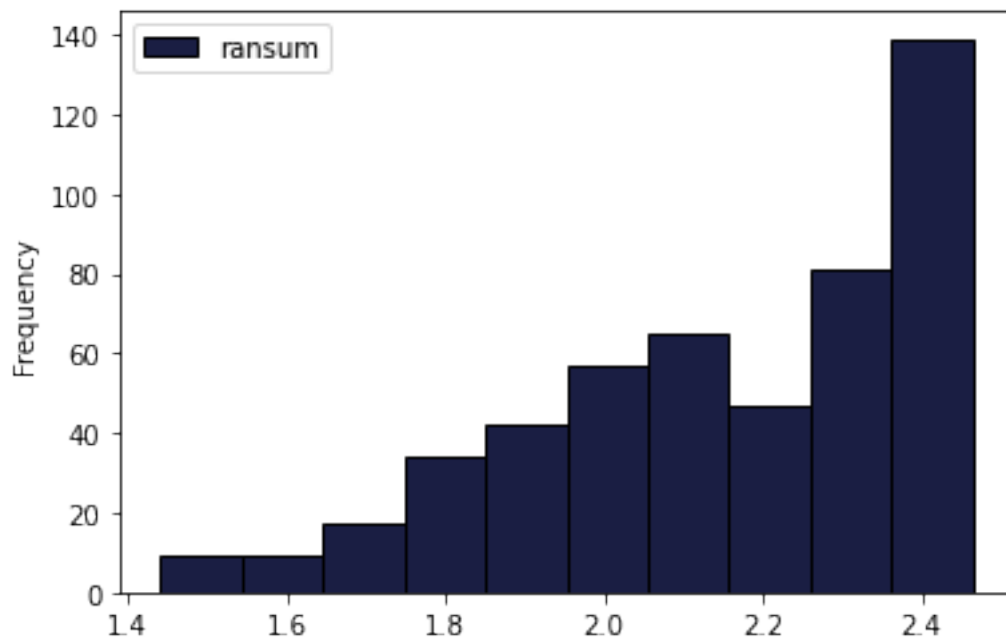
Histogram negative skew. Memiliki kurtosis leptokurtic karena bentuknya lebih ramping dari histogram yang berdistribusi normal . Terdapat outlier pada histogram

## 2.10 Ransum

```
[358]: df = pd.DataFrame(data = gandum, columns = ['ransum'])
        boxplot = df.boxplot(column = ['ransum'])
```



```
[359]: df = pd.DataFrame(data = gandum, columns = ['ransum'])
hist = df.plot.hist(bins = 10, color = '#1A1E43', edgecolor='black')
```



Uraian:



Tidak terdapat outlier pada boxplot

Lower tail menandakan nilai  $Q1 - 1.5IQR$  dan *Upper tail* menandakan nilai  $Q3 + 1.5IQR$ . Posisi kedua tail sesuai karena berdasarkan perhitungan lower tail : 1.3874295 dan upper tail : 2.9781215

Pada box ditunjukkan Q1, Q2, Q3. Q1 adalah bagian bawah dari box, Q2 adalah bagian tengah dan juga median, Q3 adalah bagian atas dari box. Posisi ketiga kuartil sesuai karena Q1: 1.983939, Q2: 2.193599, Q3: 2.381612. Jarak antara Bagian atas dan bagian bawah adalah nilai IQR.

Histogram negative skew. Memiliki kurtosis platykurtic karena bentuknya lebih lebar dari histogram yang berdistribusi normal.

### 3 Normality Test

#### 3.1 Daerah

```
[360]: df = pd.DataFrame(data = gandum, columns = ['daerah'])
stat, p = s.shapiro(df) #shapiro wilks
print("p = {:g}".format(p))
if p <= alpha: #jika p value dibawah atau sama dengan alpha, maka tidak
    ↳berdistribusi normal
    print("Tidak berdistribusi normal")
else:
    print("Berdistribusi normal")
```

p = 0.0032707

Tidak berdistribusi normal

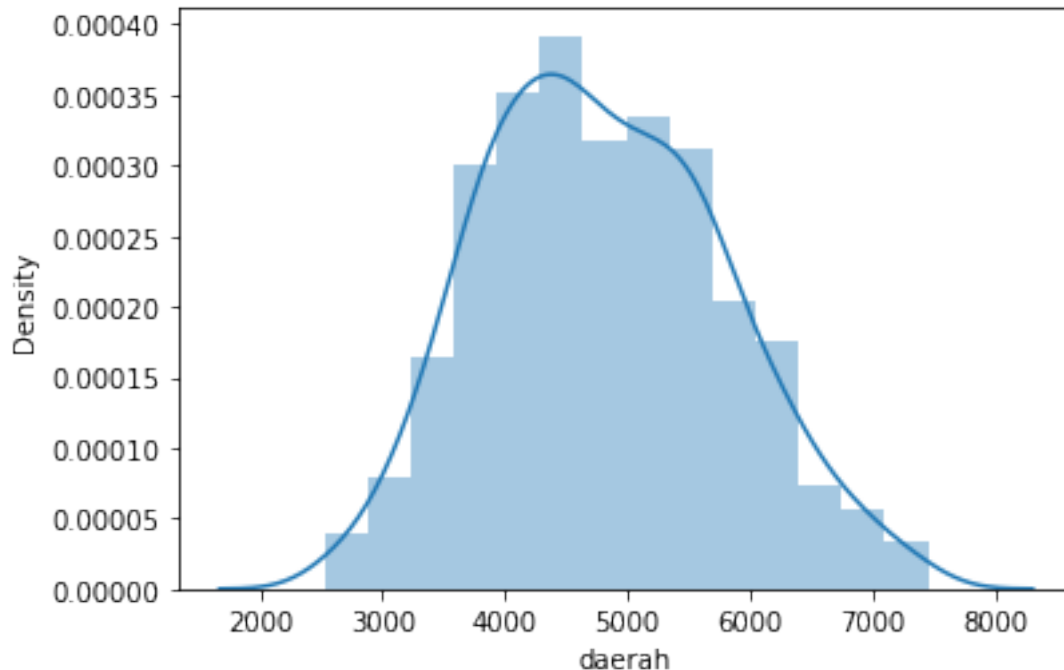
```
[361]: sns.distplot(gandum['daerah'])
```

/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2557:

FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

warnings.warn(msg, FutureWarning)

```
[361]: <matplotlib.axes._subplots.AxesSubplot at 0x7f6cfd187c90>
```



Penjelasan : Kolom daerah melalui normality test shapiro-wilks mendapatkan p-value sebesar 0.00327 yang lebih kecil dari nilai  $\alpha$  0.05 sehingga tidak bisa dianggap berdistribusi normal. Dari histogram juga terlihat bahwa distribusi data daerah tidak berbentuk bell curve.

### 3.2 Sumbu Utama

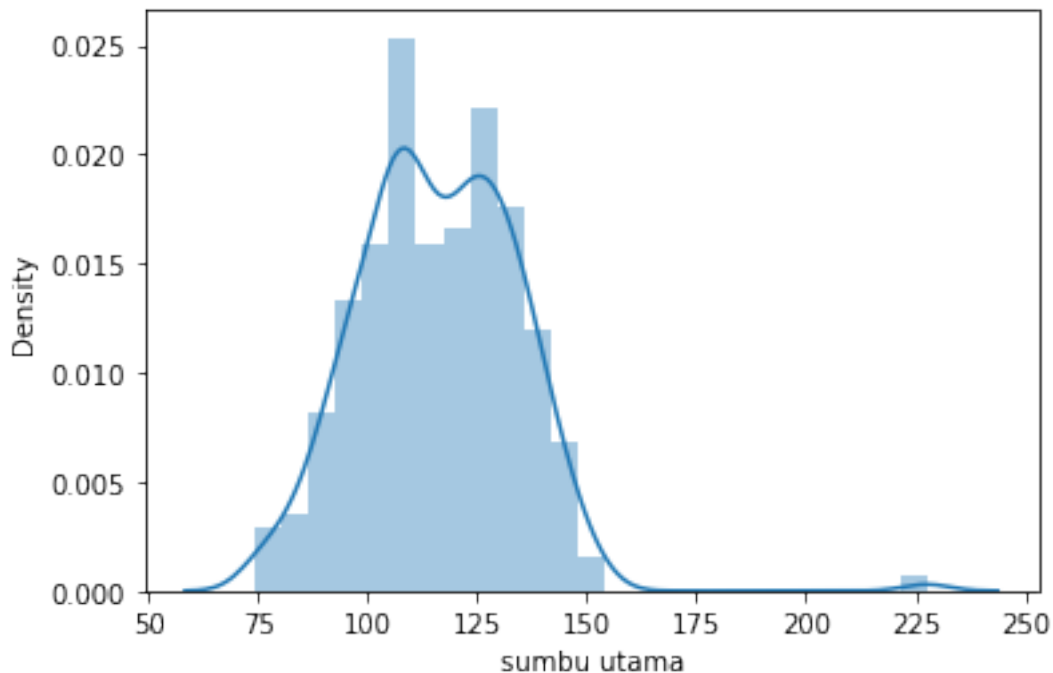
```
[362]: df = pd.DataFrame(data = gandum, columns = ['sumbu utama'])
stat, p = s.shapiro(df) #shapiro wilks
print("p = {:g}".format(p))
if p <= alpha: #jika p value dibawah atau sama dengan alpha, maka tidak
    ↳berdistribusi normal
    print("Tidak berdistribusi normal")
else:
    print("Berdistribusi normal")
```

```
p = 9.2362e-12
Tidak berdistribusi normal
```

```
[363]: sns.distplot(gandum['sumbu utama'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `histplot` (an axes-level function for
histograms).
warnings.warn(msg, FutureWarning)
```

[363]: <matplotlib.axes.\_subplots.AxesSubplot at 0x7f6cf7d2c510>



Penjelasan: Dari normality test shapiro-wilks, kolom sumbu utama mendapatkan p-value bernilai  $9.2362e-12$  yang jauh dibawah nilai  $\alpha$  0.05. Jadi kolom sumbu utama tidak berdistribusi normal. Dari histogram juga bisa dilihat distribusi data kolom sumbu utama yang bentuknya beda jauh dari bell curve.

### 3.3 Sumbu Kecil

```
[364]: df = pd.DataFrame(data = gandum, columns = ['sumbu kecil'])
stat, p = s.shapiro(df) #shapiro wilks
print("p = {:g}".format(p))
if p <= alpha: #jika p value dibawah atau sama dengan alpha, maka tidak
    ↳berdistribusi normal
    print("Tidak berdistribusi normal")
else:
    print("Berdistribusi normal")
```

```
p = 0.423456
Berdistribusi normal
```

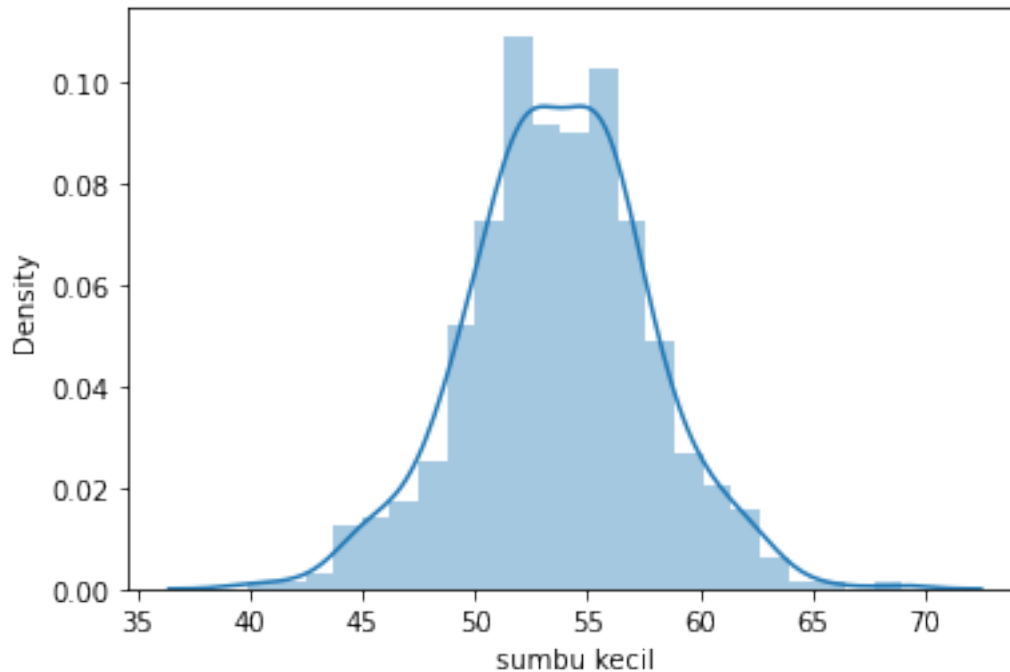
```
[365]: sns.distplot(gandum['sumbu kecil'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
```

function with similar flexibility) or `histplot` (an axes-level function for histograms).

```
warnings.warn(msg, FutureWarning)
```

```
[365]: <matplotlib.axes._subplots.AxesSubplot at 0x7f6cf7e62190>
```



Penjelasan: Dari normality test shapiro-wilks, kolom sumbu kecil mendapatkan nilai 0.423456 yang jauh di atas nilai  $\alpha$  0.05. Dari ini dapat disimpulkan kolom sumbu kecil berdistribusi normal. Dapat dilihat juga dari histogramnya bahwa distribusi data di di kolom sumbu kecil menyerupai bell curve yang merupakan penanda bahwa berdistribusi normal.

### 3.4 Keunikan

```
[366]: df = pd.DataFrame(data = gandum, columns = ['keunikan'])
stat, p = s.shapiro(df) #shapiro wilks
print("p = {:.g}".format(p))
if p <= alpha: #jika p value dibawah atau sama dengan alpha, maka tidak
    ↳berdistribusi normal
    print("Tidak berdistribusi normal")
else:
    print("Berdistribusi normal")
```

```
p = 1.31517e-22
Tidak berdistribusi normal
```

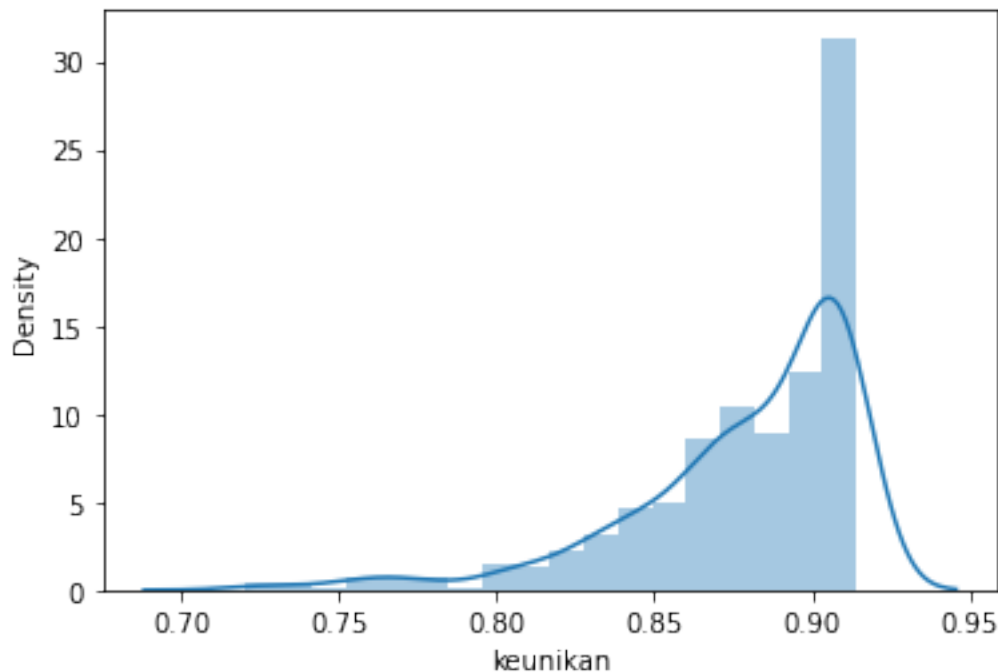
```
[367]: sns.distplot(gandum['keunikan'])
```

```

/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `histplot` (an axes-level function for
histograms).
warnings.warn(msg, FutureWarning)

```

```
[367]: <matplotlib.axes._subplots.AxesSubplot at 0x7f6cf7f87c50>
```



Penjelasan: Dari normality test shapiro-wilks, didapatkan p-value bernilai 1.31517e-22 yang jauh dibawah nilai  $\alpha$  0.05 sehingga dapat disimpulkan kolom keunikan tidak berdistribusi normal. Dari histogram juga dapat dilihat bahwa distribusinya tidak berbentuk bell curve.

### 3.5 Area Bulatan

```

[368]: df = pd.DataFrame(data = gandum, columns = ['area bulatan'])
stat, p = s.shapiro(df) #shapiro wilks
print("p = {:.g}".format(p))
if p <= alpha: #jika p value dibawah atau sama dengan alpha, maka tidak
    ↳berdistribusi normal
    print("Tidak berdistribusi normal")
else:
    print("Berdistribusi normal")

```

```

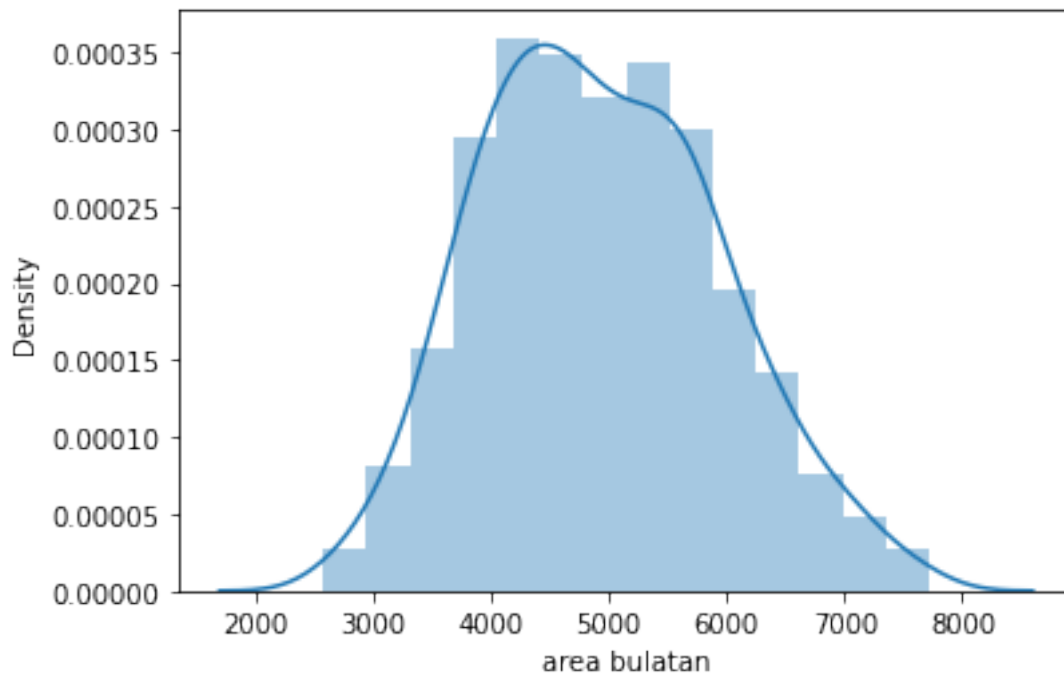
p = 0.00248471
Tidak berdistribusi normal

```

```
[369]: sns.distplot(gandum['area bulatan'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2557:  
FutureWarning: `distplot` is a deprecated function and will be removed in a  
future version. Please adapt your code to use either `displot` (a figure-level  
function with similar flexibility) or `histplot` (an axes-level function for  
histograms).  
warnings.warn(msg, FutureWarning)
```

```
[369]: <matplotlib.axes._subplots.AxesSubplot at 0x7f6cfd8b7c90>
```



Penjelasan: Dari normality test shapiro-wilks, didapatkan p-value sebesar 0.00248 yang berada dibawah nilai  $\alpha$  0.05 sehingga dapat disimpulkan kolom area bulatan tidak berdistribusi normal. Dari histogram pun dapat dilihat bahwa distribusinya tidak berbentuk bell curve.

### 3.6 Diameter

```
[370]: df = pd.DataFrame(data = gandum, columns = ['diameter'])  
stat, p = s.shapiro(df) #shapiro wilks  
print("p = {:g}".format(p))  
if p <= alpha: #jika p value dibawah atau sama dengan alpha, maka tidak  
    ↳berdistribusi normal  
    print("Tidak berdistribusi normal")  
else:  
    print("Berdistribusi normal")
```

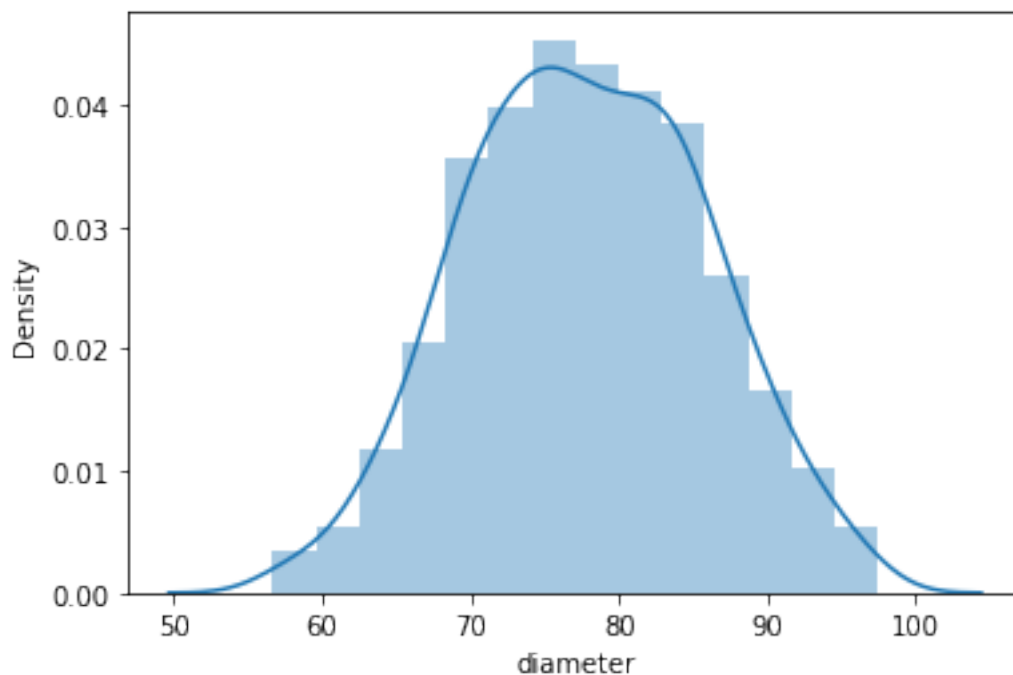
```
p = 0.118345  
Berdistribusi normal
```

```
[371]: sns.distplot(gandum['diameter'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2557:  
FutureWarning: `distplot` is a deprecated function and will be removed in a  
future version. Please adapt your code to use either `displot` (a figure-level  
function with similar flexibility) or `histplot` (an axes-level function for  
histograms).
```

```
warnings.warn(msg, FutureWarning)
```

```
[371]: <matplotlib.axes._subplots.AxesSubplot at 0x7f6cfd6346d0>
```



Penjelasan: Dari normality test shapiro-wilks, didapatkan nilai p-value 0.1118345 yang lebih dari nilai  $\alpha$  0.05 sehingga dapat disimpulkan kolom diameter berdistribusi normal. Dari histogramnya dapat dilihat distribusi datanya lumayan berbentuk bell curve yang menandakan bahwa kolom diameter berdistribusi normal.

### 3.7 Kadar Air

```
[372]: df = pd.DataFrame(data = gandum, columns = ['kadar air'])  
stat, p = s.shapiro(df) #shapiro wilks  
print("p = {:.g}".format(p))  
if p <= alpha: #jika p value dibawah atau sama dengan alpha, maka tidak  
    ↳ berdistribusi normal
```

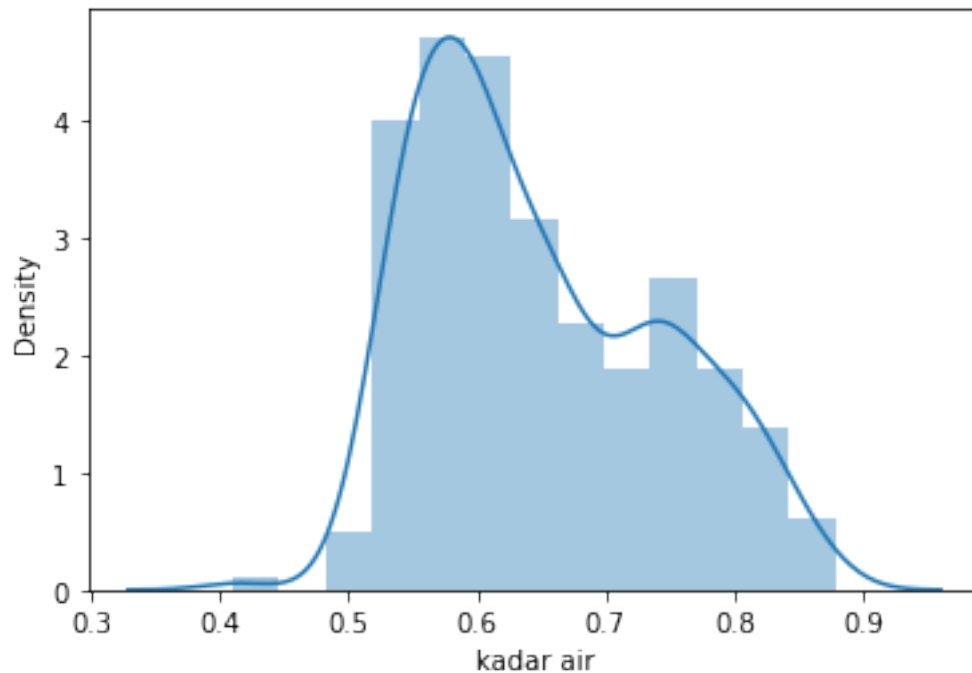
```
print("Tidak berdistribusi normal")
else:
    print("Berdistribusi normal")
```

```
p = 1.9595e-12
Tidak berdistribusi normal
```

```
[373]: sns.distplot(gandum['kadar air'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `histplot` (an axes-level function for
histograms).
    warnings.warn(msg, FutureWarning)
```

```
[373]: <matplotlib.axes._subplots.AxesSubplot at 0x7f6cfd2b1d0>
```



Penjelasan: Dari normality test shapiro-wilks, didapatkan p-value sebesar 1.9595e-12 yang berada jauh dibawah nilai  $\alpha$  0.05 sehingga dapat disimpulkan kolom kadar air tidak berdistribusi normal. Dari histogram pun dapat dilihat bahwa distribusinya tidak berbentuk bell curve



### 3.8 Keliling

```
[374]: df = pd.DataFrame(data = gandum, columns = ['keliling'])
stat, p = s.shapiro(df) #shapiro wilks
print("p = {:g}".format(p))
if p <= alpha: #jika p value dibawah atau sama dengan alpha, maka tidak
    ↳berdistribusi normal
    print("Tidak berdistribusi normal")
else:
    print("Berdistribusi normal")
```

p = 9.72839e-09

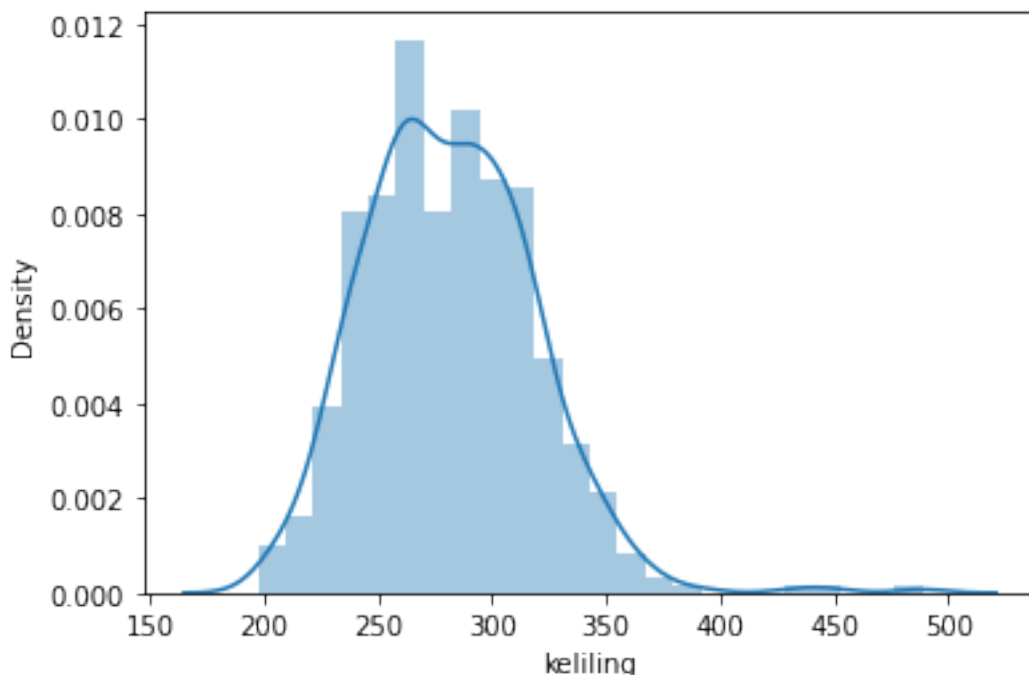
Tidak berdistribusi normal

```
[375]: sns.distplot(gandum['keliling'])
```

/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2557:  
FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

warnings.warn(msg, FutureWarning)

```
[375]: <matplotlib.axes._subplots.AxesSubplot at 0x7f6d02568b90>
```



Penjelasan: Dari normality test shapiro-wilks, didapatkan p-value sebesar 9.72839e-09 yang berada

jauh dibawah nilai  $\alpha$  0.05 sehingga dapat disimpulkan kolom keliling tidak berdistribusi normal. Dari histogram pun dapat dilihat bahwa distribusinya tidak berbentuk bell curve

### 3.9 Bulatan

```
[376]: df = pd.DataFrame(data = gandum, columns = ['bulatan'])
stat, p = s.shapiro(df) #shapiro wilks
print("p = {:g}".format(p))
if p <= alpha: #jika p value dibawah atau sama dengan alpha, maka tidak
    ↳berdistribusi normal
    print("Tidak berdistribusi normal")
else:
    print("Berdistribusi normal")
```

p = 6.89916e-26

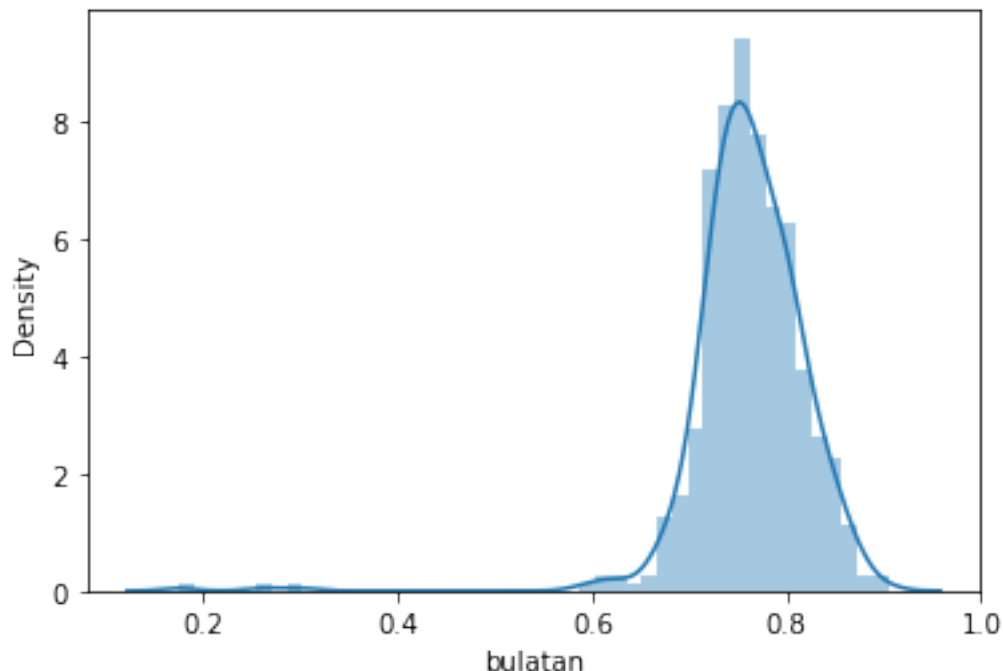
Tidak berdistribusi normal

```
[377]: sns.distplot(gandum['bulatan'])
```

/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2557:  
FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

warnings.warn(msg, FutureWarning)

```
[377]: <matplotlib.axes._subplots.AxesSubplot at 0x7f6cf7f22e50>
```



Penjelasan: Dari normality test shapiro-wilks, didapatkan p-value sebesar 6.89916e-26 yang berada jauh dibawah nilai  $\alpha$  0.05 sehingga dapat disimpulkan kolom bulatan tidak berdistribusi normal. Dari histogram pun dapat dilihat bahwa distribusinya tidak berbentuk bell curve

### 3.10 Ransum

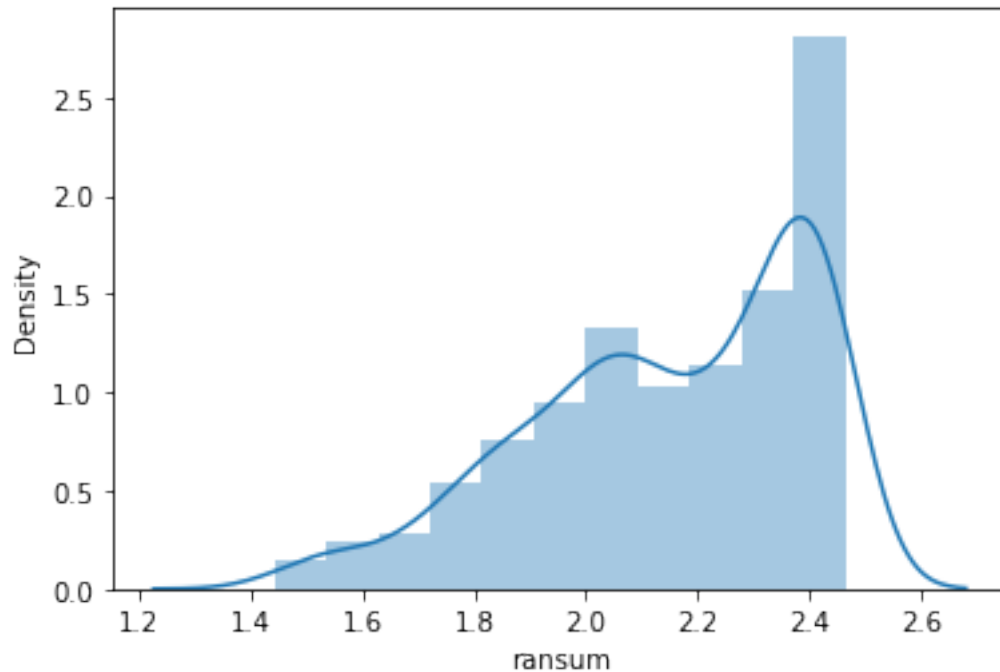
```
[378]: df = pd.DataFrame(data = gandum, columns = ['ransum'])
stat, p = s.shapiro(df) #shapiro wilks
print("p = {:g}".format(p))
if p <= alpha: #jika p value dibawah atau sama dengan alpha, maka tidak
    ↪berdistribusi normal
    print("Tidak berdistribusi normal")
else:
    print("Berdistribusi normal")
```

```
p = 6.24554e-15
Tidak berdistribusi normal
```

```
[379]: sns.distplot(gandum['ransum'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `histplot` (an axes-level function for
histograms).
  warnings.warn(msg, FutureWarning)
```

```
[379]: <matplotlib.axes._subplots.AxesSubplot at 0x7f6cfd851fd0>
```



Penjelasan: Dari normality test shapiro-wilks, didapatkan p-value sebesar 6.24554e-15 yang berada jauh dibawah nilai  $\alpha$  0.05 sehingga dapat disimpulkan kolom ransum tidak berdistribusi normal. Dari histogram pun dapat dilihat bahwa distribusinya tidak berbentuk bell curve

## 4 Test Hipotesis 1 Sampel

### 4.1 A) Nilai rata-rata Daerah di atas 4700?

#### 4.1.1 Langkah-Langkah

1. Tentukan hipotesis nol

$$H_0 : \mu = 4700$$

2. Tentukan hipotesis alternatif

$$H_1 : \mu > 4700$$

3. Tentukan tingkat signifikan

$$\alpha = 0.05$$

4. Tentukan uji statistik yang sesuai dan tentukan daerah kritis.

Uji statistik : Uji parameter populasi pengujian rata-rata satu sampel one tail test variance known

Daerah kritis :  $z > z_\alpha : z > 1.645$

5. Hitung nilai uji statistik

```
[380]: mean = 4700
rataaan = gandum['daerah'].mean()
std = gandum['daerah'].std()
n = gandum['daerah'].count() #ambil sample sepopulasi
z = zscore(rataaan, mean, std, n)
p = 1 - ztop(z)
zalpha = ptot(1-alpha)

print("Nilai z : " +str(z))
print("p-value : " + str(p))
```

Nilai z : 2.2951538242525173

p-value : 0.010862155196799894

6. Ambil keputusan

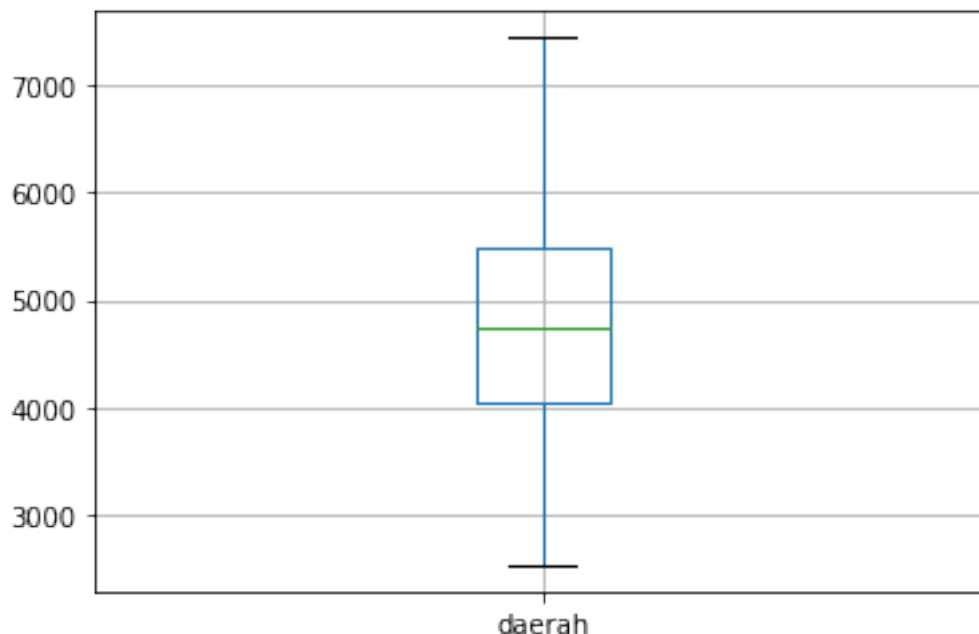
```
[381]: if z > zalpha:
    print("Tolak null hypothesis")
    print("Nilai rata-rata Daerah di atas 4700")
else:
    print("Terima null hypothesis")
    print("Nilai rata-rata Daerah tidak dapat dibilang di atas 4700")
```

Tolak null hypothesis

Nilai rata-rata Daerah di atas 4700

#### 4.1.2 Boxplot

```
[382]: df = pd.DataFrame(data = gandum, columns = ['daerah'])
boxplot = df.boxplot(column = ['daerah'])
```



## 4.2 B) Nilai Rata-rata Sumbu Utama tidak sama dengan 116?

### 4.2.1 Langkah-Langkah

1. Tentukan hipotesis nol

$$H_0 : \mu = 116$$

2. Tentukan hipotesis alternatif

$$H_1 : \mu \neq 116 \text{ (two-tailed test)}$$

3. Tentukan tingkat signifikan

$$\alpha = 0.05$$

4. Tentukan uji statistik yang sesuai dan tentukan daerah kritis.

Uji statistik : Uji parameter populasi pengujian rata-rata satu sampel two tail test variance known

Daerah kritis :  $z < -z_{\alpha/2}$  or  $z > z_{\alpha/2}$  :  $z < -1.96$  or  $z > 1.96$

5. Hitung nilai uji statistik

```
[383]: mean = 116
rataan = gandum['sumbu utama'].mean()
std = gandum['sumbu utama'].std()
n = gandum['sumbu utama'].count() #ambil sample sepopulasi
z = zscore(rataan, mean, std, n)
zalpha = ptz(1-alpha/2)
p = 2*(1-ztop(abs(z)))
print("Nilai z: " + str(z))
print("p-value: " + str(p))
```

Nilai z: 0.055247123267335795

p-value: 0.9559415872977244

6. Ambil Keputusan

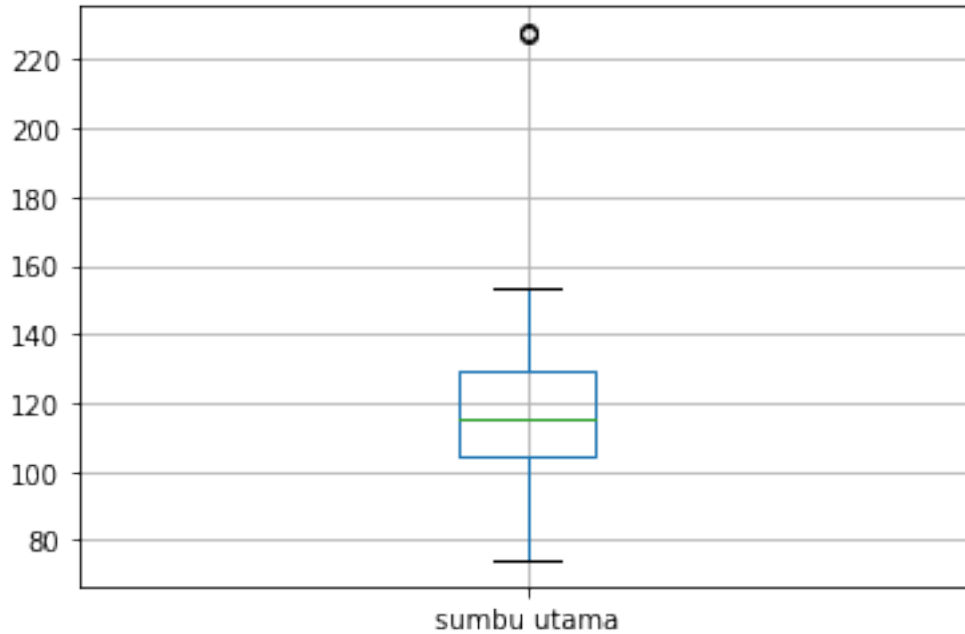
```
[384]: if z < -zalpha or z > zalpha:
    print("Null hypothesis ditolak")
    print("Rata-rata sumbu utama tidak sama dengan 116")
else:
    print("Null hypothesis gagal untuk ditolak")
    print("Rata-rata sumbu utama sama dengan 116")
```

Null hypothesis gagal untuk ditolak

Rata-rata sumbu utama sama dengan 116

### 4.2.2 Boxplot

```
[385]: df = pd.DataFrame(data = gandum, columns = ['sumbu utama'])  
boxplot = df.boxplot(column = ['sumbu utama'])
```



### 4.3 C) Nilai Rata-rata 20 baris pertama kolom Sumbu Kecil bukan 50?

#### 4.3.1 Langkah-Langkah

1. Tentukan hipotesis nol

$$H_0 : \mu = 50$$

2. Tentukan hipotesis alternatif

$$H_1 : \mu \neq 50 \text{ (two-tailed test)}$$

3. Tentukan tingkat signifikan

$$\alpha = 0.05$$

4. Tentukan uji statistik yang sesuai dan tentukan daerah kritis.

Uji statistik : Uji parameter populasi pengujian rata-ran satu sampel two tail test variance known

$$\text{Daerah kritis : } z < -z_{\alpha/2} \text{ or } z > z_{\alpha/2} : z < -1.96 \text{ or } z > 1.96$$

5. Hitung nilai uji statistik

```
[386]: mean = 50
sampel = gandum['sumbu kecil'].head(20)
rataaan = sampel.mean()
std = sampel.std()
n = sampel.count()
z = zscore(rataaan, mean, std, n)
p = 2*(1-ztop(abs(z)))
zalpha = ptoz(1 - alpha/2)
print("Nilai z: " + str(z))
print("p-value : " + str(p))
```

Nilai z: 6.478168916968886

p-value : 9.284240043427872e-11

6. Ambil Keputusan

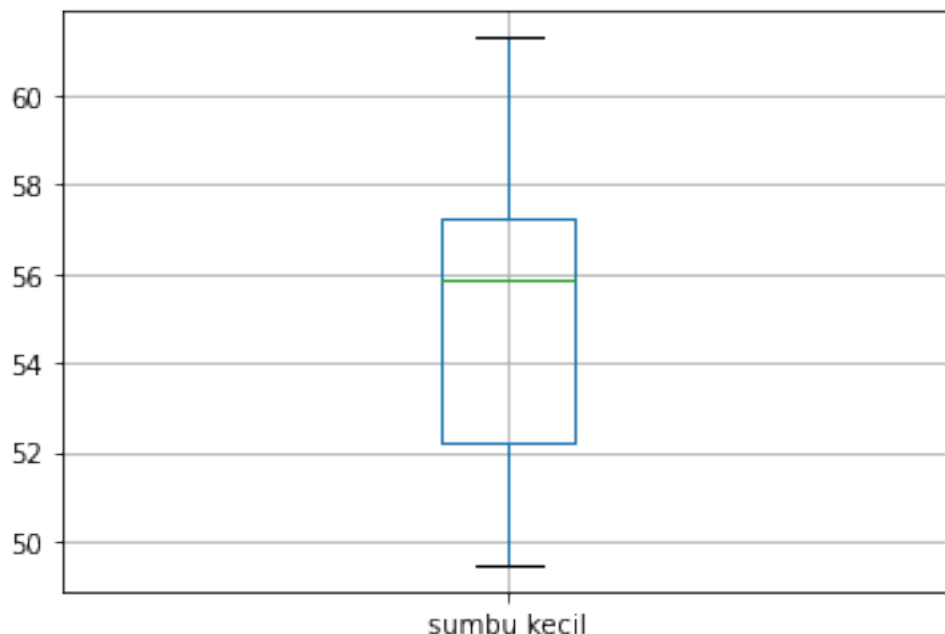
```
[387]: if z < -zalpha or z > zalpha:
    print("Null hypothesis ditolak")
    print("Nilai rata-rata 20 baris pertama kolom sumbu kecil bukan 50")
else:
    print("Null hypothesis diterima")
    print("Nilai rata-rata 20 baris pertama kolom sumbu kecil 50")
```

Null hypothesis ditolak

Nilai rata-rata 20 baris pertama kolom sumbu kecil bukan 50

#### 4.3.2 Boxplot

```
[388]: df = pd.DataFrame(data = sampel)
boxplot = df.boxplot()
```





#### 4.4 D) Proporsi nilai Diameter yang lebih dari 85, adalah tidak sama dengan 15% ?

##### 4.4.1 Langkah-Langkah

1. Tentukan hipotesis nol

$$H_0 : p = 0.15$$

2. Tentukan hipotesis alternatif

$$H_1 : p \neq 0.15 \text{ (two-tailed test)}$$

3. Tentukan tingkat signifikan

$$\alpha = 0.05$$

4. Tentukan uji statistik yang sesuai dan tentukan daerah kritis.

Uji statistik : Uji satu parameter populasi pengujian proporsi satu sampel dengan sample banyak

$$\text{Daerah kritis : } z > z_{\alpha/2} \text{ or } z < -z_{\alpha/2} : z > 1.96 \text{ or } z < -1.96$$

5. Hitung nilai statistik

```
[389]: diameter = gandum['diameter'].loc[gandum['diameter'] > 85]
p0 = 0.15
q0 = 1 - p0
zalpha = ptot(1 - alpha/2)
n = gandum['diameter'].count()
x = diameter.count()
p1 = x/n

z = (p1-p0) * math.sqrt(n)/ math.sqrt(p0*q0)
p = 2*(1- ztop(abs(z)))

print("Nilai z : " + str(z))
print("p-value : " + str(p))
```

Nilai z : 2.7553868807746587

p-value : 0.005862277168409591

6. Ambil Keputusan

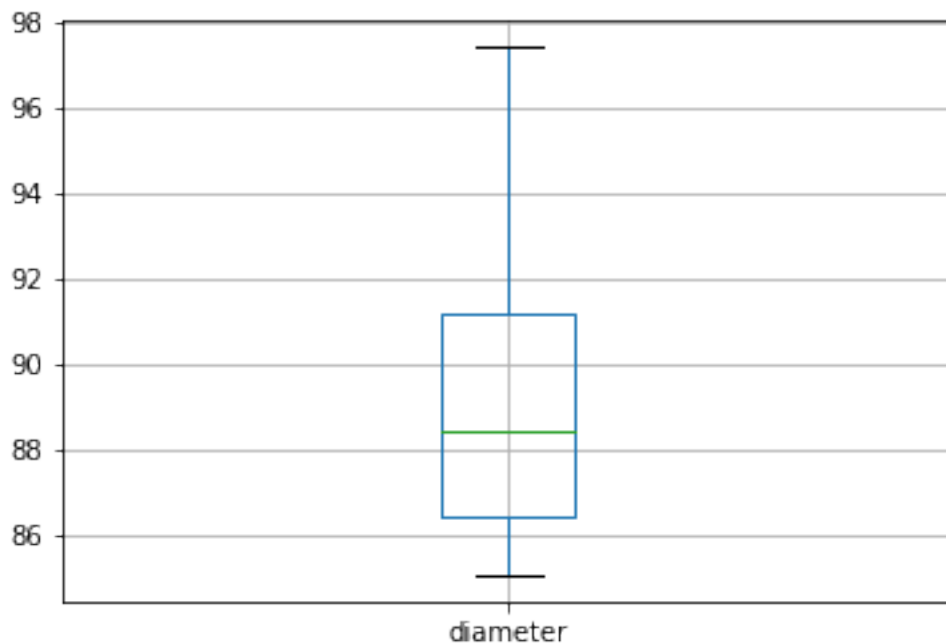
```
[390]: if z < -zalpha or z > zalpha:
    print("Null hypothesis ditolak")
    print("Proporsi nilai Diameter yang lebih dari 85 tidak sama dengan 15%")
else:
    print("Null hypothesis diterima")
    print("Proporsi nilai diameter yang lebih dari 85 sama dengan 15%")
```

Null hypothesis ditolak

Proporsi nilai Diameter yang lebih dari 85 tidak sama dengan 15%

#### 4.4.2 Boxplot

```
[391]: df = pd.DataFrame(data = diameter)
      boxplot = df.boxplot()
```



4.5 E) Proporsi nilai Keliling yang kurang dari 100, adalah kurang dari 5% ?

##### 4.5.1 Langkah-Langkah

1. Tentukan hipotesis nol

$$H_0 : p = 0.05$$

2. Tentukan hipotesis alternatif

$$H_1 : p < 0.05 \text{ (one-tailed test)}$$

3. Tentukan tingkat signifikan

$$\alpha = 0.05$$

4. Tentukan uji statistik yang sesuai dan tentukan daerah kritis.

Uji statistik : Uji satu parameter populasi pengujian proporsi satu sampel dengan sample banyak one tailed test

$$\text{Daerah kritis : } z < -z_{\alpha} : z < -1.645$$

#### 5. Hitung nilai uji statistik

```
[392]: keliling = gandum['keliling'].loc[gandum['keliling'] < 100]
p0 = 0.05
q0 = 1 - p0
zalpha = ptz(1 - alpha)
n = gandum['keliling'].count()
x = keliling.count()
p1 = x/n
z = (p1-p0) * math.sqrt(n) / math.sqrt(p0*q0)
p = ztop(z)
print("Nilai z : " + str(z))
print("p-value : " + str(p) )
```

Nilai z : -5.129891760425771

p-value : 1.4495441414387716e-07

#### 6. Ambil Keputusan

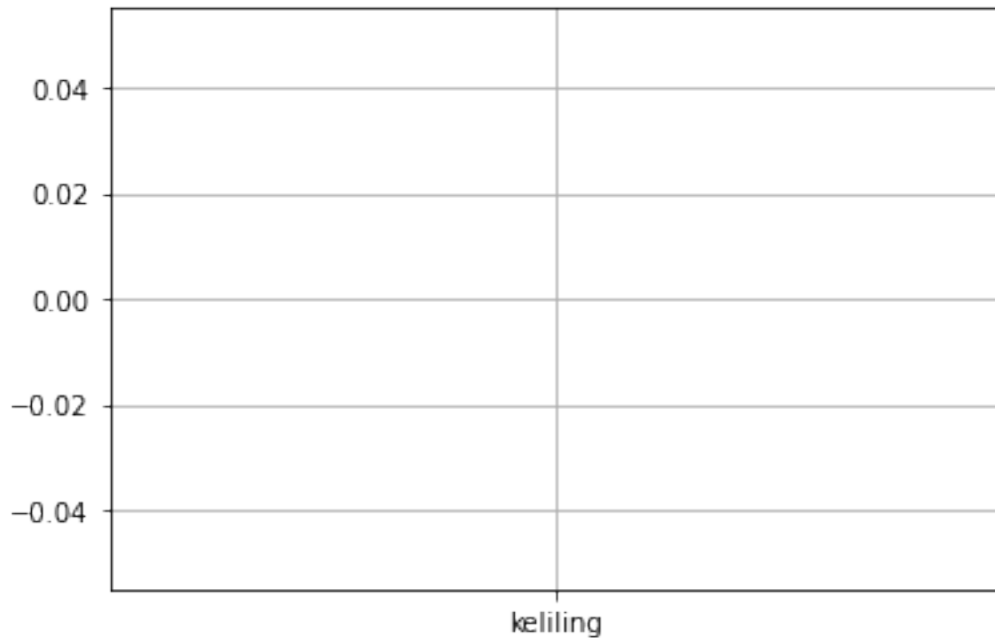
```
[393]: if z < -zalpha:
    print("Null hypothesis ditolak")
    print("Proporsi nilai Keliling yang kurang dari 100, adalah kurang dari 5%")
else:
    print("Null hypothesis diterima")
    print("Proporsi nilai Keliling yang kurang dari 100 tidak kurang dari 5%")
```

Null hypothesis ditolak

Proporsi nilai Keliling yang kurang dari 100, adalah kurang dari 5%

#### 4.5.2 Boxplot

```
[394]: df = pd.DataFrame(data = keliling)
boxplot = df.boxplot()
```



## 5 Test Hipotesis 2 Sampel

5.1 A) Data kolom AreaBulatan dibagi 2 sama rata: bagian awal dan bagian akhir kolom. Benarkah rata-rata kedua bagian tersebut sama?

### 5.1.1 Langkah-Langkah

1. Tentukan hipotesis nol

$$H_0 : \mu_1 - \mu_2 = 0$$

2. Tentukan hipotesis alternatif

$$H_1 : \mu_1 - \mu_2 \neq 0 \text{ (two-tailed)}$$

3. Tentukan tingkat signifikan

$$\alpha = 0.05$$

4. Tentukan uji statistik yang sesuai dan tentukan daerah kritis.

Uji statistik : two sample two tailed mean test known std

Daerah kritis :  $z < -z_{\alpha/2}$  or  $z > z_{\alpha/2}$  :  $z < -1.96$  or  $z > 1.96$

5. Hitung nilai uji statistik

```
[395]: areaBulat = gandum['area bulatan']
```

```
d0 = 0
```

```

zalpha = ptoz(1 - alpha/2)
bagian1 = areaBulat.head(areaBulat.size // 2)
bagian2 = areaBulat.tail(areaBulat.size // 2)

mean1 = bagian1.mean()
mean2 = bagian2.mean()

var1 = bagian1.var()
var2 = bagian2.var()
z = ztwotestmean(d0, mean1, mean2, var1, var2, bagian1.size, bagian2.size)
p = 2*(1-ztop(abs(z)))
print("Nilai z : " + str(z))
print("p-value : " + str(p))

```

Nilai z : 17.013036648485464  
 p-value : 0.0

#### 6. Ambil Keputusan

```

[396]: if z < -zalpha or z > zalpha:
        print("Null hypothesis ditolak")
        print("Rata-rata kedua bagian tidak sama")
    else:
        print("Null hypothesis diterima")
        print("Rata-rata kedua bagian sama")

```

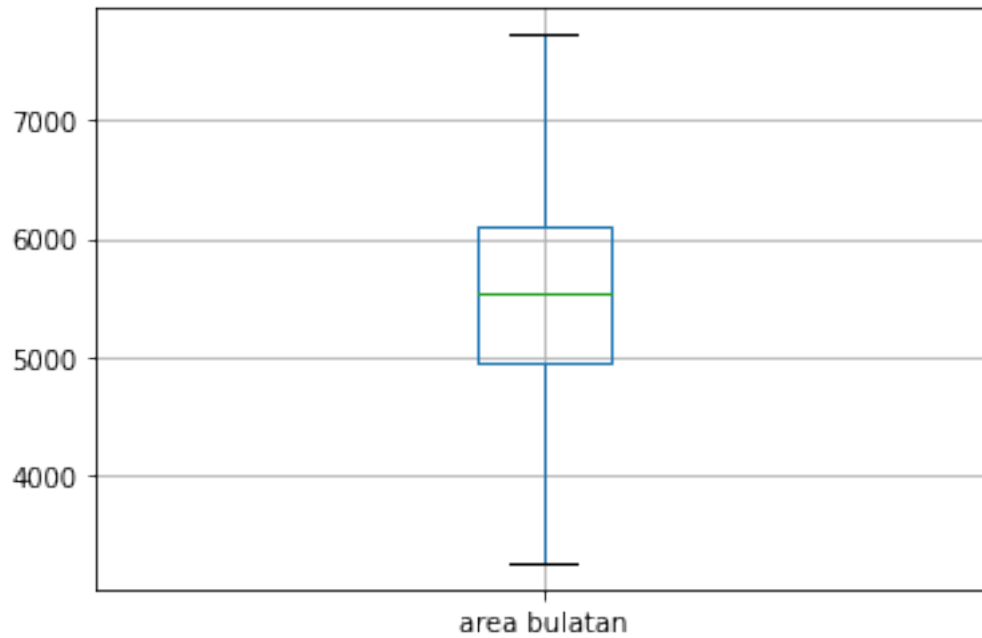
Null hypothesis ditolak  
 Rata-rata kedua bagian tidak sama

#### 5.1.2 Boxplot Bagian Awal

```

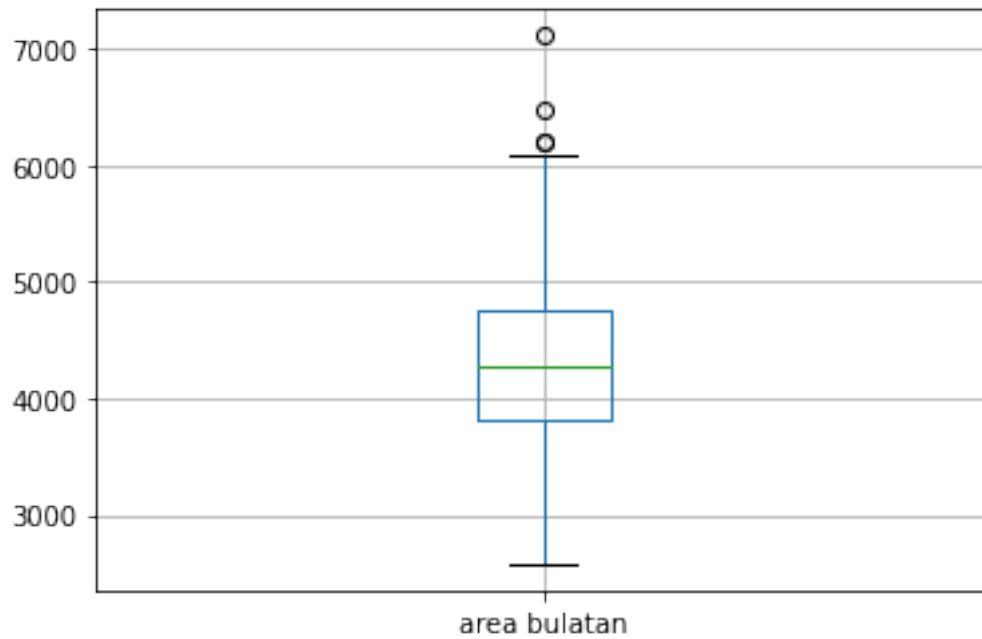
[397]: df = pd.DataFrame(data = bagian1)
        boxplot = df.boxplot()

```



### 5.1.3 Boxplot Bagian Akhir

```
[398]: df = pd.DataFrame(data = bagian2)
       boxplot = df.boxplot()
```



## 5.2 B) Data kolom Kadar Air dibagi 2 sama rata: bagian awal dan bagian akhir kolom. Benarkah rata-rata bagian awal lebih besar dari pada bagian akhir sebesar 0.2?

### 5.2.1 Langkah-Langkah

1. Tentukan hipotesis nol

$$H_0 : \mu_1 - \mu_2 = 0.2$$

2. Tentukan hipotesis alternatif

$$H_1 : \mu_1 - \mu_2 \neq 0.2 \text{ (two-tailed)}$$

3. Tentukan tingkat signifikan

$$\alpha = 0.05$$

4. Tentukan uji statistik yang sesuai dan tentukan daerah kritis.

Uji statistik : two sample two tailed mean test known std

Daerah kritis :  $z < -z_{\alpha/2}$  or  $z > z_{\alpha/2}$  :  $z < -1.96$  or  $z > 1.96$

5. Hitung nilai uji statistik

```
[399]: kadarAir = gandum['kadar air']

d0 = 0.2

zalpha = ptoz(1-alpha/2)
bagian1 = kadarAir.head(kadarAir.size // 2)
bagian2 = kadarAir.tail(kadarAir.size // 2)

mean1 = bagian1.mean()
mean2 = bagian2.mean()

var1 = bagian1.var()
var2 = bagian2.var()

z = ztwotestmean(d0, mean1, mean2, var1, var2, bagian1.size, bagian2.size)
p = 2*(1-ztop(abs(z)))
print("Nilai z : " + str(z))
print("p-value : " + str(p))
```

Nilai z : -26.903444249744958

p-value : 0.0

6. Ambil Keputusan

```
[400]: if z < -zalpha or z > zalpha:
        print("Null hypothesis ditolak")
        print("Rata-rata bagian awal tidak lebih besar dari pada bagian akhir sebesar 0.2")
```

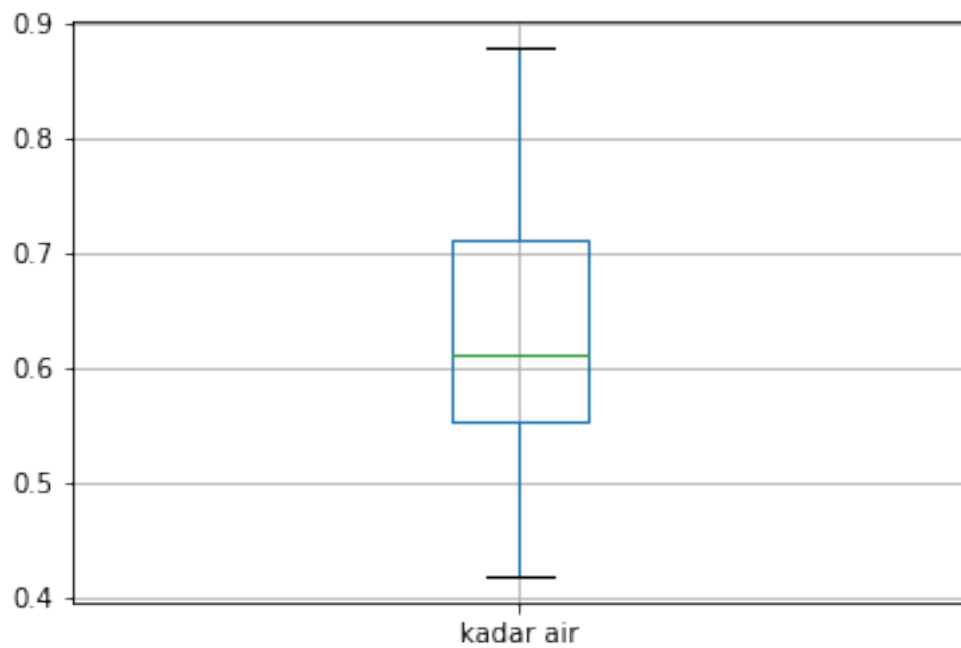
```
else:  
    print("Null hypothesis diterima")  
    print("Rata-rata bagian awal lebih besar dari pada bagian akhir sebesar 0.2")
```

Null hypothesis ditolak

Rata-rata bagian awal tidak lebih besar dari pada bagian akhir sebesar 0.2

### 5.2.2 Boxplot Bagian Awal

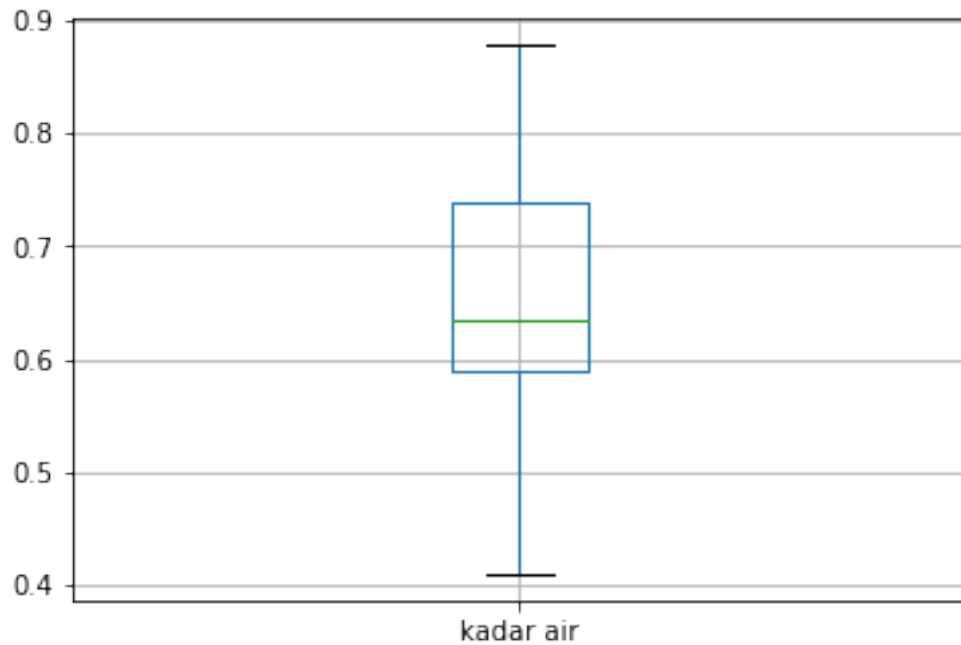
```
[401]: df = pd.DataFrame(data = bagian1)  
        boxplot = df.boxplot()
```



### 5.2.3 Boxplot Bagian Akhir

```
[402]: df = pd.DataFrame(data = bagian2)  
        boxplot = df.boxplot()
```





### 5.3 C) Rata-rata 20 baris pertama kolom Bulatan sama dengan 20 baris terakhirnya?

#### 5.3.1 Langkah-Langkah

1. Tentukan hipotesis nol

$$H_0 : \mu_1 - \mu_2 = 0$$

2. Tentukan hipotesis alternatif

$$H_1 : \mu_1 - \mu_2 \neq 0 \text{ (two-tailed)}$$

3. Tentukan tingkat signifikan

$$\alpha = 0.05$$

4. Tentukan uji statistik yang sesuai dan tentukan daerah kritis.

Uji statistik : two sample two tailed mean test known std

Daerah kritis :  $z < -z_{\alpha/2}$  atau  $z > z_{\alpha/2}$  :  $z < -1.96$  or  $z > 1.96$

5. Hitung nilai uji statistik

```
[403]: bulatan = gandum['bulatan']

bagian1 = bulatan.head(20)
bagian2 = bulatan.tail(20)
```

```

mean1 = bagian1.mean()
mean2 = bagian2.mean()

var1 = bagian1.var()
var2 = bagian2.var()

d0 = 0

zalpha = ptot(1 - alpha/2)
z = zttestmean(d0, mean1, mean2, var1, var2, bagian1.size, bagian2.size)
p = 2*(1 - ztop(abs(z)))
print("Nilai z : " + str(z))
print("p-value : " + str(p))

```

Nilai z : -3.3792268633124025  
 p-value : 0.0007268999745855265

6. Ambil keputusan

```

[404]: if z < -zalpha or z > zalpha:
        print("Null hypothesis ditolak")
        print("Rata-rata 20 baris pertama kolom bulatan tidak sama dengan 20 baris_
        ↳ terakhirnya")
    else:
        print("Null hypothesis diterima")
        print("Rata-rata 20 baris pertama kolom bulatan sama dengan 20 baris_
        ↳ terakhirnya")

```

Null hypothesis ditolak

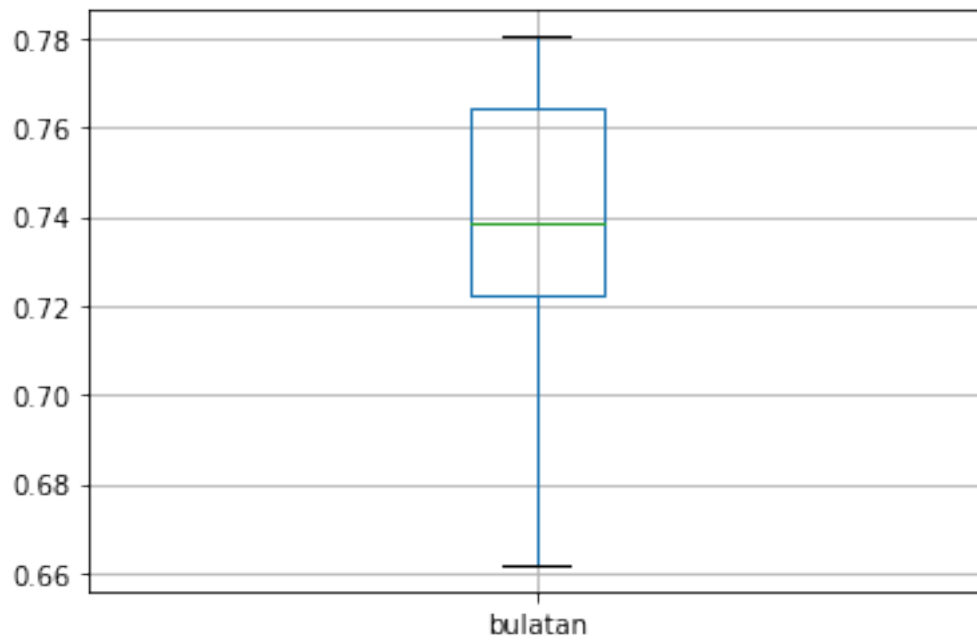
Rata-rata 20 baris pertama kolom bulatan tidak sama dengan 20 baris terakhirnya

### 5.3.2 Boxplot 20 Baris Pertama

```

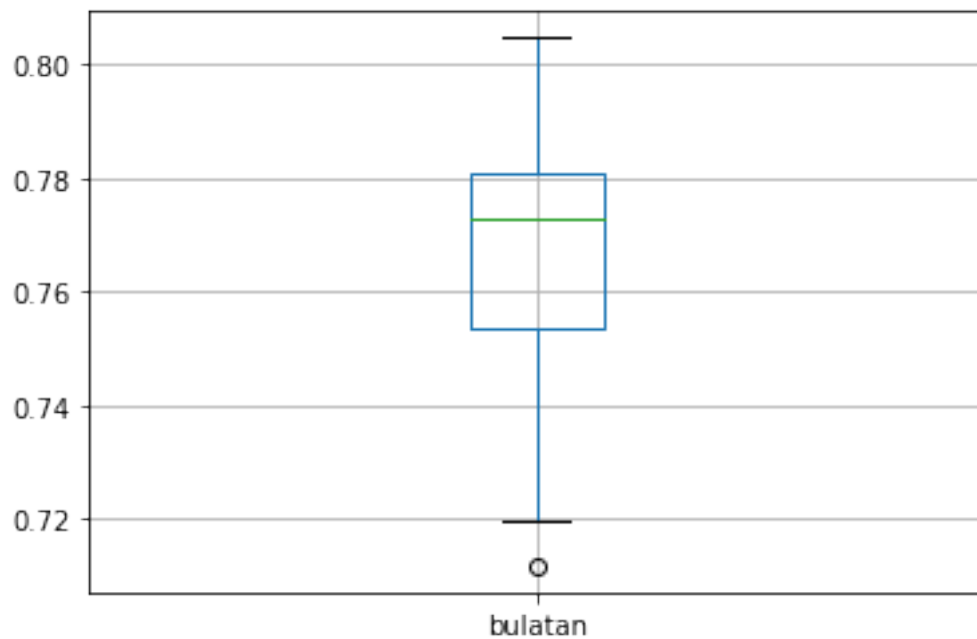
[405]: df = pd.DataFrame(data = bagian1)
        boxplot = df.boxplot()

```



### 5.3.3 Boxplot 20 Baris Terakhir

```
[406]: df = pd.DataFrame(data = bagian2)
       boxplot = df.boxplot()
```



## 5.4 D) Proporsi nilai bagian awal Ransum yang lebih dari 2, adalah lebih besar daripada, proporsi nilai yang sama di bagian akhir Ransum?

### 5.4.1 Langkah-Langkah

1. Tentukan hipotesis nol

$$H_0 : p_1 - p_2 = 0$$

2. Tentukan hipotesis alternatif

$$H_1 : p_1 - p_2 > 0 \text{ (one-tailed)}$$

3. Tentukan tingkat signifikan

$$\alpha = 0.05$$

4. Tentukan uji statistik yang sesuai dan tentukan daerah kritis.

Uji statistik : Uji statistik satu parameter populasi pengujian proporsi dua sampel one-tailed tes normal, z

$$\text{Daerah kritis : } z > z_\alpha : z > 1.645$$

5. Hitung nilai uji statistik

```
[407]: ransum = gandum['ransum']

n1 = ransum.head(ransum.size // 2)
n2 = ransum.tail(ransum.size // 2)

x1 = n1.loc[n1 > 2]
x2 = n2.loc[n2 > 2]

p1 = x1.size/n1.size
p2 = x2.size/n2.size

p = (x1.size + x2.size)/(n1.size + n2.size)
q = 1 - p

zalpha = ptot(1- alpha)

pembilang = p1 - p2
penyebut = (p*q/n1.size) + (p*q/n2.size)
z = pembilang/math.sqrt(penyebut)
p = 1-ztop(z)
print("Nilai z : " + str(z))
print("p-value : " + str(p))
```

Nilai z : 13.397486455610238

p-value : 0.0

6. Ambil Keputusan

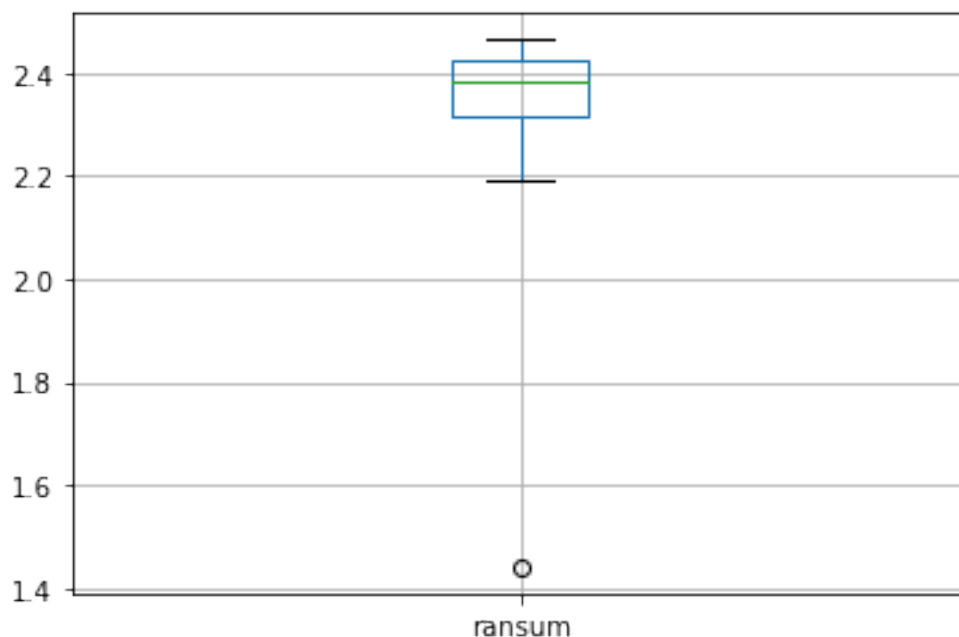
```
[408]: if z > zalpha:
        print("Null hypothesis ditolak")
        print("Proporsi nilai bagian awal Ransum yang lebih dari 2 lebih besar dari_
        ↳proporsi nilai yang sama di bagian akhir Ransum")
    else:
        print("Null hypothesis diterima")
        print("Proporsi nilai bagian awal Ransum yang lebih dari 2 sama dengan_
        ↳proporsi nilai yang sama di bagian akhir Ransum")
```

Null hypothesis ditolak

Proporsi nilai bagian awal Ransum yang lebih dari 2 lebih besar dari proporsi nilai yang sama di bagian akhir Ransum

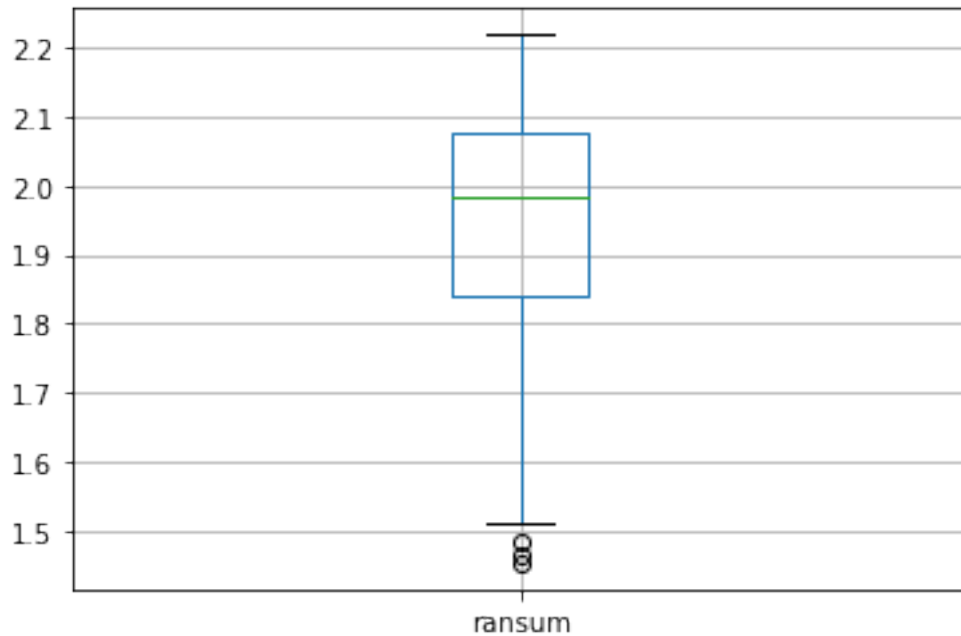
#### 5.4.2 Boxplot Bagian Awal

```
[409]: df = pd.DataFrame(data = n1)
        boxplot = df.boxplot()
```



#### 5.4.3 Boxplot Bagian Akhir

```
[410]: df = pd.DataFrame(data = n2)
        boxplot = df.boxplot()
```



## 5.5 E) Bagian awal kolom Diameter memiliki variansi yang sama dengan bagian akhirnya?

### 5.5.1 Langkah-Langkah

1. Tentukan hipotesis nol

$$H_0 : \sigma_1^2 - \sigma_2^2 = 0$$

2. Tentukan hipotesis alternatif

$$H_1 : \sigma_1^2 - \sigma_2^2 \neq 0 \text{ (two-tailed)}$$

3. Tentukan tingkat signifikan

$$\alpha = 0.05$$

4. Tentukan uji statistik yang sesuai dan tentukan daerah kritis.

Uji statistik : distribusi F

Daerah kritis :  $f < f_{1-\alpha/2}(v_1, v_2)$  or  $f > f_{\alpha/2}(v_1, v_2)$ :  $f < 0.78$  or  $f > 1.28$

5. Hitung nilai uji statistik

```
[411]: diameter = gandum['diameter']
awal = diameter.head(diameter.size // 2)
akhir = diameter.tail(diameter.size // 2)
awalvar = awal.var()
akhirvar = akhir.var()
#f = variansi besar/variansi kecil
```

```

if awalvar > akhirvar:
    f = awalvar/akhirvar
else:
    f = akhirvar/awalvar
f1 = s.f.ppf(q=1 - alpha/2, dfn=awal.size-1, dfd=akhir.size-1)
f2 = 1 / (s.f.ppf(q=1 - alpha/2, dfn=akhir.size-1, dfd=awal.size-1))
p = 2*(1-s.f.cdf(f, akhir.size-1, awal.size-1)) #find p-value of F test
    ↳ statistic
print("Nilai f : " + str(f))
print("p-value : " + str(p))

```

Nilai f : 1.0838780220421882

p-value : 0.5255791840378787

#### 6. Ambil Keputusan

```

[412]: if(f > f1 or f < f2):
        print("Null hypothesis ditolak")
        print("Bagian awal kolom Diameter tidak memiliki variansi yang sama dengan_
        ↳ bagian akhirnya")
    else:
        print("Null hypothesis gagal ditolak")
        print("Tidak cukup bukti untuk mengambil kesimpulan bahwa variansi berbeda")

```

Null hypothesis gagal ditolak

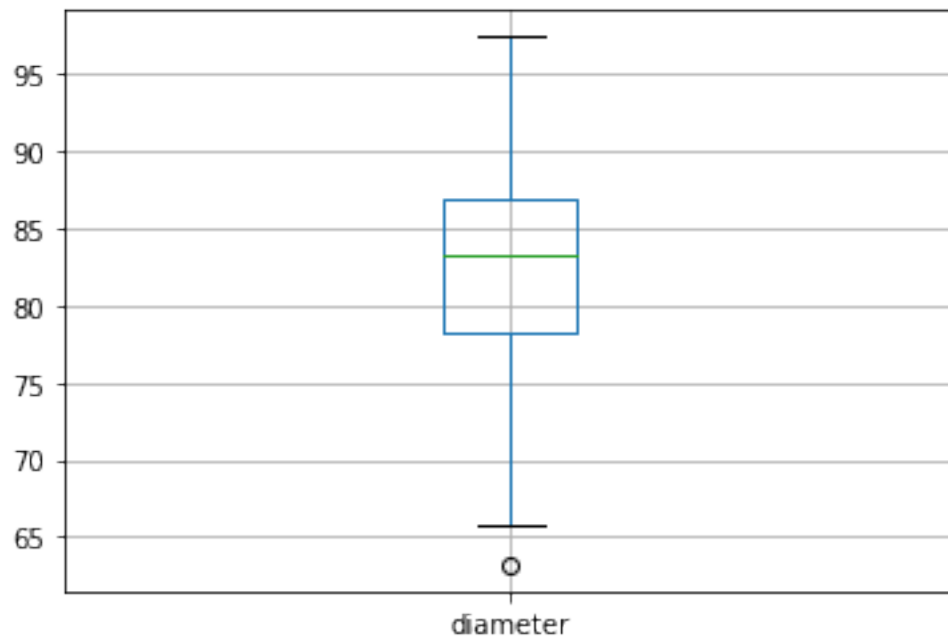
Tidak cukup bukti untuk mengambil kesimpulan bahwa variansi berbeda

#### 5.5.2 Boxplot Bagian Awal

```

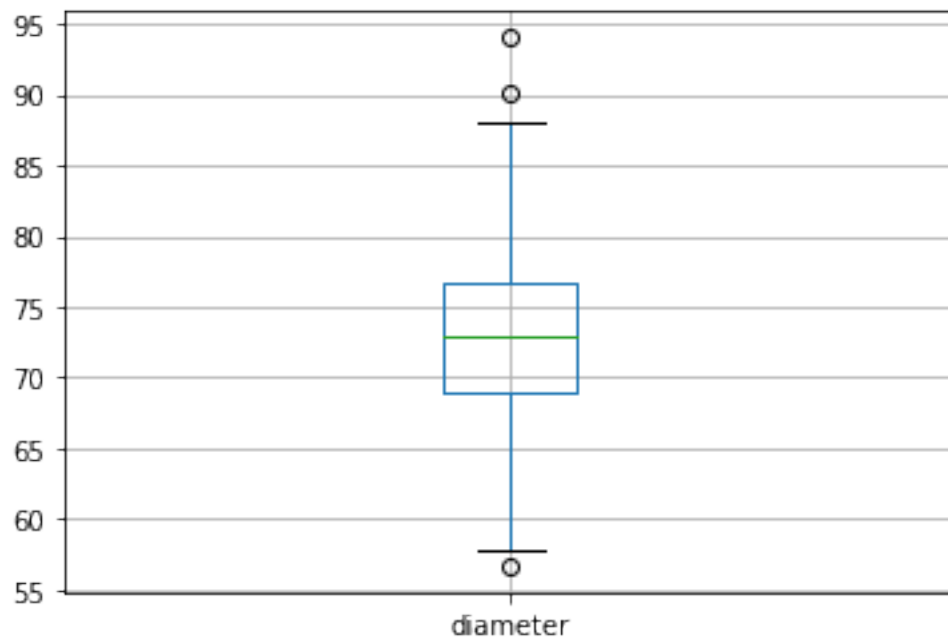
[413]: df = pd.DataFrame(data = awal)
        boxplot = df.boxplot()

```



### 5.5.3 Boxplot Bagian Akhir

```
[414]: df = pd.DataFrame(data = akhir)
       boxplot = df.boxplot()
```



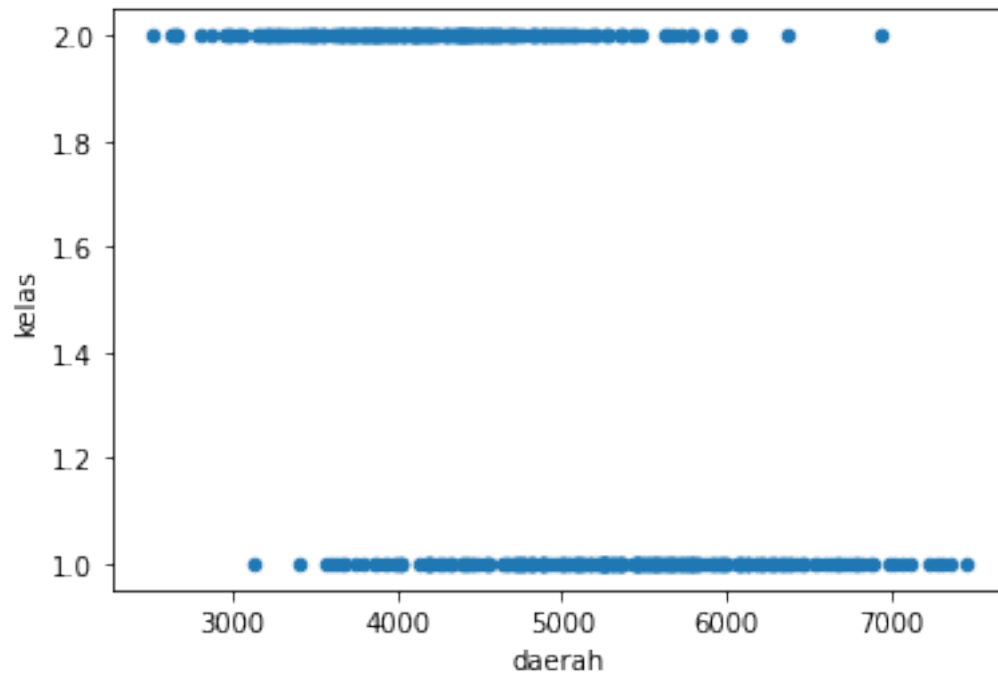


## 6 Test Korelasi

### 6.1 Daerah dengan Kelas

```
[415]: g.plot(kind='scatter', x='daerah', y='kelas')
a = g['daerah']
b = g['kelas']
print("Nilai korelasi antara daerah dan kelas: " + str(a.corr(b)))
```

Nilai korelasi antara daerah dan kelas: -0.6027466517416661



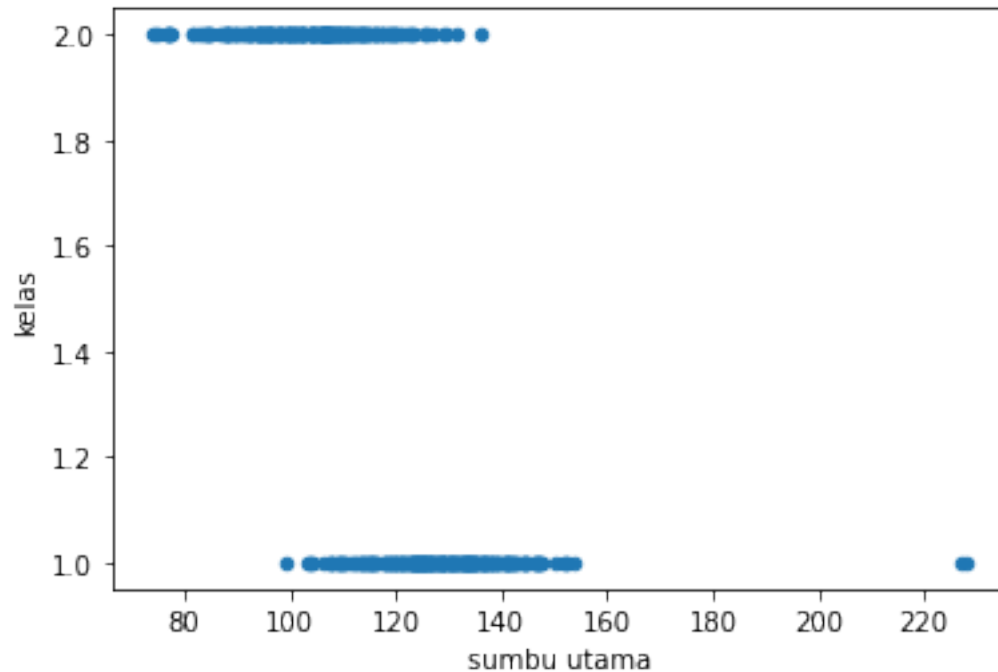
#### 6.1.1 Kesimpulan :

Daerah dan kelas memiliki korelasi yang besar dengan nilai -0.6027. Karena nilai korelasi besar dan negatif maka kedua kolom cenderung berbanding terbalik.

### 6.2 Sumbu Utama dengan Kelas

```
[416]: g.plot(kind='scatter', x='sumbu utama', y='kelas')
a = g['sumbu utama']
b = g['kelas']
print("Nilai korelasi antara sumbu utama dan kelas: " + str(a.corr(b)))
```

Nilai korelasi antara sumbu utama dan kelas: -0.7130906104204592



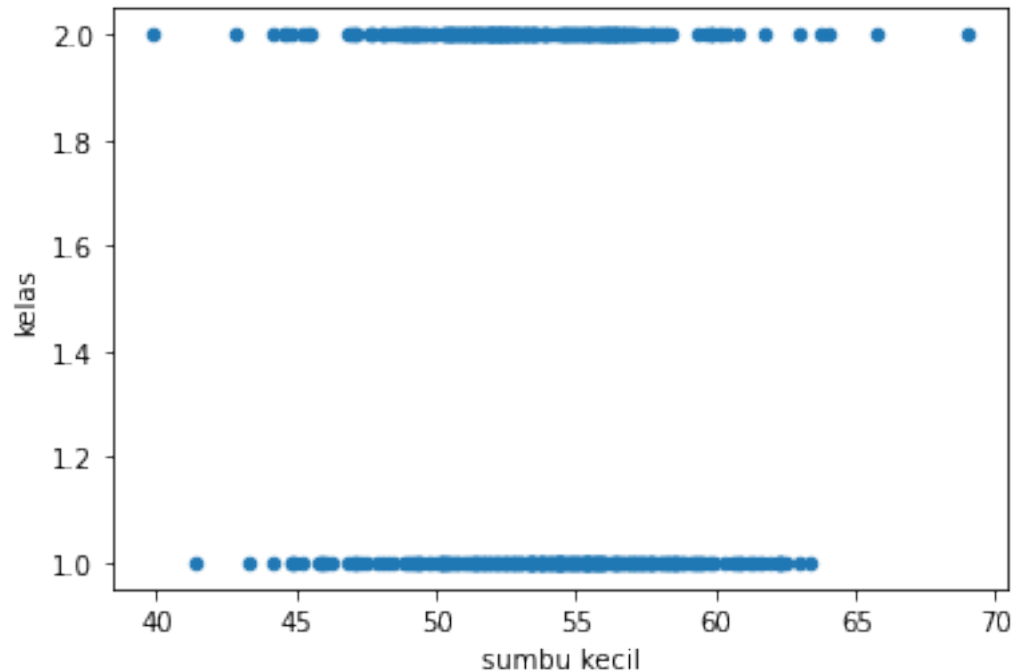
### 6.2.1 Kesimpulan :

Sumbu utama dan kelas memiliki korelasi yang besar dengan nilai -0.71309. Karena nilai korelasi besar dan negatif maka kedua kolom cenderung berbanding terbalik.

### 6.3 Sumbu Kecil dengan Kelas

```
[417]: g.plot(kind='scatter', x='sumbu kecil', y='kelas')
a = g['sumbu kecil']
b = g['kelas']
print("Nilai korelasi antara sumbu kecil dan kelas: " + str(a.corr(b)))
```

Nilai korelasi antara sumbu kecil dan kelas: -0.1529751733553502



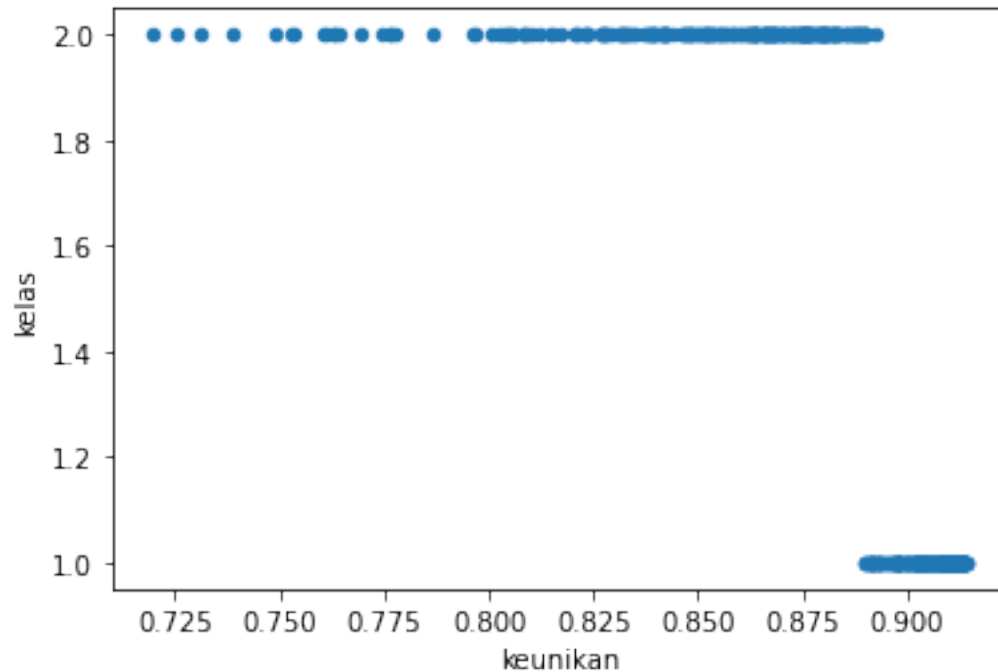
### 6.3.1 Kesimpulan :

Sumbu kecil dan kelas memiliki korelasi yang kecil dengan nilai -0.1529751. Karena nilai korelasi kecil, maka bisa dibilang bahwa nilai di kolom kelas tidak begitu mempengaruhi nilai di kolom sumbu kecil.

## 6.4 Keunikan dengan Kelas

```
[418]: g.plot(kind='scatter', x='keunikan', y='kelas')
a = g['keunikan']
b = g['kelas']
print("Nilai korelasi antara keunikan dan kelas: " + str(a.corr(b)))
```

Nilai korelasi antara keunikan dan kelas: -0.7304563686511923



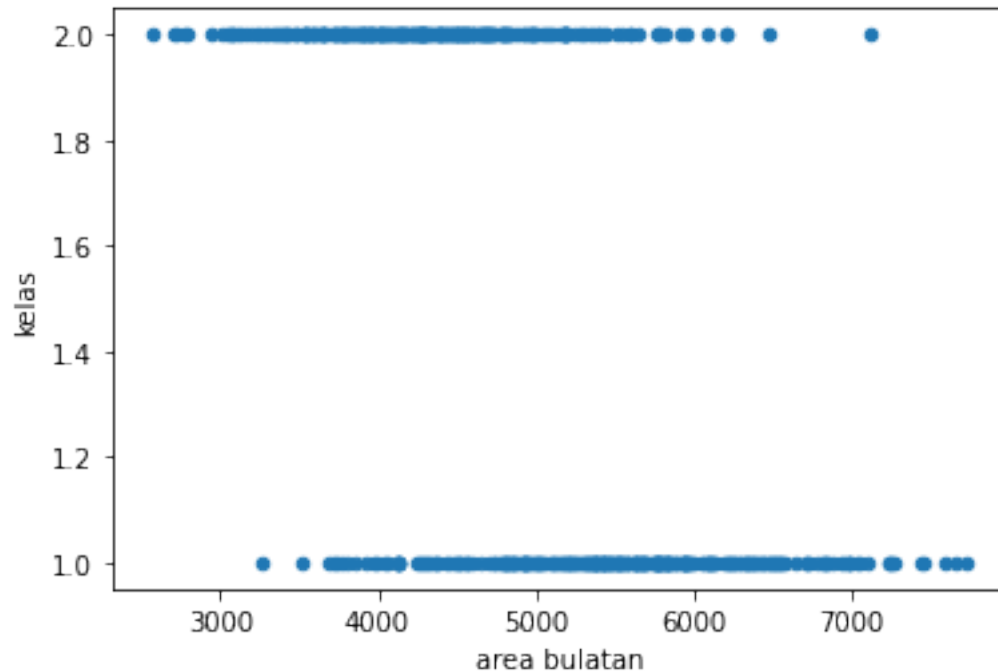
#### 6.4.1 Kesimpulan :

Keunikan dan kelas memiliki korelasi yang besar dengan nilai -0.730456. Karena nilai korelasi besar dan negatif maka kedua kolom cenderung berbanding terbalik.

### 6.5 Area Bulatan dengan Kelas

```
[419]: g.plot(kind='scatter', x='area bulatan', y='kelas')
a = g['area bulatan']
b = g['kelas']
print("Nilai korelasi antara area bulatan dan kelas: " + str (a.corr(b)))
```

Nilai korelasi antara area bulatan dan kelas: -0.6073125434153749



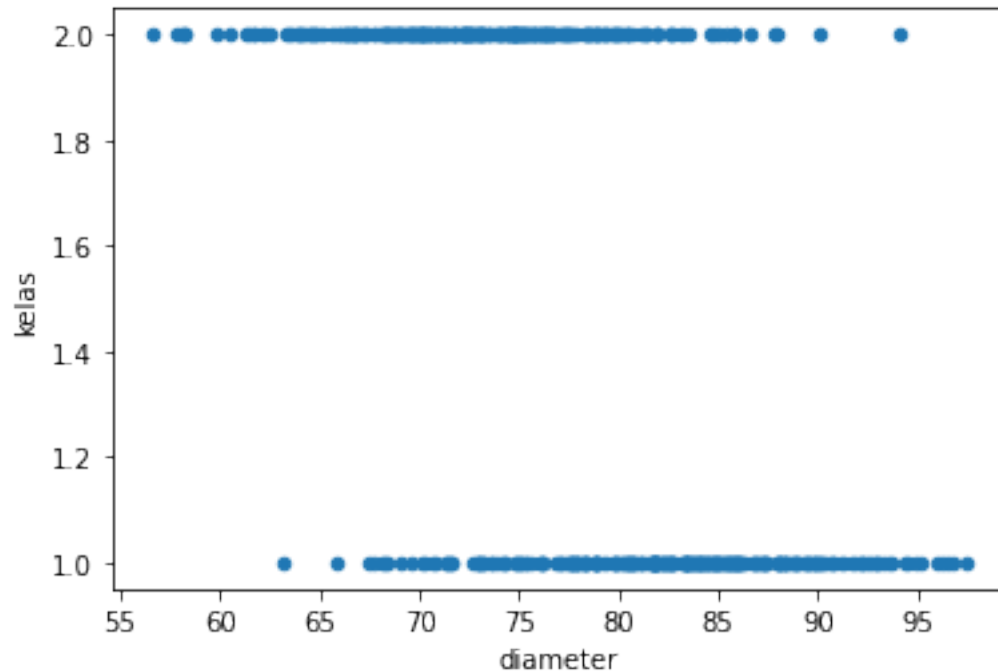
### 6.5.1 Kesimpulan :

Area bulatan dan kelas memiliki korelasi yang besar dengan nilai -0.607312. Karena nilai korelasi besar dan negatif maka kedua kolom cenderung berbanding terbalik.

## 6.6 Diameter dengan Kelas

```
[420]: g.plot(kind='scatter', x='diameter', y='kelas')
a = g['diameter']
b = g['kelas']
print("Nilai korelasi antara diameter dan kelas: " + str(a.corr(b)))
```

Nilai korelasi antara diameter dan kelas: -0.6025356896618809



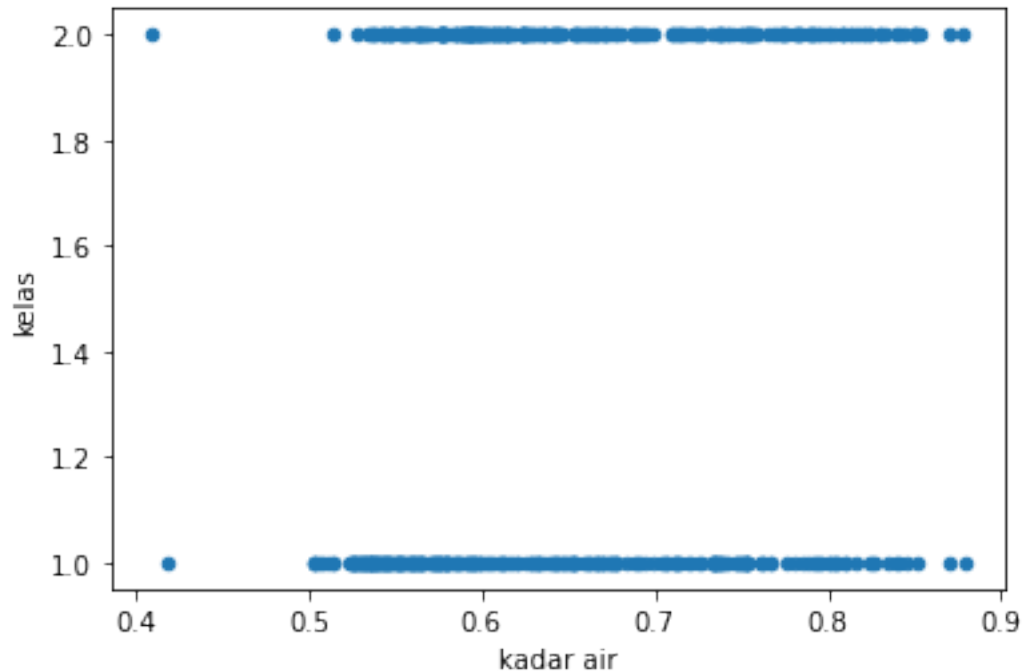
### 6.6.1 Kesimpulan :

Diameter dan kelas memiliki korelasi yang besar dengan nilai -0.6025356. Karena nilai korelasi besar dan negatif maka kedua kolom cenderung berbanding terbalik.

## 6.7 Kadar Air dengan Kelas

```
[421]: g.plot(kind='scatter', x='kadar air', y='kelas')
a = g['kadar air']
b = g['kelas']
print("Nilai korelasi antara kadar air dan kelas: " + str(a.corr(b)))
```

Nilai korelasi antara kadar air dan kelas: 0.13434422605727642



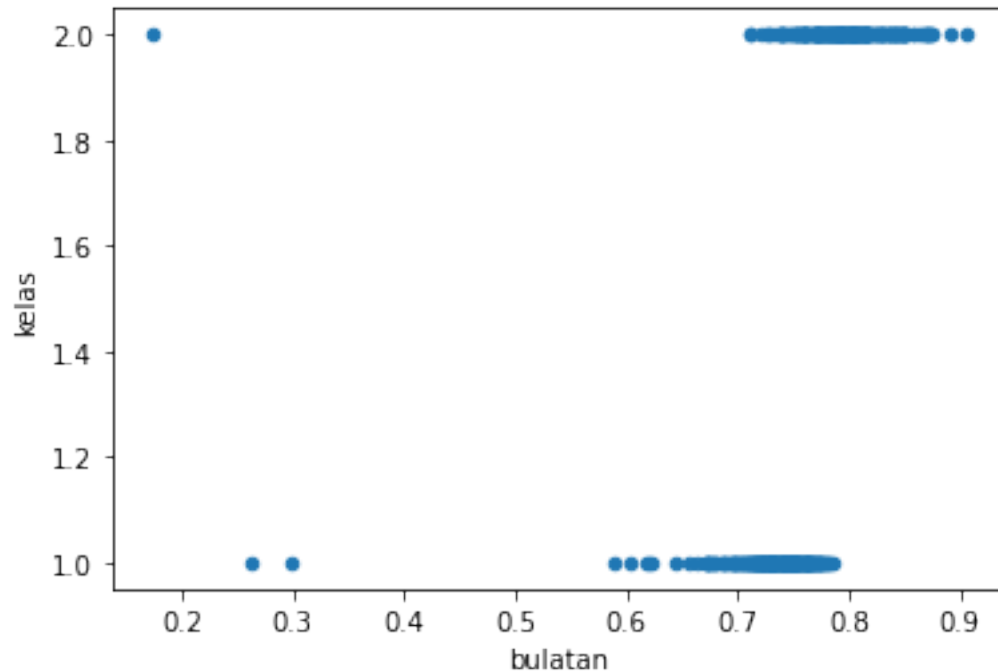
### 6.7.1 Kesimpulan :

Kadar air dan kelas memiliki korelasi yang kecil dengan nilai 0.134344. Karena nilai korelasi kecil, maka bisa dibilang bahwa nilai di kolom kelas tidak begitu mempengaruhi nilai di kolom kadar air.

## 6.8 Bulatan dengan Kelas

```
[422]: g.plot(kind='scatter', x='bulatan', y='kelas')
a = g['bulatan']
b = g['kelas']
print("Nilai korelasi antara bulatan dan kelas: " + str(a.corr(b)))
```

Nilai korelasi antara bulatan dan kelas: 0.5450045317240071



### 6.8.1 Kesimpulan :

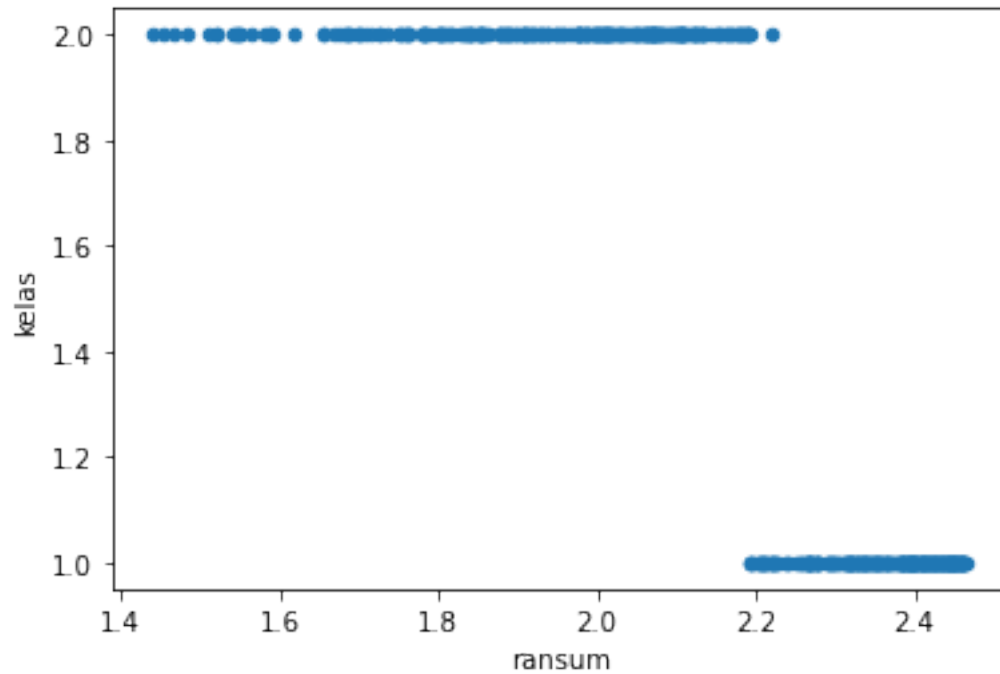
Bulatan dan kelas memiliki korelasi yang besar dengan nilai 0.545. Karena nilai korelasi besar dan positif maka kedua kolom cenderung berbanding lurus.

## 6.9 Ransum dengan Kelas

```
[423]: g.plot(kind='scatter', x='ransum', y='kelas')
a = g['ransum']
b = g['kelas']
print("Nilai korelasi antara ransum dan kelas: " + str(a.corr(b)))
```

Nilai korelasi antara ransum dan kelas: -0.8399038681287484





#### 6.9.1 Kesimpulan :

Ransum dan kelas memiliki korelasi yang besar dengan nilai -0.8399. Karena nilai korelasi besar dan negatif maka kedua kolom cenderung berbanding terbalik.