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

Nomor 1

'Conductivity',

Menulis deskripsi statistika (Descriptive Statistics) dari semua kolom pada data yang bersifat numerik, terdiri dari mean, median, modus, standar deviasi, variansi, range, nilai minimum, maksimum, kuartil, IQR, skewness dan kurtosis. Boleh juga ditambahkan deskripsi lain

```
# Mengimport Library
import pandas as pd
# Membaca data dan melihat beberapa data pertama
df = pd.read_csv('water_potability.csv', index_col=0)
df.head()
                 Hardness
                                 Solids Chloramines
                                                          Sulfate \
           На
id
                           22018.417441
               214.373394
                                                       356.886136
1
     8.316766
                                             8.059332
2
     9.092223 181.101509
                           17978.986339
                                             6.546600
                                                       310.135738
3
     5.584087 188.313324
                          28748.687739
                                             7.544869
                                                       326.678363
4
    10.223862
              248.071735
                           28749.716544
                                             7.513408
                                                       393.663396
5
     8.635849 203.361523
                          13672.091764
                                             4.563009
                                                       303.309771
                  OrganicCarbon Trihalomethanes
                                                   Turbidity
    Conductivity
Potability
id
                      18.436524
                                       100.341674
                                                    4.628771
1
      363.266516
0
                      11.558279
2
      398.410813
                                       31,997993
                                                    4.075075
0
3
      280.467916
                       8.399735
                                        54.917862
                                                    2.559708
0
4
      283.651634
                      13.789695
                                       84.603556
                                                    2.672989
0
5
      474.607645
                      12.363817
                                       62.798309
                                                    4.401425
0
# Menyatakan kolom-kolom numerik
col numeric = [
    'pH',
    'Hardness',
    'Solids',
    'Chloramines',
    'Sulfate',
```

```
'OrganicCarbon',
    'Trihalomethanes',
    'Turbidity'
1
# Memilih kolom-kolom numerik
df numeric = df[col numeric]
# Describe data + tambahan untuk kelengkapan spesifikasi
df numeric desc = df numeric.describe()
df numeric desc.loc['variance'] = df_numeric.var()
df numeric desc.loc['IQR'] = df numeric.quantile(0.75) -
df numeric.quantile(0.25)
df numeric desc.loc['range'] = df numeric.max() - df numeric.min()
df numeric desc.loc['skew'] = df numeric.skew()
df numeric desc.loc['kurtosis'] = df_numeric.kurt()
# display Descriptive Statistics
df numeric desc
                          Hardness
                                           Solids Chloramines
                   На
Sulfate \
          2010.000000 2010.000000
                                    2.010000e+03 2010.000000
count
2010.000000
             7.087193
                        195.969209
                                    2.190467e+04
                                                      7.134322
mean
333.211376
             1.572803
                         32.643166 8.625398e+03
                                                      1.585214
std
41.211111
                         73.492234
             0.227499
                                    3.209426e+02
                                                      1.390871
min
129.000000
25%
             6.090785
                        176.740657
                                    1.561441e+04
                                                      6.138326
307.626986
             7.029490
                        197.203525
                                    2.092688e+04
                                                      7.142014
50%
332.214113
75%
             8.053006
                        216.447589
                                    2.717053e+04
                                                      8.109933
359.268147
            14.000000
                        317.338124
                                    5.648867e+04
                                                     13.127000
max
481.030642
             2.473709
                       1065.576277 7.439749e+07
                                                      2.512904
variance
1698.355672
IQR
             1.962221
                         39.706932
                                    1.155612e+04
                                                      1.971607
51.641161
            13.772501
                        243.845890
                                    5.616773e+04
                                                     11.736129
range
352.030642
             0.048535
                         -0.085321
                                    5.910114e-01
                                                      0.013003
skew
0.045728
             0.626904
                          0.525480
                                    3.373203e-01
                                                      0.549782
kurtosis
0.786854
```

count	2010.000000	2010.000000	2010.000000	2010.000000
mean	426.476708	14.357940	66.400717	3.969497
std	80.701872	3.325770	16.081109	0.780471
min	201.619737	2.200000	8.577013	1.450000
25%	366.619219	12.122530	55.949993	3.442882
50%	423.438372	14.323286	66.482041	3.967374
75%	482.209772	16.683562	77.294613	4.514663
max	753.342620	27.006707	124.000000	6.494749
variance	6512.792113	11.060746	258.602066	0.609135
IQR	115.590553	4.561031	21.344620	1.071781
range	551.722883	24.806707	115.422987	5.044749
skew	0.268012	-0.020220	-0.051383	-0.032266
kurtosis	-0.237206	0.031018	0.223017	-0.049831