Sample program for Factor Analysis

Import libraries

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from factor_analyzer import FactorAnalyzer
```

Read CSV data

In [2]:
```
df = pd.read_csv('dm-end1-3.csv', delimiter=',', skiprows=0, header=
0)
print(df.shape)
print(df.info())
display(df.head())
```

```
(100, 4)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 4 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   sepal length    100 non-null    float64
 1   sepal width     100 non-null    float64
 2    petal length   100 non-null    float64
 3    petal width    100 non-null    float64
dtypes: float64(4)
memory usage: 3.2 KB
None
```

	sepal length	sepal width	petal length	petal width
0	5.8	4.0	1.2	0.2
1	5.1	2.5	3.0	1.1
2	6.6	3.0	4.4	1.4
3	5.4	3.9	1.3	0.4
4	7.9	3.8	6.4	2.0

Factor analysis

In [3]:
```
fa = FactorAnalyzer(n_factors=2, rotation='varimax', method='ml')
#fa = FactorAnalyzer(n_factors=2, rotation='varimax', method='minres
')
fa.fit(df.values)
```

Out[3]:
```
FactorAnalyzer(bounds=(0.005, 1), impute='median', is_corr_matrix=
False,
               method='ml', n_factors=2, rotation='varimax', rotat
ion_kwargs={},
               use_smc=True)
```

Correlation matrix (相関行列)

In [4]:
```
df_corr = df.corr(method='pearson')
display(df_corr)
```

	sepal length	sepal width	petal length	petal width
sepal length	1.000000	-0.168730	0.877020	0.824843
sepal width	-0.168730	1.000000	-0.470254	-0.417026
petal length	0.877020	-0.470254	1.000000	0.963373
petal width	0.824843	-0.417026	0.963373	1.000000

Eigenvalues (固有値)

In [5]:
```
eigen_org, eigen_cf = fa.get_eigenvalues()
ser_eigen_org = pd.Series(eigen_org)
ser_eigen_cf = pd.Series(eigen_cf)
print(ser_eigen_org)
#print(ser_eigen_cf)
```

```
0    2.971179
1    0.868739
2    0.138571
3    0.021511
dtype: float64
```

Ans.1

2.97
0.87

Loadings (因子負荷量)

In [8]:
```
loadings = fa.loadings_
df_loadings = pd.DataFrame(loadings, index=df.columns,
                           columns=['Factor1','Factor2'])
display(df_loadings)
```

	Factor1	Factor2
sepal length	0.995881	0.056843
sepal width	-0.131062	-0.680056
petal length	0.850871	0.520716
petal width	0.797756	0.544999

Ans. 2

0.851 (petal length, factor 1)

Uniquenesses (独自因子の割合)

In [9]:
```
uniqueness = fa.get_uniquenesses()
ser_uniqueness = pd.Series(uniqueness, index=df.columns)
print(ser_uniqueness)
```

```
sepal length     0.004991
sepal width      0.520347
 petal length    0.004873
 petal width     0.066561
dtype: float64
```

Ans. 3

petal width's uniqueness:
0.067

Contribution of each factor (各共通因子の寄与)

- Variance: 因子負荷量平方和(因子寄与)
- Proportion Var: 寄与率
- Cumulative Var: 累積寄与率

In [10]:
```
fa_var = fa.get_factor_variance()
df_fa_var = pd.DataFrame(fa_var,
                         index=['var', 'prop_var', 'cum_var'],
                         columns=['Factor1', 'Factor2'])
display(df_fa_var)
```

	Factor1	Factor2
var	2.369352	1.033875
prop_var	0.592338	0.258469
cum_var	0.592338	0.850807

Ans.4

0.851