

MANOVA

- IRIS
- We are going to use the width and length columns as depedent variabbles.
- Besides, the species column is used as the independent variable.
- MANOVA by the Statsmodel library.

```
In [ ]: # import Librarieres
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from statsmodels.multivariate.manova import MANOVA
```

```
In [ ]: # Load the dataset
phool = sns.load_dataset('iris')
phool.head()
```

Out[]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

```
In [ ]: phool.columns
```

Out[]: Index(['sepal_length', 'sepal_width', 'petal_length', 'petal_width', 'species'], dtype='object')

```
In [ ]: # Apply MANOVA
manova_phool = MANOVA.from_formula('sepal_length + sepal_width + petal_length + petal_width ~ species', data=phool)
print(manova_phool.mv_test())
```

Multivariate linear model						
=====						

Intercept	Value	Num DF	Den DF	F Value	Pr > F	

Wilks' lambda	0.0170	4.0000	144.0000	2086.7720	0.0000	
Pillai's trace	0.9830	4.0000	144.0000	2086.7720	0.0000	
Hotelling-Lawley trace	57.9659	4.0000	144.0000	2086.7720	0.0000	
Roy's greatest root	57.9659	4.0000	144.0000	2086.7720	0.0000	

species	Value	Num DF	Den DF	F Value	Pr > F	

Wilks' lambda	0.0234	8.0000	288.0000	199.1453	0.0000	
Pillai's trace	1.1919	8.0000	290.0000	53.4665	0.0000	
Hotelling-Lawley trace	32.4773	8.0000	203.4024	582.1970	0.0000	
Roy's greatest root	32.1919	4.0000	145.0000	1166.9574	0.0000	
=====						

- How to read MANOVA.\ The p-value to consider in this case is that of Wilk's lambda, relative to output variable (Species). As we can see even in this case it is significant

ASSIGNMENTS

1. How to save the results in a tabulated format or a picture?

2. How to interpret Manova?

Solutions to ASSIGNMENTS

1. How to save the results in a tabulated format or a picture?

```
In [ ]: x = pd.DataFrame((manova_phool.mv_test().results['species'] ['stat']))
x
```

Out[]:

	Value	Num DF	Den DF	F Value	Pr > F
Wilks' lambda	0.023439	8	288.0	199.145344	0.0
Pillai's trace	1.191899	8.0	290.0	53.466489	0.0
Hotelling-Lawley trace	32.47732	8	203.40239	582.197018	0.0
Roy's greatest root	32.191929	4	145	1166.957433	0.0

2. How to interpret Manova?

- F Value – This is the F statistic for the given predictor and test statistic.
- Num DF – This is the number of degrees of freedom in the model.
- Den DF – This is the number of degrees of freedom associated with the model errors. Note that there are instances in MANOVA when the degrees of freedom may be a non-integer (here, the DF associated with Hotelling-Lawley Trace is a non-integer) because these degrees of freedom are calculated using the mean squared errors, which are often non-integers.
- Pr > F – This is the p-value associated with the F statistic of a given effect and test statistic. The null hypothesis that a given predictor has no effect on either of the outcomes is evaluated with regard to this p-value. For a given alpha level, if the p-value is less than alpha, the null hypothesis is rejected. If not, then we fail to reject the null hypothesis. In this example, we reject the null hypothesis that group has no effect on useful, difficulty or importance scores at alpha level .05 because the p-values are all less than .05.