

- Statistique multivariée
- Analyse des données
- Apprentissage statistique à partir des données
 - Supervisé
 - Non supervisé
- Reconnaissance des formes statistique
- Fouille de données ou *Data mining*

Exemple des Mouchérons (1)

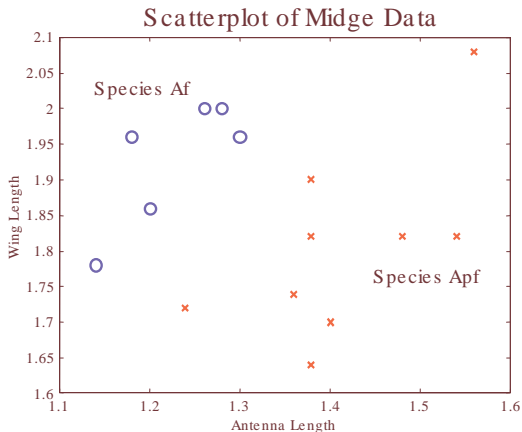
- Découverte de 2 espèces de moucheron (1981)
- Difficile de les distinguer
- On cherche à le faire avec des caractéristiques externes simples
- Données :
 - 9 mouchérons Af et 6 mouchérons Apf
 - Longueurs de l'aile et de l'antenne en mm

Exemple des Mouchérons (2)

0	1.38	1.64
0	1.40	1.70
0	1.24	1.72
0	1.36	1.74
0	1.38	1.82
0	1.48	1.82
0	1.54	1.82
0	1.38	1.90
0	1.56	2.08
1	1.14	1.78
1	1.20	1.86
1	1.18	1.96
1	1.30	1.96
1	1.26	2.00
1	1.28	2.00

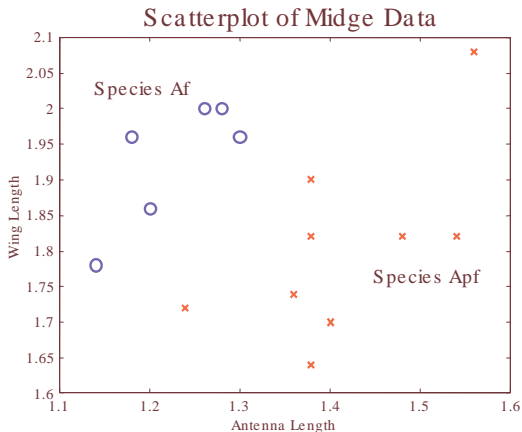
Exemple des Moucheron (2)

0	1.38	1.64
0	1.40	1.70
0	1.24	1.72
0	1.36	1.74
0	1.38	1.82
0	1.48	1.82
0	1.54	1.82
0	1.38	1.90
0	1.56	2.08
1	1.14	1.78
1	1.20	1.86
1	1.18	1.96
1	1.30	1.96
1	1.26	2.00
1	1.28	2.00



Exemple des Moucheron (2)

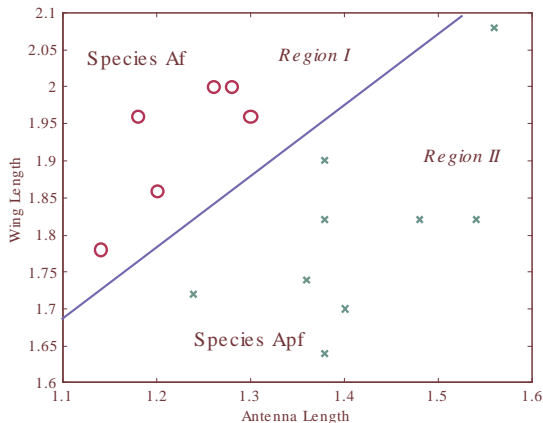
0	1.38	1.64
0	1.40	1.70
0	1.24	1.72
0	1.36	1.74
0	1.38	1.82
0	1.48	1.82
0	1.54	1.82
0	1.38	1.90
0	1.56	2.08
1	1.14	1.78
1	1.20	1.86
1	1.18	1.96
1	1.30	1.96
1	1.26	2.00
1	1.28	2.00



Comment distinguer les 2 groupes ?

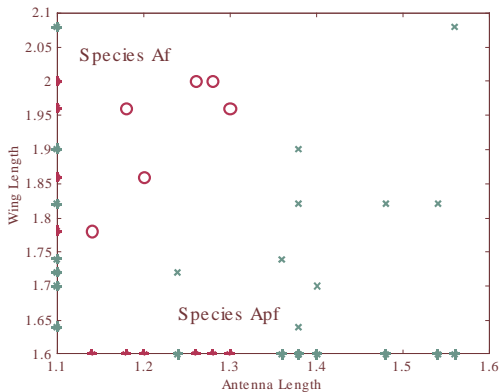
Exemple des Moucheron (3)

0	1.38	1.64
0	1.40	1.70
0	1.24	1.72
0	1.36	1.74
0	1.38	1.82
0	1.48	1.82
0	1.54	1.82
0	1.38	1.90
0	1.56	2.08
1	1.14	1.78
1	1.20	1.86
1	1.18	1.96
1	1.30	1.96
1	1.26	2.00
1	1.28	2.00



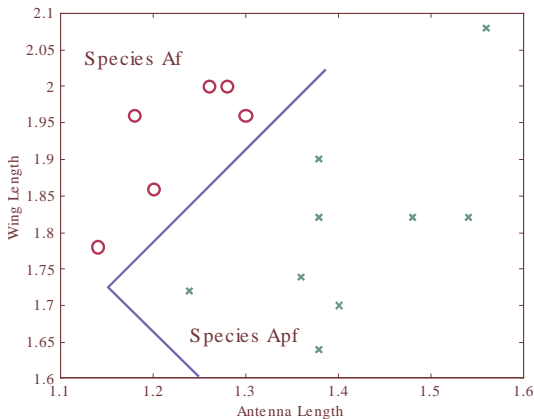
Il est facile visuellement de tracer une ligne

Exemple des Moucheron (4)



Aucune des 2 variables ne peut faire le travail seul

Exemple des Moucheron (5)



On peut voir cela comment un changement de variables

Exemple des Moucheron (6)

- **Variable intéressante** : d=aile-antenne
- Mais aussi : aile/antenne
- Intérêt du **graphe de dispersion** (scatter plot)
- Mais que faire si on a plus de 2 variables ?
- Problème de **discrimination** et ensemble d'**apprentissage**
- **Validité** des résultats sur la population totale ?
 - Nécessité de la Statistique :
 - Vecteur aléatoire : (aile, antenne)
 - Loi jointe, lois marginales, lois conditionnelles, ...

- Méthodes non supervisées :
 - Méthodes de représentation : ACP, AFC et MDS
 - Classification à base de modèles : modèle de mélange
- Méthodes supervisées

- Probabilités, analyse de données et statistique, Saporta, G., Technip, Paris (2006)
- A First Course in Multivariate Statistics, Flury, Springer (1997)
- The elements of statistical learning, Hastie, Tibshirani, Friedman, Springer (2001)
- Pattern recognition, Duda, Hart et Stork, Wiley (2000)