

# Statistiques appliquées à la biologie et l'environnement



Daniel Petit  
UMR INRA 1061  
Université de Limoges

Université de Biskra, janvier 2012


# Plan

- Techniques de base
- Régressions multiples
- Analyses des peuplements
- Climagrammes

Logiciel puissant et cher

**PAST version 1.91**

<http://folk.uio.no/ohammer/past>



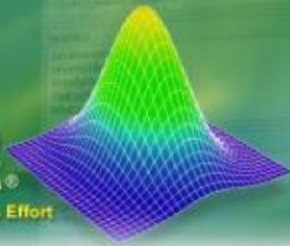
Øyvind Hammer, D.A.T. Harper (May 2009).

Copyright Hammer and Harper 1999-2009.

Hammer, Ø., Harper, D.A.T. and Ryan, P.D. 2001.  
PAST: Paleontological Statistics software package for educatic  
and data analysis. Palaeontologia Electronica 4(1):9 pp.

Logiciel convivial, évolutif  
et gratuit

SYSTAT 32-bit UNICODE English  
Version No. 12.00.00  
Standalone  
Daniel Petit  
UMR 1061 INRA  
Serial No. 797072375



**SYSTAT 12**  
For Windows®  
More Statistics. More Graphs. Less Effort

**SYSTAT**®

Copyright © SYSTAT Software, Inc. 2007  
All Rights Reserved

**Installed Products**  
SYSTAT 12 Version 12.00.00  
Monte Carlo Version 1.00.00  
Quality Analysis Version 1.00.00

# Techniques de base

## Catégories de variables

Variable qualitative : couleur, catégorie, condition de culture ou d'élevage, espèce...

Variable quantitative : mesure de poids, longueur, concentration, fréquence...


Variable semi-quantitative : abondance, classe d'abondance

## Nombre de variables à traiter


2-3 variables : statistiques univariées

Plus de 5 variables : statistiques multivariées

# Statistiques univariées


1 variable quantitative + 1 variable qualitative  Analyse de variance  
Régression logistique

2 variables quantitatives  Analyse de corrélation  
Régression

n variables quantitatives  Régression multiple

# Statistiques multivariées

n variables quantitatives  Analyse en Composantes  
Principales ACP

n variables qualitatives  Analyse Factorielle des  
Correspondances AFC  
Detrended Correspondance DC  
Multidimensional Scaling MDS

# Stratégie d'analyse

1. Analyses multivariées	Dégrossissage Visualisation des tendances	Statistiques descriptives
2. Analyses univariées	Affinage	Statistiques inférencielles

# AFC : avec données d'abondance

Pelouse  
écorchée

Pelouse à  
Fétuque

Pelouse à  
Brachypode

Clairière  
sèche

espèces	code	PE1	PE2	PE3	PE4	PE5	PE6	PE7	PE8	PE9	PE10	PF1	PF2	PF3	PF4	PF5	PF6	PF7	PF8	PF9	PF10	PB1	PB2	PB3	PB4	PB5	PB6	PB7	PB8	PB9	PB10	PS1	PS2	PS3	PS4	PS5	PS6	PS7	PS8	PS9	PS10	
Omocestus petraeus	Opetr	1	3	1	0	2	1	0	1	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Calliptamus barbarus	Cbarb	0	0	3	1	1	2	2	0	0	4	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	3	0	1	3	0	0	2	0	
Euchorthippus pulvinatus	Epulv	0	0	2	1	3	2	0	5	2	1	5	3	3	5	6	3	3	8	8	6	1	1	3	6	1	0	2	1	0	0	0	0	1	0	0	0	0	0	0	0	
Oedipoda caerulea	Ocaer	0	0																																							
Pezotettix giornai	Pgior	0	0																																							
Euchorthippus declivus	Edecl	0	0																																						1	
Chorthippus biguttulus	Cbigu	0	0																																						0	
Stenobothrus lineatus	Sline	0	0																																						0	
Gomphocerrius rufus	Gomp	0	0																																						0	
Tylopsis liliifolia	Tlili	0	0																																						1	
Platycleis grisea	Pgris	0	0																																						0	
Platycleis tessellata	Ptess	0	0																																						0	
Ruspolia nitidula	Rnitid	0	0																																						0	
Phaneroptera nana	Pnana	0	0																																						3	
Mantis religiosa	Mreli	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	1	1	0	1	0	1	0	1	0	0	1	1	0	0	1	1	2	0	2
Gryllus campestris	Gcamp	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0

*Annales Scientifiques du Limousin, 2000, n° 11 : 49-59*

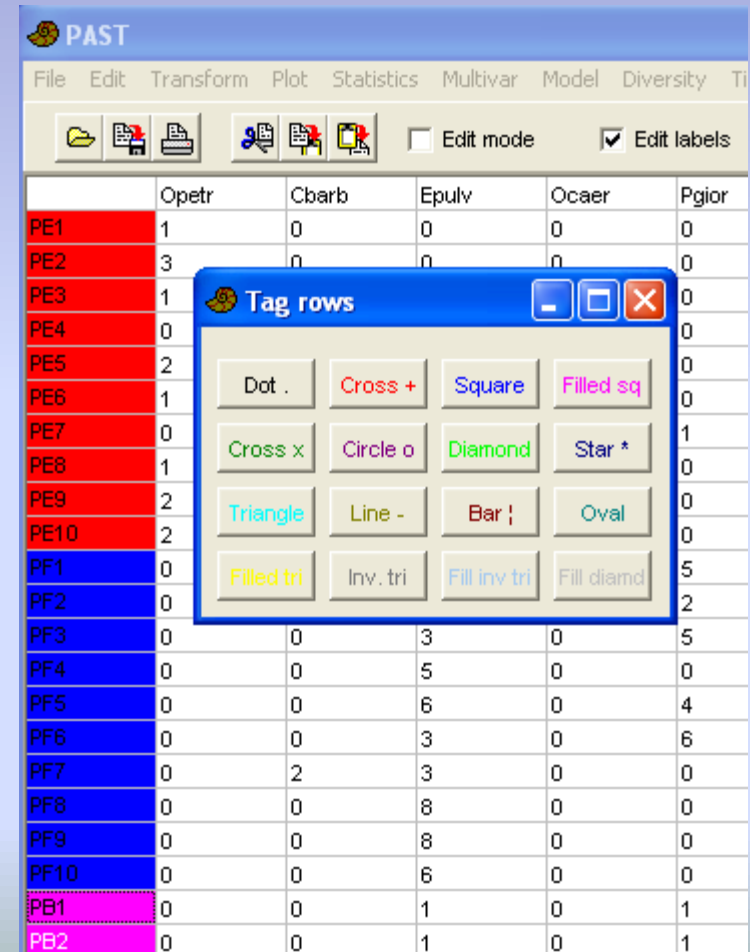
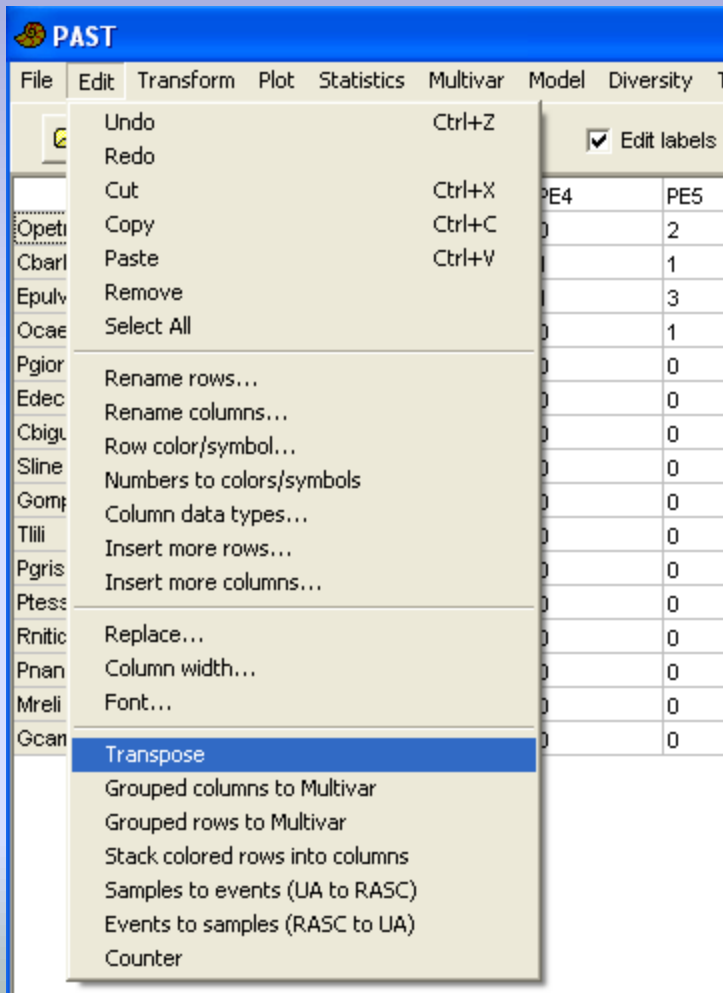
## ANALYSE DES PEUPELEMENTS D'ORTHOPTÉROÏDES SUR LES PELOUSES CALCICOLES DE LA CÔTE PELÉE DE CHASTEАUX

Picaud F. \*, Dufour, S., Decaux, E. et Petit, D.

*Laboratoire de Biologie des Populations  
Faculté des Sciences de Limoges, 123, Av. A. Thomas  
87060 Limoges Cedex*

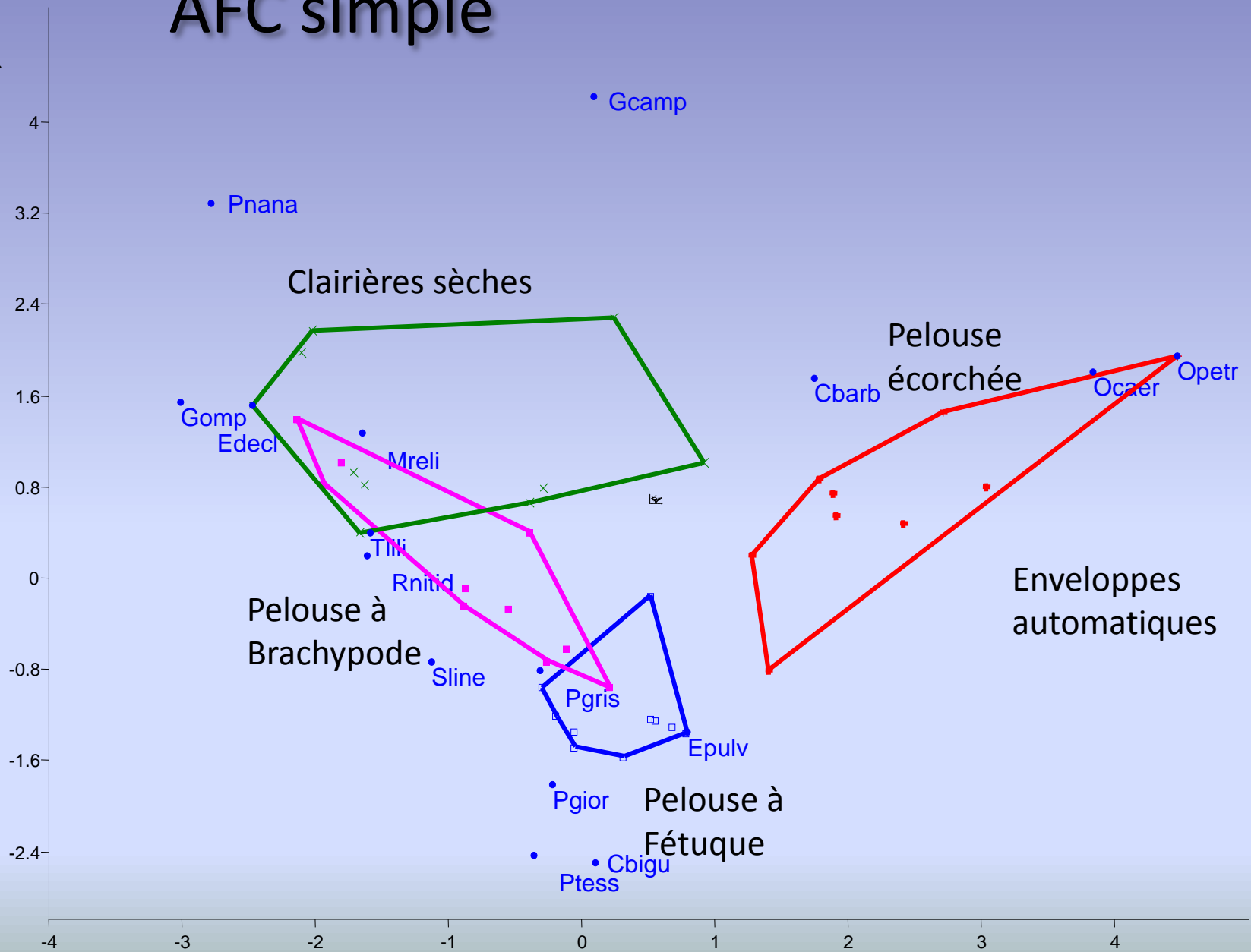
1. Transposer la matrice =  
intervertir les lignes et les  
colonnes

2. Marquer les intitulés de  
lignes par des couleurs =  
définition des groupes



# AFC simple

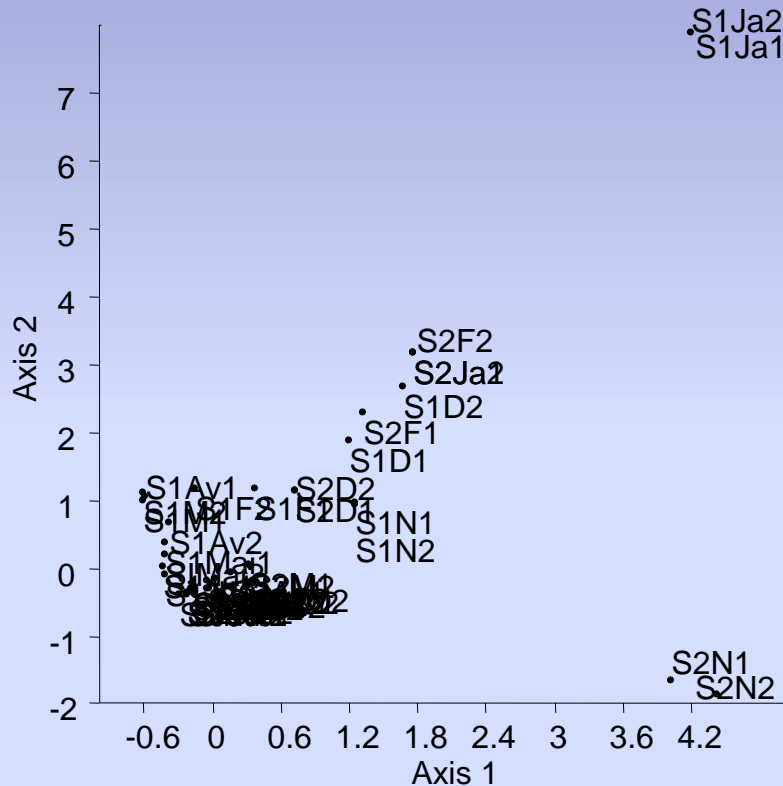
Axis 2 : 15,8 %



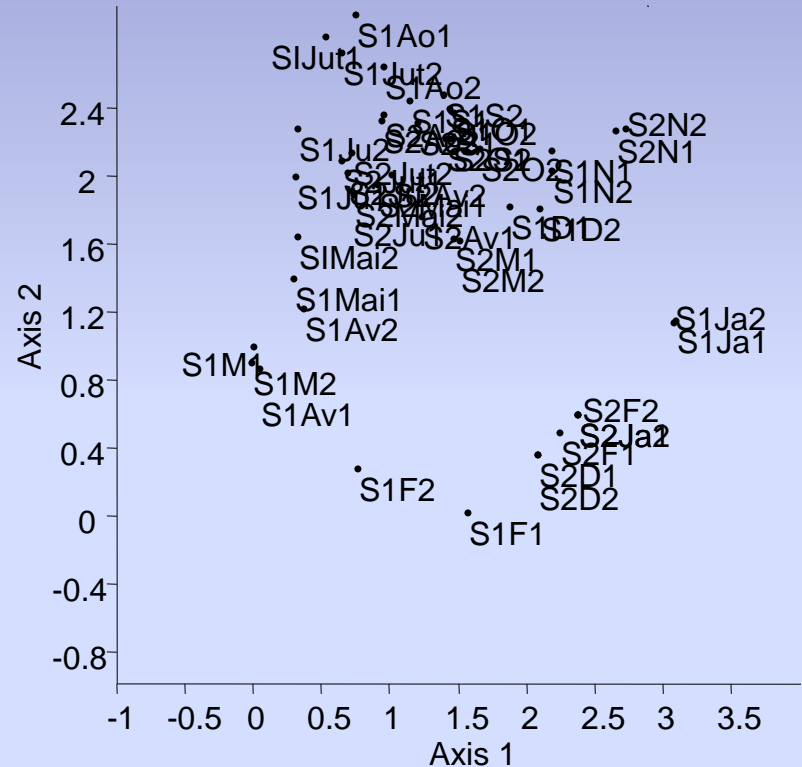
Axis 1 : 18,5 %



# AFC ou DCA (Detrended Correspondence Analysis)



AFC simple : superposition à cause de points éloignés



L'option « detrending » permet de mieux disperser le nuage de points

Relevés de l'entomofaune associée au Peuplier noir (Z. Djazouli *et al.*, en révision)

# Classification hiérarchique

## 1. obtenir les coordonnées des facteurs lignes et colonnes sur les premiers axes

### Correspondence Analysis

Axis	Eigenvalue	% of total
1	0.66387	18.586
2	0.566556	15.862
3	0.432114	12.098
4	0.361639	10.125
5	0.330211	9.2448
6	0.258149	7.2273
7	0.217233	6.0818

### Correspondence Analysis scores

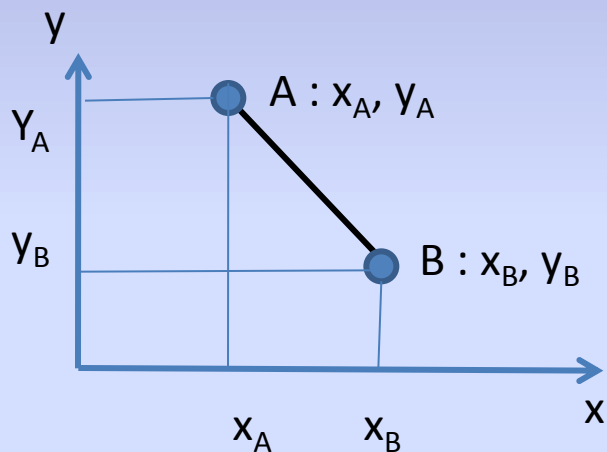
	Axis 1	Axis 2	Axis 3	Axis 4	Axis
Opetr	2.0199	1.0152	0.62136	-1.1495	-0.84
Cbarb	0.73659	0.91988	-0.21056	1.1181	0.548
Epulv	0.28373	-0.623	-0.14486	-0.19499	0.314
Ocaer	1.7245	0.94662	0.3901	0.048261	-0.39
Pgior	-0.19211	-0.84834	0.28063	0.72837	-1.01
Edecl	-1.2576	0.80279	1.3479	-0.066906	0.283
Cbigu	-0.037709	-1.1887	0.20671	0.58222	-0.86
Sline	-0.62041	-0.31477	-0.19862	-0.092505	-0.94
Gomp	-1.5097	0.81245	3.0438	-0.44262	1.083
Tilli	-0.83725	0.24414	0.617	-0.053223	-0.03
Pgris	-0.15148	-0.40538	-0.19461	-0.12452	0.712
Ptess	-0.17103	-1.2097	0.11263	1.2096	-2.02
Rnitid	-0.76061	0.093415	-0.90321	-1.0989	0.030
Pnana	-1.3129	1.6278	-1.6047	-0.23142	-1.00
Mreli	-0.7764	0.62952	-0.8265	-0.18312	-0.18
Gcamp	0.040388	2.091	-1.9231	2.5129	0.463
PE1	2.479	1.3487	0.94525	-1.9115	-1.47
PE2	2.479	1.3487	0.94525	-1.9115	-1.47

Combien d'axes faut-il considérer ?

# Classification hiérarchique

## 2. Coordonnées sur 3 axes factoriels

Choix de l'indice de similarité  
= distance euclidienne

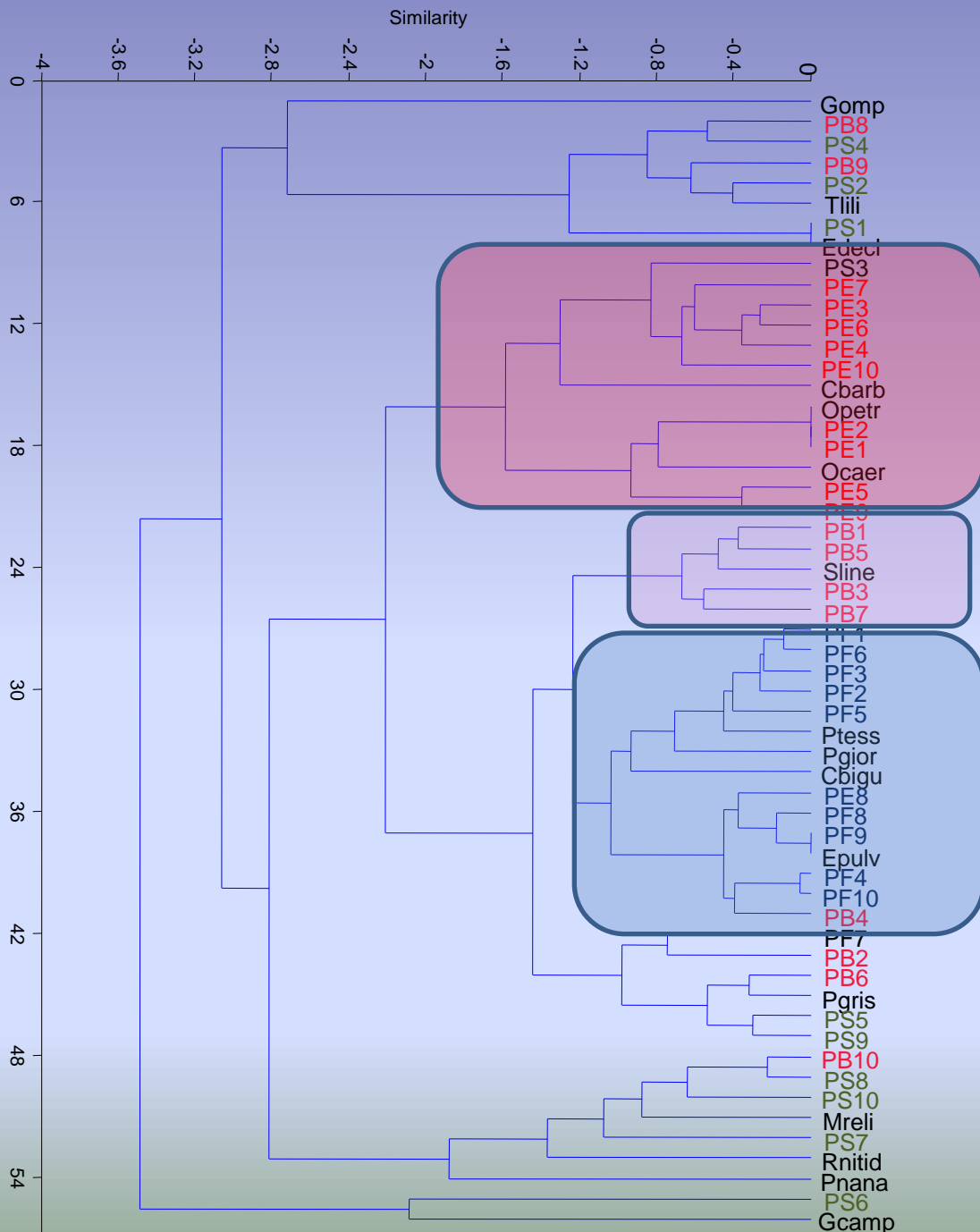


$$D_{AB} = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2}$$

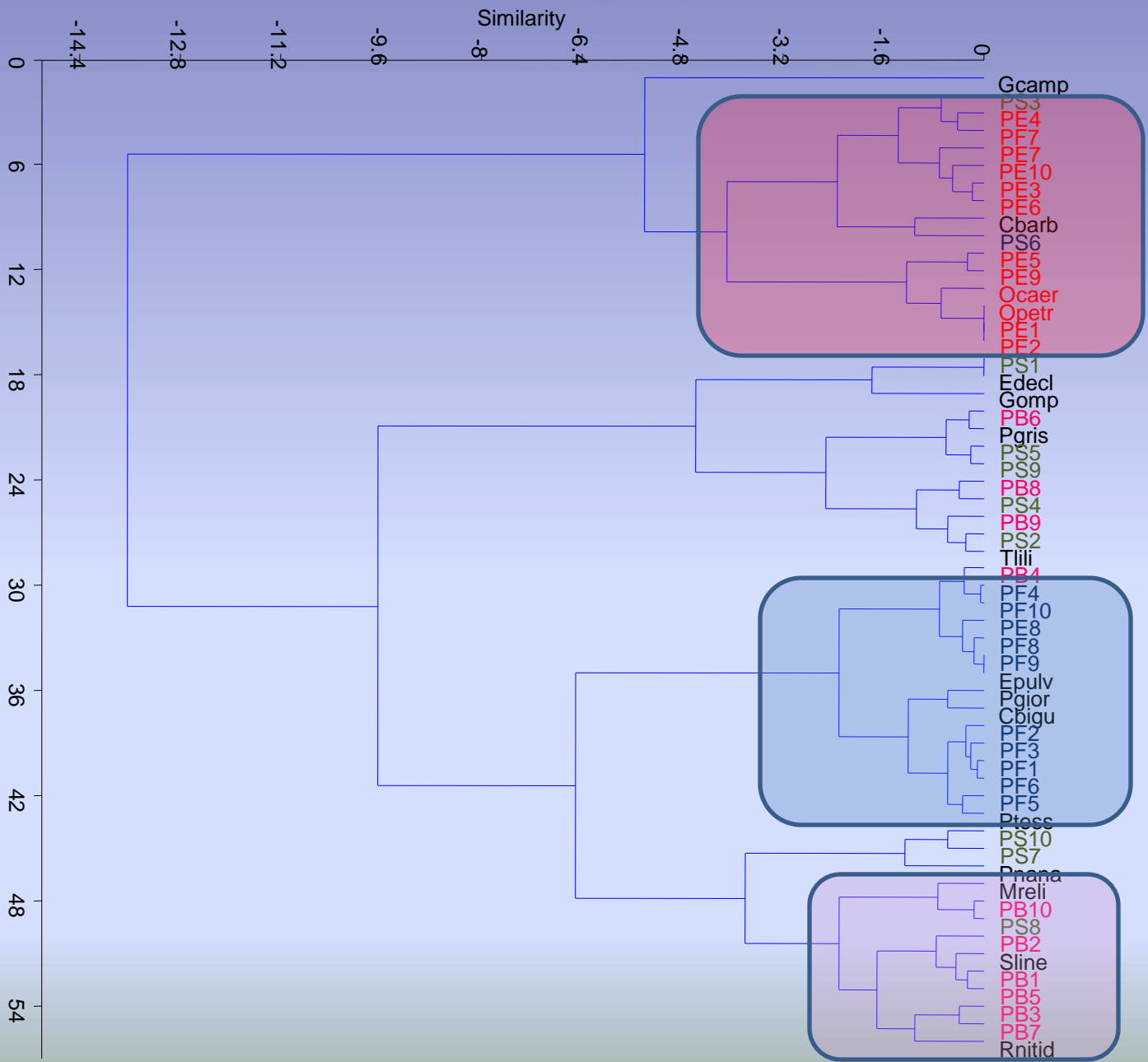
Méthode d'agglomération

- Paired group =UPGMA
- Ward's method

## Paired group =UPGMA



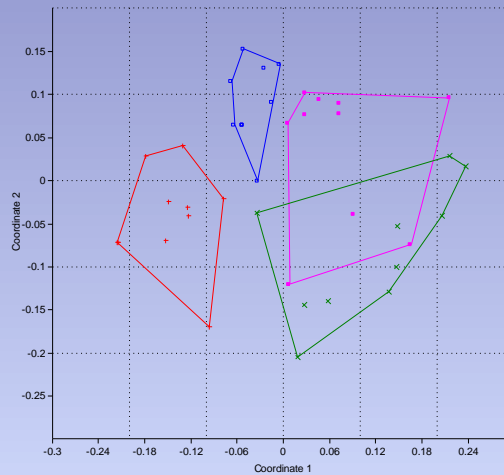
# Ward's method : minimisation de la variance intra-groupe



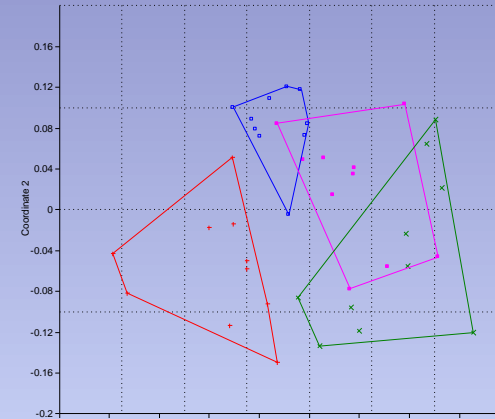
# Non-metric Multidimensional scaling

- Les calculs ne se font pas sur les valeurs absolues mais sur les rangs
- Augmente le % de représentation sur les 2 premiers axes
- Ne permet pas de représenter à la fois les facteurs lignes et colonnes
- Meilleure mesure de similarité ou distance ?  
Regarder le diagramme de Shepard sur les rangs observés et préservés dans la projection:  $y = x$

Rho de Spearman

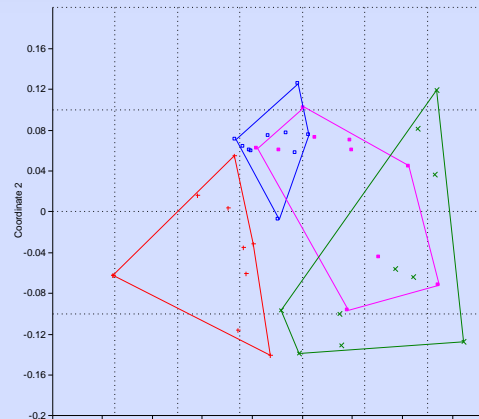


Bray-Curtis



$$Bray - Curtis_{jk} = \frac{\sum_{i=1}^s |x_{ij} - x_{ik}|}{\sum_{i=1}^s (x_{ij} + x_{ik})}$$

Morista

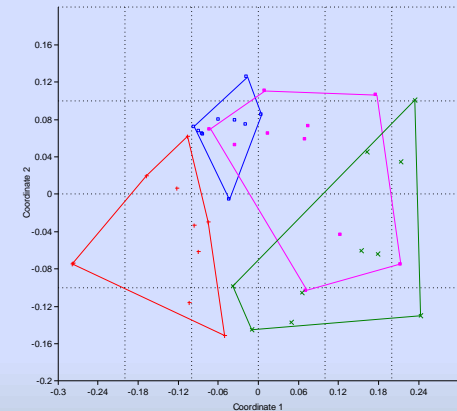


$$\lambda_1 = \frac{\sum_{i=1}^s (x_{ij}(x_{ij} - 1))}{\sum_{i=1}^s x_{ij} (\sum_{i=1}^s x_{ij} - 1)}$$

$$\lambda_2 = \frac{\sum_{i=1}^s (x_{ik}(x_{ik} - 1))}{\sum_{i=1}^s x_{ik} (\sum_{i=1}^s x_{ik} - 1)}$$

$$Morista_{jk} = \frac{2 \sum_{i=1}^s (x_{ij} x_{ik})}{(\lambda_1 + \lambda_2) \sum_{i=1}^s x_{ij} \sum_{i=1}^s x_{ik}}$$

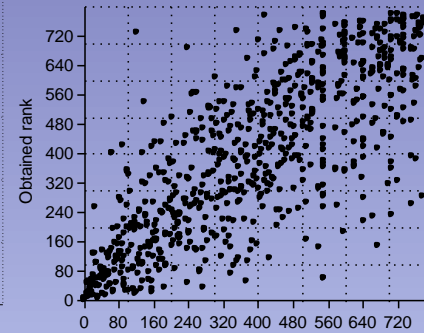
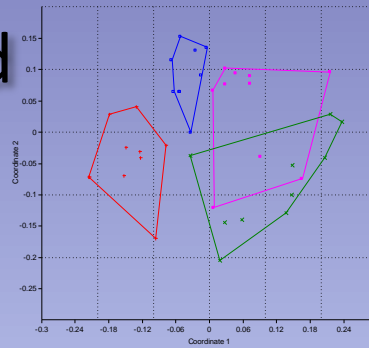
Chord



$$Chord_{jk} = \sqrt{2 - 2 \frac{\sum_{i=1}^s (x_{ij} x_{ik})}{\sqrt{\sum_{i=1}^s x_{ij}^2 \sum_{i=1}^s x_{ik}^2}}}$$

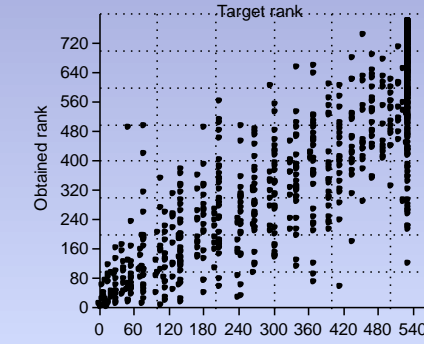
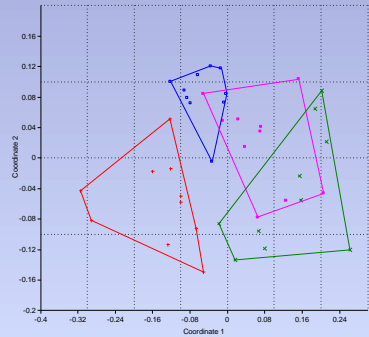
# Diagrammes de Shepard

Rho de Spearman



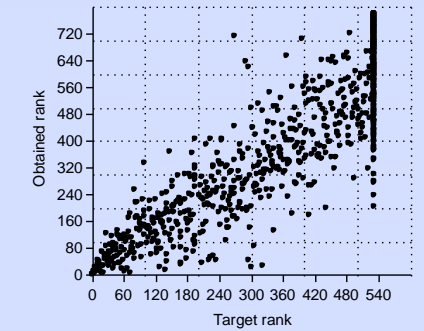
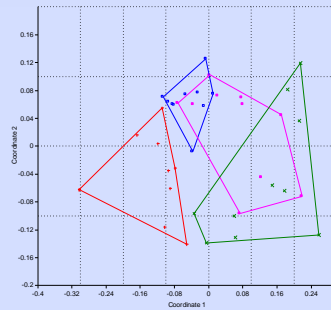
Stress = 0.3004

Bray-Curtis



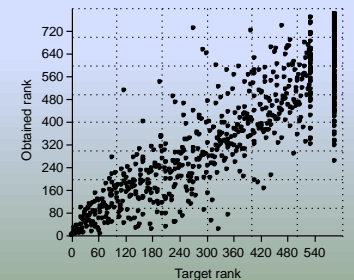
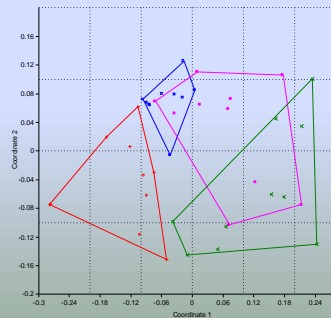
Stress = 0.2801

Morisita



Stress = 0.2528

Chord



Stress = 0.2357

# ACP : avec variables quantitatives

site\$	Age	haut	Rec	Gram	nbesp_V	div_vgt	div_ins	Dens_ins	nbesp_Ins
Ma61	0.50	40	0.50	11.25	15	3.67	2.56	53.00	9
Ma62	0.50	40	0.90	21.72	11	2.90	1.16	28.00	3
Te62	1.00	50	0.80	60.35	12	2.28	2.22	61.60	8
Ma71	1.50	30	0.50	16.00	10	2.81	1.14	26.00	3
Ma73	1.50	30	0.50	16.00	10	2.81	2.55	36.00	7
Ma74	1.50	55	0.50	13.00	9	2.25	1.49	11.00	3
Pa51	1.50	23	0.60	17.14	8	2.95	2.12	53.30	5
Pa52	1.50	20	0.80	34.72	11	3.17	1.87	25.00	4
Pa53	1.50	15	0.80	30.67	12	3.32	1.84	20.00	4
Pb61	1.50	25	0.70	30.00	7	2.35	1.90	120.00	6
Pb63	1.50	40	0.90	78.26	7	1.49	2.00	56.00	5
Te71	2.00	25	1.00	37.00	13	3.41	1.37	39.00	3
Te72	2.00	60	1.00	47.00	17	3.39	1.74	25.00	4
Pa61	2.50	5	0.60	15.00	11	2.27	2.16	91.70	7

Ann. soc. entomol. Fr. (n.s.), 2007, 43 (1) : 69-79

ARTICLE

## Primary succession of Orthoptera on mine tailings: role of vegetation

FRÉDÉRIQUE PICAUD <sup>(1)</sup> & DANIEL PIERRE PETIT <sup>(2)\*</sup>

<sup>(1)</sup> Société Entomologique du Limousin, av. Baudin, 87000 Limoges, France. Email: picaud@educagri.fr

<sup>(2)</sup> UMR INRA1061, Faculté des Sciences et Techniques, 123, av. A. Thomas, 87060, Limoges Cedex France

\* Corresponding author

Mg62	8.00	30	1.00	89.01	7	1.02	1.40	97.00	5
Mg63	8.00	45	1.00	94.44	6	1.38	0.76	25.00	2
Mg71	9.00	20	0.90	40.00	16	3.00	1.19	44.00	3
Mg72	9.00	15	0.95	60.00	16	2.44	1.46	78.00	3

Eigenvalue	% variance
2.84324	31.592
2.49855	27.762
1.42657	15.851
1.00952	11.217
0.611219	6.7913
0.32128	3.5698
0.145425	1.6158
0.0939108	1.0435
0.0502733	0.55859

Matrix:  
 Var-covar  
 Correlation

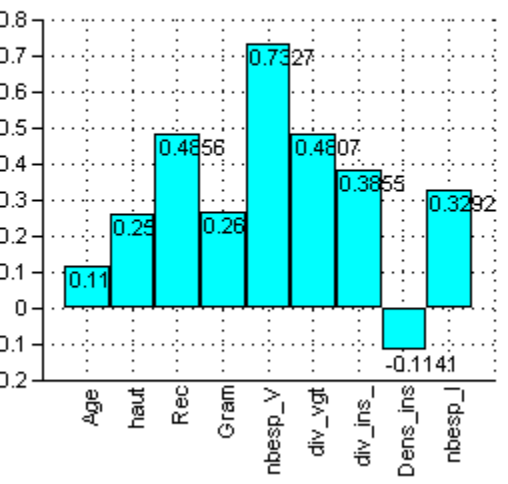
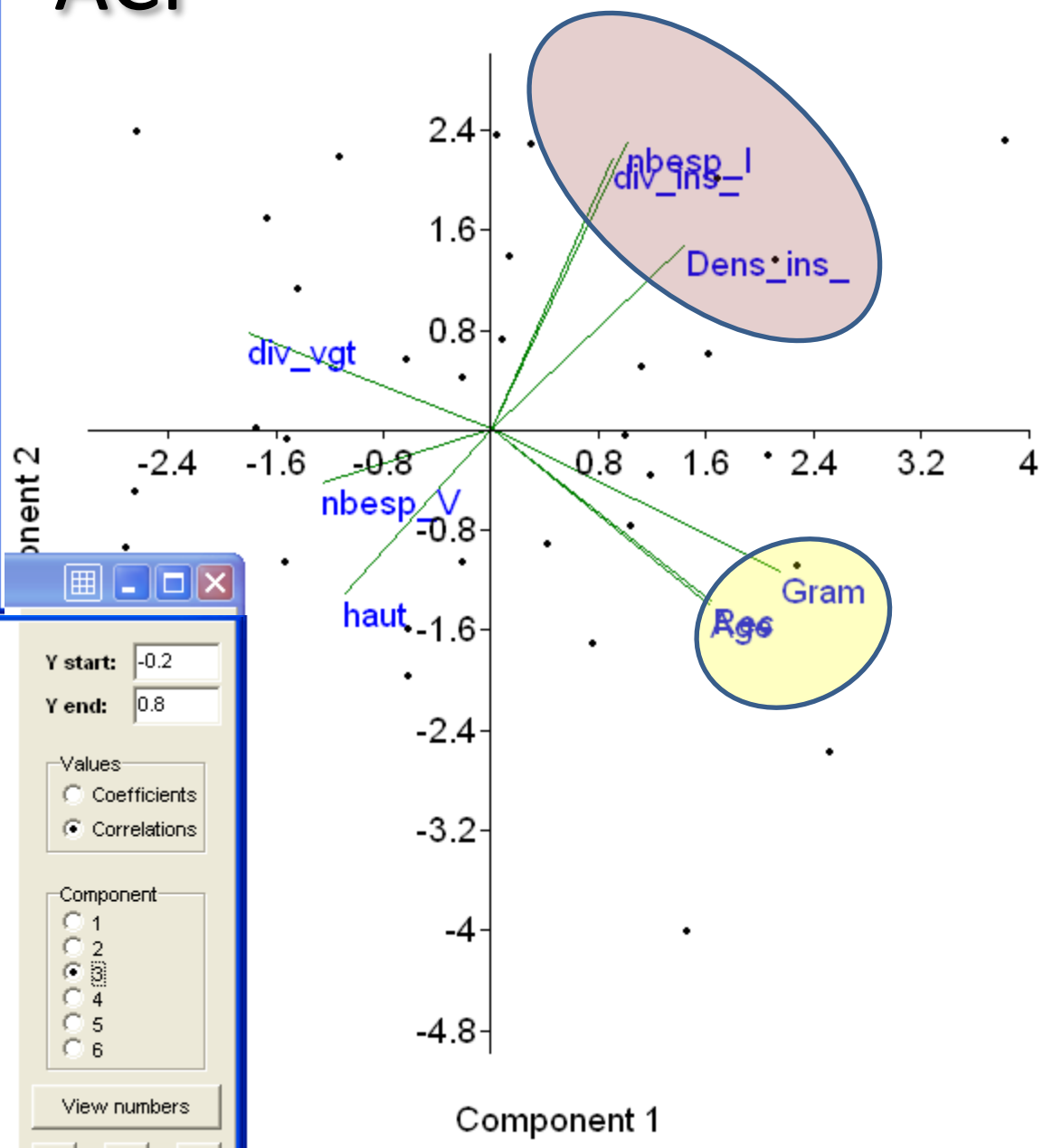
SVD  
 Shape PCA

Boot II:

Jolliffe cut-off: 0.7

View scatter  
 View loadings  
 Scree plot  
 Shape deform (2D)

# ACP



Y start:   
 Y end:

Values:  
 Coefficients  
 Correlations

Component:  
 1  
 2  
 3  
 4  
 5  
 6

View numbers

X start:  
 X end:  
 Y start:  
 Y end:

Row  
 Min.  
 95%  
 Con  
 Bipl  
 Eige

Axes:  
 1+2  
 1+3  
 2+3

View



# Analyse de corrélation

P-values, probabilité associée, risque d'erreur

Correlation \ p(uncorr)

	Age	haut	Rec	Gram	nbesp_V	div_vgt	div_ins_	Dens_ins_	nbesp_I
Age		0.22089	0.00042347	0.00017233	0.98202	0.025365	0.29571	0.40781	0.26563
haut	-0.21553		0.86766	0.85267	0.039607	0.86973	0.07408	0.0030928	0.086306
Rec	0.57078	0.029681		4.0339E-09	0.71302	0.12217	0.93882	0.97217	0.86685
Gram	0.60073	0.033077	0.81615		0.25441	0.00044193	0.5286	0.4713	0.57225
nbesp_V	0.004015	0.35462	0.065457	-0.20097		0.0001241	0.55246	0.22905	0.5521
div_vgt	-0.38301	0.029213	-0.27024	-0.56928	0.61096		0.5904	0.13318	0.84608
div_ins_	-0.18469	-0.31031	0.013674	0.11192	-0.10555	0.095676		0.047121	7.5012E-14
Dens_ins_	0.14667	-0.49242	-0.0062153	0.12781	-0.21185	-0.26282	0.34289		0.00081216
nbesp_I	-0.19639	-0.29857	-0.029864	0.10036	-0.10564	0.034576	0.91099	0.54714	

Correlation statistic

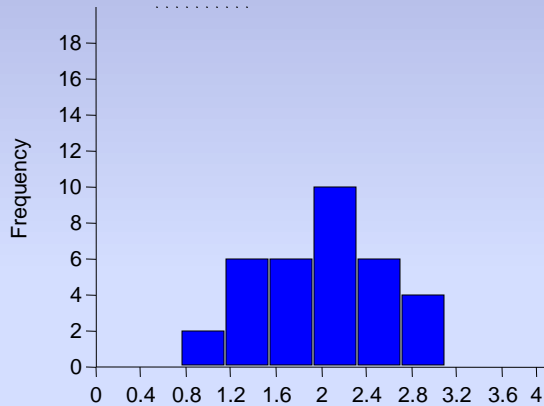
- Linear correlation coefficient
- Spearman's D
- Spearman's rs
- Kendall's tau

r de Pearson, coefficient de corrélation linéaire

Conforte les observations de l'ACP avec indication du risque d'erreur

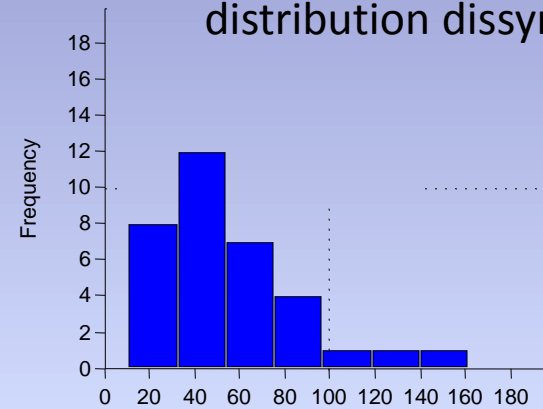
# Distribution d'une variable quantitative

Appréciation visuelle de la distribution par histogramme



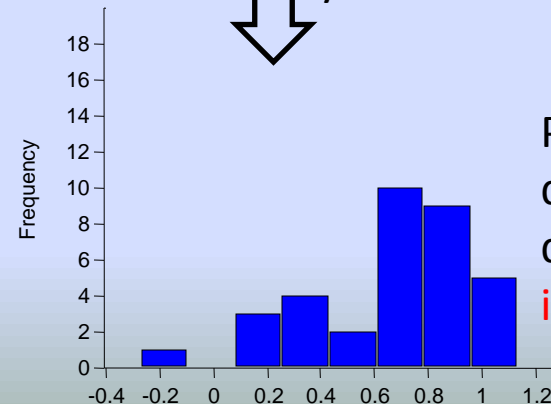
Variable diversité (H') des insectes :  
distribution symétrique

Variable densité des insectes :  
distribution dissymétrique



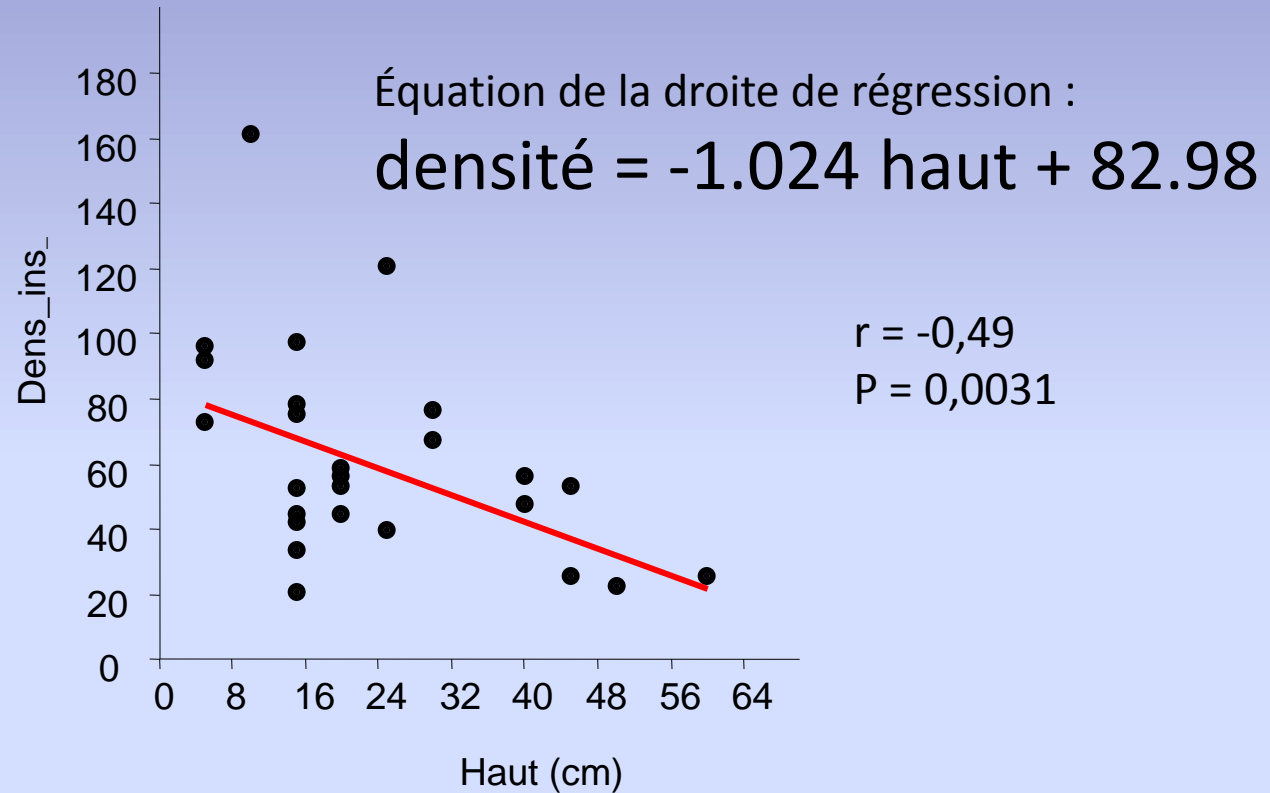
Transformation logarithmique

$$y = \ln x$$



Pas  
d'amélioration  
de la symétrie  
**inutile**

# Graphe de corrélation



# Régression multiple

## Hétérocarpie des Astéraceae du Maroc

- 45 localités
- Type biologique
- Type de dissémination :  
hétérocarpie ou homocarpie



*Hyoseris radiata*



*Leontodon taraxacoides subsp. longirostris*



# Bilan des prospections

	Cardueae	Lactuceae
nombre de taxa	78	82
nombre d'espèces hétérocarpes	3	21
nombre hétéroc thérophytes	3	14
nombre hétéroc hémicryptophytes	0	7

**Problème** : forte liaison entre hétérocarpie et type biologique thérophyte  
Nécessité d'enlever l'effet thérophyte pour tester l'influence climatique

# Comment extraire l'information ?

	m°C	Préc	Théro	Hétéro Théro	Hétéro NonThéro	Hétéroc	Total Card-lact
Essaouira	9.6	234.5	8	3	1	4	5
Kenitra	6.9	604.7	12	5	2	7	14
Sk-el-arba	6.3	641.2	8	3	1	4	6
Casablanca	8.6	481.2	15	4	1	5	12
Tanger	9.9	892.4	19	6	3	9	10
Arbaoua	6.7	846.3	9	3	1	4	4
Tiznite	5.6	154	7	1	1	2	6
Berrechid	4.1	368	14	4	1	5	13
Taroudant	5.1	212	4	1	1	2	1
Ouezzane	5.2	948.1	10	4	1	5	9

La variable à expliquer y (hétérocarpie) dépend d'une combinaison linéaire de plusieurs variables quantitatives x1, x2, x3, ...

$$y = a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + \dots + b$$

# Régressions multiples (SYSTAT 12.0)

Variable : nombre d'espèces hétérocarpes et thérophytes

	Coefficient	p-value
Constant	-0.24187	0.33657
THERO	0.23196	0.00000
m°C	0.08666	0.02683
PREC mm	0.00096	0.03671

Nb espèces hétérocarpes = 0.23196 THERO + 0.08666 m°C + 0.00096 PREC - 0.24187

Variable : nombre d'espèces hétérocarpes et non-thérophytes

	Coefficient	p-value
Constant	-0.53456	0.04823
NON-THERO	0.08387	0.00202
m°C	0.09376	0.00182
PREC mm	0.00138	0.00123

Nb espèces hétérocarpes = 0.08387 NON-THERO + 0.09376 m°C + 0.00138 PREC - 0.53456