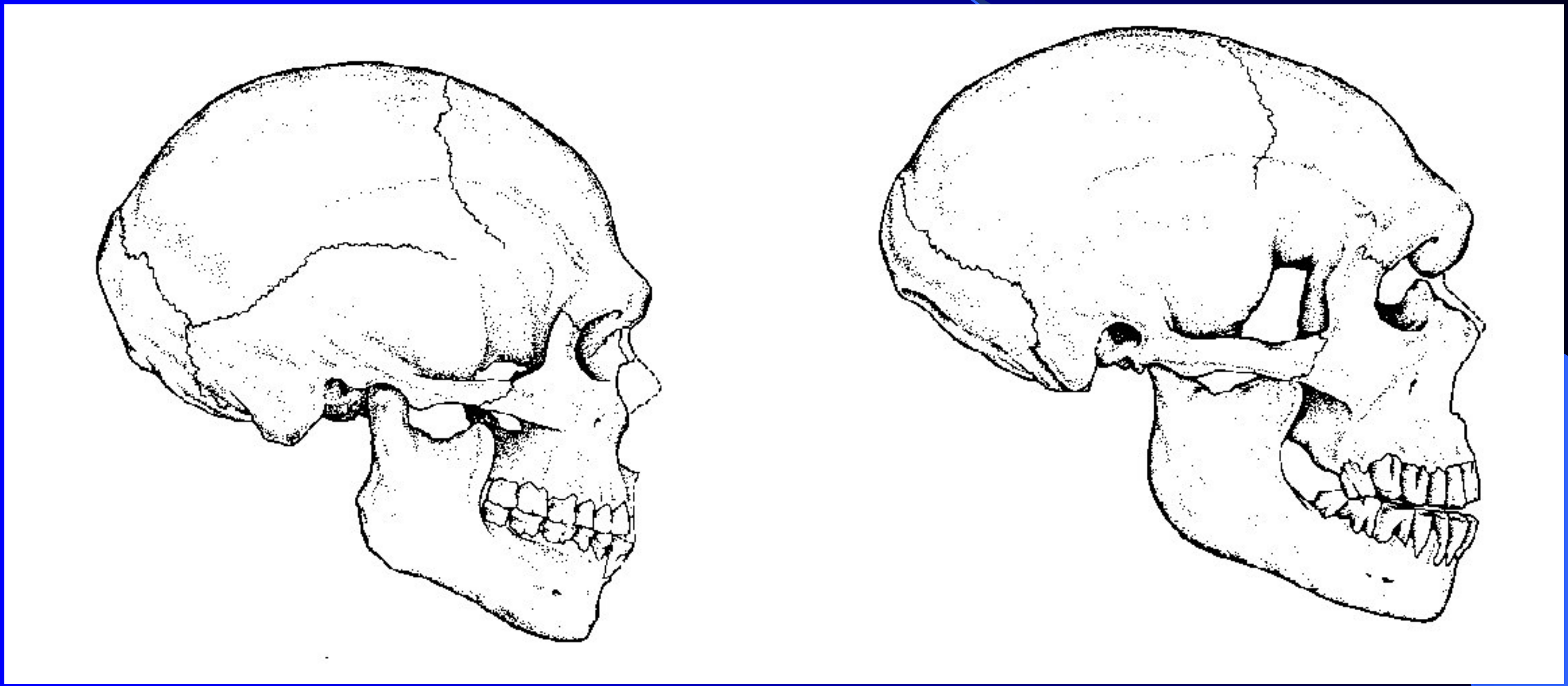
Analyses Multivariées

Analyses Multivariées

Statistiques Multidimensionnelles

Fréquentes en anatomie



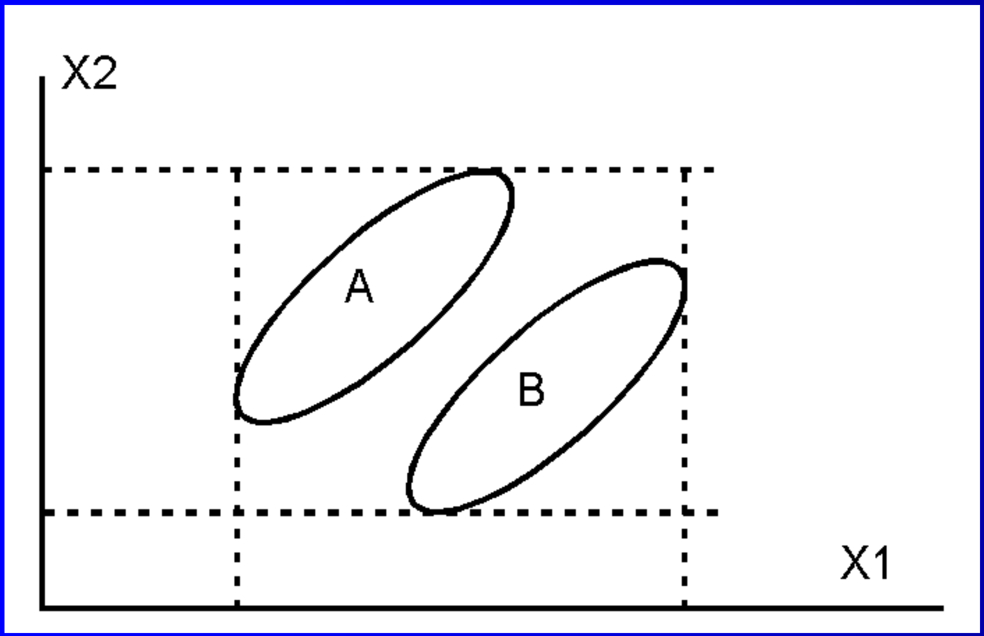
A cause de la complexité des structures

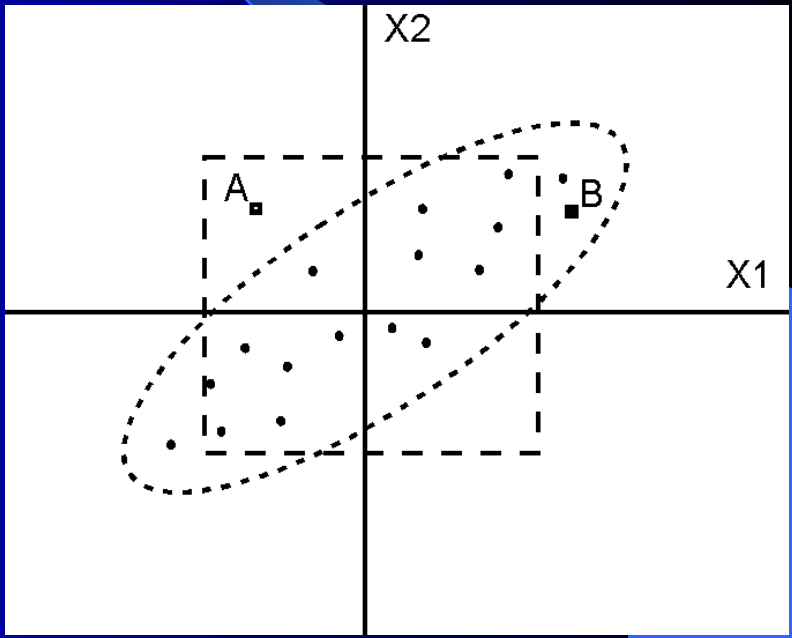
Le multivariable est différent de la somme d'univariables

Plus on utilise de variables:

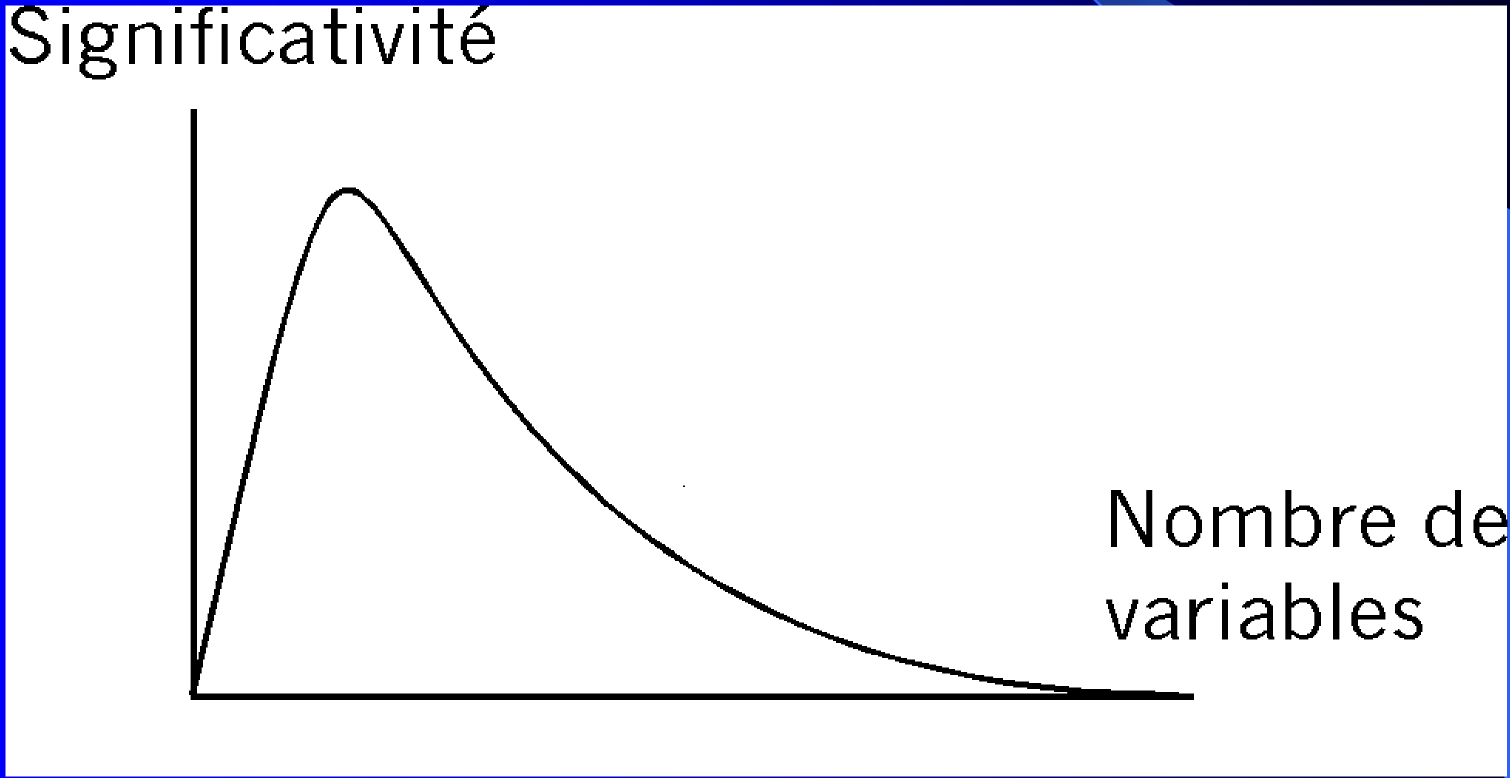
- Plus la covariance augmente
- Plus la puissance des tests diminue

La Covariance





La Significativité

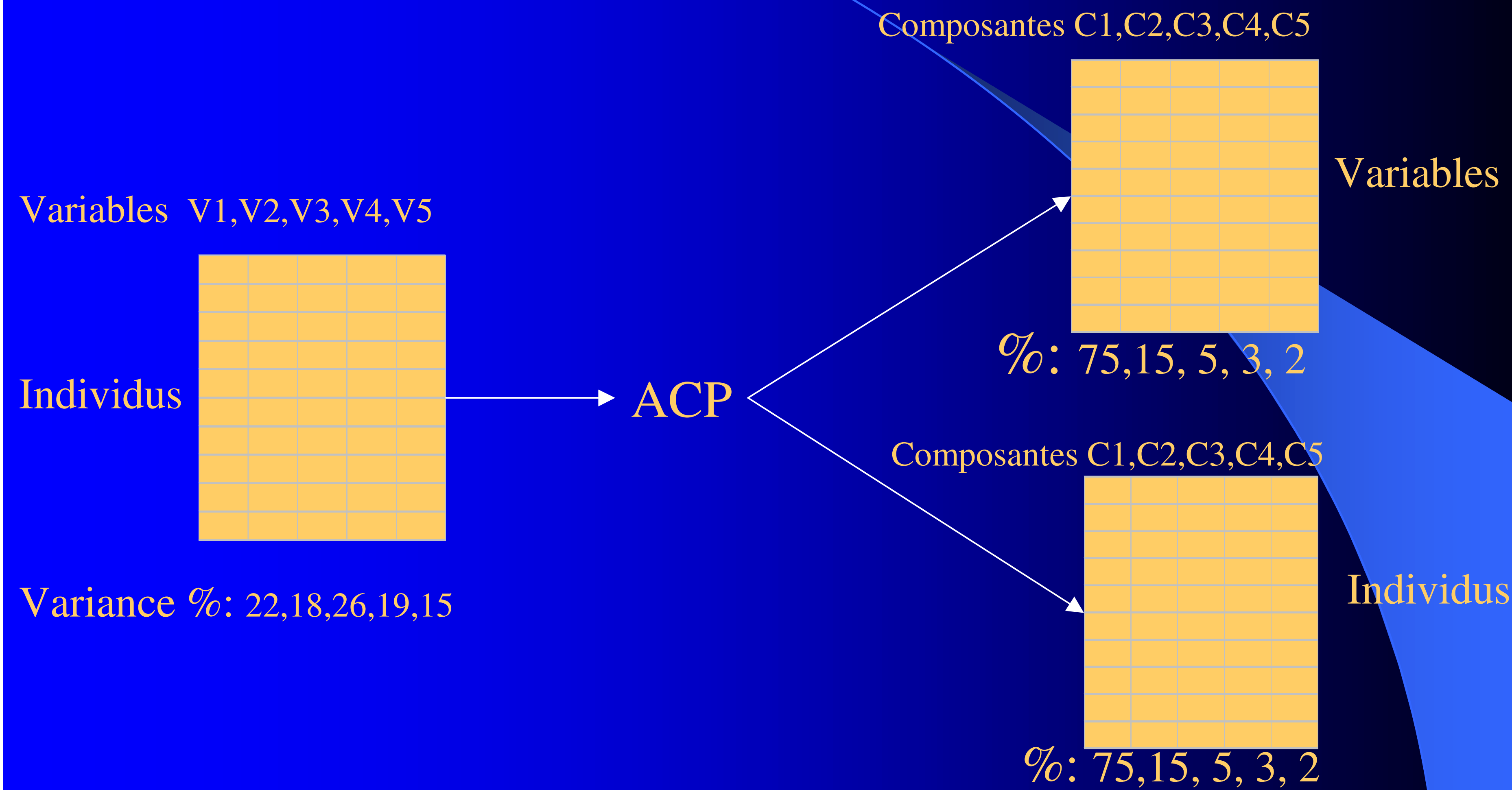


La solution

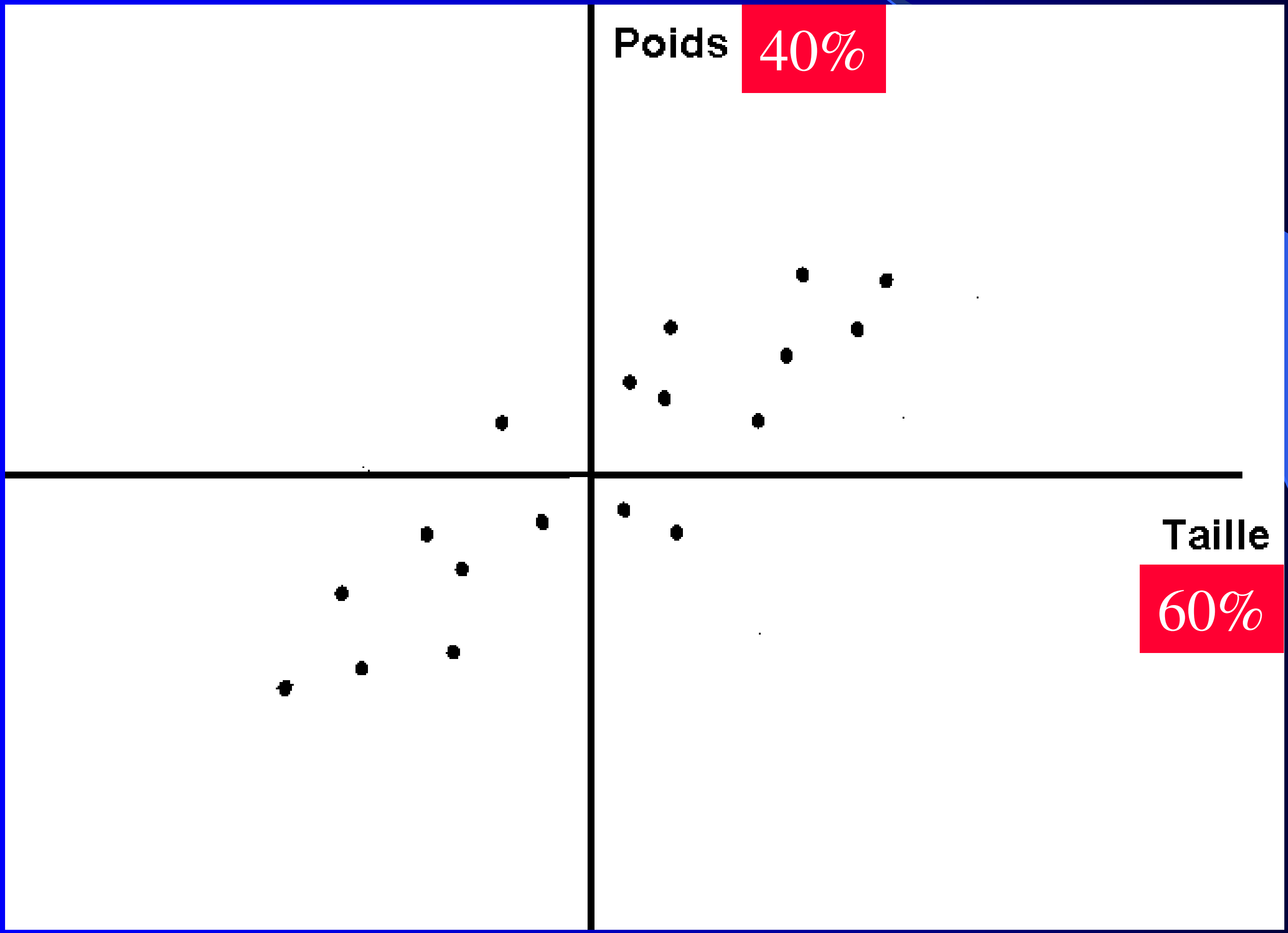
Calculer des variables composites qui synthétisent le maximum d'information avec peu de variables

Analyse en Composantes Principales (ACP)
Décomposition en valeurs singulières
Analyse discriminante
Analyse en coordonnées Principales
Analyse Factorielle des correspondances
.....

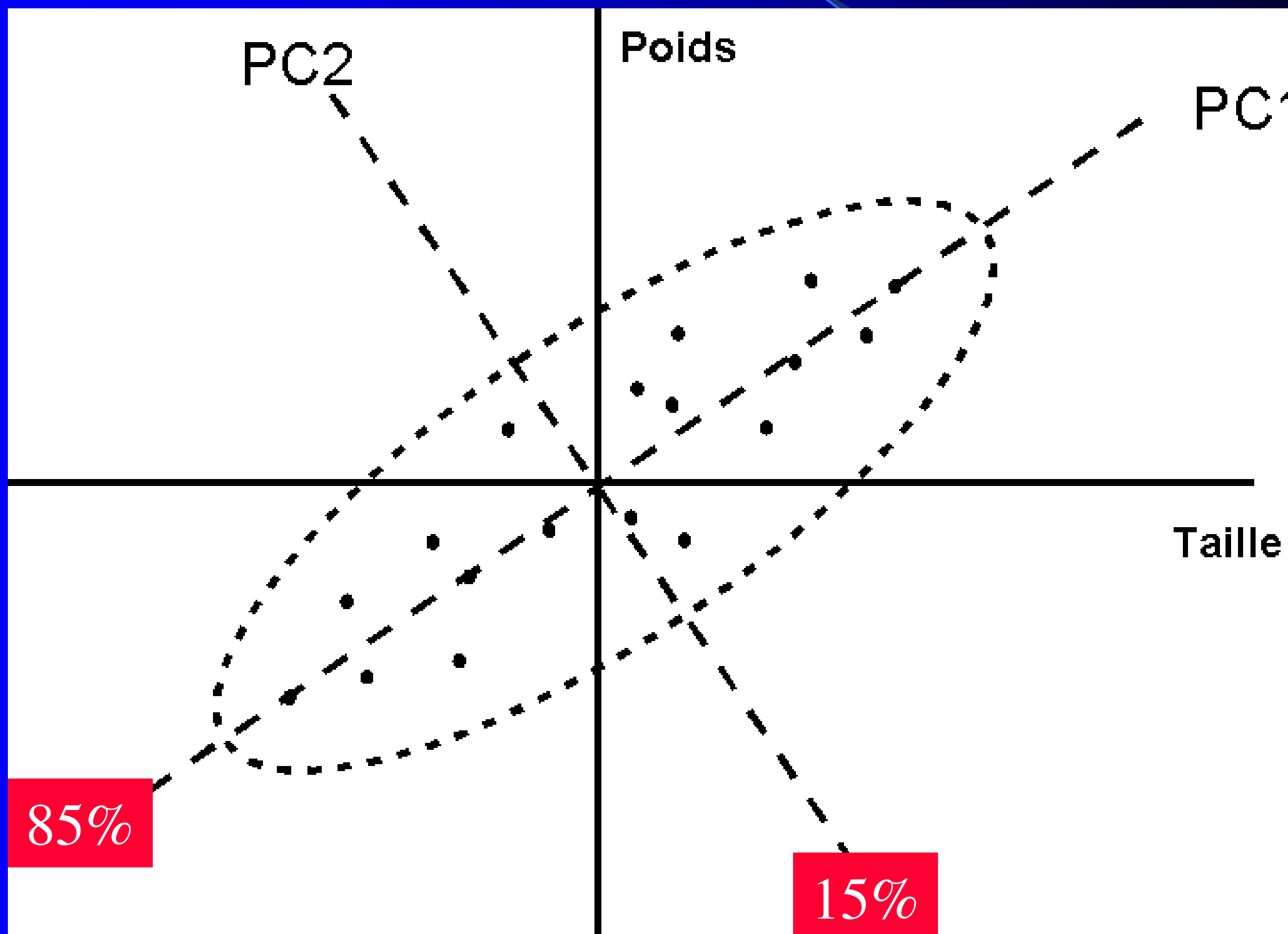
Transformation des données



Exemple de la taille et du poids



Résultats de l'ACP



Propriétés des composantes principales

1. Classées par ordre de variance décroissante

2. Indépendantes entre elles (orthogonalité)

Des combinaisons linéaires

$$PC1 = a.Taille + b.Poids = 85\%$$

$$PC2 = c.Taille - d.Poids = 15\%$$

Réification des composantes

PC1 = Robustesse = a.Taille + b.Poids

PC2 = Sveltesse = c.Taille − d.Poids

Le rapport Signal / Bruit

Premières composantes

Variation structurée

Variabilités individuelles

Composantes ultérieures

Deux espaces complémentaires

1. L'espace des variables pour analyser les relations des variables entre elles.

2. L'espace des individus pour étudier les regroupements d'individus entre eux.

Espace des variables

Correlations	X1	X2	X3	X4	X5	X6
X1 (longueur crâne)	1	0.584	0.615	0.601	0.57	0.6
X2 (largeur crâne)		1	0.576	0.53	0.526	0.555
X3 (humerus)			1	0.94	0.875	0.878
X4 (cubitus)				1	0.877	0.886
X5 (femur)					1	0.924
X6 (tibia)						1

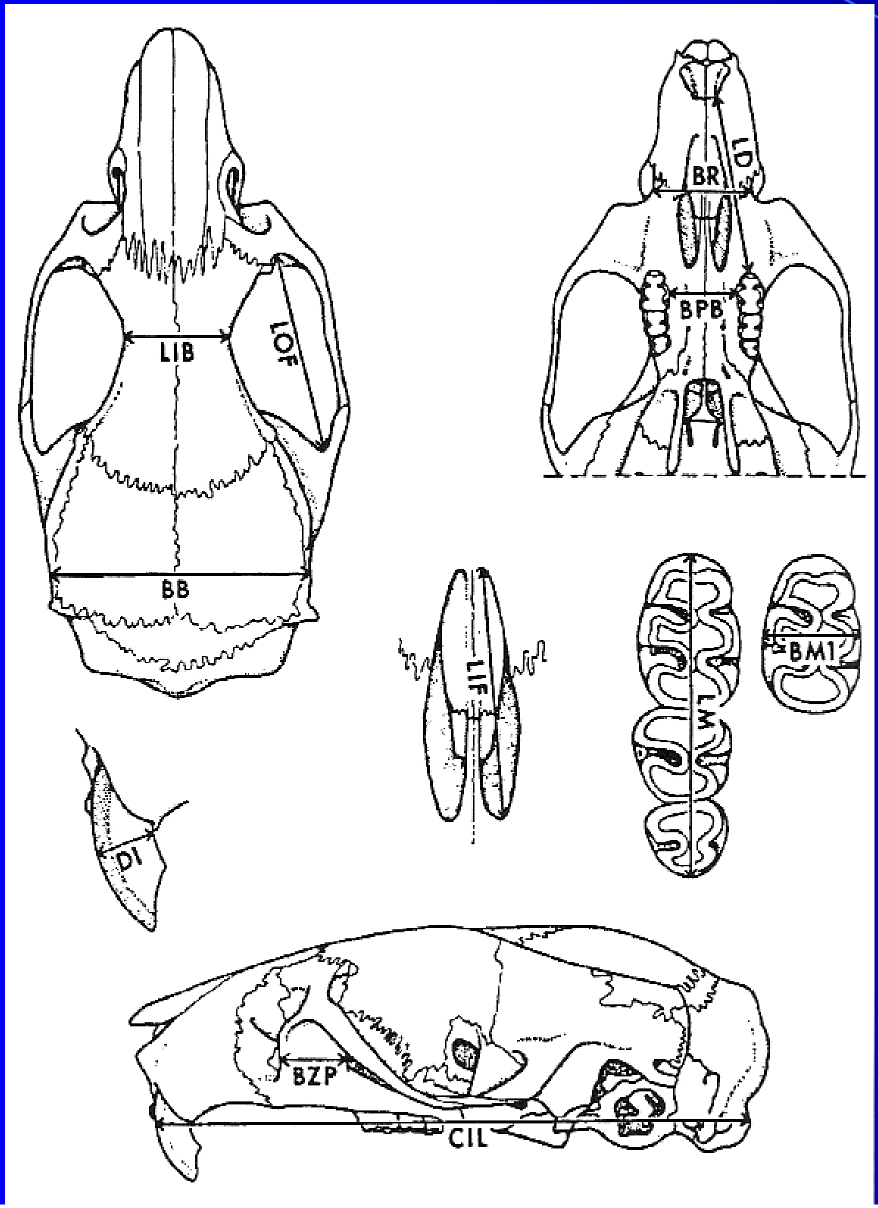
Analyse des données « White Leghorn Fowl » (S. Wight 1954)

276 femelles X1=Longueur crâne, X2=largeur, X3= longueur humérus, X4= longueur cubitus, X5=longueur femur, X6=longueur tibia

ACP sur matrice de corrélation

Poids des variables	PC1	PC2	PC3	PC4	PC5	PC6
Longueur crâne (X1)	0.35	0.53	0.76	-0.04	0.02	0.01
Largeur crâne (X2)	0.33	0.69	-0.64	-0.01	0.01	0.03
Humerus (X3)	0.44	-0.19	-0.05	0.53	0.18	0.67
Cubitus (X4)	0.44	-0.25	0.02	0.48	-0.15	-0.71
Femur (X5)	0.44	-0.28	-0.06	-0.49	0.65	-0.13
Tibia (X6)	0.44	-0.22	-0.05	-0.48	-0.69	0.17
Variance	4.58	0.71	0.41	0.17	0.08	0.05
Pourcentages	76%	12%	7%	3%	1%	1%

Espace des individus



Etude du rongeur tropical ZYGODONTOMYS à partir d'échantillons issus de 15 localités

Leslie F. MARCUS †
American Museum of Natural History

La matrice de covariance

Variance-covariance matrix × 100

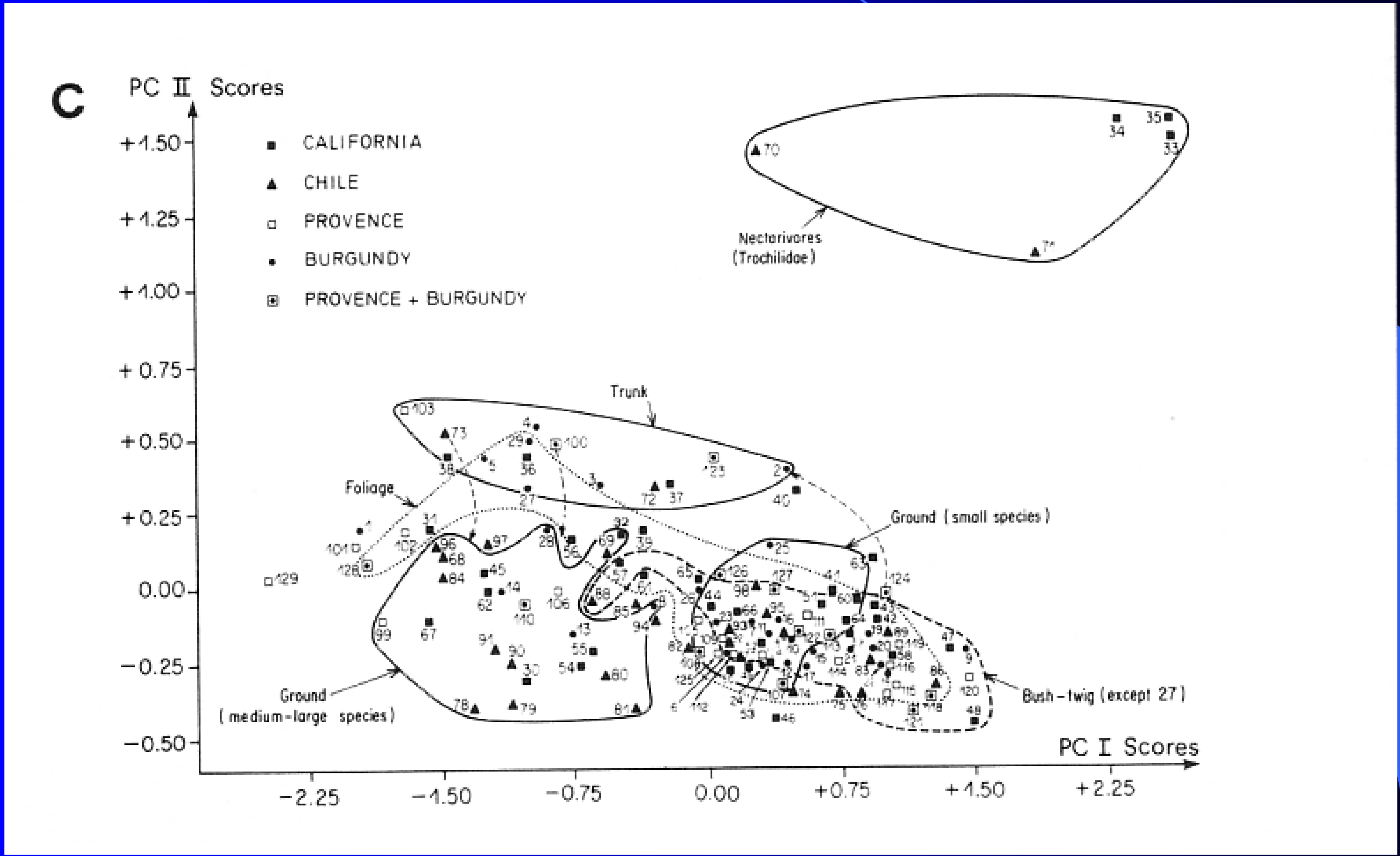
	CIL	LD	LM	BM1	LIF	BR	BPB	BZP	LIB	BB	DI	LOF
CIL	.3711	.4696	.0575	.0255	.3636	.4254	.4928	.4695	.2241	.1185	.4994	.3447
LD	.4696	.6464	.0292	.0217	.4979	.5381	.6486	.5877	.2795	.1599	.6016	.4360
LM	.0575	.0292	.1575	.0831	.0174	.0951	.0445	.1031	.0826	-.004	.1066	.0444
BM1	.0255	.0217	.0831	.1762	.0196	.0992	.0280	.0504	.0491	-.001	.0618	.0167
LIF	.3636	.4979	.0174	.0196	.5453	.4044	.4974	.4836	.2195	.1449	.4663	.3486
BR	.4254	.5381	.0951	.0992	.4044	.7536	.6705	.5822	.3343	.1396	.6263	.3995
BPB	.4928	.6486	.0445	.0280	.4974	.6705	.9362	.7023	.3226	.1832	.7267	.4462
BZP	.4695	.5877	.1031	.0504	.4836	.5822	.7023	.9718	.2893	.1298	.6402	.4282
LIB	.2241	.2795	.0826	.0491	.2195	.3343	.3226	.2893	.3093	.0800	.2869	.1952
BB	.1185	.1599	-.004	-.001	.1449	.1396	.1832	.1298	.0800	.1021	.1601	.1239
DI	.4994	.6016	.1066	.0618	.4663	.6263	.7267	.6402	.2869	.1601	.9293	.4712
LOF	.3447	.4360	.0444	.0167	.3486	.3995	.4462	.4282	.1952	.1239	.4712	.3686

L'ACP

Table 3. Principal component analysis of logarithms of Dividive data.

PC	Eivenvalue	Cumulative%
1	0.456573	72.85
2	0.036284	78.64
3	0.033649	84.00
4	0.025552	88.08
5	0.023436	91.82
6	0.013521	93.98
7	0.012521	95.98
8	0.009079	97.42
9	0.006374	98.44
10	0.005768	99.36
11	0.003131	99.86
12	0.000833	100.00

Ajout des données complémentaires



Groupements en fonction de l'habitat

Matrice de covariance ou de corrélation?

$$\mathrm{cov} = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{n}$$

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})(x_i - \bar{x}) \sum (y_i - \bar{y})(y_i - \bar{y})}}$$

En Pratique avec Scilab

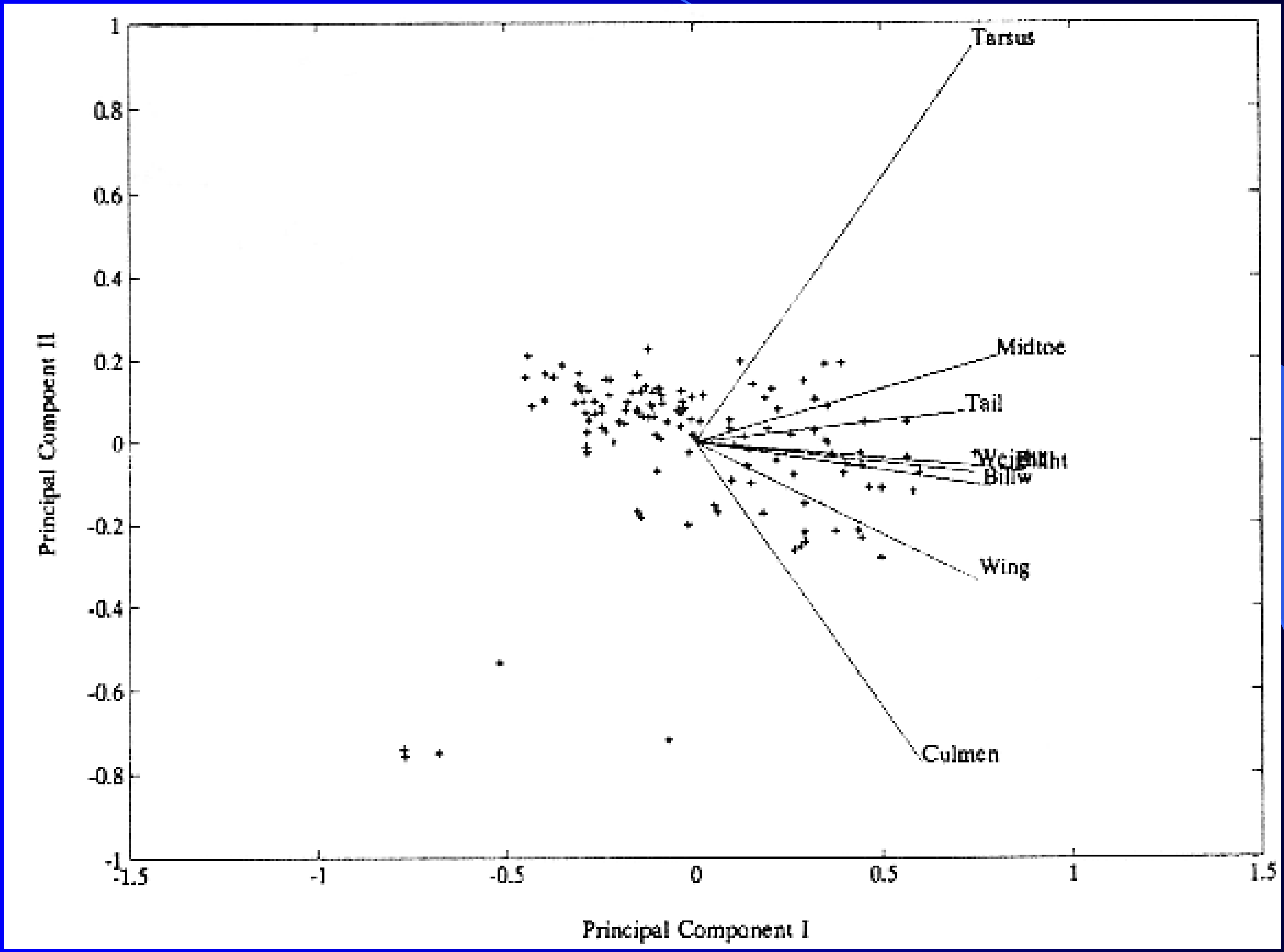
La décomposition en valeurs singulières:

$$[U\ S\ V] = svd(X)$$

U*S=Position des individus

V = coefficients des composantes

Le Biplot

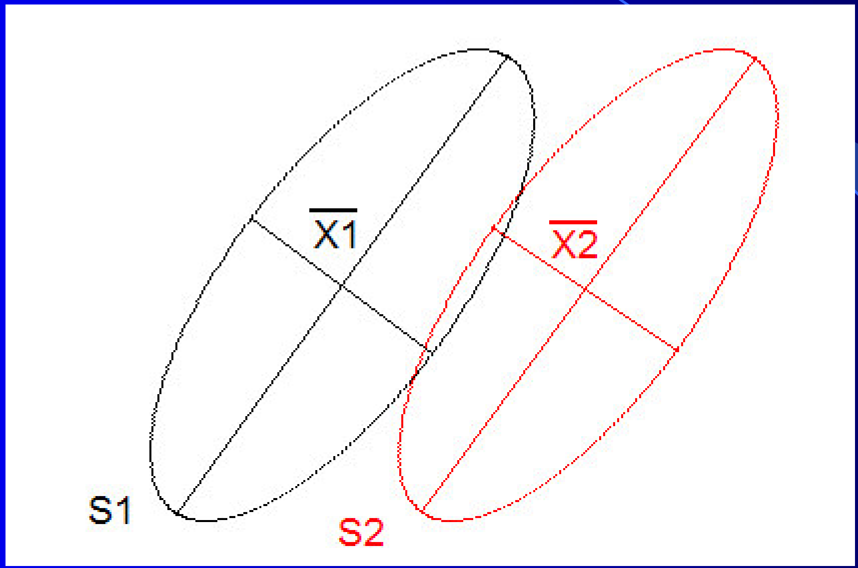


Données initiales: 8 mesures morphologiques enregistrées chez 126 espèces d'oiseaux méditerranéens

Les tests statistiques en analyse multivariable

- Le T^2
- Les tests de régression - discrimination
- Les tests de permutation

Le T² : un test t multivariable



$$T^2 = \frac{n_1 \cdot n_2}{n_1 + n_2} \cdot D^2$$

$$D^2 = (\overline{X}_1 - \overline{X}_2) \cdot inv(\overline{S}) \cdot (\overline{X}_1 - \overline{X}_2)$$

Distance de Mahalanobis

Du T² au F

$$F = \frac{N - g - p + 1}{(N - g) \cdot p} \cdot T^2$$

Avec N le nombre total d'observations dans les g groupes et p le nombre de variables

Calcul du F avec une fonction discriminante (2 groupes)

Utilisation d' EXCEL:

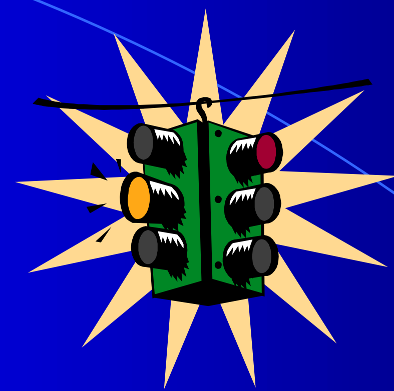
Calcul d'une régression multiple
avec codage binaire (0,1)
de la variable dépendante

L’analyse discriminante

Une méthode à utiliser
avec précautions

Association avec une variable extérieure

Testée par régression multiple
et donc un test F



L’analyse des données est différente des calculs de significativité