



Analyses Factorielles et Régressions Logistiques réalisées à partir de données récoltées dans le cadre de l'Observatoire National des Maladies du Bois de la Vigne



Frédéric BERTRAND (ULP Strasbourg)

Myriam MAUMY (ULP Strasbourg)

Nathalie KOBES (DRAF-SRPV Alsace)

Lionel FUSSLER (ULP Strasbourg)

Serge SAVARY (INRA Bordeaux)

Introduction

- Interdiction des traitements à l'arsénite de soude en 2001
➡ inquiétudes au sein de la profession viticole
- Mise en place de l'observatoire national des maladies du bois

Protocole et données

■ Protocole :

- Nombre minimum de 25 parcelles par vignoble
- 300 ceps/parcelle, repérés et marqués, répartis en 10 placettes de 30 ceps choisies aléatoirement
- Unique critère de choix de la parcelle : cépage

■ Données :

- 701 parcelles au total
- 11 régions
- 26 cépages

Protocole et données

Ce protocole a été proposé par le ministère chargé de l'agriculture et validé par le groupe technique national auquel appartiennent l'Institut Technique de la Vigne et du Vin et l'INRA. Il reprend les ordres de grandeur des nombres d'observations qui sont habituellement réalisées lors des expérimentations du même type.

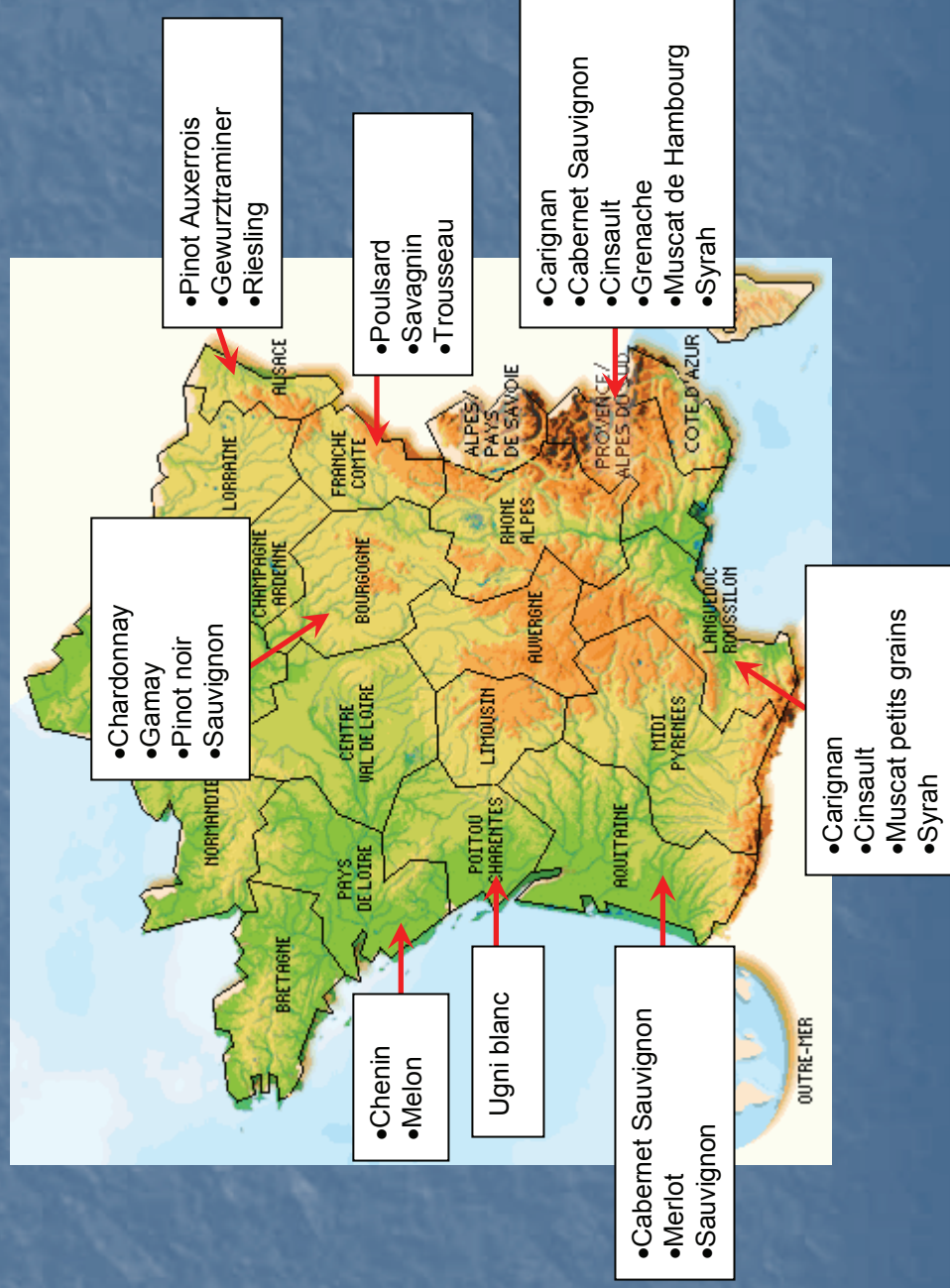
En complément, des informations sur les caractéristiques de la parcelle et les pratiques culturales ont été recueillies par enquête auprès des exploitants viticoles dont les parcelles avaient été choisies aléatoirement.



Questionnaire d'enquête

Fiche de parcelle pour l'Observatoire des Maladies du bois	
Vignoble de	Organisme(s) :
EXPLOITANT	
Nom Prénom : Adresse :	Raison sociale : Tél. : Fax : Mail :
PARCELLE	
Référence :	Cépage (clone) :
Commune :	Porte-greffe (clone) :
Lieu dit :	Année de plantation :
Type de taille :	Surface :
Type de sol :	Densité de plantation :
	Vigueur : forte, moyenne ou faible
TRAITEMENTS D'HIVER	
Arsénite en 1999 : oui ou non	Autre :
Arsénite en 2000 : oui ou non	Autre :
Arsénite en 2001 : oui ou non	Autre :
PROPHYLAXIE RÉALISÉE EN 2004	
Date de taille : Sarmants :	Indiquer au moins la quinzaine broyage ou enlèvement
Pratiques contre les maladies du bois :	Prétaillage : Nettoyage bois mort :
égourmandage, perçage des souches, traitement des plaies de taille, environnement de la parcelle, ...	
Date de début de la mise en œuvre de mesures prophylactiques	Indiquer l'année
AUTRES OBSERVATIONS	

Carte des régions et cépages

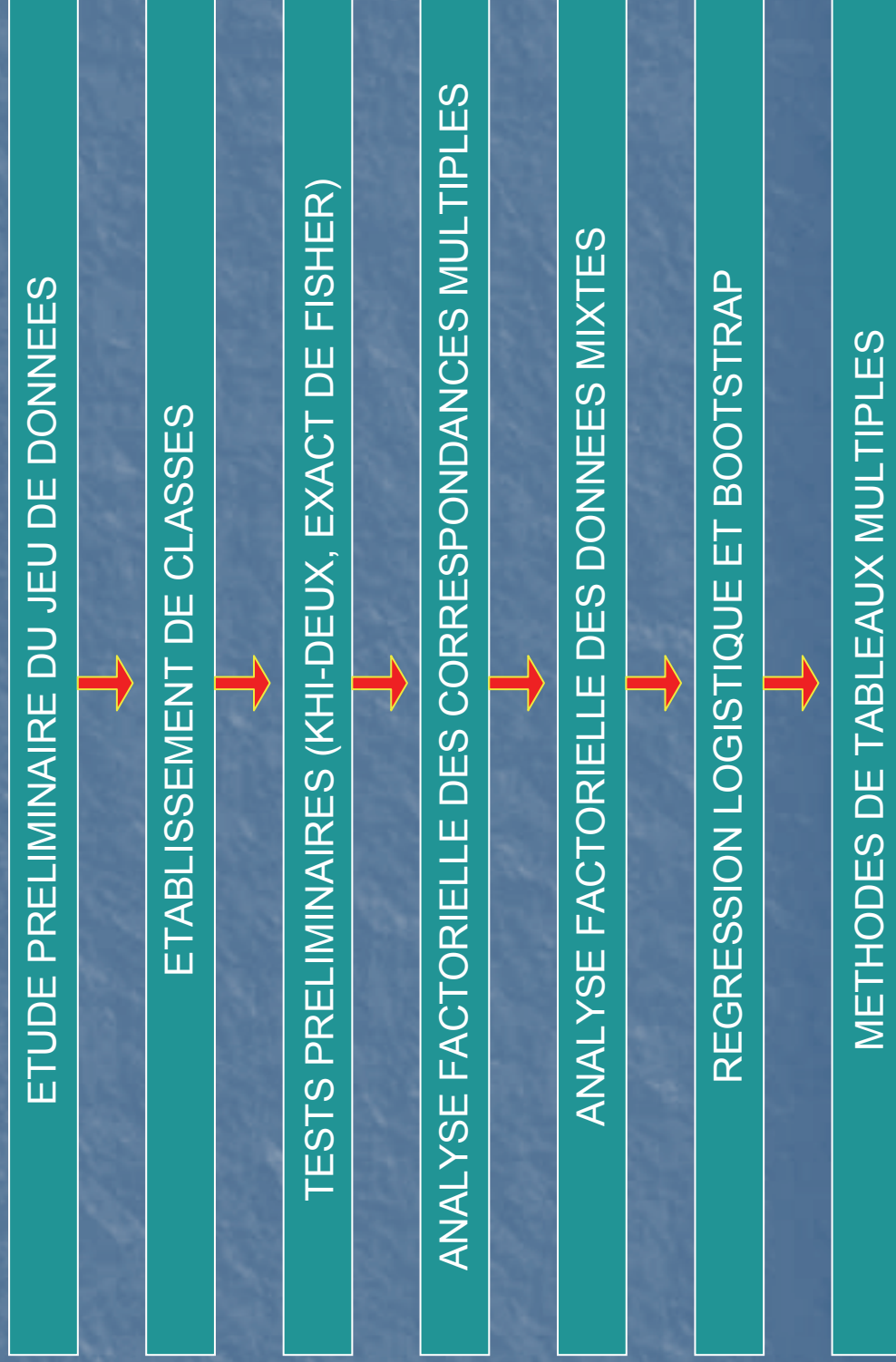




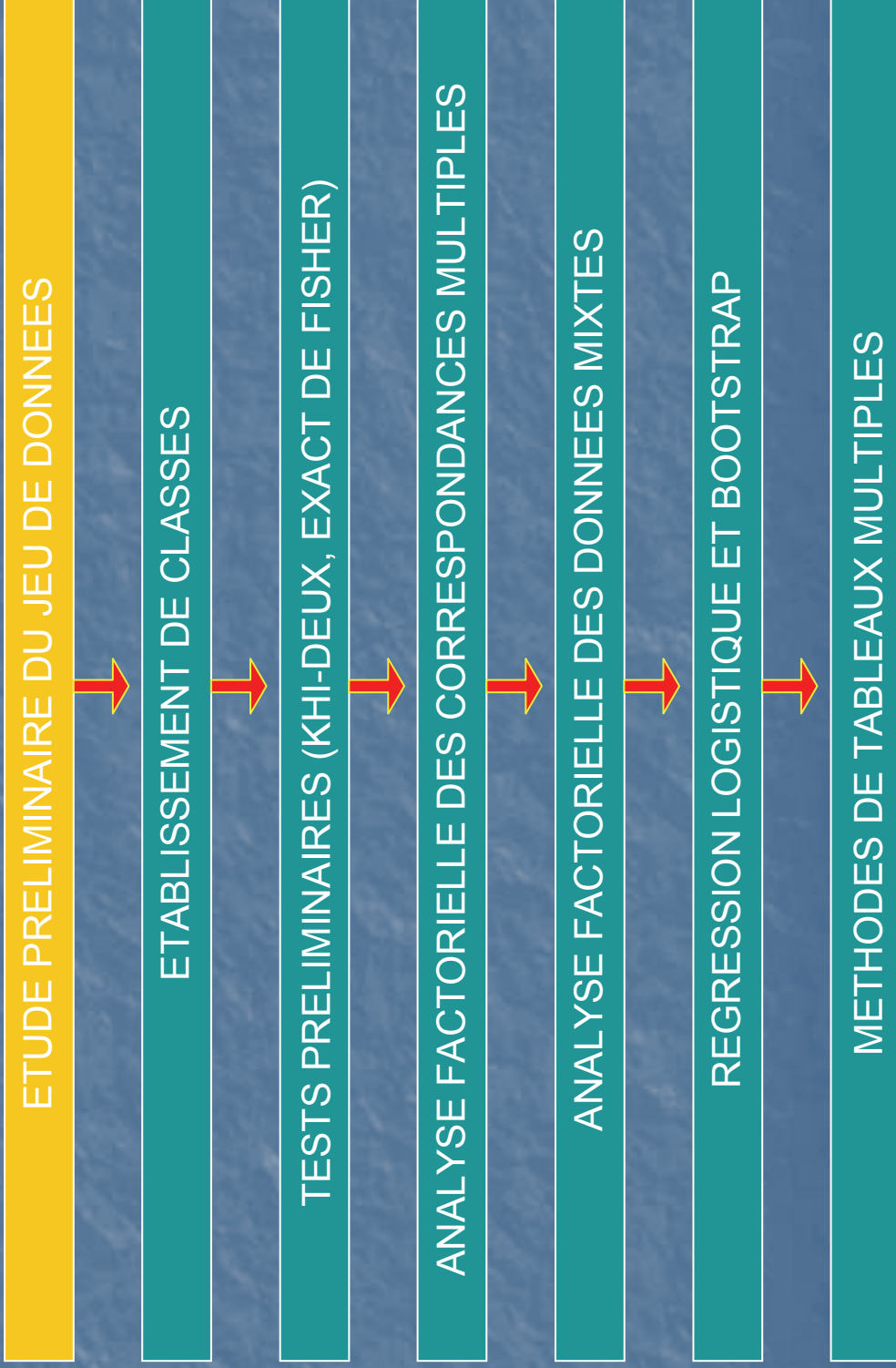
Jeu de données

Région	Cépage	Porte- greffe	Âge	Prétailla ge	Sarments	Enlèvement bois morts	Traitement arsénite de soude en 1999	Traitement arsénite de soude en 2000	Traitement arsénite de soude en 2001	eutypiose	esca/BDA
Alsace	Pinot Auxerrois	SO4	29	non	broyage	oui	non	non	non	3,17%	0,00%
Alsace	Pinot Auxerrois	3309C	33	non	broyage	oui	non	non	non	0,00%	1,54%
Alsace	Pinot Auxerrois	SO4	21	non	broyage	oui	non	non	non	0,00%	0,00%
Alsace	Gewurztraminer	3309C	17	oui	broyage	oui	non	non	non	0,67%	0,00%
Alsace	Gewurztraminer	3309C	21,5	oui	broyage	oui	non	non	non	0,67%	2,00%
Alsace	Gewurztraminer		18	non	broyage	oui	non	non	non	0,00%	0,00%
Alsace	Gewurztraminer	34EM	15	oui	broyage	oui	non	non	non	0,00%	0,67%
Alsace	Gewurztraminer		35	non	enlèvement	oui	non	non	non	0,00%	7,00%
Alsace	Gewurztraminer	SO4	41	oui	broyage	oui	non	non	non	0,00%	0,67%
Alsace	Gewurztraminer	3309C	23				non	non	non	1,00%	3,00%
Alsace	Gewurztraminer	161-49					oui	oui	non	0,00%	1,33%
Alsace	Riesling		25	oui	broyage	oui	non	non	non	0,00%	2,33%
Alsace	Riesling	3309C		oui	broyage	oui	non	non	non	1,33%	0,00%
Alsace	Riesling	SO4	26	non	broyage	oui	non	non	non	1,00%	1,67%
Alsace	Riesling	3309C	13	oui	broyage	oui	non	non	non	0,00%	0,00%
Alsace	Riesling	3309C	51				non	non	non	0,67%	2,67%
Alsace	Riesling	SO4	18	non	broyage	oui	non	non	non	0,00%	3,33%
Alsace	Riesling	161-49	21				oui	oui	non	0,00%	0,00%

Organigramme de l'étude statistique



Organigramme de l'étude statistique



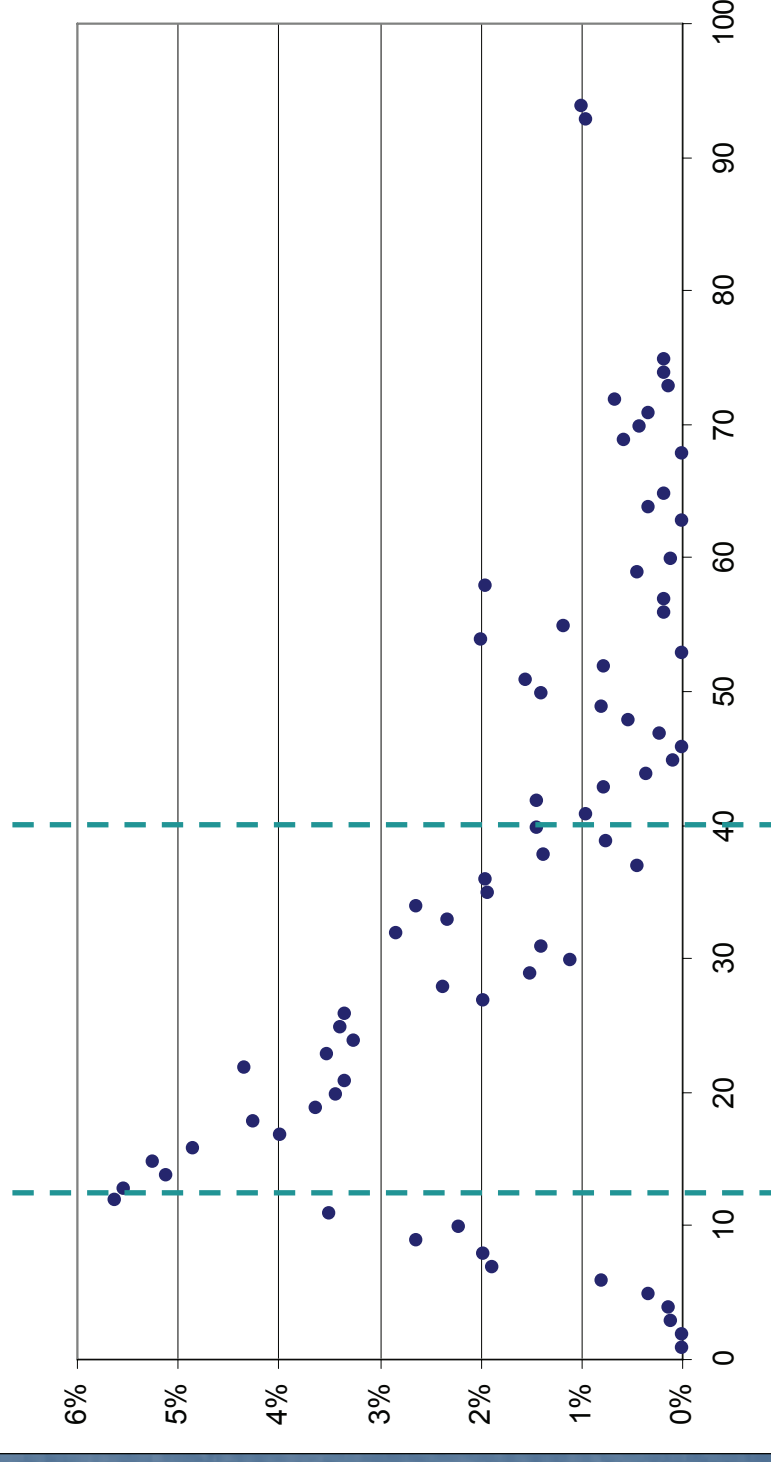
Etude préliminaire (1/5)

Description des variables

- Variables à expliquer:
 - Incidence de l'Eutypiose
 - Incidence de l'Esca/BDA
 - Taux de ceps improductifs
- Variables explicatives:
 - Qualitatives: cépage, région,...
 - Quantitatives: âge de la parcelle, densité de plantation,...

Etude préliminaire (2/5) Age de la parcelle

Taux d'expression de l'esca/BDA en fonction de l'âge de la parcelle



Augmentation entre 0 et 12 ans
 Diminution entre 12 et 40 ans
 Variations irrégulières après 40 ans dues à la faible quantité de données

Etude préliminaire (3/5)

Traitements à l'arsénite de soude entre 1999 et 2001

	Aucun traitement réalisé entre 1999 et 2001	Au moins une application entre 1999 et 2001
Taux d'expression de l'esca/BDA	3,47%	2,64%

Les parcelles traitées au moins une fois expriment moins la maladie, mais la différence est-elle significative ?

Etude préliminaire (4/5)

Sensibilité des cépages à l'Eutypiose

Tolérant(s)	Sensible(s)	Très sensible(s)	Extrêmement sensible(s)
Carignan	Chardonnay	Cabernet Sauvignon	Ugni Blanc
Gewurztraminer	Cinsault	Chenin	
Melon	Grenache	Gamay	
Merlot	Muscat Petits Grains	Sauvignon	
Poulsard	Pinot Auxerrois		
Riesling	Pinot Noir		
Savagnin	Syrah		
Moyenne	1,29%	3,56%	20,20%
Médiane	0,00%	2,17%	17,19%

Le calcul des comparaisons multiples grâce au test de Kruskal-Wallis a permis d'établir ces classes.

Etude préliminaire (5/5)

Sensibilité des cépages à l'Esca/BDA

Tolérant(s)	Sensible(s)	Très sensible(s)
Carignan	Chardonnay	Cabernet Sauvignon
Grenache	Gamay	Chenin
Merlot	Melon	Cinsault
Muscat Petits Grains	Pinot Auxerrois	Gewurztraminer
Pinot Noir		Poulsard
Syrah		Riesling
		Savagnin
		Sauvignon
		Trousseau
		Ugni Blanc
Moyenne	0,91%	4,53%
Médiane	0,31%	2,66%

Le calcul des comparaisons multiples grâce au test de Kruskal-Wallis a permis d'établir ces classes.

Tendance générale (comparaison 2003-2005 sur 286 parcelles)

L'Eutypiose a été plus fréquent en 2005

Prévalence moyenne

2003 = 54,4%

2004 = 49,0%

2005 = 60,3%

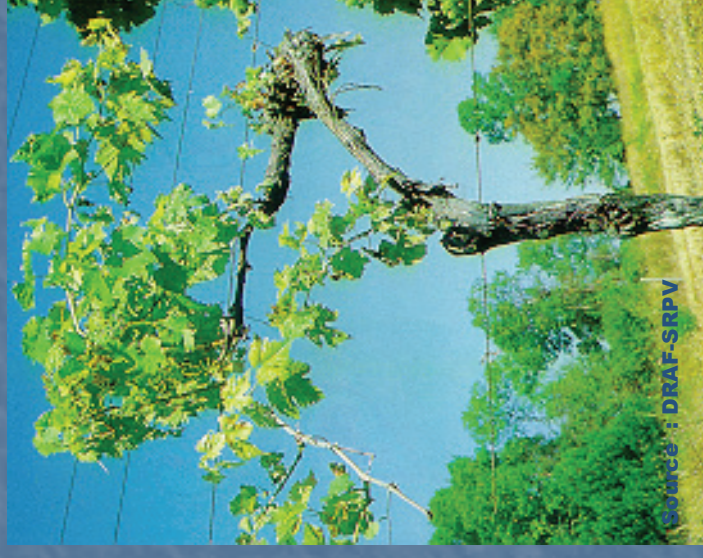
L'expression des symptômes foliaires d'Eutypiose est toujours stable

Incidence moyenne

2003 = 3,5 %

2004 = 3,5 %

2005 = 3,6 %



Tendance générale (comparaison 2003-2005 sur 494 parcelles)

**L'Esca et le BDA ont été plus fréquents
en 2005**

Prévalence moyenne

2003 = 61,1 %
2004 = 69,8 %
2005 = 79,9 %

**L'expression des symptômes foliaires
d'Esca et de BDA est en augmentation**

Incidence moyenne

2003 = 1,7 %
2004 = 2,7 %
2005 = 3,3 %



Source : ONAF-SPV

Organigramme de l'étude statistique



Etablissement de classes

Variables explicatives majoritairement
qualitatives

➡ Analyse Factorielle des
Correspondances Multiples

➡ Classes pour les variables
quantitatives

Classes pour les variables quantitatives

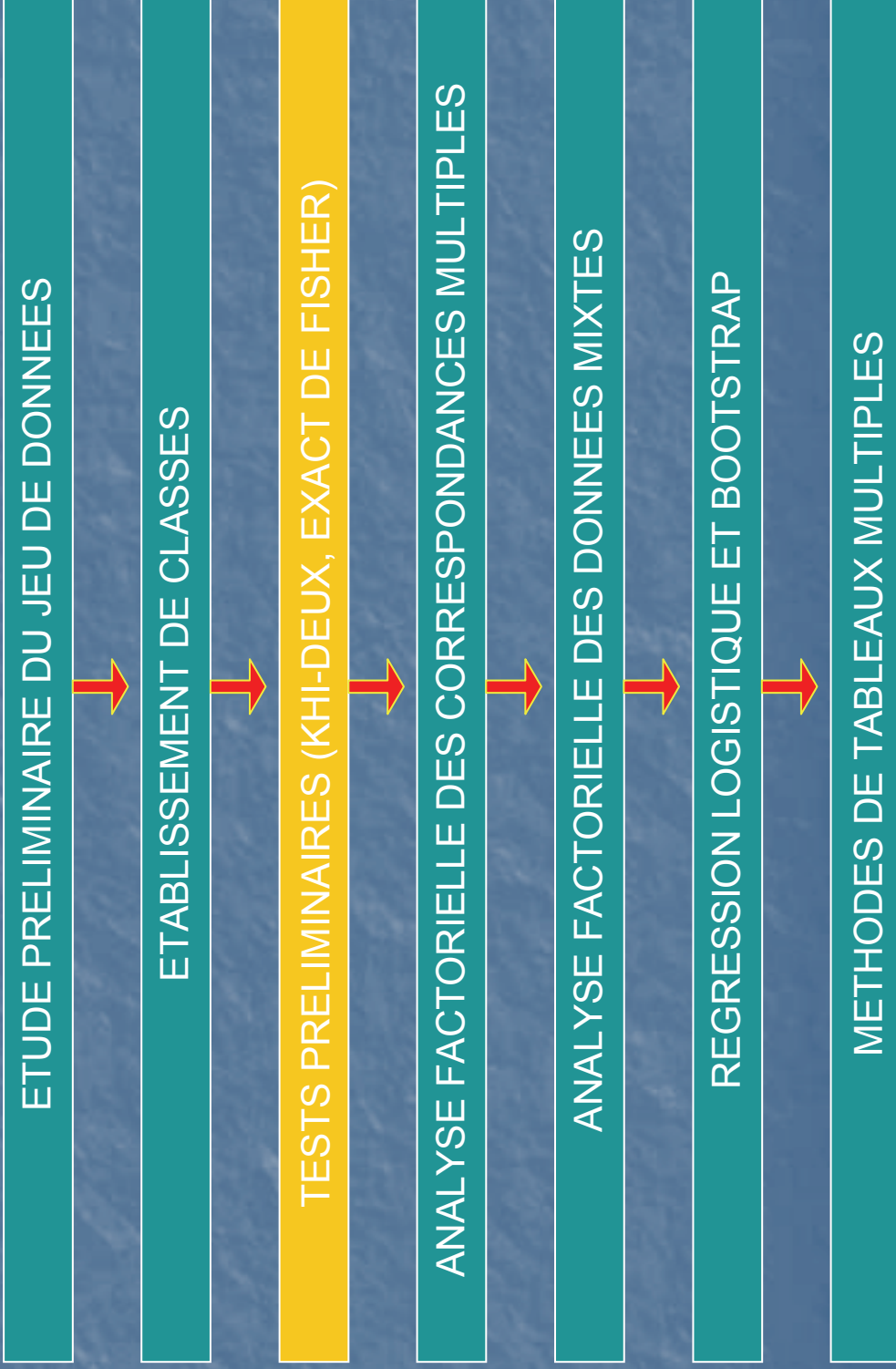
Variables à expliquer

Évolution	Incidences (Eutypiose, Esca/BDA) et pourcentage (mortalité)	Eutypiose	Esca/BDA	Mortalité des plants
	aucune	Euty0 : = 0%	Esca0 : = 0%	Mort0 : < 3%
	faible	Euty1 : ≤ 2%	Esca1 : ≤ 3%	Mort1 : < 10%
	forte	Euty2 : > 2%	Esca2 : > 3%	Mort2 : ≥ 10%

Âge de la parcelle

Âge de la parcelle	Age1	Age2	Age3	Age4
	de 0 à 15 ans	de 16 à 25 ans	de 26 à 40 ans	plus de 40 ans

Organigramme de l'étude statistique



Tests préliminaires

Analyse bivariée



Tests d'indépendance (Khi-deux,
test exact de Fisher,...)

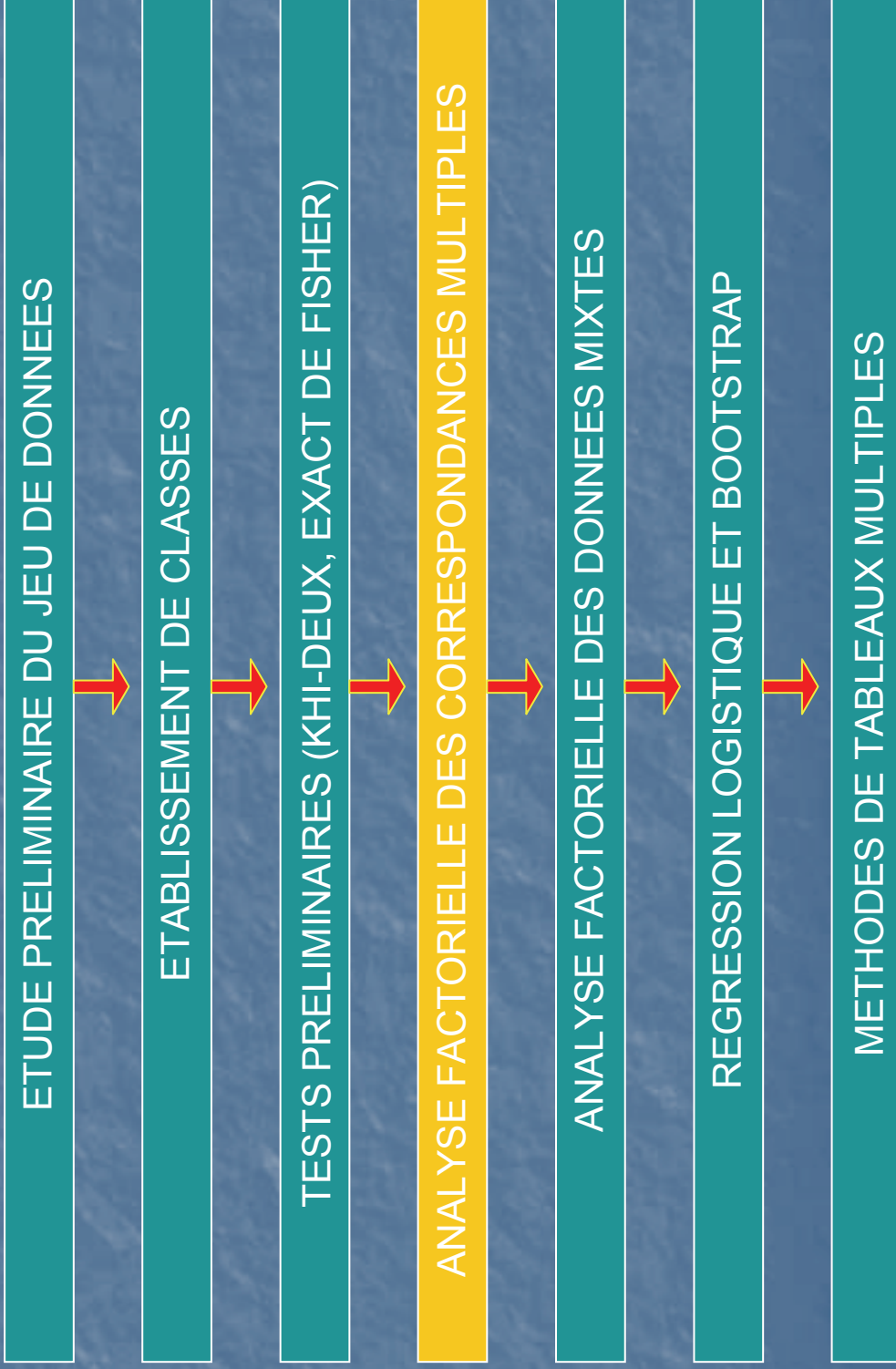


Première sélection parmi les
variables explicatives

Variables conservées et exclues

Variables conservées dans l'analyse	Variables supprimées de l'analyse
Région	Vigueur
Cépage	Clone cépage
Âge de la parcelle	Clone porte-greffe
Nombre de traitements à l'arsénite de sodium	Type de sol
Porte-greffe	Enlèvement des bois morts
Devenir des sarments	Date de début de taille
Prétaillage	Date de fin de taille
Type de taille	

Organigramme de l'étude statistique



Analyse Factorielle des Correspondances Multiples

- Théorie :

Les données :

- n unités
- r variables qualitatives (ou rendues qualitatives par codage en classes) possédant chacune p_k modalités $k=1,\dots, r$.

Définition : Nous appelons tableau disjonctif total associé aux données observées, le tableau

$$D = (d_{i,j}, i = 1, \dots, n, j = 1, \dots, p)$$

Définition : Nous appelons tableau de Burt associé aux données observées, le tableau

$$B = (b_{j,j'}, j = 1, \dots, p, j' = 1, \dots, p)$$

où $b_{j,j'} = \sum_{i=1}^n d_{i,j} d_{i,j'}, j = 1, \dots, p, j' = 1, \dots, p$.



Analyse Factorielle des Correspondances Multiples

Interprétation : L'interprétation des résultats se fait à l'aide des projections, des cosinus carrés et des contributions.

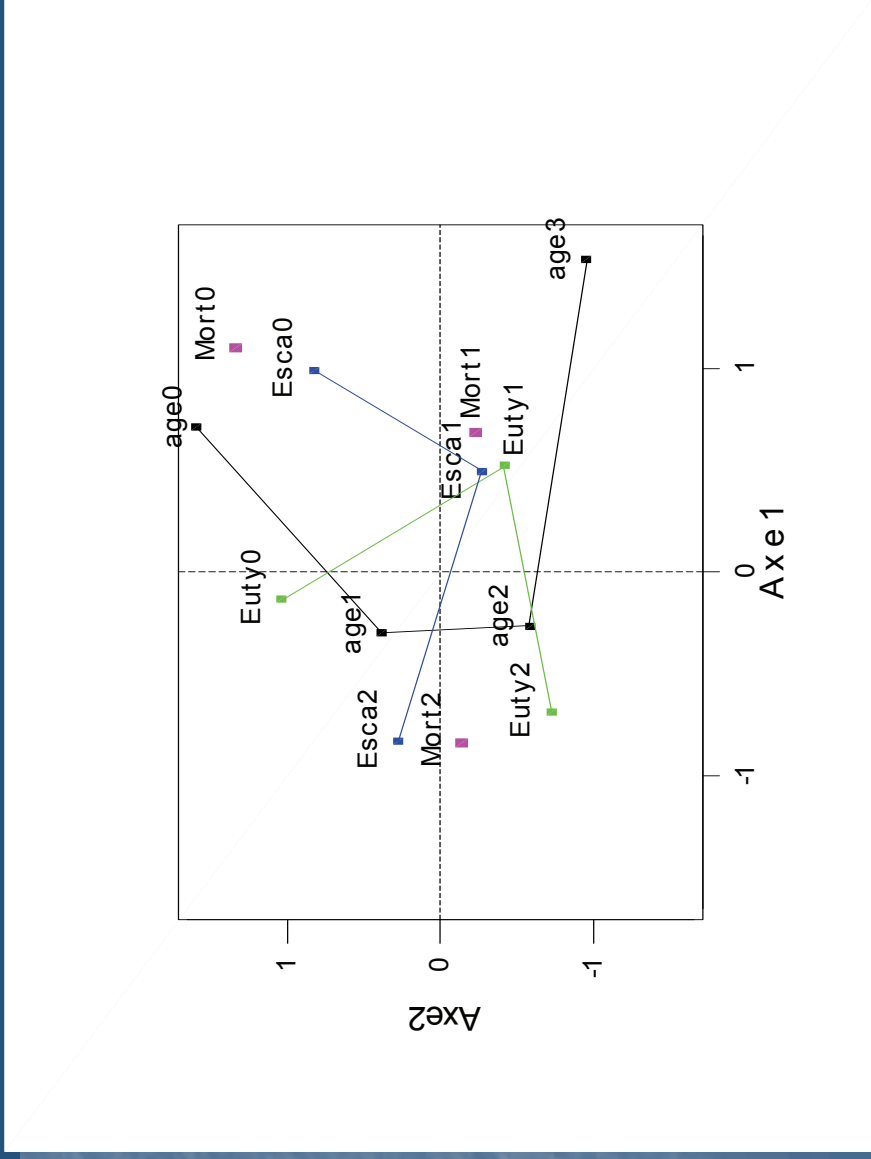
$$F_{\alpha}(j^B) = F_{\alpha}^{*}(j^B) = \frac{1}{r} \sum_{j=1}^p \frac{d_{i,j} d_{i,j'} v_{\alpha,j'}}{d_{.,j}} \quad j=1, \dots, p, \alpha=1, \dots, p-1,$$

$$Q_{\alpha}(j^B) = Q_{\alpha}^{*}(j^B) = \frac{F_{\alpha}^{*2}(j^B)}{nr \sum_{j'=1}^p \frac{b_{j,j'}^2}{d_{.,j} d_{.,j'}}} \quad j=1, \dots, p, \alpha=1, \dots, p-1,$$

$$C_{\alpha}(j^B) = C_{\alpha}^{*}(j^B) = \frac{1}{nr \lambda_{\alpha}^2 d_{.,j}} \left(\sum_{i=1}^n d_{i,j} \frac{1}{r} \sum_{j=1}^p d_{i,j} v_{\alpha,j} \right)^2 \quad j=1, \dots, p, \alpha=1, \dots, p-1.$$

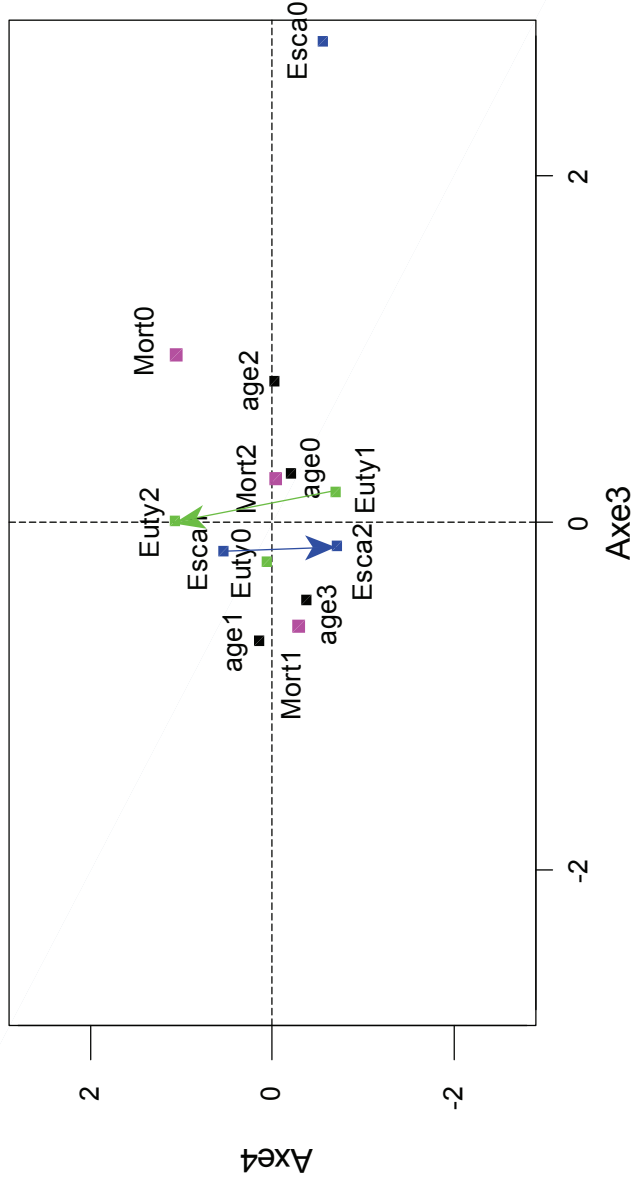
Analyse Factorielle des Correspondances Multiples

■ Résultats : « ACM principale »



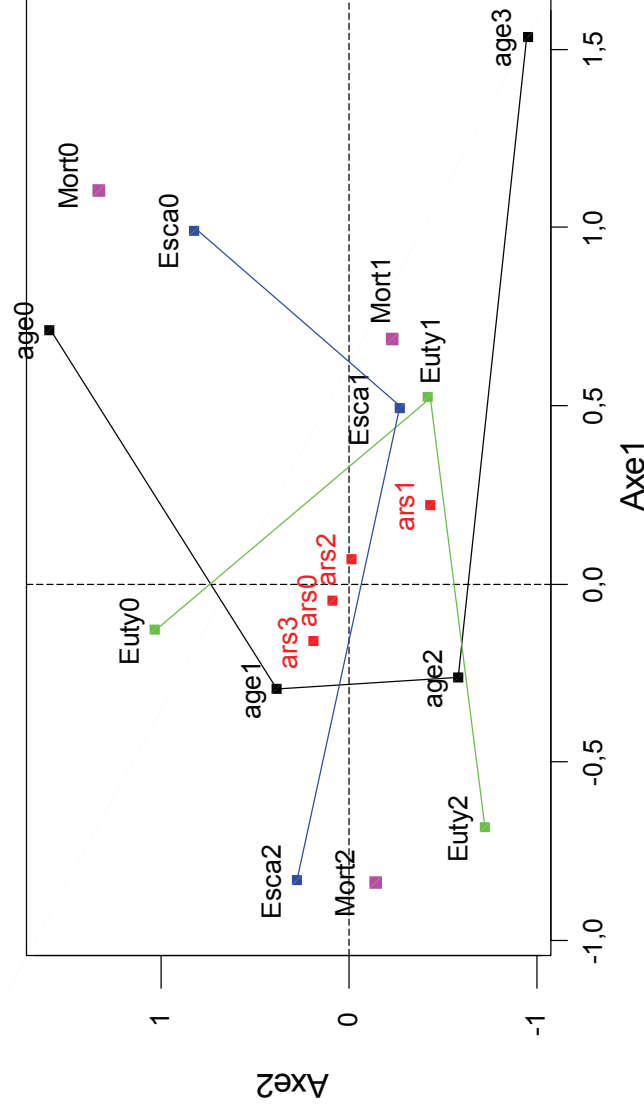
Analyse Factorielle des Correspondances Multiples

■ Résultats : « ACM principale »

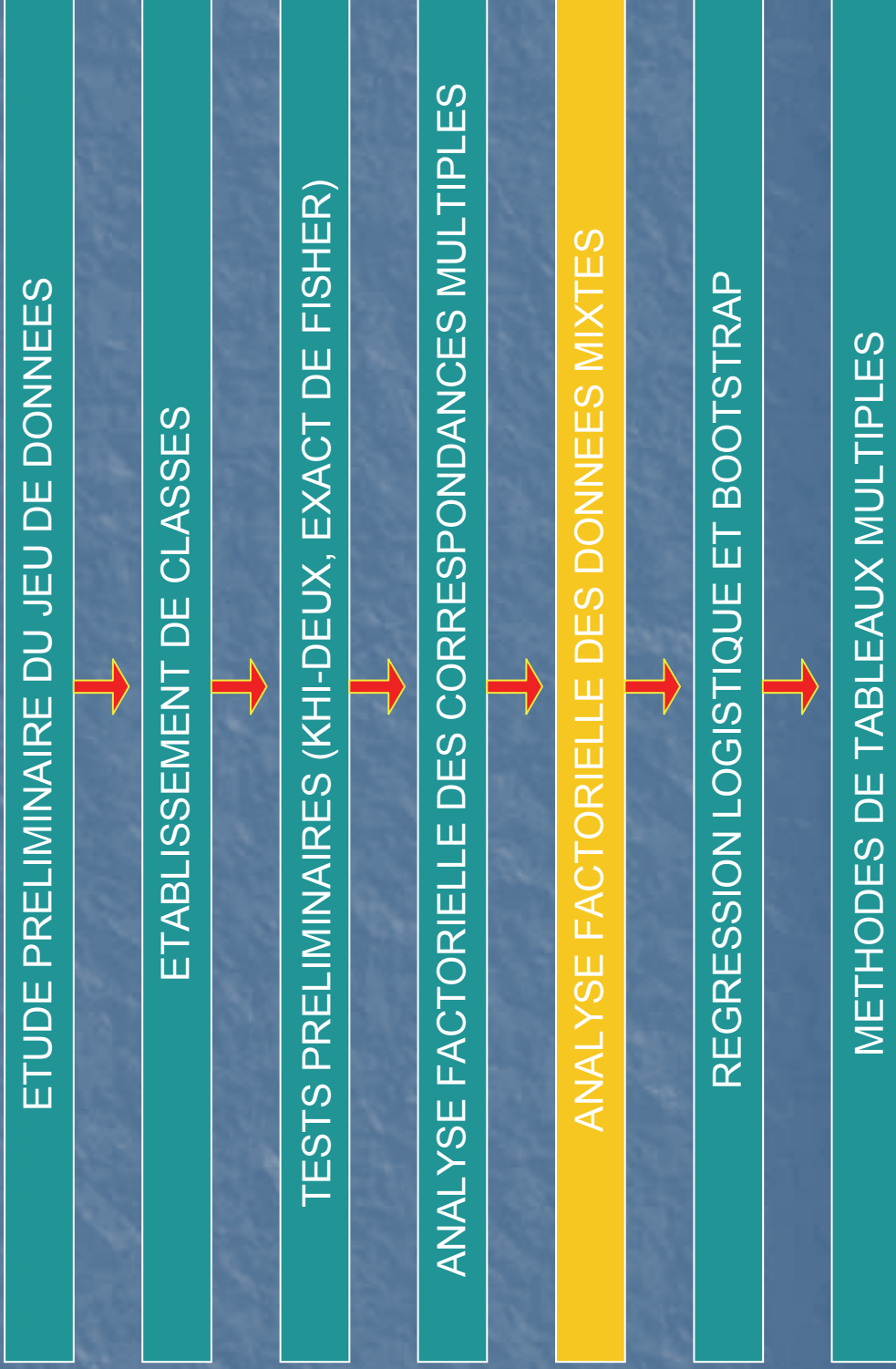


Analyse Factorielle des Correspondances Multiples

■ Résultats : Arsénite de soude en supplémentaire



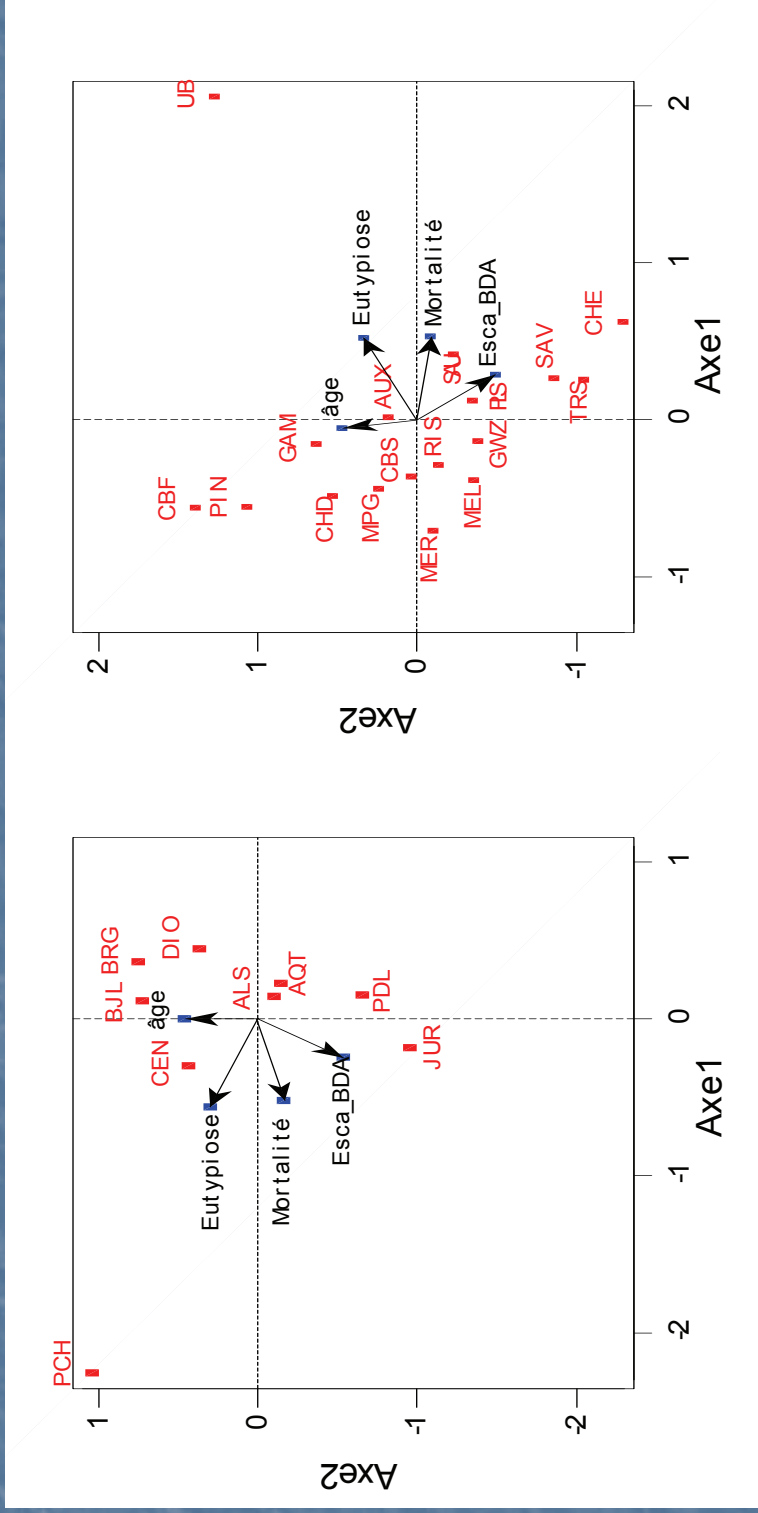
Organigramme de l'étude statistique



Analyse factorielle de données mixtes

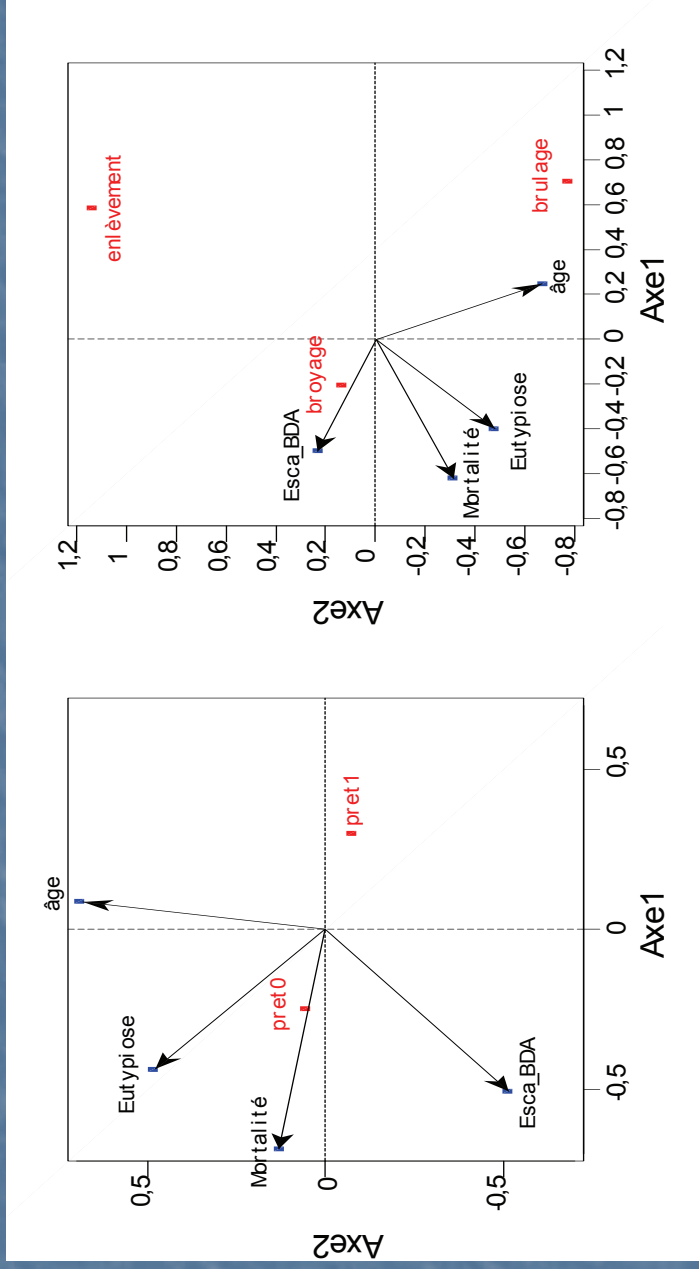
- Afin de pouvoir effectuer une analyse des correspondance multiples les données sur l'incidence de l'eutypiose, celles sur l'incidence de l'esca/BDA ainsi que celles sur la mortalité des plants ont dues être regroupées en classes.
-  Possibles réductions de l'information disponible.
- Technique à même de combiner données qualitatives et quantitatives sous forme d'analyse factorielle : l'analyse factorielle de données mixtes.
- Même schéma d'étude des différentes variables à notre disposition que lors des ACM : tronc commun de trois variables, l'incidence de l'eutypiose, l'incidence de l'esca/BDA, la mortalité des plants (variables quantitatives) et d'une variable supplémentaire (variable qualitative).

Analyse factorielle de données mixtes



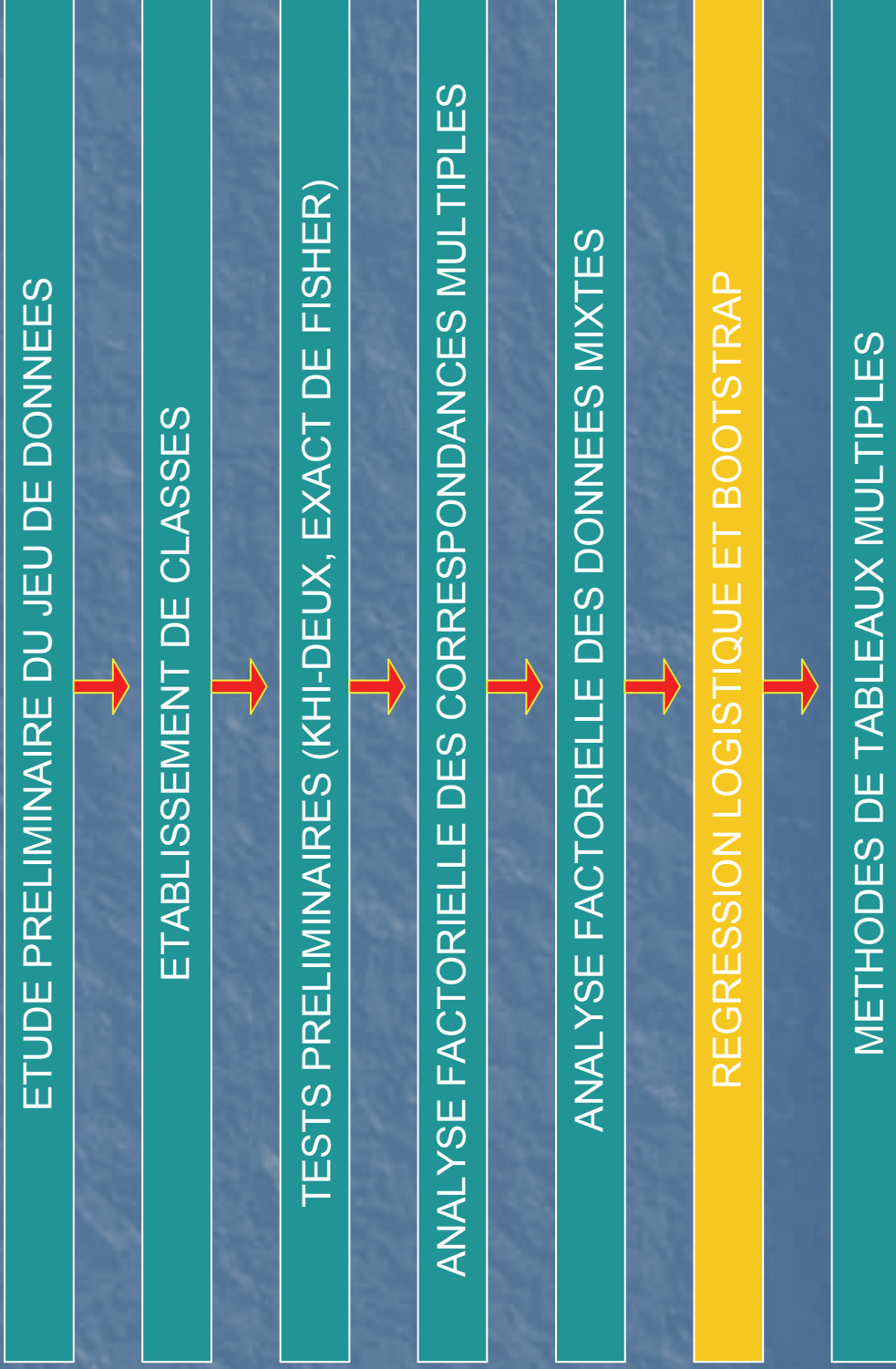
- Représentations graphiques des deux premiers axes factoriels des analyses factorielles de données mixtes avec pour variables qualitatives respectivement la région et le cépage

Analyse factorielle de données mixtes



- L'analyse des données mixtes permet de mettre en évidence une relation entre le prétaillage et le devenir des sarments, deux mesures prophylactiques, qui ne semblaient pas avoir d'effet sur les maladies et la mortalité lors des ACM.

Organigramme de l'étude statistique



Régression logistique

Deux types de régressions logistiques utilisés:

- Régressions logistiques binaires
 ➡ Apparition des symptômes ou de la mortalité
- Régressions logistiques ordinales
 ➡ Intensification des symptômes ou évolution de la mortalité

Régression logistique

Incidences (eutypiose, esca/BDA) et pourcentage (mortalité)		Eutypiose	Esca/BDA	Mortalité des plants
Apparition	non	0%	0%	< 3%
	oui	> 0%	> 0%	≥ 3%
Évolution	aucune	0%	0%	< 3%
	faible	≤ 2%	≤ 3%	< 10%
	forte	> 2%	> 3%	≥ 10%

Régression logistique

Théorie:

- Régression logistique binaire

$$P(Y = 1) = \frac{\exp(\alpha + \sum \beta_i X_i)}{1 + \exp(\alpha + \sum \beta_i X_i)}$$

$$\text{soit } \text{logit}(P(Y = 1)) = \alpha + \sum \beta_i X_i$$

avec β_j les paramètres à estimer et X_j les covariables ou cofacteurs

- Régression logistique ordinaire

$$\text{Soit } \theta_j = \frac{P(Y \in C_1 \cup \dots \cup C_j)}{1 - P(Y \in C_1 \cup \dots \cup C_j)}.$$

$$\text{Alors } \ln(\theta_j) = \alpha_j - \sum \beta_i X_i$$

avec α_j sera une constante associée au modèle, β_j les paramètres à estimer et X_j les covariables ou cofacteurs

Régression logistique

Résultats : Incidence de l'eutypiose

Tests

Test	Valeur de la statistique	Degré(s) de liberté	p-valeur	Décision
Pentes nulles	230,225	29	0,000	au moins un coefficient est non nul
χ^2 de Pearson	313,309	395	0,999	le modèle est adéquat
χ^2 de déviance	266,755	395	1,000	le modèle est adéquat
Adéquation du cofacteur Région	30,656	10	0,001	le cofacteur Région est significatif
Adéquation du cofacteur Cépage (Région)	62,946	18	0,000	le cofacteur Cépage (Région) est significatif

Régression logistique

Résultats : Incidence de l'eutypiose

➤ Intensification

Apparition des symptômes

$$\begin{aligned} \ln(\theta_1) = & 1,4228 + 0,05139 \text{Âge} - 1,8186I_{\{BRG\}} - 2,6303I_{\{CEN\}} - 1,7580I_{\{DIO\}} - 3,9616I_{\{JUR\}} \\ & - 2,951I_{\{PAC\}} - 2,5961I_{\{GWZ(ALS)\}} - 2,9440I_{\{RIS(ALS)\}} - 4,170I_{\{MER(AQT)\}} \\ & + 4,420I_{\{PIN(BRG)\}} + 4,5591I_{\{CHE(CEN)\}} + 2,565I_{\{SYR(DIO)\}}. \end{aligned}$$

Intensification des symptômes

$$\begin{aligned} \ln(\theta_2) = & -2,1155 + 0,05139 \text{Âge} - 1,8186I_{\{BRG\}} - 2,6303I_{\{CEN\}} - 1,7580I_{\{DIO\}} - 3,9616I_{\{JUR\}} \\ & - 2,951I_{\{PAC\}} - 2,5961I_{\{GWZ(ALS)\}} - 2,9440I_{\{RIS(ALS)\}} - 4,170I_{\{MER(AQT)\}} \\ & + 4,420I_{\{PIN(BRG)\}} + 4,5591I_{\{CHE(CEN)\}} + 2,565I_{\{SYR(DIO)\}}. \end{aligned}$$

Régression logistique

Résultats : Incidence de l'eutypiose
Critères de bon classement

Critères	
Concordants (%)	89,0
D de Somers	0,78
Gamma de	0,79
Goodman-Kruskal	0,51
Tau-a de Kendall	

Régression logistique

Résultats : Taux de mortalité

➤ Évolution

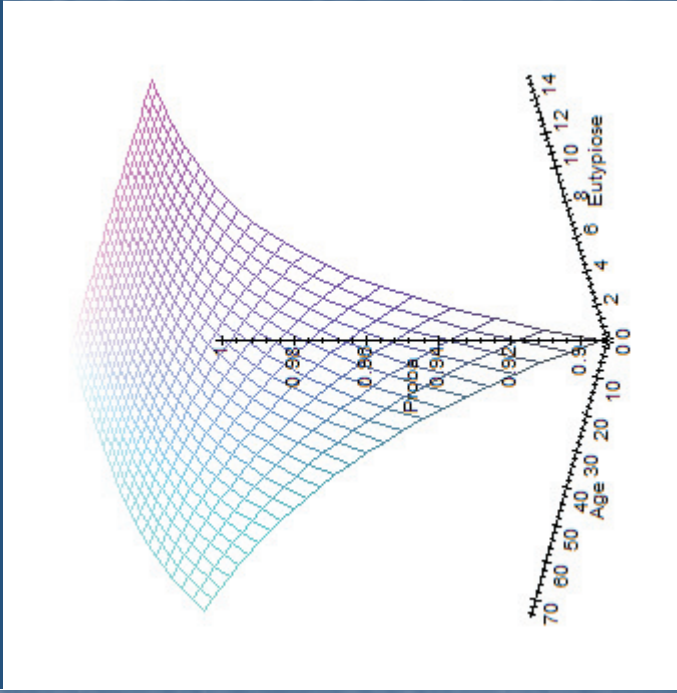
Passage d'un taux faible à un taux moyen

$$\ln(\theta_1) = 2,1396 + 26,862Eutypiose + 0,02727\hat{Age} - 1,9419I_{\{CBF\}} - 2,0555I_{\{CHD\}} - 3,857I_{\{MER\}} - 1,5149I_{\{MPG\}} - 2,1319I_{\{PIN\}}$$

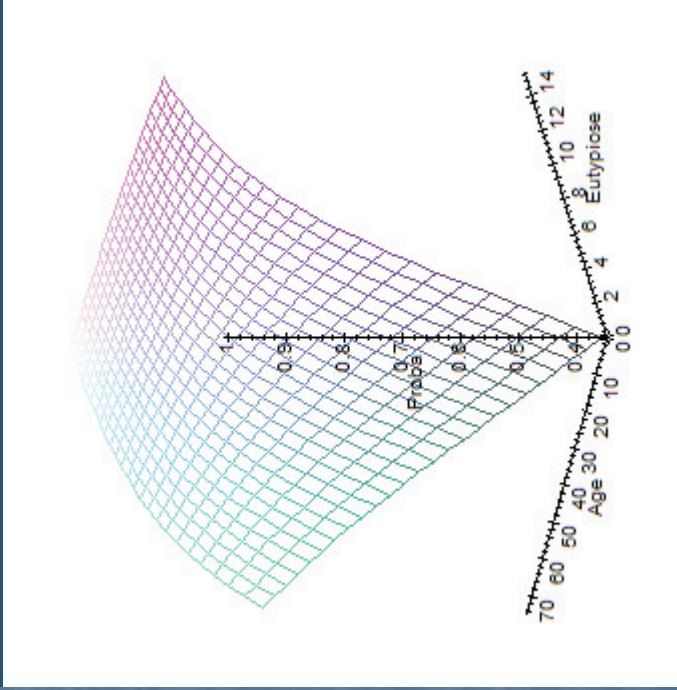
Passage d'un taux moyen à un taux fort

$$\ln(\theta_2) = -0,5729 + 26,862Eutypiose + 0,02727\hat{Age} - 1,9419I_{\{CBF\}} - 2,0555I_{\{CHD\}} - 3,857I_{\{MER\}} - 1,5149I_{\{MPG\}} - 2,1319I_{\{PIN\}}$$

Passage d'un taux de mortalité faible à moyen



Passage d'un taux de mortalité moyen à fort



Modèles hiérarchiques

- Etant donné les associations complexes existant entre les deux variables « région » et « cépage », nous avons été amenés à introduire des modèles hiérarchiques.

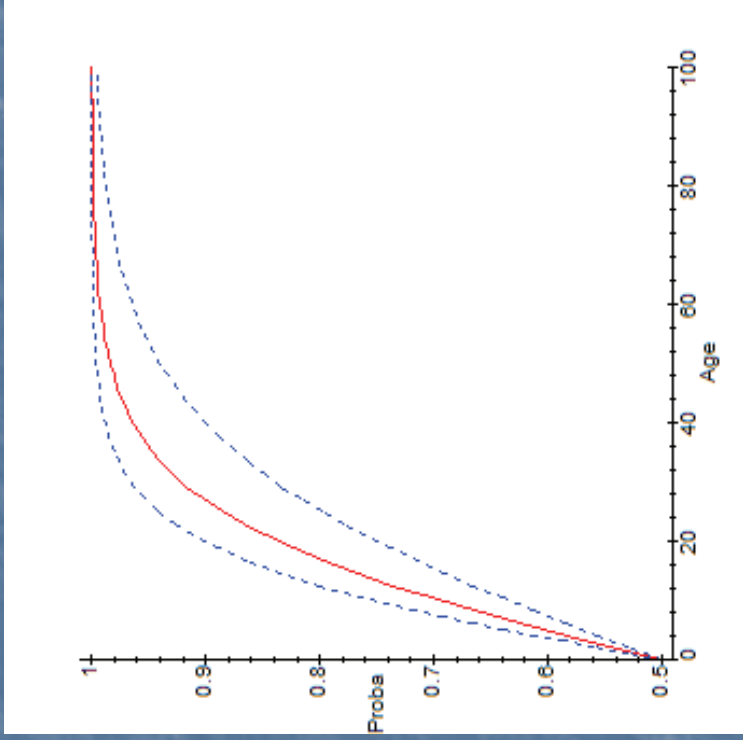
Utilisation du bootstrap

- Les estimations par intervalle des paramètres des différents modèles obtenues dans le cadre de l'approximation asymptotique classique sont-ils précis ?
- Une façon de répondre à cette question est l'utilisation de la technique dite du bootstrap. L'idée ici est de trouver des intervalles de confiance pour ces paramètres par bootstrap.

Exemple de résultats

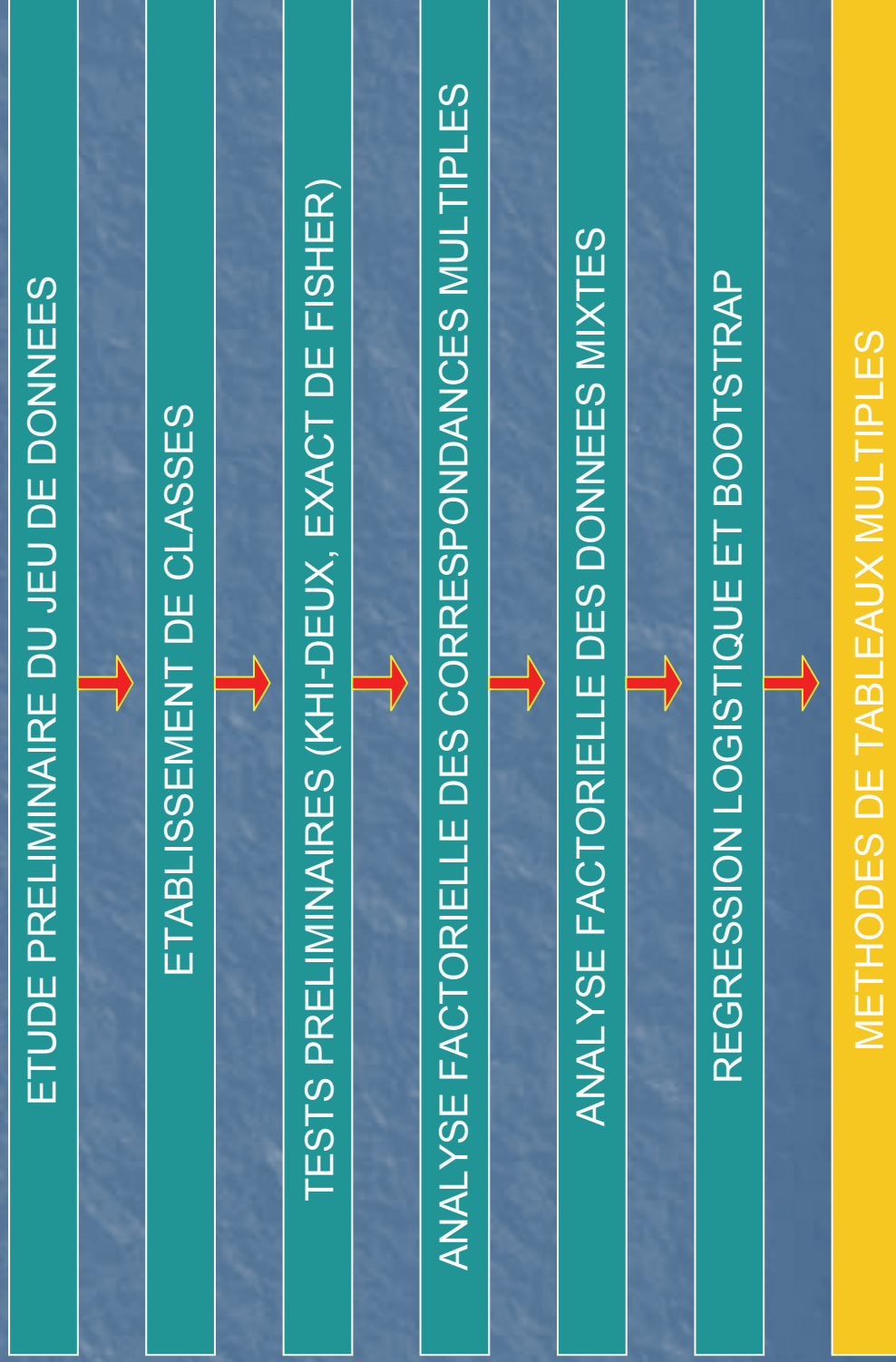
Modèle	Paramètre	Estimation	Intervalle de confiance
Intensification			
logit(Mortalité) ~ Eutypiose + âge + Cépage	Constante1	2,140	[1,021 ; 3,169]
	Constante2	-0,573	[-1,707 ; 0,559]
	Eutypiose	26,862	[12,830 ; 32,010]
	âge	0,027	[0,005 ; 0,054]
	CBF	-1,942	[-3,620 ; -0,508]
	CHD	-2,056	[-3,282 ; -0,799]
	MER	-3,857	[-19,734 ; -1,013]
	MPG	-1,515	[-2,874 ; -0,309]
	PIN	-2,132	[-3,532 ; -0,799]

Intervalles de confiance



Probabilité d'avoir au moins un pied symptomatique (eutyposis, phase d'apparition) dans une parcelle du réseau en fonction de l'âge de la vigne avec limites de confiance à 95%

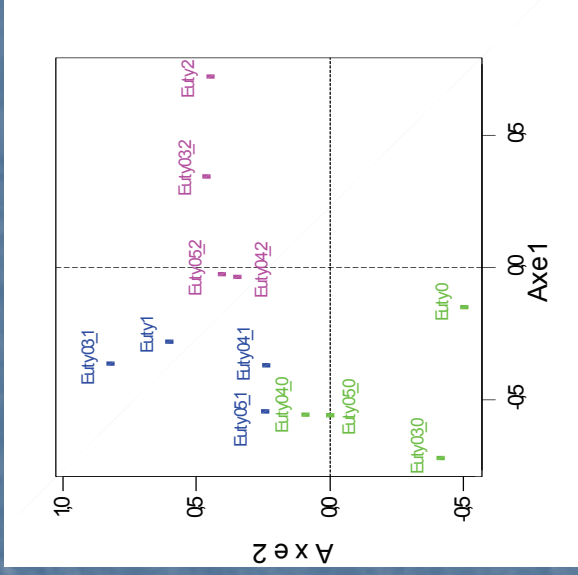
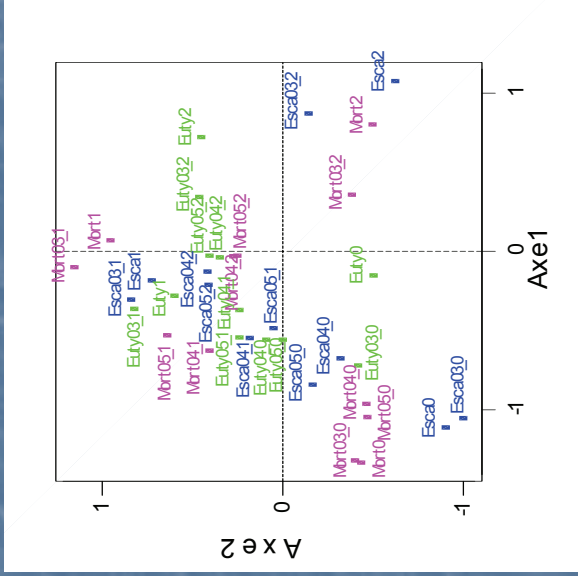
Organigramme de l'étude statistique



Méthodes de tableaux multiples

- Chaque tableau de données représente une année d'observation.
- Observer si une année occupe une place prépondérante dans l'explication de l'évolution globale des maladies et de la mortalité.
- Evolution temporelle des deux maladies du bois de la vigne et de la mortalité des plants.
- Déceler si une des trois années exprime des comportements radicalement différents des deux autres en tenant compte de la variabilité inter annuelle, par exemple due aux conditions climatiques : l'année 2003 a été marquée par un été caniculaire.

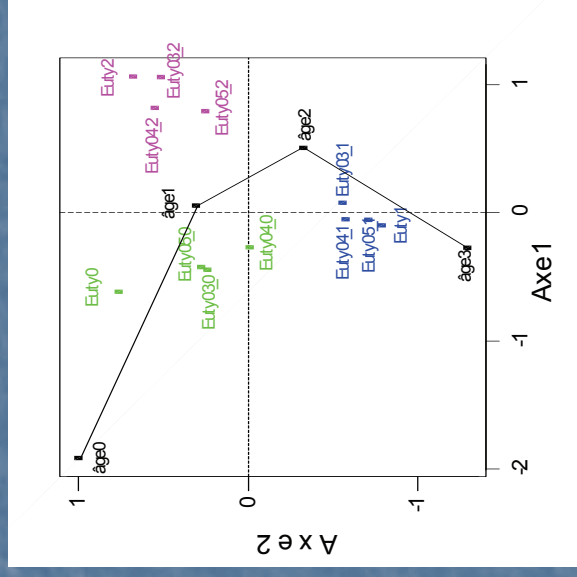
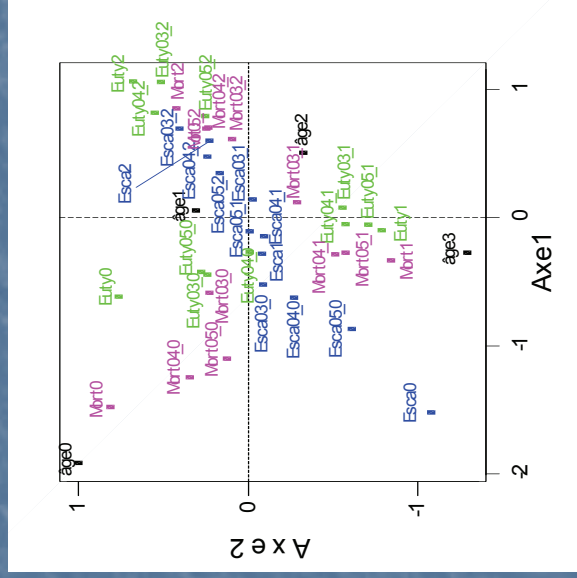
Incidence des maladies en fonction de l'année



■ Deux premiers axes factoriels présentant les modalités annuelles et moyennes sans tenir compte de l'âge de la parcelle

■ Deux premiers axes factoriels présentant les modalités annuelles et moyennes uniquement pour l'eutypiose (sans tenir compte de l'âge de la parcelle)

Incidence des maladies en fonction de l'année et de l'âge



■ Deux premiers axes factoriels présentant les modalités annuelles et moyennes en tenant compte de l'âge de la parcelle

■ Deux premiers axes factoriels présentant les modalités annuelles et moyennes uniquement pour l'eutypiose (en tenant compte de l'âge de la parcelle)

Conclusion

- Complémentarité des différentes méthodes statistiques.
- Elaboration d'une méthodologie statistique performante pour traiter des données quantitatives, qualitatives et dynamiques.
- L'analyse des correspondances multiples nous a permis d'établir un panorama des associations entre les variables les plus importantes.
- L'analyse factorielle de données mixtes nous a confirmé les résultats de l'analyse des correspondances multiples et a dégagé d'autres relations entre les variables.
- La régression logistique nous a permis de modéliser ces relations, nos variables à expliquer étant de type binaire ou ordinal.
- L'analyse des correspondances multiples multi-tableau a rendu possible la prise en compte du facteur temps et ainsi de juger de l'évolution des maladies du bois de la vigne et de la mortalité des ceps.
- Avec l'apport des données des trois années à venir de nouvelles techniques pourront être mises en oeuvre, notamment pour mieux appréhender la question de l'évolution de ces trois maladies.