

EXPLOITATION DES DONNÉES DES ENQUÊTES  
EN ÉPIDÉMIOLOGIE DES MALADIES ANIMALES  
L'ANALYSE MULTIFACTORIELLE



J.P. TILLON

Ministère de l'Agriculture, Direction de la Qualité, Services Vétérinaires  
Station de Pathologie Porcine B.P. 9 - 22440 Ploufragan

RESUME

*Outil statistique, l'analyse multifactorielle permet à l'épidémiologiste de simplifier un tableau de données en le réduisant aux influences essentielles. Interprétées comme "facteurs de risque" certaines de ces influences à caractère causal permettent d'élaborer une stratégie de la prévention sanitaire basée sur l'estimation du risque d'apparition d'un trouble. La mise à disposition des utilisateurs d'une "carte" des facteurs de risque est la principale retombée à usage pratique des enquêtes épidémiologiques visant la recherche étiologique.*

A l'issue d'une période d'enquête épidémiologique, l'observateur se voit confronté à une multitude d'informations de toute nature. Sa tâche s'est trouvée simplifiée depuis l'avènement des ordinateurs de grande capacité, permettant le stockage de l'information et la mise en oeuvre d'un traitement statistique approprié. Ce travail est généralement réalisé par un statisticien-informaticien, dont le rôle intervient dès la conception du questionnaire d'enquête. L'épidémiologiste, quant à lui, suit le déroulement de l'exploitation des données en l'orientant en fonction de sa propre expérience; il donne aux résultats une forme définitive qui doit satisfaire aux exigences statistiques et se prêter à une présentation intelligible. L'analyse multifactorielle, et certaines de ses variantes, se prête particulièrement à l'usage pour des enquêtes épidémiologiques visant la recherche étiologique.

SCHEMA GENERAL ET ROLE DE L'EPIDEMIOLOGISTE DANS L'EXPLOITATION DES DONNEES

Dans le cas le plus habituel, l'exploitation des données d'une enquête épidémiologique à visée étiologique tend à faire apparaître plusieurs facteurs (= variables) épidémiologiques qui par leur action propre et leur inter-actions rendent compte, dans des proportions variables, des phénomènes étudiés et permettent d'en dégager un modèle explicatif. Chacun des éléments soulignés correspond à une étape de l'exploitation des données :

- 1 - identification des variables en cause
- 2 - étude de leurs inter-actions
- 3 - valeur statistique de l'agencement des variables
- 4 - édition d'un modèle explicatif

Tout au long de ce processus, l'épidémiologiste est appelé à donner son avis. Il faut admettre en effet qu'il dispose "a priori" d'hypothèses concernant les variables présumées responsables : on ne peut imaginer entreprendre une enquête épidémiologique à visée étiologique sans idées précon-

gues. L'enquête réalisée s'apparente ainsi à une expérimentation dont les résultats confirmeront ou infirmeront les hypothèses de départ.

L'intervention de l'épidémiologiste tout au long du processus d'exploitation des données consiste à orienter le statisticien vers des variables d'influence vraisemblable en laissant de côté (au moins momentanément) certaines variables dont l'importance n'est peut-être qu'apparente (phénomènes de redondance).

Pour illustrer notre propos, nous allons considérer une application des méthodes d'analyse factorielle à l'exploitation des données d'une enquête épidémiologique.

#### APPLICATION DE L'ANALYSE FACTORIELLE DES CORRESPONDANCES A L'INTERPRETATION D'UNE ENQUETE EPIDEMIOLOGIQUE SUR LA PATHOLOGIE RESPIRATOIRE EN ELEVAGE PORCIN

##### Présentation de l'enquête :

103 élevages porcins naisseurs-engraisseurs ont participé, dans la période d'Avril à Octobre 1980, à une enquête sur la pathologie respiratoire en engraissement et ses conséquences sur la croissance des animaux. 635 informations sont disponibles pour chaque élevage (voir figure 1 et figure 2) soient 635 x 103 informations au total.

##### Classification des élevages selon les résultats :

Les informations disponibles sur les variables résultantes (148 pour ETAT DE SANTE et 51 pour PERFORMANCES) sont considérées : celles qui sont jugées particulièrement représentatives du problème étudié sont sélectionnées et interprétées comme les "objectifs" de l'enquête, c'est-à-dire l'ensemble des critères sur lesquels les élevages seront jugés simultanément. Les objectifs retenus sont les suivants :

- fréquence et étendue des lésions de pneumonie [PNEU 1 à 4]
- fréquence et étendue des lésions d'atrophie des cornets nasaux [GRO 1 à 4]
- fréquence et étendue des lésions de polysérite [SER 1 à 4]
- présence ou absence d'abcès pulmonaires [ABC 1 OU 2]
- vitesse de croissance ou âge à 100 kg [AGE 1 à 4]

Des critères sélectionnés au départ, tels que le classement commercial des carcasses ou l'indice de consommation sont écartés (données non comparables pour les élevages, renseignements manquants...).

Un tableau de données, décrivant tous les élevages en fonction de leurs résultats (c'est-à-dire de la classe d'appartenance) vis-à-vis des objectifs est constitué (figure 3) : son interprétation ne peut être réalisée sans le recours à une méthode statistique appropriée, l'analyse factorielle des correspondances.

Le problème auquel nous devons apporter une solution est le suivant : nous disposons de 5 informations pour chacun des élevages que nous devons classer. Si nous n'en avons que 2 ou 3, il nous serait possible de proposer une représentation graphique dans un système à 2 ou 3 dimensions (repère d'axes orthonormés par exemple) et de visualiser les rapprochements entre élevages de coordonnées voisines sur les 3 axes. La méthode d'analyse factorielle des correspondances permet, en créant un nouveau système d'axes (axes = facteurs), de représenter de manière simplifiée le tableau de données constitué de n lignes (les observations) et de m colonnes (les variables). Les axes factoriels correspondent à des "lignes de force" à la contribution desquelles s'associent les variables qui distinguent le plus les

FIGURE 1 - COLLECTE DES DONNEES (ENQUÊTE "PATHOLOGIE RESPIRATOIRE 1980")

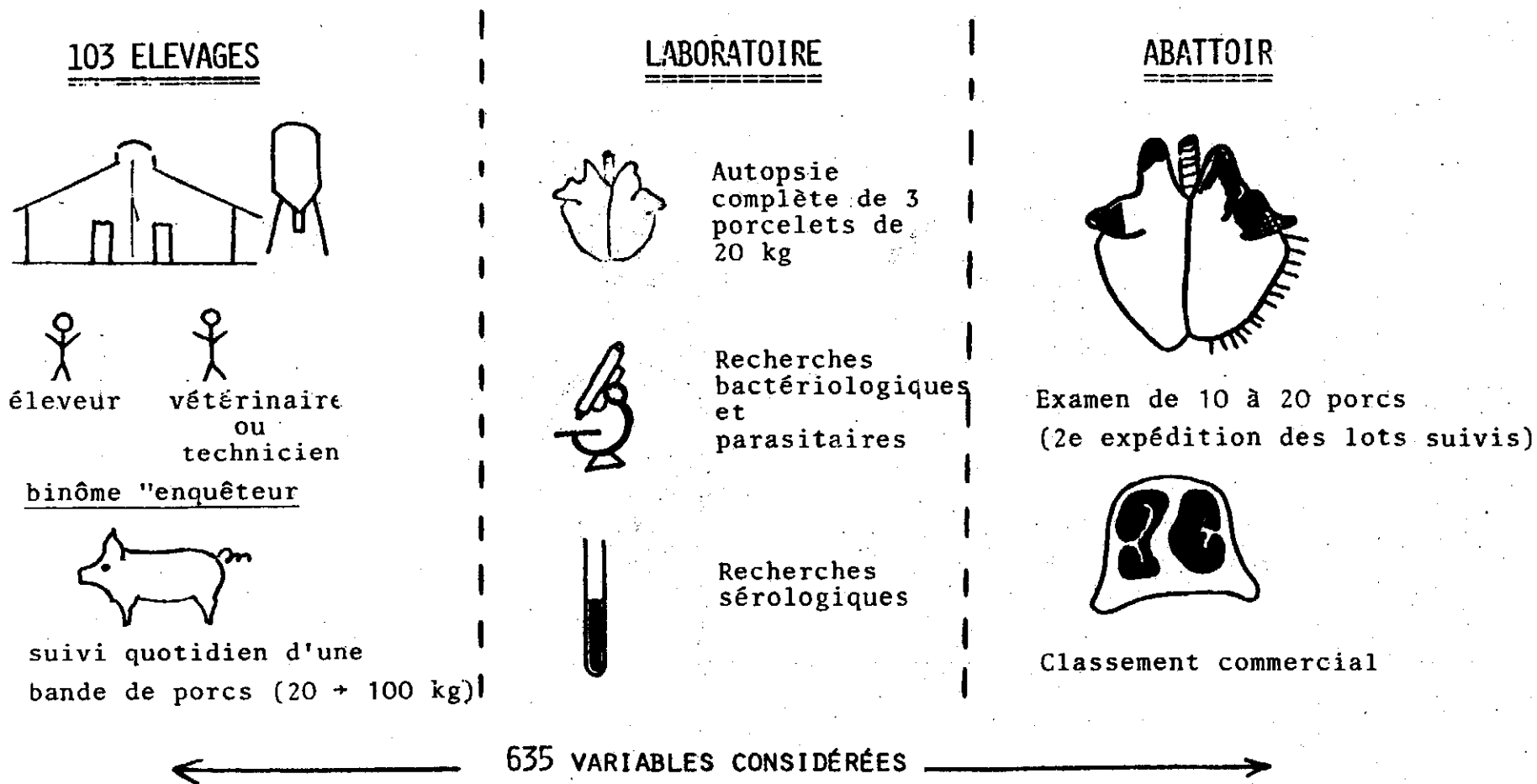
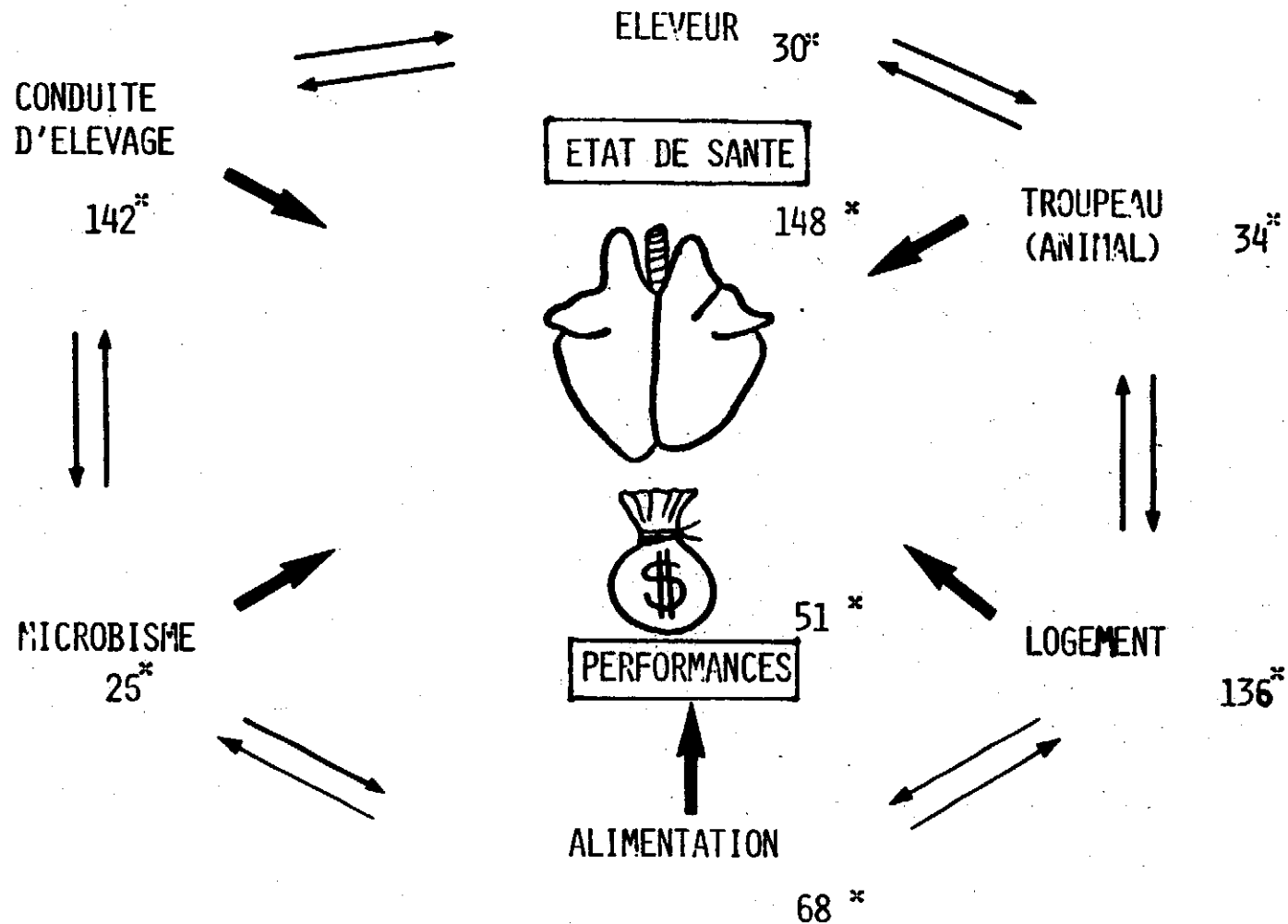


FIGURE 2 - STRUCTURATION DES DONNÉES (ENQUÊTE "PATHOLOGIE RESPIRATOIRE 1980")

103 ÉLEVAGES NAISSEURS-ENGRAISSEURS



\*Nombre d'informations recueillies pour décrire la variable TOTAL = 635

FIGURE 3 - TABLEAU DE DONNÉES

COLONNES = VARIABLES

LIGNES  
=  
OBSERVATIONS  
=  
ELEVAGES

	PNEU1	PNEU 2	PNEU3	PNEU4	GRO 1	GRO 2	GRO 3	GRO 4...
N°1	X				X			
N°2		X					X	
N°3	X				X			
N°4			X				X	
N°5		X				X		
N°6				X				X
N°7				X			X	
N°8		X				X		

...

...

observations et qui ont, entre elles, un certain niveau de corrélation. Ces axes, dont le nombre n'est pas limité, sont classés en fonction de leur inertie c'est-à-dire de la part d'explication de la variance qu'ils représentent.

On choisit alors un système de 2 axes de référence (généralement les deux premiers axes factoriels en raison de leur inertie plus élevée) dont la réalisation est due à des variables particulières dont on connaît la contribution propre ; la représentation des observations est donc réalisée sur un plan qui visualise simultanément les variables prépondérantes et les observations. Les regroupements entre variables et entre observations sont confirmés par une méthode statistique complémentaire, la classification hiérarchique ascendante, qui tient compte des huit premiers axes factoriels.

La figure 4 nous rapporte, sur les facteurs 1 et 2, la représentation plane des variables "objectifs" de l'enquête sur la pathologie respiratoire. L'axe des X (inertie 20,8) illustre une opposition entre des variables affectées d'un chiffre 1 (= absence de ..) ou 2 (= peu de ..) et des variables affectées d'un chiffre 3 (= beaucoup de..) ou 4 (= très nombreux ..). Le fait que les variables affectées du même chiffre soient très proches confirme l'existence de corrélations entre ces variables, qui varient donc dans le même sens : ainsi peut-on dire que les élevages dont les porcs n'ont pas de lésions de pneumonie n'ont pas non plus de lésions de rhinite ni de lésions de polysérite, et ont un âge d'abattage peu élevé (c'est-à-dire que la croissance des animaux est bonne). Il en est de même pour les 3 autres groupes. L'axe des Y (inertie 10,9) oppose deux groupes de variables que la seule projection sur l'axe des X confondrait : le groupe 3 et le groupe 4. Il semble que cette séparation soit due à la présence d'abcès pulmonaires (ABC 2) sur les porcs du groupe 4, le distinguant ainsi du groupe 3 et des autres groupes (ABC 1 = pas d'abcès).

La figure 5 représente les observations (c'est-à-dire les élevages) sur le système d'axes que nous venons de présenter. Les élevages se classent selon leur degré de réalisation des objectifs de l'enquête, leur situation sur le plan traduisant leur profil particulier. Ce profil peut être résumé par un symbole selon l'appartenance à un groupe particulier (groupe 1 : pas de problèmes sanitaires, bonne croissance ; .. groupe 5 : lésions les plus importantes, mauvaisé croissance).

Ainsi, au terme de cette phase de l'exploitation des données, est-il possible d'attribuer à chaque élevage un symbole résumant sa situation en regard des objectifs de l'enquête.

#### Recherche des "facteurs de risque" :

Il s'agit de rechercher, parmi les variables causales (6 variables, soit 436 informations), celles qui sont susceptibles de prédire l'appartenance des élevages à l'un ou à l'autre des groupes de résultats que nous venons de présenter.

Dans un premier temps, le tableau de données 103 x 436 est réduit en fonction de l'étude des corrélations et des contingences entre variables : seules sont conservées les variables causales ayant une relation avec le problème étudié, soit 60 environ. Par la suite, des séries d'analyses factorielles sont réalisées jusqu'à ce qu'apparaissent les variables qui, à elles seules, résument l'essentiel de l'explication de la répartition des élevages selon leurs résultats ou leur groupe d'appartenance vis-à-vis des objectifs de l'enquête.

Ces variables (au nombre de 11 dans l'enquête considérée-voir tableau n°1) se présentent comme des facteurs de risque de la pathologie respiratoire : leur agencement, sous la forme d'une représentation plane (figure 6) réalise une "carte de risque" sur laquelle les élevages se positionnent en fonction de leurs résultats prévisibles (figure 7).

FIGURE N°4 - ANALYSE DES OBJECTIFS - POSITION DES VARIABLES  
(ENQUÊTE "PATHOLOGIE RESPIRATOIRE 1980")

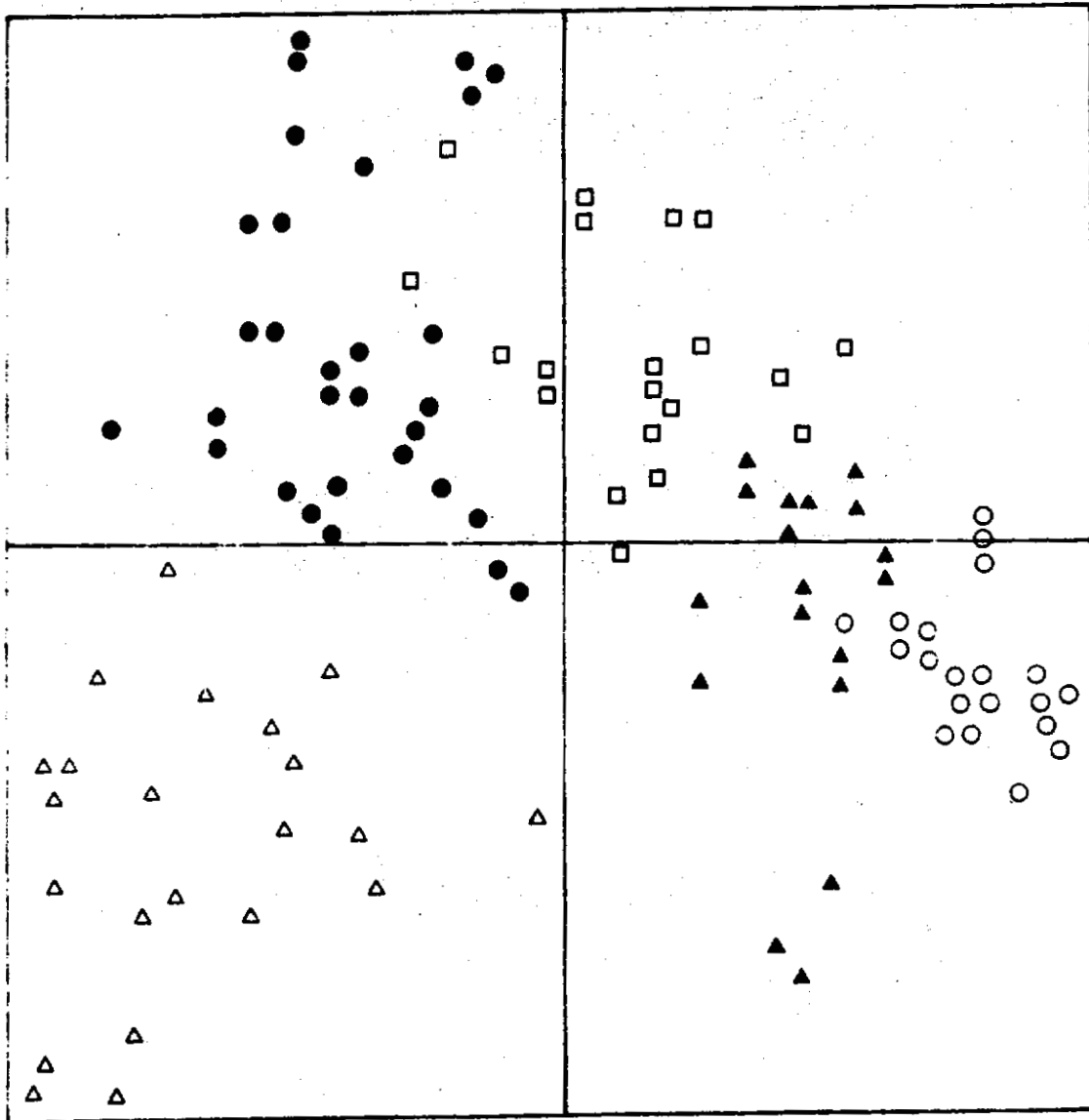
ANAFAC N°1 - Plan 1-2 Graphe n°1

	10,9
AGE3 GRO3 SER3 PNE3	ABC1 AGE2 SER1
	20,8
PNE4 SER4 GRO4 ABC2 AGE4	GRO2 PNE2 SER2 PNE1 GRO1 AGE1

- PNE 1 → 4 : note lésionnelle de pneumonie
- GRO 1 → 4 : note lésionnelle de rhinite
- SER 1 → 4 : note lésionnelle de polysérite
- AGE 1 → 4 : âge à l'abattage à 100 kg
- ABC 1 : pas d'abcès pulmonaires
- ABC 2 : abcès pulmonaires

FIGURE N°5 - ANALYSE DES OBJECTIFS - POSITION DES OBSERVATIONS  
(ENQUÊTE "PATHOLOGIE RESPIRATOIRE 1980")

ANAFAC N°1 - Plan 1-2 Graphe n°2



Légende

Groupe 1	...	○	pas de problèmes respiratoires	
"	2	...	▲	très peu de problèmes respiratoires
"	3	...	□	peu de problèmes respiratoires
"	4	...	●	beaucoup de problèmes respiratoires
"	5	...	△	troubles respiratoires nombreux et graves



TABLEAU N°1 - Les 11 variables causales prépondérantes dans l'estimation

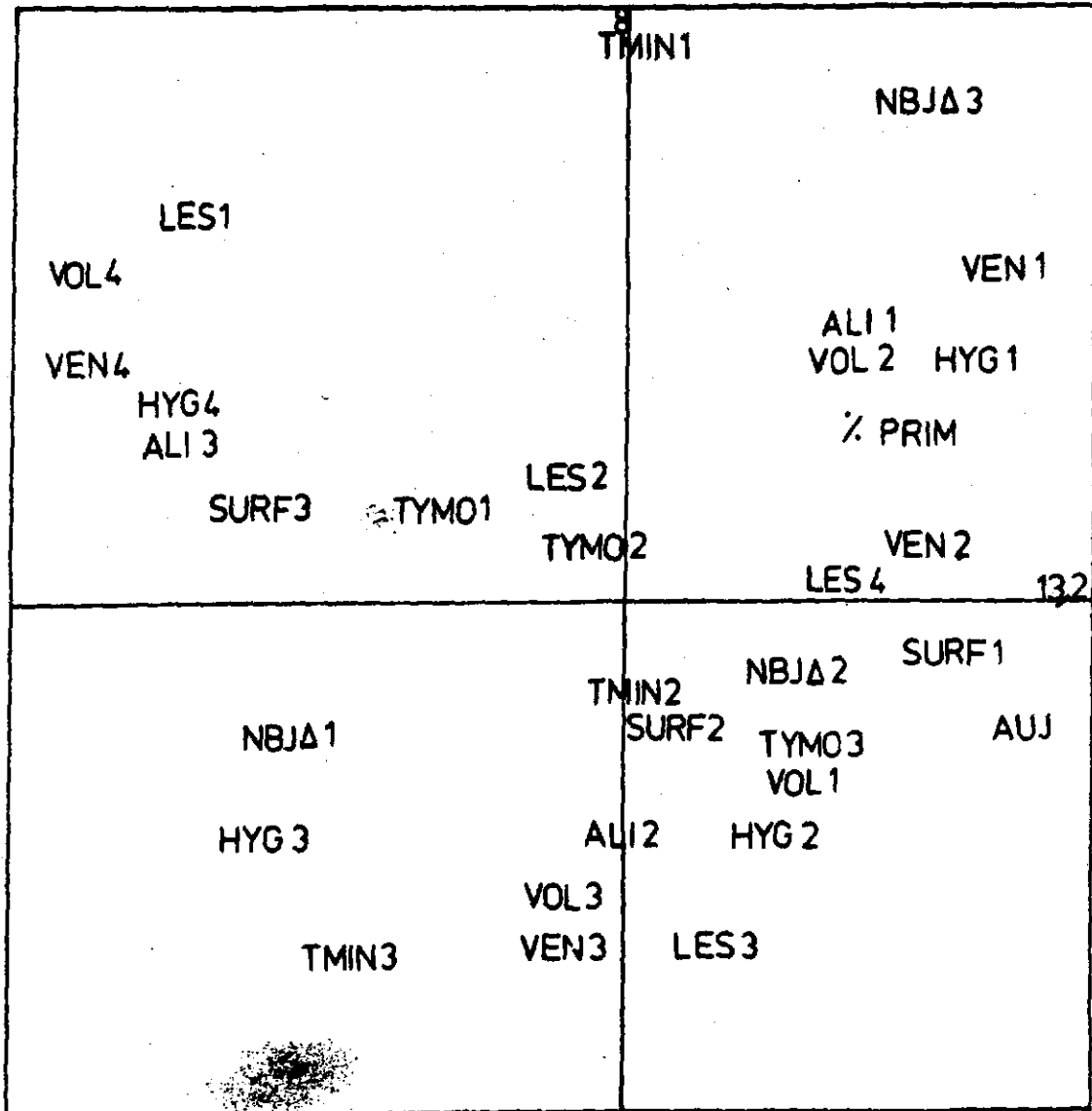
de l'impact de la pathologie respiratoire en engraissement.

Tableau de mise en classes

Les variables (et abréviations)	Numéro de classe	1	2	3	4
Volume par porc en engraissement (m <sup>3</sup> ) [VOL]		< 2,4	2,4 à 2,70	2,71 à 3	> 3
Nombre de jours pendant lesquels les écarts de température ambiante sont > à 6° (sur 100 jours) [NBJ Δ]		< 15 J	15 à 30	> 30	/
Moyenne des températures minimales au cours des 2 premiers mois [TMIN]		- 15°	15° à 19°	19°	/
Surface disponible au sol par porc (m <sup>2</sup> ) [SURF]		< 0,60	0,60 à 0,70	> 0,70	/
Qualité de la ventilation (NH <sub>3</sub> , courant d'air) note [VENT]		6	7 à 8	9	> 9
Pourcentage de primipares [% PRIM]		≤ 30%	> 30%	/	/
Nombre de types morphologiques pour 30 porcs [TYMO]		< 6	6 à 10	> 10	/
Episode de maladie d'Aujeszky en engraissement [AUJ]		oui	non	/	/
Conduite hygiénique (nett., désinf., VS) [HYG]		v.s. > 8 j	5 j v.s. à 8 j	nett. ± désinf.	conduite en continu
Niveau alimentaire (14 repas/sem.) ration quotidienne à 35 kg + 50 kg [ALIM]		< 1,8	1,8 à 2,1	> 2,1	/
Qualité sanitaire porcelet (8-11 sem.) (lésions poumons + péricarde + groin) [LES]		pas de lésions	quelques lésions	lésions nombreuses	lésions très nombreuses

FIGURE N°6 - ETUDE DES CAUSES - POSITION DES VARIABLES\*  
(ENQUÊTE "PATHOLOGIE RESPIRATOIRE 1980")

ANAFAC N°2 - Plan 1-2 Graphe n°1

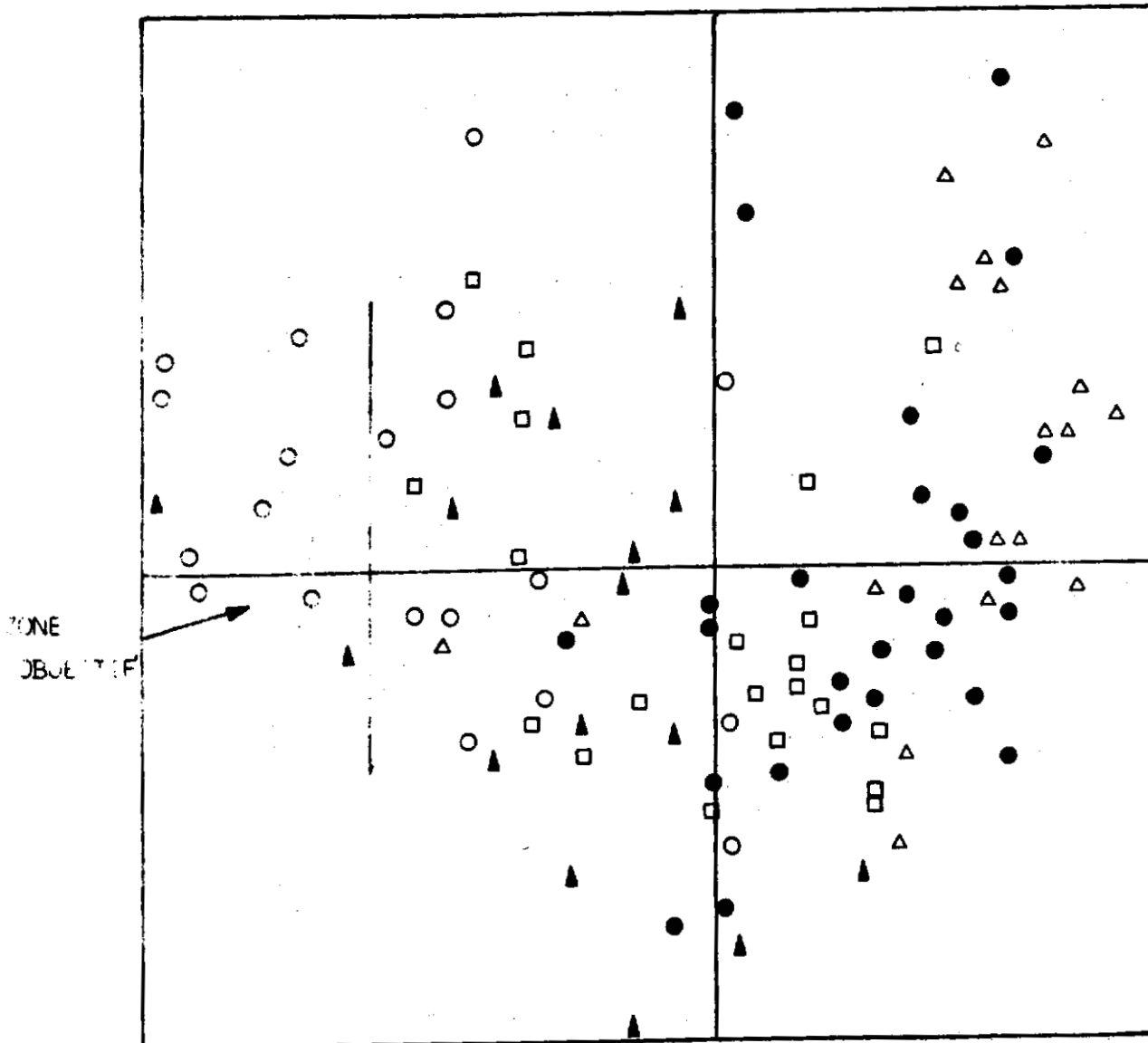


\* voir légende au tableau n°1

# FIGURE N°7 - RÉPARTITION DES ÉLEVAGES SELON LE RISQUE D'APPARITION DES TROUBLES RESPIRATOIRES

(ENQUÊTE "PATHOLOGIE RESPIRATOIRE 1980")

ANAFAC N°3 Plan 1-2 Graphe n°3



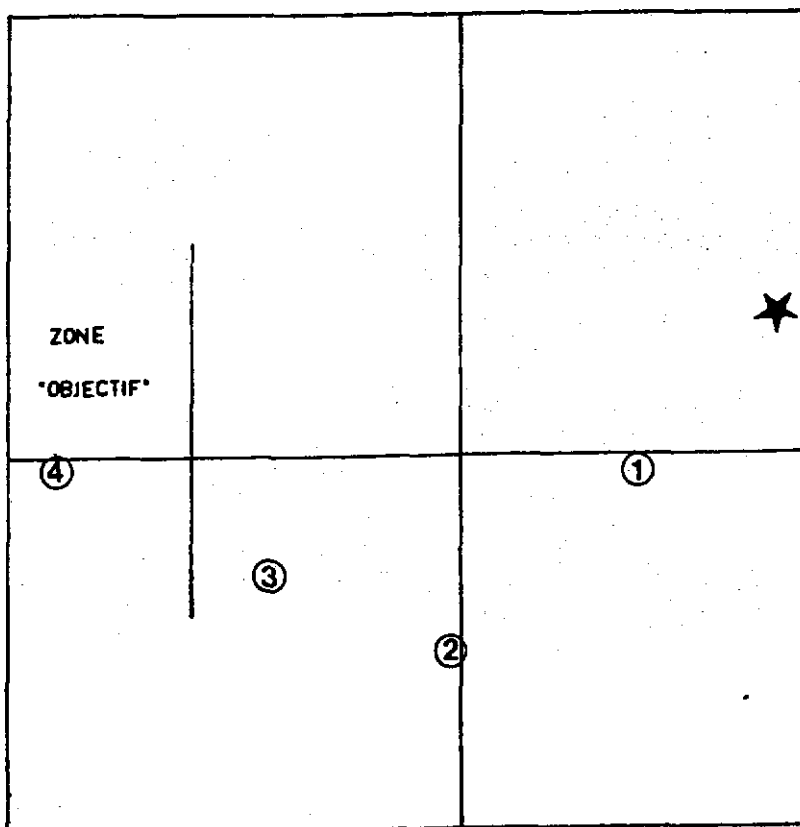
- Legende
- |         |     |     |                                |   |
|---------|-----|-----|--------------------------------|---|
| Group 1 | ... | ○   | pas de problèmes respiratoires |   |
| "       | 2   | ... | ▲                              | très peu de problèmes respiratoires       |
| "       | 3   | ... | □                              | peu de problèmes respiratoires            |
| "       | 4   | ... | ●                              | beaucoup de problèmes respiratoires       |
| "       | 5   | ... | △                              | troubles respiratoires nombreux et graves |

## L'ANALYSE FACTORIELLE ET LA PREVENTION DES MALADIES

En réalisant une simplification du contenu du tableau de données et une interprétation simplifiée des résultats d'une enquête épidémiologique, l'analyse factorielle et sa principale retombée, la "carte des facteurs de risque", permet à des utilisateurs confrontés à des situations pathologiques sur le terrain de disposer d'informations de référence ainsi que d'une méthode de travail : en modifiant la situation d'un élevage en regard des principaux facteurs de risque, on a la possibilité de soustraire un élevage à une pathologie particulière. Cette démarche, qui peut mettre en oeuvre une simulation de situation (figure 8) est à la base de l'établissement d'un programme de prévention efficace.

FIGURE N°8 - POSITION D'UN ÉLEVAGE SUR LA "CARTE" DES FACTEURS DE RISQUE ET SIMULATION D'UNE EVOLUTION POSSIBLE (PROGRAMME DE PRÉVENTION)

ANAFAC N° 2 - Graphe n°6



### REFERENCE BIBLIOGRAPHIQUE

Benzecri J.P. - L'analyse des données I - La taxinomie (631 p.) - II - L'analyse des correspondances (616 p.) - 1976 Dunod. ed. Paris.