

Symbolic Analysis of Hierarchical-Structured Data. Application to Veterinary epidemiology

C. Fablet¹, E. Diday², S. Bougeard¹, C. Toque³ & L. Billard⁴

¹ French agency for food, environmental and occupational health safety (Anses), France

² University of Paris Dauphine, France

³ SYROKKO, France

⁴ University of Georgia, Athens, USA



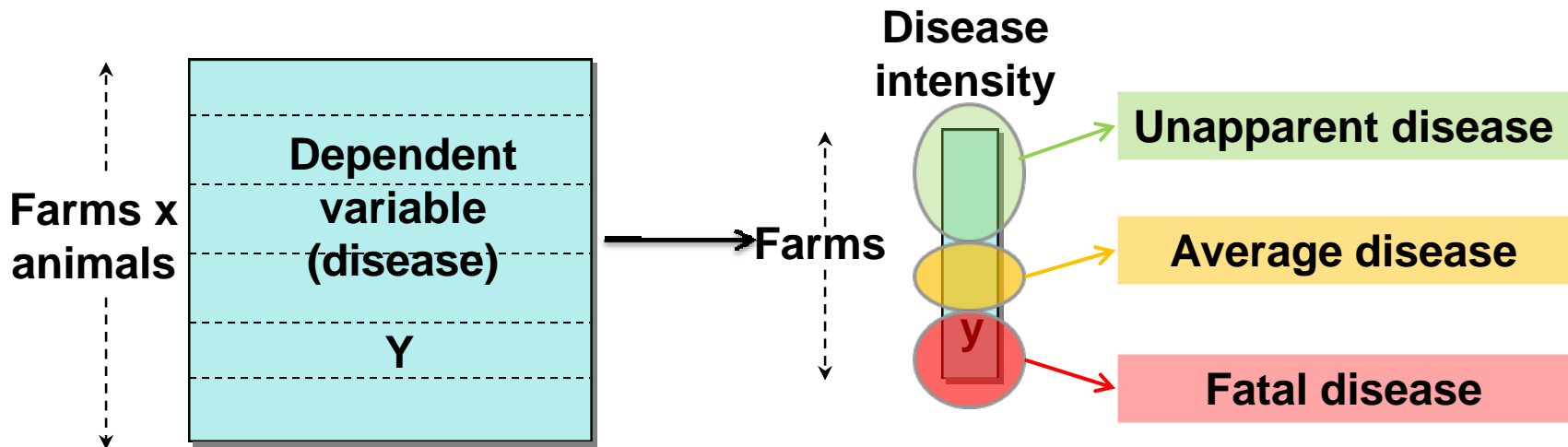
19th International Conference on Computational Statistics, Paris, August 22-27, 2010

Context of veterinary epidemiological surveys

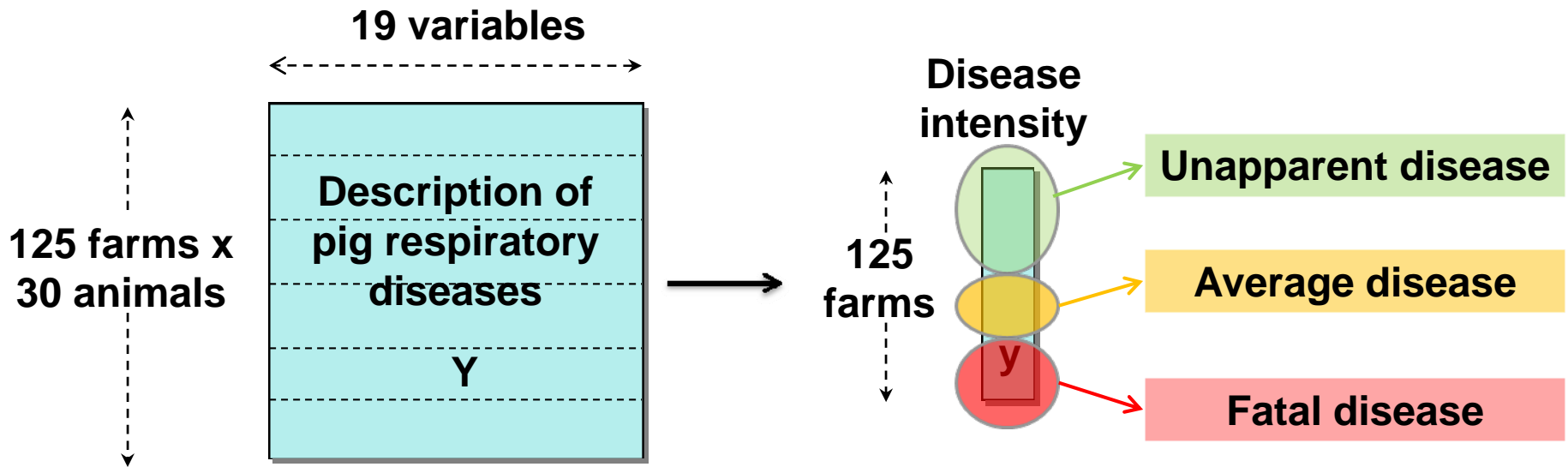
Statistical issue

1. Description of the relationships between the dependent variables → variable selection,
2. Summary of the dependent variables into an overall single variable (*i.e.* the disease),

... with a hierarchical structure of observations (P animals each within N farms).



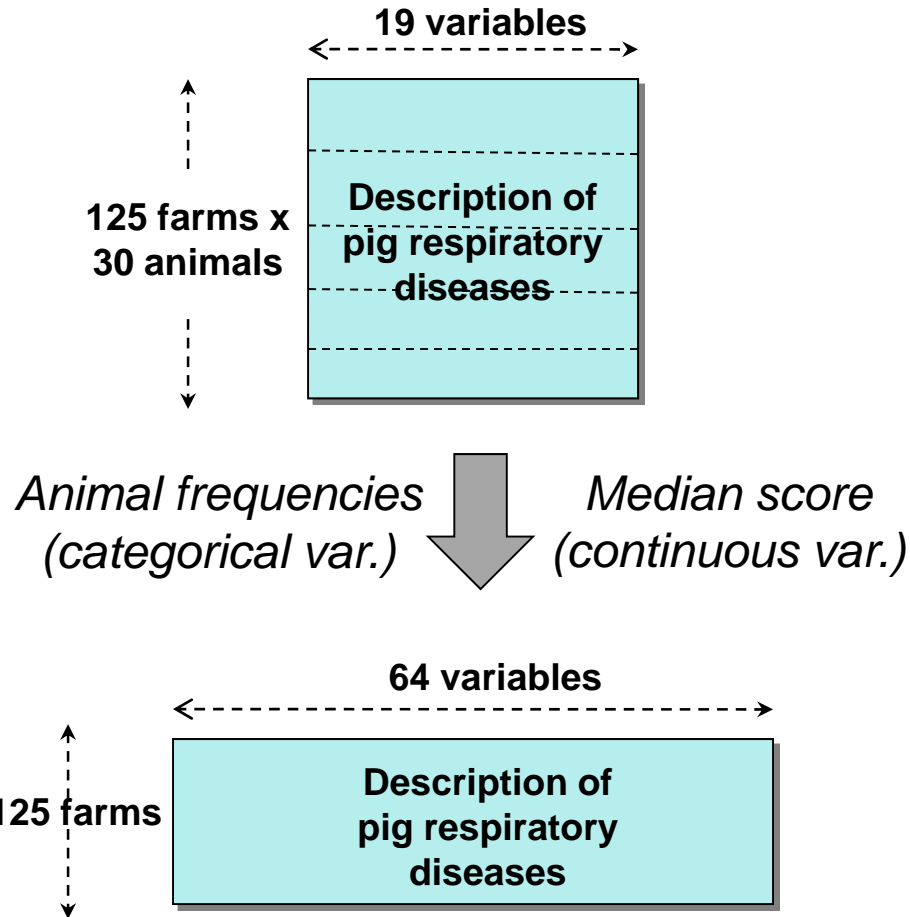
Dataset: Study of pig respiratory diseases



- Pneumonia (0 → 28), pleuritis (0 → 4),
- Lung abscess (0/1), lung nodules (0/1), healing from pneumonia (0/1),
- Hypertrophy of lung lymph nodes (0 → 3), pericarditis (0/1),
- Frequency of coughs at 16 and 22 weeks of age.

Step 1: Variable synthesis

Classical procedure



Symbolic procedure

- Categorical variable: histogram of the frequencies based on 30 animals,
- Continuous variable: histogram which keep the data variation.

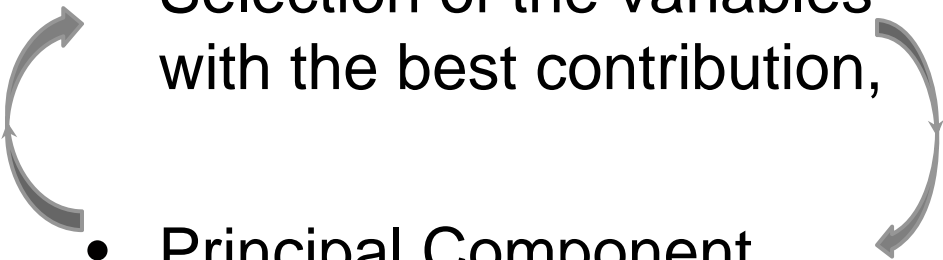
Step 1: Variable synthesis (symbolic results)



SYR software with the TABSYR & STATSYR modules

Step 2: Variable selection

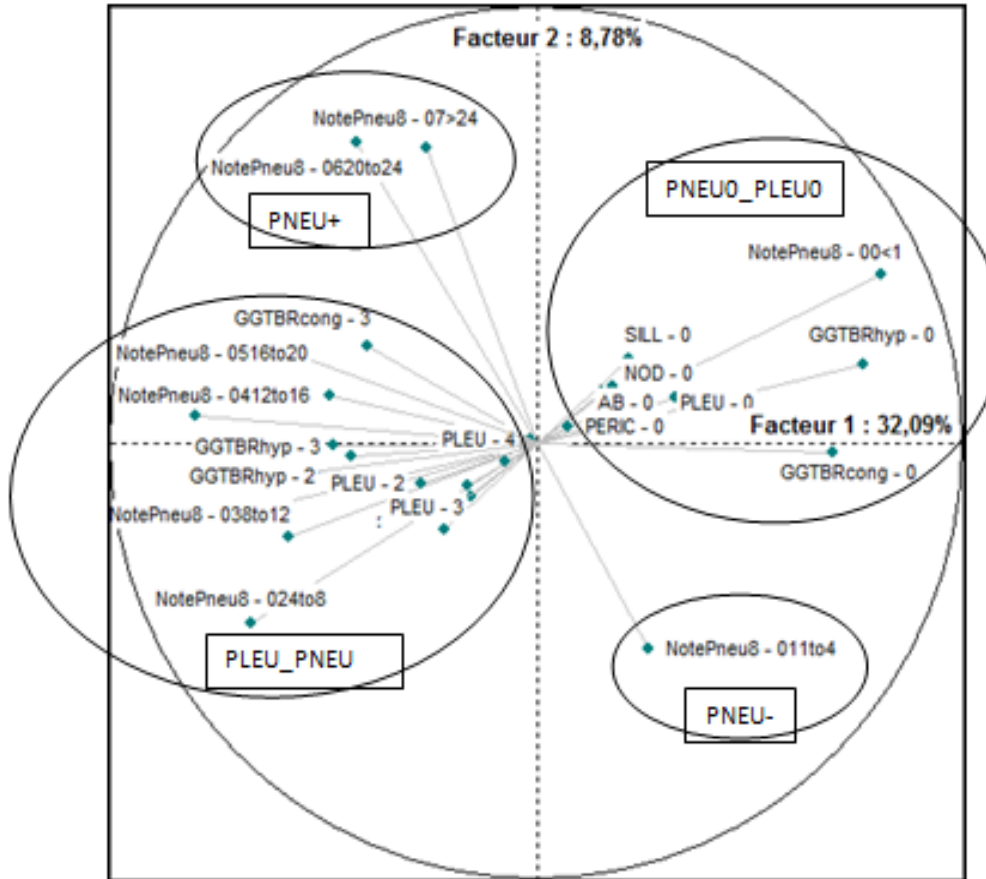
Classical procedure

- Principal Component Analysis of the 64 variables,
 - Selection of the variables with the best contribution,
 - Principal Component Analysis of the selected variables.
- 

Symbolic procedure

- Symbolic Principal Component Analysis of the 19 variables,
 - ‘Global’ variable selection (best var. contribution)
 - ‘Quadrants’ variable selection (best var. correlation),
 - Final symbolic PCA representation of the selected ‘bins’ variables.

Step 2: Variable selection (symbolic results)



Symbolic PCA of the 8 'bins' selected var.

- Var. group PNEU+: severe pneumonia,
- Var. group PLEU_PNEU: average level of pleuritis and pneumonia,
- Var. group PLEU0_PNEU0: few lung lesions,
- Var. group PNEU-: light pneumonia lesions.

SYR software with the ACPSYR module

Step 3: Individual clustering

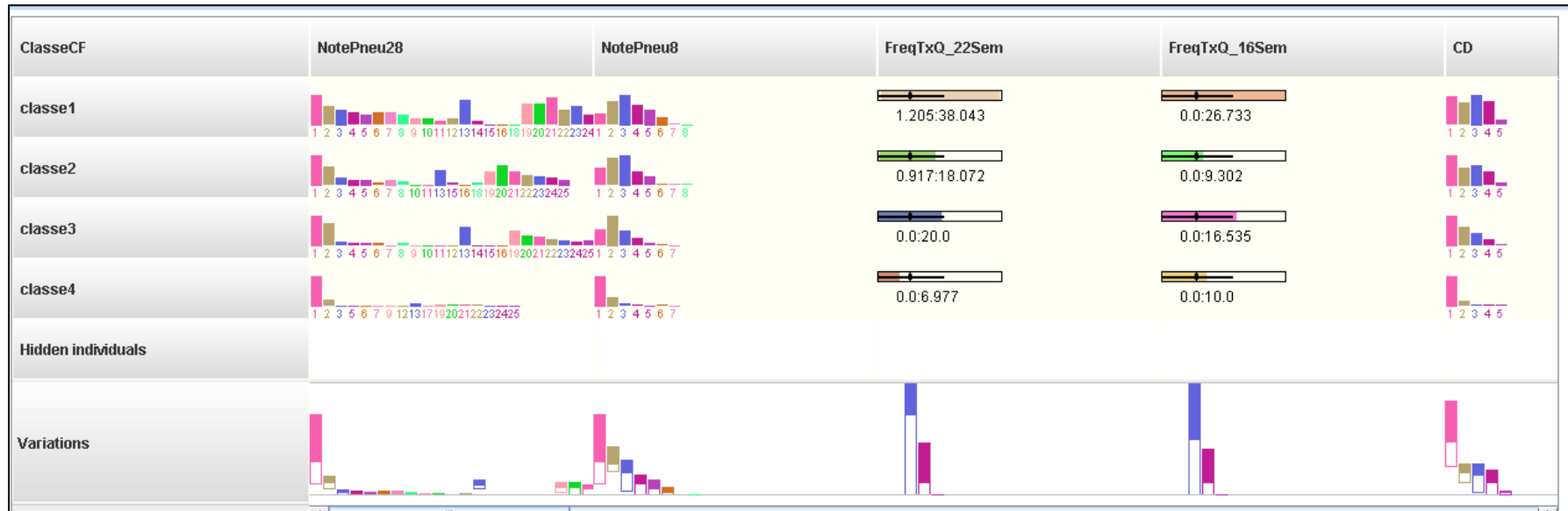
Classical procedure

- Hierarchical Ascendant Classification (Ward criterion)
- Cluster description
 - Comparison of the variable means (& standard deviations) of each cluster, with the variable means on the whole sample.

Symbolic procedure

- Symbolic partitioning (inertia criterion)
- Cluster description
 - Variables sorted in order of overall discriminant power,
 - Cluster description with the most discriminant variables (or variable modalities).

Step 3: Individual clustering (symbolic results)



CATEGORIES WHERE INDIVIDUAL classe2 IS THE HIGHEST Close

category	variable	proba/mean	proba	range	range map	opposite individual
21	NotePneu28	4.000000	0.010000	0.010000		classe1
24	NotePneu28	4.000000	0.010000	0.010000		classe1
07>24	NotePneu8	4.000000	0.010000	0.010000		classe1
3	PLEU	3.294118	0.140000	0.130000		classe1
1	AB	3.111111	0.070000	0.070000		classe4
1	NOD	2.909091	0.080000	0.070000		classe1
2	PLEU	2.857143	0.250000	0.230000		classe4
1	PERIC	2.666667	0.080000	0.070000		classe1
1	PLEU	2.193548	0.170000	0.150000		classe4
8	NotePneu28	1.777778	0.040000	0.040000		classe4
4	NotePneu28	1.777778	0.120000	0.110000		classe4
1	GGTBRhyp	1.731959	0.420000	0.360000		classe4
1	SILL	1.611940	0.270000	0.180000		classe4
6	NotePneu28	1.500000	0.060000	0.050000		classe4
024to8	NotePneu8	1.493976	0.310000	0.270000		classe4
1	APD	1.411765	0.180000	0.140000		classe4

SYR software with the CLUSTSYR module

Conclusion & perspectives

Conclusion

- Symbolic analysis to process hierarchical-structured data without reducing information,
- Relevant and useful methods for veterinary epidemiological surveys (competes with GEE including a random measurement effect),
- Available software (SYR).

Perspectives

- Other symbolic methods available for various aims,
- Extension to multiblock modelling (hierarchical-structured observations and variables).

Symbolic Analysis of Hierarchical-Structured Data. Application to Veterinary epidemiology

C. Fablet¹, E. Diday², S. Bougeard¹, C. Toque³ & L. Billard⁴

¹ French agency for food, environmental and occupational health safety (Anses), France

² University of Paris Dauphine, France

³ SYROKKO, France

⁴ University of Georgia, Athens, USA



19th International Conference on Computational Statistics, Paris, August 22-27, 2010