

# **Journée d'Etudes**

## **L'analyse statistique des données catégorielles en Sciences Humaines et Sociales**

### **Présentation générale**

Jean-Marc Bernard

Denis Corroyer

Yannick Savina

Laboratoire de Psychologie Environnementale  
CNRS UMR 8069 & Université Paris 5

Boulogne-Billancourt, France

6 Juin 2006

# Merci d'être venus!

## □ Quatrième éditions de ces Journées d'études

L'analyse statistique des  
données catégorielles en  
Sciences Humaines et Sociales

## □ Editions précédentes

- 2003 - Dynamique des Proximités Urbaines
- 2004 - Trajectoires de Vie et Rapport à l'Environnement
- 2005 - Risques et Environnement: Les outils des Sciences Sociales

# Laboratoire de Psychologie Environnementale

□ **Unité CNRS UMR-8069**, Section 39, SHS

□ **Directeur**: Michel-Louis Rouquette

□ **Axes de recherche**:

- Axe 1 - Cadres de vie: Qualité environnementale, besoins spécifiques et bien-être  
Resp.: Gabriel Moser
- Axe 2 - Environnements critiques: pratiques, conduites et représentations  
Resp.: Michel-Louis Rouquette)
- Axe 3 - Evolution de la relation personne - environnement: Développement individuel, transitions et phases critiques Resp.: Alain Legendre & Eugénia Ratiu
- Axe 4 - Développements méthodologiques - Conception de méthodes et d'outils pour l'analyse statistique des données Resp.: Jean-Marc Bernard & Denis Corroyer

# Axe “Développements méthodologiques”

## □ Besoins méthodologiques

- Enquêtes
- Questionnaires, données d'observation souvent catégorielles
- Données expérimentales

## □ Thèmes de recherche

- Analyse des données (Anal. Corresp., ACM)
- ANOVA
- Inférence bayésienne
- Analyse des tableaux de contingence
- Analyse implicative
- Logiciels statistiques
- Système Expert en Statistique

# Thème de cette journée

## □ Données catégorielles

- Données d'observation (questionnaires)
- Données expérimentales

## □ Tableaux de contingence

- TC simple: deux variables  $2 \times 2$ ,  $J \times K$
- TC multiples: nombreuses variables  $J$ ,  $K$ ,  $L$ , *etc.*
- Questions symétriques: association ou liaison
- Questions asymétriques: comparaison, implication
- Recherche des proximités / dissimilarités entre individus ou entre modalités

# Buts d'une analyse statistique

## □ Résumer les données

- Définition d'indices résumés pertinents
- Réduction de la dimensionnalité
- Décomposition les données en aspects essentiels et secondaires

## □ Généraliser les conclusions

- Si on avait recueilli 1 ou 2 observations supplémentaires (stabilité)
- Si on avait recueilli 2 fois plus de données (problème d'inférence prédictif)
- Si on avait pu observer la population entière (problème d'inférence paramétrique)

## □ Echantillon et population

# Description et Inférence

## □ Méthodes descriptives

- Tris croisés
- Indices de liaison entre modalités, entre variables
- Analyse des correspondances, simple et multiple
- Analyse booléenne, analyse implicative
- Classification, Analyse des similitudes

## □ Méthodes inductives

- Méthodes fréquentistes (tests)
- Méthodes bayésiennes
- Modélisation log-linéaire

# Aspects pratiques

## □ Déroulement de la journée

- 4 exposés le matin
- 4 exposés l'après-midi
- Chaque exposé: 45 minutes, avec 5-10 minutes de discussion
- Fin prévue à 17h30. Mais discussion générale en plus pour ceux qui seront encore vaillants

## □ Pauses Café

- De 10h45 à 11h15
- De 15h30 à 16h00
- Cafétéria: RdC

## □ Déjeuner

- Restaurant universitaire (CROUS)  
Tickets en vente à la pause du matin
- Restaurants extérieurs  
Liste disponible



## Avant l'entrée en piste ...

Henry ROUANET

Claude FLAMENT

Jean-Pierre BENZECRI