

# Introduction aux modèles à variables latentes

Gilbert Saporta  
Conservatoire National des Arts et Métiers

[gilbert.saporta@cnam.fr](mailto:gilbert.saporta@cnam.fr)  
<http://cedric.cnam.fr/~saporta>

- Expliquer des phénomènes observés par l'intervention de facteurs cachés peut relever des superstitions, croyances, magies...
- En sciences sociales, de nombreux concepts ne peuvent être observés directement comme l'engagement religieux...
  - Mais un haut niveau d'engagement conduit à une fréquentation élevée des lieux de culte, à prier plus souvent, à se conduire de façon plus conforme à l'orthodoxie...

- Si chaque **variable manifeste** est causée par une ou plusieurs variables latentes, les covariations entre variables manifestes s'expliquent par la relation de chacune avec les **variables latentes**
- **Hypothèse fondamentale**: les variables manifestes sont indépendantes conditionnellement aux variables latentes

# Les différents modèles

	Variables latentes	
Variables manifestes	<b>qualitative</b>	<b>quantitatives</b>
<b>qualitatives</b>	Classes latentes	Traits latents
<b>quantitatives</b>	Profils latents	Analyse factorielle

**Traits latents** : en psychométrie Item Response Theory  
**cas particulier** : modèles de Rasch

**Profils latents** : modèles de mélange particuliers

# Comparaison avec les méthodes exploratoires

	Variables latentes	
Variables manifestes	<b>qualitative</b>	<b>quantitatives</b>
<b>qualitatives</b>	Classes latentes	Traits latents
<b>quantitatives</b>	Profils latents	Analyse factorielle

modèles

	Composantes ou variables synthétiques	
Variables observées	<b>qualitative</b>	<b>quantitatives</b>
<b>qualitatives</b>	classification	ACM
<b>quantitatives</b>		ACP

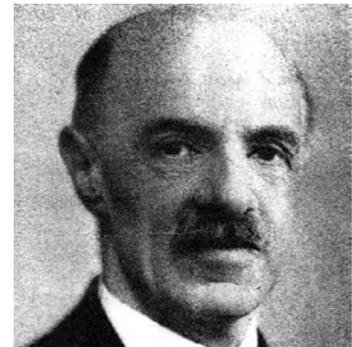
« analyse des données »

# 1. L'analyse factorielle

- « Factor analysis » ou analyse en facteurs communs et spécifiques
- Charles Spearman (1904) : **modèle unifactoriel**

$$x_i = \lambda_i f + u_i \quad i = 1, \dots, p$$

- $f$  facteur commun,  $u_i$  facteur spécifique,  $\lambda_i$  saturation ou « factor loading »



# Exemple

- Corrélation entre résultats scolaires  
Classics, english, french

$$R = \begin{pmatrix} 1 & 0.83 & 0.78 \\ 0.83 & 1 & 0.67 \\ 0.78 & 0.67 & 1 \end{pmatrix} \quad \begin{cases} x_1 = 0.983 f + u_1 \\ x_2 = 0.844 f + u_2 \\ x_3 = 0.794 f + u_3 \end{cases}$$

## L.L. Thurstone: modèle plurifactoriel

$$x_i = \sum_{j=1}^k \lambda_{ij} f_j + u_i \quad i = 1, \dots, p$$



- Modèle de régression multiple multilinéaire sur variables inobservables
- Hypothèses :
  - $x_i$  centrés (réduits)
  - $f_j$  centrés réduits indépendants entre eux
  - $u_i$  centrés indépendants entre eux  $V(u_i) = \psi_i$
  - $u_i$  indépendants des  $f_j$

Note: Garnett (1919) précéda Thurstone

# Formulation matricielle

$$\mathbf{x} = \mathbf{\Lambda f} + \mathbf{u}$$

- Décomposition de la variance

$$V(x_i) = \sigma_i^2 = \sum_{j=1}^k \lambda_{ij}^2 + \psi_i = h_i^2 + \psi_i$$


Communauté + spécificité

$$\mathbf{\Sigma} = \mathbf{\Lambda \Lambda'} + \mathbf{\Psi}$$

# Indépendance conditionnelle

- Matrice de variance des manifestes et des facteurs

$$\begin{pmatrix} \Sigma & \Lambda \\ \Lambda' & \mathbf{I} \end{pmatrix}$$

$$\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \quad \Sigma_{11/2} = \Sigma_{11} - \Sigma_{12} (\Sigma_{22})^{-1} \Sigma_{21}$$

$$\Sigma_{x/f} = \Sigma - \Lambda\Lambda' = \Psi$$

**Matrice diagonale: l'AF explique les corrélations**

# Indétermination des facteurs

- Modèle surparamétré
- Invariance par transformation orthogonale

$$\mathbf{x} = \Lambda \mathbf{f} + \mathbf{u} = \Lambda \mathbf{G} \mathbf{G}' \mathbf{f} + \mathbf{u}$$

$$\Sigma = \Lambda \Lambda' + \Psi = \Lambda \mathbf{G} \mathbf{G}' \Lambda' + \Psi$$

- Solution (?): contrainte

$$\Lambda' \Psi^{-1} \Lambda \text{ diagonale}$$

# Cadre gaussien

- Hypothèses:  $\mathbf{f} \sim N_k(\mathbf{0}; \mathbf{I})$      $\mathbf{x} / \mathbf{f} \sim N_p(\Lambda \mathbf{f}; \Psi)$
- Conséquences:

$$\mathbf{x} \sim N_p(\mathbf{0}; \Lambda \Lambda' + \Psi)$$

$$\mathbf{f} / \mathbf{x} \sim N_k\left(\left(\mathbf{I} + \Lambda' \Psi^{-1} \Lambda\right)^{-1} \Lambda' \Psi^{-1} \mathbf{x}; \left(\Lambda' \Psi^{-1} \Lambda + \mathbf{I}\right)^{-1}\right)$$

- Les facteurs sont indépendants *a posteriori* également si:

$$\Lambda' \Psi^{-1} \Lambda \text{ diagonale}$$

# Estimation des scores ou mesure en facteurs

- **Scores de Thompson**

Espérance de  $f$  *a posteriori*

$$\hat{\mathbf{f}} = \left( \mathbf{I} + \mathbf{\Lambda}'\mathbf{\Psi}^{-1}\mathbf{\Lambda} \right)^{-1} \mathbf{\Lambda}'\mathbf{\Psi}^{-1}\mathbf{x}$$

Rem: estimateur ridge

- **Scores de Bartlett**

$\mathbf{x} = \mathbf{\Lambda}\mathbf{f} + \mathbf{u}$   $f$  coefficients d'un modèle linéaire!

Moindres carrés généralisés

$$\hat{\mathbf{f}} = \left( \mathbf{\Lambda}'\mathbf{\Psi}^{-1}\mathbf{\Lambda} \right)^{-1} \mathbf{\Lambda}'\mathbf{\Psi}^{-1}\mathbf{x}$$

# Degré de liberté

- Nombre d'équations  $\Sigma = \Lambda\Lambda' + \Psi$  :  $p(p+1)/2$
- Nombre de contraintes  $\Lambda'\Psi^{-1}\Lambda = \mathbf{D}$  :  $k(k-1)/2$
- Nombre d'inconnues  $\lambda_{ij}$  et  $\psi_i$  :  $pk+p$

$$s = \frac{p(p+1)}{2} + \frac{k(k-1)}{2} - p(k+1) = \frac{1}{2} \left[ (p-k)^2 - (p+k) \right]$$

- $s=0$  solution unique mais..
- $s<0$  indétermination
- $s>0$  plus d'équations que d'inconnues: le cas utile!  
Représentation parcimonieuse

- $s=0$  solution artificielle ,  $\Psi_i$  peut être négatif (Heywood case)
- Cas exacts:
  - $p=3$   $k=1$
  - $p=6$   $k=3$
  - $p=10$   $k=6$
  - $p=15$   $k=10$
  - $p=21$   $k=15$
  - ...

# Estimation des paramètres du modèle

- **Méthode des facteurs principaux**

- 1ère estimation des  $h_i^2 = 1 - \psi_i = \sum \lambda_{ij}^2$   
par  $R^2(x_i; x_1, \dots, x_{i-1}, x_{i+1}, \dots) \sim R^2(x_i; f_1^j, \dots, f_k)$

D'où estimation de  $\Psi$

- Diagonalisation de R-  $\Psi$  Matrice de corrélation  
«réduite »

$$\mathbf{R} - \Psi \simeq \Lambda \Lambda'$$

- Réestimation de  $\Psi$  etc.

- **Maximum de vraisemblance**

# Rotation des facteurs

- Indétermination:  $\mathbf{GG}'=\mathbf{I}$   $\Delta=\Lambda\mathbf{G}$
- Recherche de « structures simples »
  - Chaque variable corrélée avec peu de facteurs
  - Chaque facteur avec peu de variables
- Rotation « varimax »
  - Maximise la somme des variances des carrés des saturations intra-colonnes, normalisée par la communauté

$$\sum_{i=1}^p V_j \left( \frac{\delta_{ij}^2}{h_i^2} \right)$$

- Nombreuses autres méthodes
  - **Orthogonales**: quartimax, orthomax, parsimax
  - **Obliques**: biquartimin, oblumin, quartimin ...  
fournit des facteurs corrélés après rotation.  
« *Oblique rotations often produce more useful patterns than do orthogonal rotations. However, a consequence of correlated factors is that there is no single unambiguous measure of the importance of a factor in explaining a variable* »

# Un exemple: enquête « tolérance », BVA ,1994



2 France

Un sondage B.V.A.

## Les Français sont-ils tolérants ?

Enfin un sondage surprenant : les Français rejettent le racisme et plébiscitent les échanges culturels. Ils acceptent sans états d'âme les minorités et leurs modes de vie, que celles-ci soient raciales, politiques ou culturelles. Nos questions dessinent une vaste zone de tolérance mais aussi un terrain de débat, où la France se partage. Sur le vote des Européens, par exemple, le libre-échange ou même l'antiaméricanisme. Sur les vingt et une propositions qui nous ont servi d'indicateurs, seules quatre suscitent un réflexe de fermeture. Mieux vaut ne pas fumer dans les lieux publics et enlever son voile islamique à l'école.

Décembre 1994

Les étrangers en provenance de l'Union Européenne et vivant en France depuis 5 ans ou plus devraient avoir le droit de vote aux élections municipales

Il faut ouvrir des points de distribution gratuite de seringues propres pour les toxicomanes

Les Américains nous inondent de leurs films et de leurs programmes de télévision pour assurer leur impérialisme culturel

Pour combattre le chômage, il faut d'abord protéger les producteurs français en interdisant l'accès des produits en provenance des pays à bas salaires.

Les mariages entre gens de races différentes sont à éviter

Les magazines pornographiques et les films X offrent des distractions inoffensives à ceux et celles qui aiment cela

Il faut interdire le port du voile islamique dans les écoles publiques

Jacques Toubon a raison de vouloir éliminer les mots étrangers à la radio et à la télévision ou dans les publicités

Les immigrés rentrent notre pays grâce à leur travail

La liberté d'expression ne devrait pas être reconnue aux extrémistes des deux bords.

Les accords de libre-échange, comme le Gatt sont une source de prospérité réciproque

L'homosexualité est un mode de vie que la société devrait accepter

La police devrait avoir le droit de pratiquer des contrôles d'identité à son initiative

La culture française s'enrichit de ses échanges avec les cultures étrangères

On devrait interdire complètement de fumer dans les lieux publics

Les femmes devraient revenir à leur rôle traditionnel dans la société

Les marques d'alcool et de tabac ne devraient pas avoir le droit de faire de la publicité

A la radio, on entend trop de rock et de rap

Nous devons faire plus d'efforts pour améliorer la situation des minorités en France

On ne devrait pas permettre aux restaurants de fast-food, du genre Mac Donald, de s'établir dans le centre de nos villes

Je ne me sens pas grand chose en commun avec les gens appartenant à d'autres races

Il est normal que les médias parlent de la vie privée des hommes politiques

NUM . IDEN - LIBELLE	EFFECTIF	POIDS	MOYENNE	ECART-TYPE	MINIMUM	MAXIMUM
18 . eurv - vote des etr.europ.	1049	1049.00	2.94	1.61	1.00	5.00
19 . srin - seringues.grat.toxic	1049	1049.00	2.11	1.44	1.00	5.00
20 . usa - imp,rialisme cult.US	1049	1049.00	3.20	1.55	1.00	5.00
21 . nimp - plus de produit étr.	1049	1049.00	3.73	1.39	1.00	5.00
22 . mari - mariages de races #	1049	1049.00	2.19	1.49	1.00	5.00
23 . porn - magaz.porno & film X	1049	1049.00	2.96	1.50	1.00	5.00
24 . voil - interdire le voile	1049	1049.00	4.11	1.32	1.00	5.00
25 . toub - Toubon a raison d'in	1049	1049.00	2.75	1.57	1.00	5.00
26 . immi - les immigrés renforc	1049	1049.00	3.23	1.48	1.00	5.00
27 . lexp - la lib. d'express. n	1049	1049.00	2.52	1.47	1.00	5.00
28 . lech - libr.éch.=prospérité	1049	1049.00	3.00	1.37	1.00	5.00
29 . homo - accept.homosexualité	1049	1049.00	2.38	1.46	1.00	5.00
30 . iden - droit control d'iden	1049	1049.00	3.44	1.56	1.00	5.00
31 . cult - éch.enrich.cult.FRAN	1049	1049.00	1.99	1.22	1.00	5.00
32 . fume - interdt fumer Lx.Pub	1049	1049.00	3.87	1.50	1.00	5.00
33 . femm - retour femm. tradit.	1049	1049.00	2.28	1.56	1.00	5.00
34 . alco - ps de pub alc.&tabac	1049	1049.00	3.13	1.59	1.00	5.00
35 . rock - trop de rock et rap	1049	1049.00	2.88	1.51	1.00	5.00
36 . mino - pour minorités en FR	1049	1049.00	1.85	1.10	1.00	5.00
37 . fast - interd. fast.f.en FR	1049	1049.00	2.45	1.43	1.00	5.00
38 . race - rien en com race étr	1049	1049.00	2.11	1.33	1.00	5.00
39 . hpol - vie privee hom.pol.	1049	1049.00	3.93	1.40	1.00	5.00

MATRICE DES CORRELATIONS

	eurv	srin	usa	nimp	mari	porn	voil	toub	immi	lexp	lech	homo	iden	cult	fume	femm	alco	rock
eurv	1.00																	
srin	.19	1.00																
usa	-.06	-.05	1.00															
nimp	.21	.01	.01	1.00														
mari	.25	.14	-.04	.20	1.00													
porn	.08	.14	-.09	.00	.18	1.00												
voil	.19	.01	.05	.16	.17	.01	1.00											
toub	.18	.20	.05	.16	.27	.18	.08	1.00										
immi	.32	.14	-.08	.26	.26	.01	.20	.10	1.00									
lexp	.03	.02	.05	.15	.14	.04	.06	.05	.10	1.00								
lech	.19	.10	-.02	.18	.10	.05	.08	.01	.16	-.03	1.00							
homo	.19	.28	-.08	.10	.27	.30	.06	.26	.20	.10	.09	1.00						
iden	.20	.09	.03	.26	.26	.07	.15	.21	.21	.12	.09	.25	1.00					
cult	.32	.18	-.10	.16	.21	.08	.08	.11	.28	.03	.20	.20	.20	1.00				
fume	.00	.02	.02	.10	.10	.10	.08	.18	.01	.08	-.03	.10	.15	-.01	1.00			
femm	.13	.23	-.05	.18	.29	.11	-.02	.31	.17	.17	.09	.26	.25	.22	.12	1.00		
alco	-.04	-.03	.06	.06	.05	.14	.07	.11	-.06	.08	-.05	.08	.04	.01	.22	.10	1.00	
rock	.13	.10	.13	.11	.20	.13	.10	.22	.08	.12	.04	.17	.15	.05	.06	.21	.10	1.00
mino	.13	.14	-.11	.01	.11	.02	.07	.04	.18	.00	.15	.13	.09	.20	-.03	.03	-.06	.03
fast	.07	.04	.07	.11	.14	.12	.01	.17	.06	.17	.06	.07	.06	.05	.06	.21	.11	.18
race	.24	.11	-.03	.22	.31	.09	.16	.19	.26	.13	.11	.18	.21	.20	.09	.26	.06	.13

	mino	fast	race
mino	1.00		
fast	.03	1.00	
race	.12	.10	1.00

# ACP

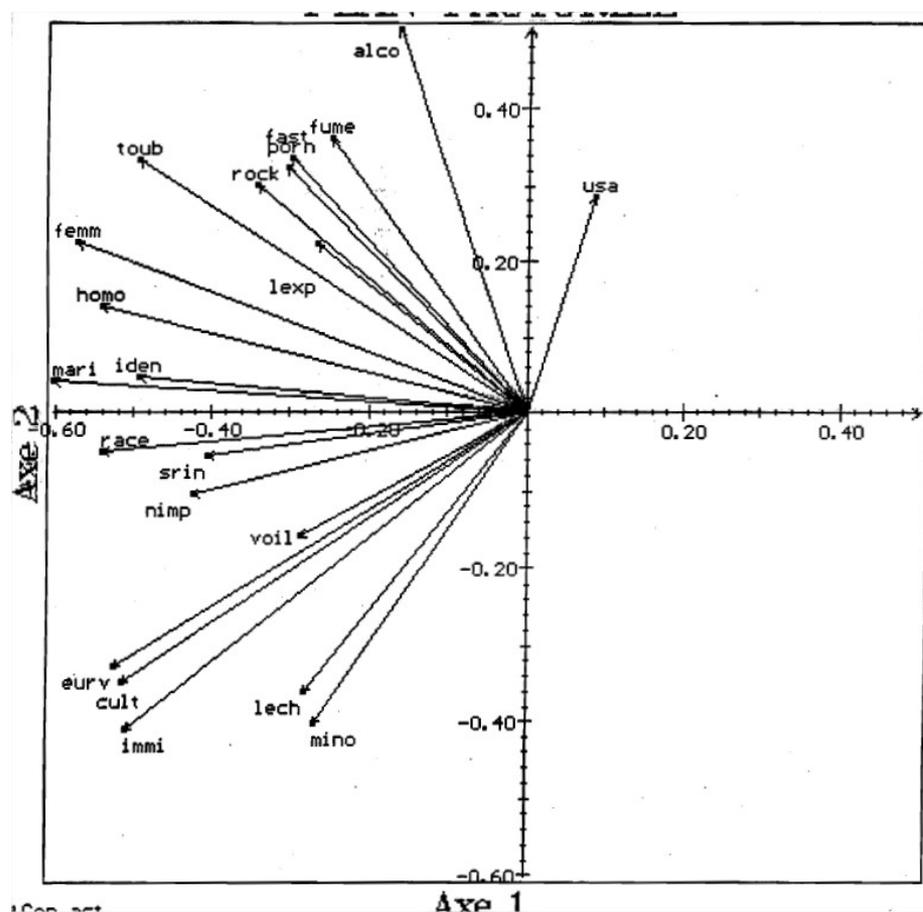
HISTOGRAMME DES 21 PREMIERES VALEURS PROPRES

NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	3.6396	17.33	17.33	*****
2	1.7228	8.20	25.54	*****
3	1.4060	6.70	32.23	*****
4	1.1294	5.38	37.61	*****
5	1.0587	5.04	42.65	*****
6	1.0045	4.78	47.43	*****
7	.9669	4.60	52.04	*****
8	.9123	4.34	56.38	*****
9	.8669	4.13	60.51	*****
10	.8311	3.96	64.47	*****
11	.8089	3.85	68.32	*****
12	.7806	3.72	72.04	*****
13	.7608	3.62	75.66	*****
14	.7182	3.42	79.08	*****
15	.7072	3.37	82.45	*****
16	.6786	3.23	85.68	*****
17	.6577	3.13	88.81	*****
18	.6265	2.98	91.79	*****
19	.6172	2.94	94.73	*****
20	.5766	2.75	97.48	*****
21	.5296	2.52	100.00	*****

# ACP

## COORDONNEES DES VARIABLES SUR LES AXES 1 A 5

VARIABLES	CORRELATIONS VARIABLE-FACTEUR				
	1	2	3	4	5
IDEN - LIBELLE COURT	1				
<b>VARIABLES ACTIVES</b>					
eurv - vote des etr.europ.	-.53	-.34	-.08	-.01	.20
srin - seringues.grat.toxic	-.38	-.09	.47	-.17	.17
usa - impérialisme cult.US	.06	.32	-.37	-.36	.50
nimp - plus de produit étr.	-.45	-.03	-.46	.06	-.14
mari - mariages de races #	-.61	.04	-.01	.02	-.04
porn - magaz.porno & film X	-.31	.25	.48	.18	.12
voil - interdire le voile	-.29	-.09	-.44	.29	.39
toub - Toubon a raison d'in	-.50	.32	.15	-.04	.20
immi - les immigrés renforc	-.52	-.38	-.20	.02	-.09
lexp - la lib. d'express. n	-.26	.26	-.25	-.20	-.47
lech - libr., ch.=prospérité	-.29	-.36	-.07	-.14	.17
homo - accept.homosexualit,	-.55	.07	.39	.11	.06
iden - droit contrôl d'iden	-.52	.03	-.19	.15	-.01
cult - éch.enrich.cult.FRAN	-.49	-.37	.08	.02	-.07
fume - interdt fumer Lx.Pub	-.23	.42	-.07	.51	-.02
femm - retour femm. tradit.	-.56	.21	.14	-.20	-.31
alco - ps de pub alc.&tabac	-.15	.51	-.04	.36	.04
rock - trop de rock et rap	-.38	.32	-.04	-.35	.31
mino - pour minorités en FR	-.26	-.39	.17	-.02	.05
fast - interd. fast.f.en FR	-.29	.33	-.07	-.44	-.16
race - rien en com race étr	-.54	-.05	-.16	.07	-.15
<b>VARIABLES ILLUSTRATIVES</b>					
hpol - vie privée hom.pol.	.12	.00	-.01	.01	.01



# Analyse factorielle

The FACTOR Procedure  
Initial Factor Method: Principal Factors

## Factor Pattern

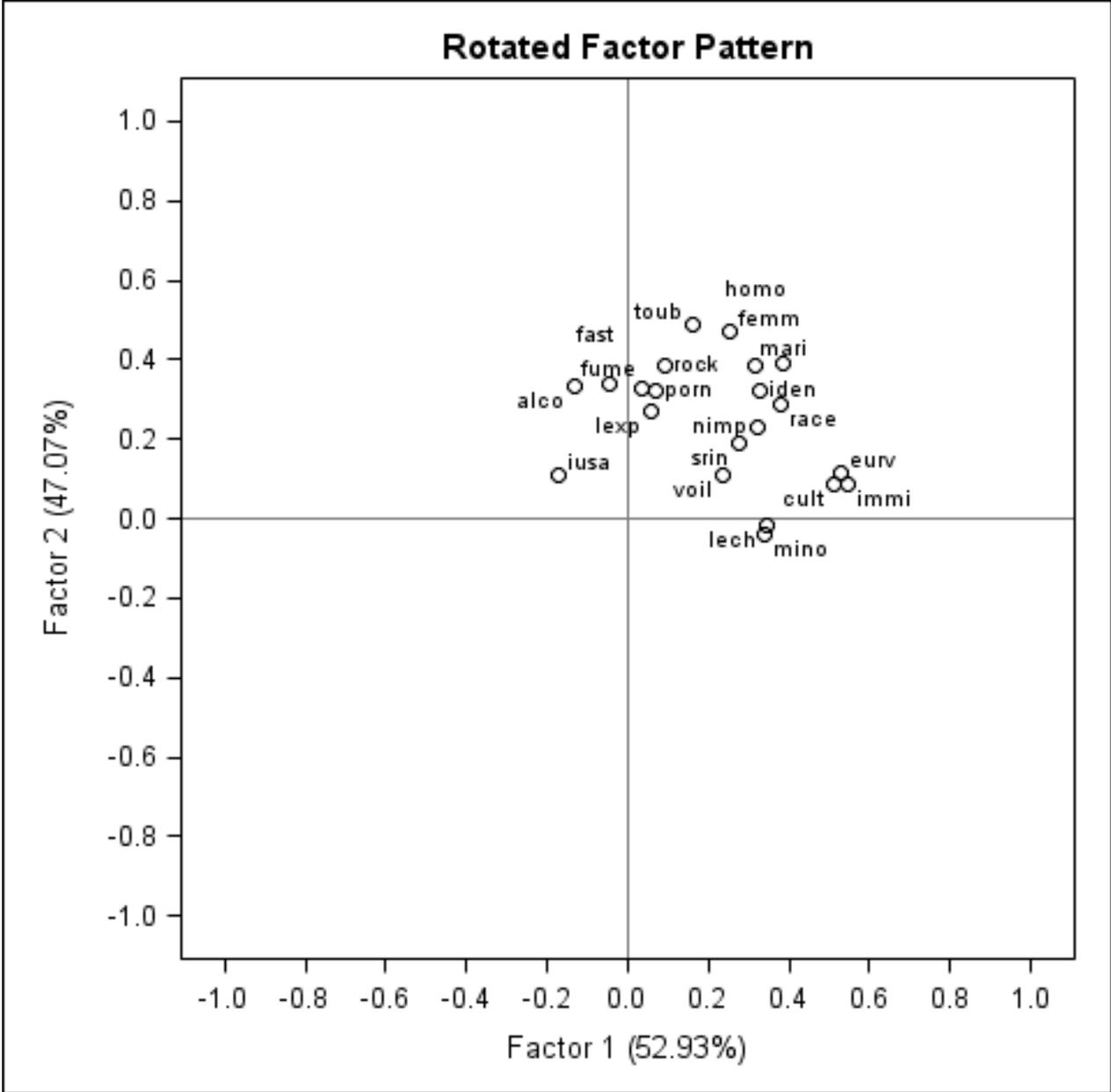
	Factor1	Factor2
eurv	0.47035	-0.26836
srin	0.33299	-0.04558
iusa	-0.05396	0.19390
nimp	0.39327	-0.04073
mari	0.54837	0.03256
porn	0.26712	0.19369
voil	0.25147	-0.07639
toub	0.44609	0.25582
immi	0.46354	-0.30168
lexp	0.22231	0.16444
lech	0.24538	-0.24173
homo	0.49328	0.07325
iden	0.46092	0.02007
cult	0.43889	-0.27355
fume	0.19117	0.28081
femm	0.50374	0.18306
alco	0.12225	0.33824
rock	0.32746	0.22628
mino	0.22388	-0.25621
fast	0.24632	0.22143
race	0.47472	-0.03765

Orthogonal Transformation Matrix

	1	2
1	0.74512	0.66693
2	-0.66693	0.74512

Rotated Factor Pattern

	Factor1	Factor2
eurv	<b>0.52945</b>	0.11373
srin	0.27852	0.18812
iusa	-0.16952	0.10849
nimp	0.32020	0.23193
mari	<b>0.38689</b>	<b>0.38999</b>
porn	0.06986	<b>0.32248</b>
voil	0.23832	0.11079
toub	0.16178	<b>0.48813</b>
immi	<b>0.54659</b>	0.08437
lexp	0.05598	0.27079
lech	<b>0.34406</b>	-0.01647
homo	0.31870	<b>0.38356</b>
iden	0.33005	0.32235
cult	<b>0.50946</b>	0.08888
fume	-0.04484	0.33673
femm	0.25326	<b>0.47237</b>
alco	-0.13449	0.33357
rock	0.09309	<b>0.38700</b>
mino	0.33769	-0.04160
fast	0.03586	<b>0.32927</b>
race	0.37883	0.28855



# ACP versus Analyse factorielle

- description
- solution unique
- explique la variance
- orienté individus
- modèle
- indétermination
- explique les corrélations
- orienté variables

Résultats très proches si les variances spécifiques sont égales, ou si les  $p-k$  dernières valeurs propres de l'ACP sont presque nulles

## 2. L'analyse en classes latentes

- **Classes et traits latents :**
  - "Latent class analysis was developed mainly within the social and political sciences, whereas latent trait models have a clear psychometrical background." (Heinen, 1996, ix)
  - Deux méthodes développées par Lazarsfeld des années 1940 aux années 1960 sous le nom commun de structures latentes. Cf. *Latent Structure Analysis* avec N.W. Henry (1968)
  - Contexte: études psycho-sociales menées par le département de la guerre américain sur le personnel militaire pendant la deuxième guerre mondiale.

# Des « trace lines » aux classes latentes

## Premier exemple

Let us look at specific test items. Suppose in connection with studying the effect of an indoctrination film we want to test the “ethnocentrism” of American soldiers. By this we mean vaguely the extent to which they think that only Americans are of value and that the actions and ideals of the war allies are of little value or importance for the war effort. Suppose the following three items have been proposed for inclusion in the test:

1. I believe that our European allies are much superior to us in strategy and fighting morale. Yes \_\_\_\_\_ No \_\_\_\_\_
2. The majority of all equipment used by all the allies comes from American lend lease shipment. True \_\_\_\_\_ False \_\_\_\_\_
3. Neither we nor our allies could win the war if we didn't have each other's help. Agree \_\_\_\_\_ Disagree \_\_\_\_\_

All three items were included in the test because they are related and partially express ethnocentrism the way we have conceived it to begin with. This fact can be formulated more precisely in the following way: We assume that the probability  $y$ , that respondents check the first alternative in the  $i$ th question, is a function  $f_i(x)$  of their position on the continuum  $x$ , their degree of ethnocentrism. This formulation can be graphically represented in Figure 1. The graphical picture of our functions  $f_i(x)$  we shall call the *trace line* of item  $i$ .

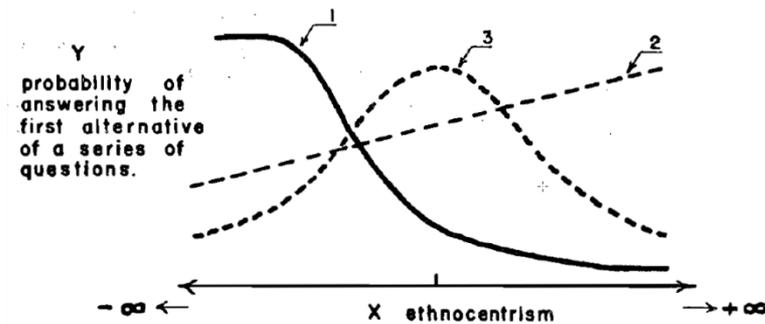


Figure 1. Trace lines of 3 items in a test of “ethnocentrism.”

Lazarsfeld, Paul F. [1950a], *The Logical and Mathematical Foundations of Latent Structure Analysis*, dans Stouffer, S.A., et al. chapter 10, 362-412  
*Measurement and Prediction, Volume IV of The American Soldier: Studies in Social Psychology in World War II.*  
Princeton University Press.

# Des « trace lines » aux classes latentes

## Deuxième exemple

## Un modèle à deux classes:

Computed Latent Structure for Attitude toward the Army

Latent Class Frequencies	Item 1 Probability +	Item 2 Probability +	Item 3 Probability +	Item 4 Probability +
424.3	.9240	.6276	.5704	.5125
575.7	.4324	.1871	.1008	.0635

Manifest Data of Four Items on Attitude toward the Army

In general how do you feel the Army is run?	Do you think when you are discharged you will [have] a favorable attitude toward the Army?	In general do you feel you yourself have gotten a square deal from the Army?	Do you feel that the Army is trying its best to look out for the welfare of enlisted men?	Count
+	+	+	+	75
+	+	+	-	69
+	+	-	+	55
+	-	+	+	42
-	+	+	+	3
+	+	-	-	96
+	-	+	-	60
+	-	-	+	45
-	+	+	-	16
-	+	-	+	8
-	-	+	+	10
+	-	-	-	199
-	+	-	-	52
-	-	+	-	25
-	-	-	+	16
-	-	-	-	229

## Classes latentes et classes sociales:

**American sociologists have been fascinated from the very beginning with the idea of social class. Whether social status is enabled by membership in a class of persons or should be conceptualized as a continuous variable has been hotly, warmly and coolly debated throughout this century. Lazarsfeld's latent class analysis was taken up by the class proponents as a method that would allow the scientific assessment of a person's class membership. (Paul Henry, Latent Structure Analysis at Fifty, 1999)**

# Classes latentes

- Equivalent de **l'analyse factorielle** dans le cas entièrement **qualitatif** : les  $p$  variables observées sont qualitatives (souvent dichotomiques) et on postule l'existence d'une variable latente également qualitative à  $k$  modalités (les classes latentes).
- Une méthode de **classification** où les classes sont telles que les variables observées sont indépendantes dans chaque classe
- Un modèle particulier de **mélange** de distributions.

- **Hypothèse d'indépendance**  
**« locale » (ou conditionnelle)**

$$f(\mathbf{x}) = \sum_{j=1}^k \pi_j \prod_{i=1}^p p_{ij}^{x_i} (1 - p_{ij})^{1-x_i}$$

**Un modèle de mélange**

- **1950:** ni test d'ajustement, ni méthode convenable d'estimation
- **1968:** Identifiabilité, méthode des **moments**
- **1974:** Leo Goodman identifie le modèle de classes latentes comme un **modèle log-linéaire**.

Ensuite, les classes latentes comme données manquantes conduisent à de nouvelles interprétations et à adapter des algorithmes d'estimation de type **EM**

# le modèle théorique dans le cas dichotomique

- $p$  variables observées dichotomiques  $X_1, X_2, \dots, X_p$  prenant des valeurs 0 ou 1
- $Y$  variable latente à  $k$  classes,
- $p_{ij}$  la probabilité (conditionnelle) que  $X_i=1$  pour un individu de la classe latente  $j$ .
- $\pi_j$  probabilité *a priori* d'appartenir à la classe latente  $j$

- Hypothèse d'indépendance conditionnelle:

$$f(\mathbf{x}) = \sum_{j=1}^k \pi_j \prod_{i=1}^p p_{ij}^{x_i} (1 - p_{ij})^{1-x_i}$$

- probabilité *a posteriori* qu'un individu de vecteur  $\mathbf{x}$  appartienne à la classe latente  $j$

$$h(j / \mathbf{x}) = \pi_j \prod_{i=1}^p p_{ij}^{x_i} (1 - p_{ij})^{1-x_i} / f(\mathbf{x})$$

# L'estimation des paramètres

maximum de vraisemblance avec EM:

$$l = \sum_{h=1}^n \ln \left( \sum_{j=1}^k \pi_j \prod_{i=1}^p p_{ij}^{x_{ih}} (1 - p_{ij})^{1-x_{ih}} \right)$$

$$\sum_{j=i}^k \pi_j = 1 \qquad \phi = l + \lambda \sum_{j=i}^k \pi_j$$

$$\frac{\partial \phi}{\partial \pi_j} = \sum_{h=1}^n \left( \prod_{i=1}^p p_{ij}^{x_{ih}} (1 - p_{ij})^{1-x_{ih}} / f(\mathbf{x}_h) \right) + \lambda = \sum_{h=1}^n \frac{g(\mathbf{x}_h / j)}{f(\mathbf{x}_h)} + \lambda$$

$$\frac{\partial \phi}{\partial p_{ij}} = \sum_{h=1}^n \pi_j \frac{\partial g(\mathbf{x}_h / j)}{\partial p_{ij}} / f(\mathbf{x}_h)$$

$$\frac{\partial g(\mathbf{x}_h / j)}{\partial p_{ij}} = \frac{\partial}{\partial p_{ij}} \exp\left(\sum_{i=1}^p (x_{ih} \ln(p_{ij}) + (1 - x_{ih}) \ln(1 - p_{ij}))\right) =$$

$$g(\mathbf{x}_h / j) \left\{ \frac{x_{ih}}{p_{ij}} - \frac{1 - x_{ih}}{1 - p_{ij}} \right\} = (x_{ih} - p_{ij}) g(\mathbf{x}_h / j) / p_{ij} (1 - p_{ij})$$

- D'où:

$$\frac{\partial \phi}{\partial p_{ij}} = \left( \pi_j / p_{ij} (1 - p_{ij}) \right) \sum_{h=1}^n (x_{ih} - p_{ij}) g(x_h / j) / f(x_h)$$

- en introduisant les probabilités *a posteriori* :

$$h(j / x_h) = \pi_j g(x_h / j) / f(x_h)$$

- en annulant les dérivées :

$$\sum_{h=1}^n h(j / x_h) = -\lambda \pi_j \quad \lambda = -n$$

$$\sum_{h=1}^n (x_{ih} - p_{ij}) h(x_h / j) / p_{ij} (1 - p_{ij}) = 0$$

$$\hat{\pi}_j = \sum_{h=1}^n h(j / \mathbf{x}_h) / n$$

$$\hat{p}_{ij} = \sum_{h=1}^n x_{ih} h(j / \mathbf{x}_h) / n \hat{\pi}_j$$

$$h(j / \mathbf{x}) = \pi_j \prod_{i=1}^p p_{ij}^{x_i} (1 - p_{ij})^{1-x_i} / f(\mathbf{x})$$

# Ajustement et choix de modèles

- Le test du  $G^2$

$$G^2 = 2 \sum_x n(x) \ln\left(\frac{n(x)}{nf(x)}\right)$$

- nombre de paramètres à estimer:

$k - 1$  probabilités  $\pi_j$  et  $kp$  probabilités conditionnelles  $p_{ij}$

=  $k(p+1) - 1$  paramètres.

- $G^2$  est à comparer à un khi-deux à  $\nu = 2^p - k(p+1) + 1$  degrés de liberté
- Tests d'hypothèses emboîtées sur  $k$
- En général si  $p$  est grand, les  $2^p$  réponses possibles ont des effectifs insuffisants voire nuls: loi du khi-deux inapplicable.
- Simulations, ou bootstrap pour la loi de  $G^2$

- Recherche de modèles **parcimonieux**, minimisant AIC ou BIC
- Critères AIC d'Akaïké, ou BIC de Schwartz :

$$\text{AIC} = -2\ln(L) + 2(k(p+1) - 1)$$

$$\text{BIC} = -2\ln(L) + \ln(n) \cdot (k(p+1) - 1)$$

**vraisemblance pénalisée**

- Remarque: théories sous-jacentes différentes pour AIC (information) et BIC (stat. bayésienne)

# Exemple (Bartholomew et Knott)

enquête sur les Attitudes Sociales Britanniques en 1990, 1077 répondants avec 10 questions binaires d'opinions sur les attitudes sexuelles.

Sur les 1024 possibilités de réponse, seules 147 ont été observées:

1	90	0110011100
2	11	0110011000
3	9	0110111000
4	117	0110000000
5	18	0100000100
6	93	0100000000
7	19	0111111100
8	35	0010000000
9	21	0110001100
10	6	0111111110

- X1 Devrait-on rendre le divorce plus facile ?
- X2 Est ce que vous soutenez les lois contre la discrimination sexuelle ?
- X3 Opinion sur le sexe prénuptial : pas du tout opposé..... toujours opposé.
- X4 Opinion sur le sexe extra- marital
- X5 Opinion sur les relations sexuelles entre personnes de même sexe .
- X6 Doit-on permettre aux homosexuels d'enseigner dans les écoles ?
- X7 Doit-on permettre aux homosexuels d'enseigner dans l'enseignement supérieur ?
- X8 Doit -on permettre aux homosexuels d'occuper des fonctions officielles ?
- X9 Un couple de lesbiennes devrait-il avoir le droit d'adopter des enfants ?
- X10 Un couple d'homosexuels mâles devrait-il avoir le droit d'adopter des enfants ?

<b>Nb. de classes</b>	<b>AIC</b>	<b>BIC</b>
2	9328	9432
3	8946	9105
4	<b>8850</b>	<b>9064</b>
5	8852	9121

$$\hat{\pi}_1 = 0.4611 \quad \hat{\pi}_2 = 0.0139 \quad \hat{\pi}_3 = 0.4169 \quad \hat{\pi}_4 = 0.1081$$

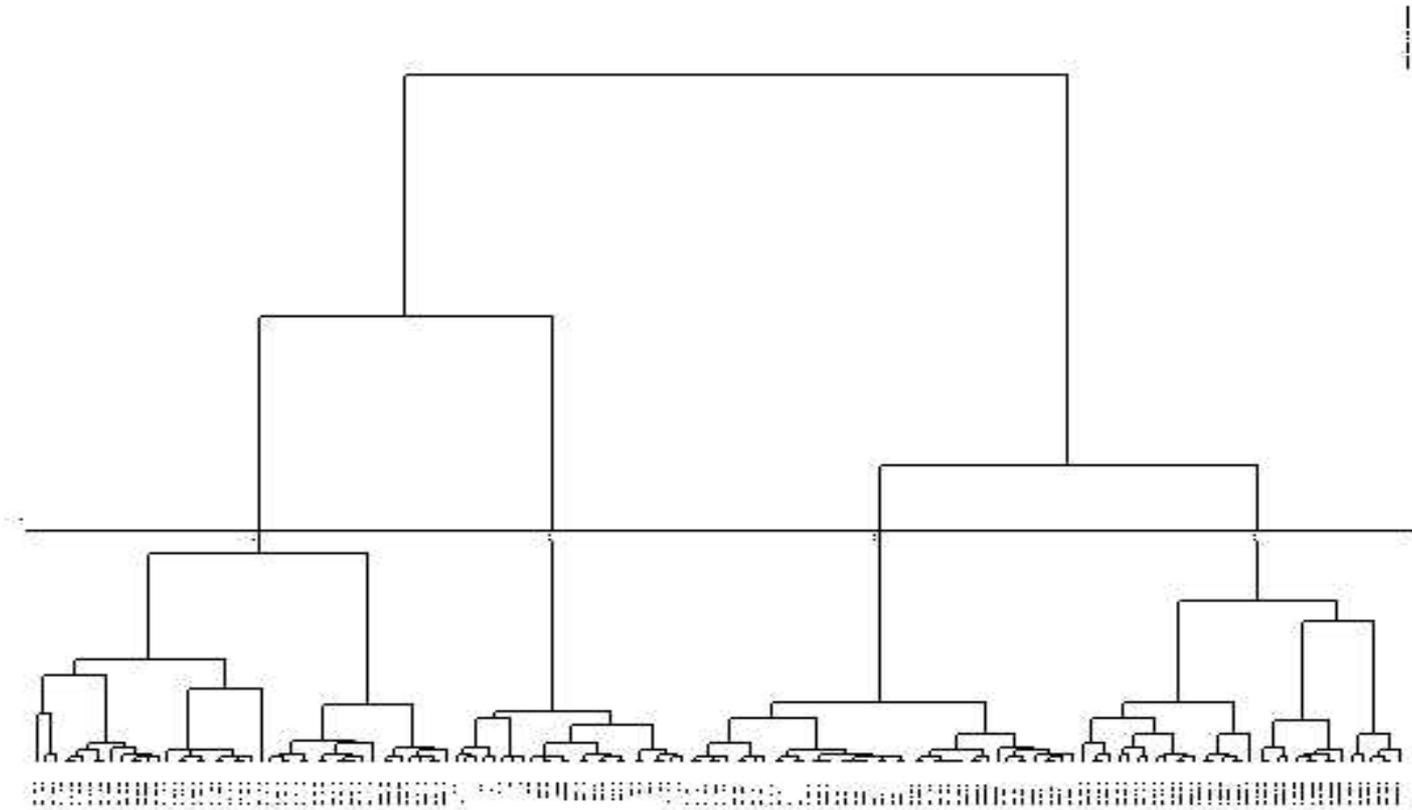
## estimations des probabilités de donner la réponse oui(1) à chacune des 10 questions , conditionnellement aux classes latentes.

	classe 1	classe 2	classe 3	classe 4
X1	0.1360	0.0667	0.0947	0.2144
X2	0.7656	0.6000	0.8712	0.9246
X3	0.6319	0.8667	0.8620	0.9635
X4	0.0822	0.2667	0.1319	0.3089
X5	0.0681	0.6000	0.3822	0.8299
X6	0.0081	0.0000	0.8721	1.0000
X7	0.0589	0.2000	0.9829	1.0000
X8	0.2077	0.2667	0.9141	1.0000
X9	0.0463	1.0000	0.1071	0.9790
X10	0.0000	1.0000	0.0000	0.8505

la classe 1 est non-permissive, la classe 2 (de très faible effectif) se distingue par une grande permissivité pour l'adoption par des homosexuels mais est très négative sur les items 6, 7 et 8. La classe 3 est permissive sur tout sauf l'adoption, la classe 4 est permissive à peu près sur tous les items.

OBS.FREQ.	E(FREQ)	LAT. CLASS	RESPONSE VECTOR
90	114.608	3	0110011100
11	10.825	3	0110011000
9	6.661	3	0110111000
117	125.255	1	0110000000
18	19.173	1	0100000100
93	72.966	1	0100000000
19	10.831	3	0111111100
35	38.337	1	0010000000
21	18.862	3	0110001100
6	4.348	4	0111111110
14	16.949	3	0010011100
1	2.737	3	0111001100
2	0.315	3	0111001110
15	11.223	1	0111000000
11	9.161	1	0110100000
1	1.583	3	0010101100
3	1.549	3	0110101000
32	18.344	3	0100011100
1	0.142	1	1011000100
27	33.119	1	0110000100
8	7.973	4	0110011111
95	71.011	3	0110111100
7	3.887	3	0100001100
40	38.911	4	0110111111
2	2.252	3	0100011110
13	15.139	3	0110011110

# Comparaison avec une classification classique après ACM



# Après coupure en 4 classes et consolidation par nuées dynamiques

	CLASSE HIER. 1	CLASSE HIER. 2	CLASSE HIER. 3	CLASSE HIER. 4	ENSEMBLE
CLASSE LAT. 1	3	0	67	421	491
CLASSE LAT. 2	0	14	0	1	15
CLASSE LAT. 3	419	0	42	0	461
CLASSE LAT. 4	6	99	5	0	110
ENSEMBLE	428	113	114	422	1077

# 3. Traits latents

- “Latent trait models” , “ item response theory ”
- Recherche de variables latentes continues
- conditionnellement à un vecteur  $\mathbf{y}$  de  $q$  variables latentes, les  $p$  variables manifestes (en général dichotomiques) sont des Bernoulli indépendantes.

$$P(x_i = 1) = \pi_i(\mathbf{y}) \qquad \ln\left(\frac{\pi_i(\mathbf{y})}{1 - \pi_i(\mathbf{y})}\right) = \alpha_{i0} + \sum_{j=1}^q \alpha_{ij} y_j$$

- modèle “ logit-probit ” si la distribution *a priori* de  $\mathbf{y}$  est multinormale centrée réduite
- Le modèle de Rasch est celui où  $q=1$  (une seule variable latente) avec des pentes  $\alpha_{ij}$  toutes égales.
- Les  $\alpha_{ij}$  s'interprètent comme des scores pour les catégories des variables manifestes.
- Bartholomew a montré que les scores obtenus par l'ACM sont une approximation au premier ordre des  $\alpha_{ij}$ .  
Voir Aitkin et al. (RSA 1987,3,53-82)

# Sortie partielle du modèle à deux facteurs

## POSTERIOR ANALYSIS:

OBS	EXPECT	E(Z1/X)	SD1	E(Z2/X)	SD2	+RESP	RESPONSE PATTERN
29	27.8	-1.33	0.74	-0.74	0.85	0	0000000000
93	75.6	-1.27	0.71	-0.48	0.81	1	0100000000
5	3.4	-1.12	0.72	-1.00	0.85	1	1000000000
9	8.7	-1.05	0.70	-0.75	0.81	2	1100000000
3	1.5	-0.91	0.69	-0.96	0.81	1	0001000000
18	13.7	-0.90	0.63	-0.01	0.66	2	0100000100
3	2.2	1.66	0.44	-0.38	0.77	8	0111011111
2	0.1	1.69	0.44	-0.79	0.69	8	1011011111
1	0.7	1.86	0.45	-0.36	0.81	8	1010111111
40	27.9	1.86	0.44	0.19	0.89	8	0110111111
12	6.4	1.92	0.45	-0.03	0.87	9	1110111111
1	1.6	1.96	0.46	-0.23	0.85	8	0011111111
18	16.5	2.03	0.48	0.13	0.90	9	0111111111
5	4.1	2.10	0.50	-0.11	0.88	10	1111111111

# ACM

HISTOGRAMME DES 10 PREMIERES VALEURS PROPRES

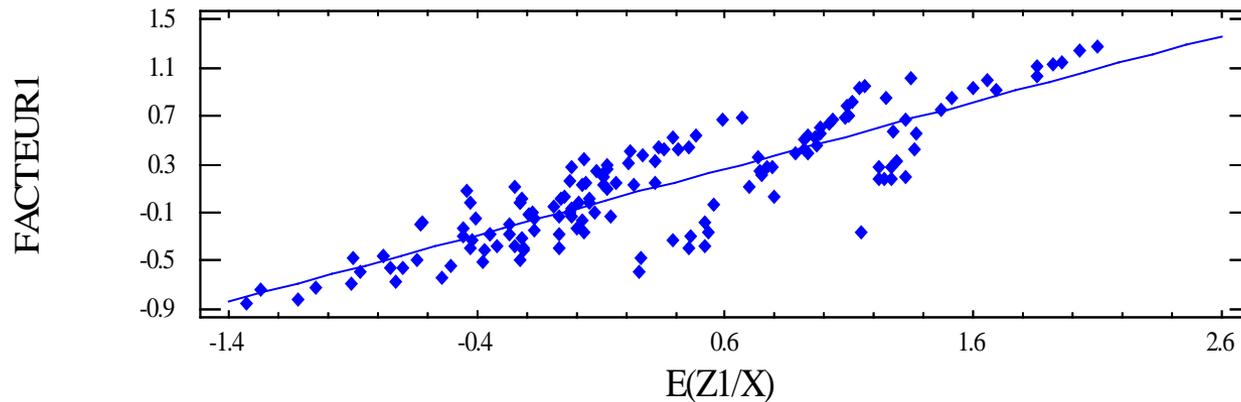
NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	0.3384	33.84	33.84	*****
2	0.1425	14.25	48.09	*****
3	0.1085	10.85	58.94	*****
4	0.1005	10.05	68.99	*****
5	0.0874	8.74	77.73	*****
6	0.0787	7.87	85.61	*****
7	0.0617	6.17	91.77	*****
8	0.0366	3.66	95.44	*****
9	0.0287	2.87	98.31	*****
10	0.0169	1.69	100.00	****

## COORDONNEES ET VALEURS-TEST DES MODALITES

AXES 1 A 5

MODALITES			VALEURS-TEST					COORDONNEES					
IDEN - LIBELLE	EFF.	P.ABS	1	2	3	4	5	1	2	3	4	5	DISTO.
2 . Devrait-on rendre le divorce plus facile?													
AA_1 - C3=0	97	941.00	-1.6	-9.9	-19.0	14.7	-19.4	-0.02	-0.12	-0.22	0.17	-0.23	0.14
AA_2 - C3=1	50	136.00	1.6	9.9	19.0	-14.7	19.4	0.13	0.80	1.52	-1.18	1.56	6.92
3 . Soutenez-vous les lois contre la discrimination sexuelle?													
AB_1 - C4=0	54	189.00	-7.6	5.8	-6.7	-26.5	-12.5	-0.50	0.38	-0.44	-1.75	-0.83	4.70
AB_2 - C4=1	93	888.00	7.6	-5.8	6.7	26.5	12.5	0.11	-0.08	0.09	0.37	0.18	0.21
4 . Opinion sur le sexe pré-nuptial?													
AC_1 - C5=0	32	251.00	-14.7	-1.2	-16.4	-3.6	5.0	-0.81	-0.07	-0.91	-0.20	0.28	3.29
AC_2 - C5=1	115	826.00	14.7	1.2	16.4	3.6	-5.0	0.25	0.02	0.28	0.06	-0.08	0.30
5 . Opinion sur le sexe extra-marital?													
AD_1 - C6=0	102	937.00	-9.4	-9.8	-17.3	-2.1	19.1	-0.11	-0.12	-0.20	-0.03	0.23	0.15
AD_2 - C6=1	45	140.00	9.4	9.8	17.3	2.1	-19.1	0.74	0.77	1.37	0.17	-1.51	6.69
6 . Relations sexuelles entre personnes de même sexe?													
AE_1 - C7=0	92	766.00	-21.9	-3.7	-0.5	-6.4	-1.2	-0.42	-0.07	-0.01	-0.12	-0.02	0.41
AE_2 - C7=1	55	311.00	21.9	3.7	0.5	6.4	1.2	1.05	0.18	0.02	0.31	0.06	2.46
7 . Les homosexuels devraient-ils enseigner dans les écoles?													
AF_1 - C8=0	84	565.00	-27.1	12.0	2.3	4.5	-1.3	-0.79	0.35	0.07	0.13	-0.04	0.91
AF_2 - C8=1	63	512.00	27.1	-12.0	-2.3	-4.5	1.3	0.87	-0.38	-0.07	-0.14	0.04	1.10
8 . Les homosexuels devraient-ils enseigner dans le Supérieur?													
AG_1 - C9=0	63	487.00	-26.9	13.1	1.0	4.8	-1.0	-0.90	0.44	0.03	0.16	-0.03	1.21
AG_2 - C9=1	84	590.00	26.9	-13.1	-1.0	-4.8	1.0	0.75	-0.36	-0.03	-0.13	0.03	0.83
9 . Occupation de fonctions officielles par des homosexuels?													
AH_1 - C10=0	56	443.00	-24.5	11.4	1.4	7.6	1.4	-0.89	0.42	0.05	0.28	0.05	1.43
AH_2 - C10=1	91	634.00	24.5	-11.4	-1.4	-7.6	-1.4	0.62	-0.29	-0.04	-0.19	-0.04	0.70
10 . Droit d'adopter des enfants pour un couple de lesbiennes?													
AI_1 - C11=0	91	877.00	-19.4	-20.1	9.8	-0.3	-1.3	-0.28	-0.29	0.14	0.00	-0.02	0.23
AI_2 - C11=1	56	200.00	19.4	20.1	-9.8	0.3	1.3	1.24	1.28	-0.62	0.02	0.08	4.39
11 . Droit d'adopter de enfants pour un couple d'hommes?													
AJ_1 - C12=0	122	963.00	-19.1	-21.0	9.5	-2.0	-2.5	-0.20	-0.22	0.10	-0.02	-0.03	0.12
AJ_2 - C12=1	25	114.00	19.1	21.0	-9.5	2.0	2.5	1.70	1.86	-0.84	0.18	0.23	8.45

- Corrélation de 0.934 entre  $E(Z_1/X)$  et le premier facteur de l'ACM.



# CONCLUSIONS

- Peu enseignés en France, les modèles à variables latentes méritent une attention particulière
- Intellectuellement stimulants
- Les résultats d'ACM ou de classification peuvent servir d'initialisation aux modèles à variables latentes
- Mais: problèmes d'identification, d'existence des variables latentes, de convergence des algorithmes...
- Analyse des données ou modèles à variables latentes?

# Appendice: Paul Félix Lazarsfeld 1901-1976

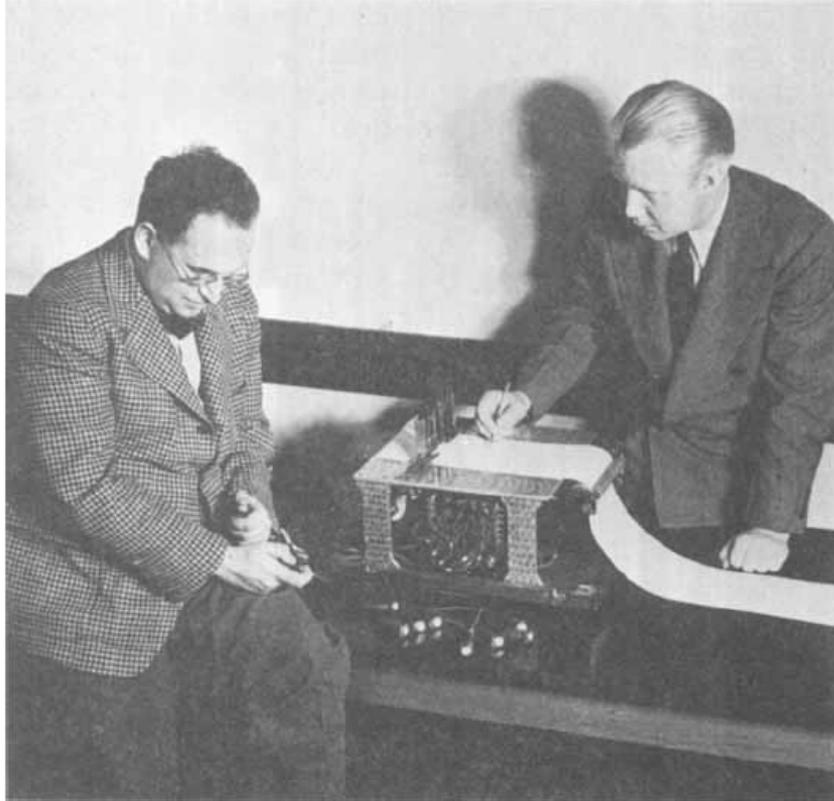


- «L'un des quatre ou cinq sociologues de sa génération qui auront le plus fortement marqué l'histoire de leur discipline» (R.Boudon), le promoteur de la pensée mathématique en sociologie
- Né à Vienne en 1901. Etudes de droit, d'économie et de mathématiques à l'université de Vienne. Thèse en mathématiques appliquées en 1924 : «Über die Berechnung der Perihelbewegung des Merkur aus der Einsteinschen Gravitationstheorie»
- Bifurque sur les enquêtes psycho-sociales. Son étude sur les chômeurs de Marienthal (édition française de 1981 préfacée par Pierre Bourdieu) lui vaut une invitation aux États-Unis en 1933 grâce à une bourse Rockefeller. A l'expiration de sa bourse, il retourne en Europe

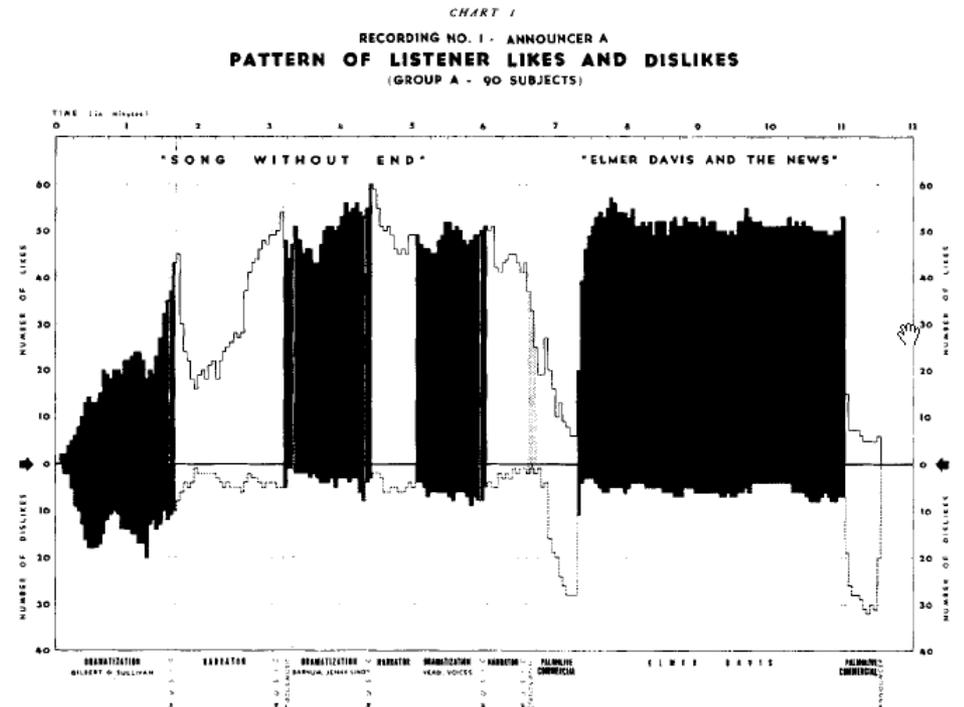
- En 1934, le Parti socialiste d'Autriche est déclaré illégal. La plupart des membres de sa famille sont arrêtés. Il perd son poste. Décide d'employer les dollars qui lui restent à l'achat d'un billet de troisième classe pour les États-Unis. Il s'y établit et devint «**un marxiste en congé**»
- Crée à Newark un institut fonctionnant sur contrats: études sur la consommation pour l'industrie laitière, ou Cosmopolitan. Pour faire croire que son institut avait un personnel plus nombreux, Lazarsfeld publiait également sous un pseudonyme
- Professeur de sociologie pendant près de 30 ans à l'université Columbia, nommé en 1962 premier titulaire du titre de « **Quetelet Professor of Social Science** » qu'il considérait, comme le véritable fondateur de la sociologie, plutôt qu'Auguste Comte. Il prit sa retraite en 1969.

## Principaux travaux : opinion et mass media

- Nouveau financement de la Fondation Rockefeller: crée en 1937 un centre de recherches à l'université de Princeton consacré à l'étude des effets sociaux de la radio
- Avec Frank Stanton, chef du dépt. des études d'audience de CBS (futur président de CBS de 1946 à 1971), met au point en 1938 le premier appareil d'analyse de la satisfaction des auditeurs (avec des boutons *satisfait, pas satisfait*) le «**Lazarsfeld-Stanton Program Analyzer**» appelé familièrement *Little Annie*
- L'Office of Radio Research, transféré à Columbia avec Lazarsfeld, devint en 1944 le Bureau of Applied Social Research, renommé Center for the Social Sciences en 1976, puis *Lazarsfeld Center for the Social Sciences* pour le vingtième anniversaire de son décès



Paul Lazarsfeld and Frank Stanton with an early model of their program analyzer, circa 1938



## Contributions aux études d'opinion

- Paul F. Lazarsfeld étudia les déterminants des votes et élaborera pour sa première grande étude électorale, lors des élections présidentielles de 1940, une méthode d'enquête (baptisée «**panel**») consistant à mener, tout au long de la campagne électorale, des interviews répétées auprès d'un échantillon représentatif d'électeurs. *The People's Choice* [1944]
- A l'occasion des élections de 1948, Paul F. Lazarsfeld proposa d'utiliser des **chaines de Markov** pour modéliser les changements d'opinion dans un panel. Il en confia la résolution à T.W.Anderson [1954]

- Créateur de 4 instituts de recherche, véritable entrepreneur scientifique et propagandiste de l'approche empirique en sociologie, P.F.Lazarsfeld considérait qu'il n'avait eu que 4 idées originales dans sa vie dont l'analyse des **panels** et les **structures latentes**

# Lazarsfeld et la sociologie française

- Collaboration avec Raymond Boudon
- Professeur «émérite» de l'Université René Descartes en 1972, sur proposition de Jean Stoetzel (professeur de sociologie et fondateur de l'IFOP)
- Relations avec P.Bourdieu

relations interindividuelles.

Par la suite, Paul Lazarsfeld exporte en Europe son modèle empiriste d'approche de la sociologie. Michaël Pollak rappelle ainsi : « *Infatigable, Lazarsfeld propageait sa conception de la sociologie partout en Europe : en 1948, il conseillait le gouvernement norvégien dans la création d'un Institut de recherches sociales. En 1958, il était le premier sociologue occidental à donner des cours à Varsovie et à présider une conférence de l'UNESCO sur les techniques de sondages (...)* »<sup>19</sup>. Il se rend même en France à plusieurs reprises à partir des années 1960. Pierre Bourdieu relate le magistère alors exercé par le sociologue américain d'origine autrichienne dans une confrontation qu'il eut avec lui peu après la publication de *L'Amour de l'art, les Musées et leur Public* (1966), écrit en collaboration avec Alain Darbel : « *L'histoire de ma confrontation, à première vue désespérée avec Paul Lazarsfeld dont on a peine à imaginer aujourd'hui l'empire à la fois scientifique qu'il exerçait sur la sociologie mondiale, a trouvé quelque chose comme un dénouement heureux en ce jour de la fin des années soixante où il nous avait littéralement convoqués Alain Darbel et moi à l'hôtel des Ambassadeurs (...) pour nous dire ses critiques du modèle mathématique de la fréquentation des musées que nous venions de publier dans L'Amour de l'art. Alors à l'apogée de sa renommée, il était arrivé avec un exemplaire du livre grossièrement griffonné à l'encre bleue et un gros cigare et avait pointé non sans brutalité ce qu'il donnait pour d'impardonnables erreurs. (...) Ses corrections accordées, Paul Lazarsfeld déclara avec quelque solennité qu'il n'avait jamais fait aussi bien aux Etats-Unis* »<sup>20</sup>.

sm

T.Gay, l'indispensable de la sociologie, 2006, Studyrama

# Références

- Aitkin, M.; Francis, B.; Raynal, N.: Une étude comparative d'analyses des correspondances ou de classifications et des modèles de variables latentes ou de classes latentes , **Revue de Statistique Appliquée**, 35, 2, p. 53-81, 1987
- Bartholomew, D.J. et Knott, M.: **Latent Variable Models and Factor Analysis** , Kendall's Library of Statistics 7: 2nd Edition, J.Wiley, 1999.
- Celeux, G.: Classification et modèles. **Revue de Statistique Appliquée**, 36, 4, p. 43-57, 1988
- Chavent, M., Kuentz, V., Saracco, J.: Analyse en facteurs: présentation et comparaison des logiciels SAS, SPAD et SPSS, **Revue Modulad**, 37, 1-30, 2007
- Everitt, B.S.: **An Introduction to Latent Variable Models** , Chapman & Hall, London, 1984.
- Fine, J.: Problèmes d'indétermination en analyse en facteurs et analyse en composantes principales optimale. **Revue de Statistique Appliquée**, 41, 4 , p. 45-72 , 1993
- Hagenaars J.A., McCutcheon A.L. (eds.): **Applied latent class analysis**, Cambridge University Press, 2002.
- Collins L., Lanza S.: **Latent class and latent transition analysis**, Wiley, 2010
- Lautman, J. et Lécuyer, B.P. (éditeurs): **Paul Lazarsfeld (1901-1976), la sociologie de Vienne à New-York**, Paris, L'Harmattan, 1998.
- Lazarsfeld, P.F. et Henri, N.W.: **Latent Structure Analysis** , Houghton Mifflin, Boston, 1968.
- McCutcheon, A.L.: **Latent Class Analysis** , (Sage University Paper Series on Quantitative Applications in the Social Sciences n 64) , SAGE publications, 1987.

# Logiciels et sites web:

- logiciel " Latent Gold " distribué par Statistical Innovations:  
<http://statisticalinnovations.com/products/latentgold.html>
- Proc LCA pour SAS:  
<http://methodology.psu.edu/downloads/proclcalta>
- " Latent Class Analysis Website ", par John Uebersax :  
<http://www.john-uebersax.com/stat/index.htm>