Modélisation Multiniveaux (MultiLevel Modelling)

- Modèles statistiques plus complexes
- Structures qui génèrent des observations Y dépendantes
- Distinction entre niveaux (modalités): influence E(Y) ou Var(Y)
  - niveaux fixés …………. influence potentielle sur moyenne Y
  - niveaux aléatoires …… influence potentiele sur variance Y

terminologie

- Modèles avec des effects aléatoires
- Modèles avec facteurs emboitées
- Modèles avec composantes variances
- Modèles avec coefficients aléatoires
- Modèles avec parcelles divisées (split plot)
- Modèles mixtes : facteurs fixées + facteurs aléatoires
- Modèles pour données longitudinales (temps = facteur)
- Modèles pour mesures répétées

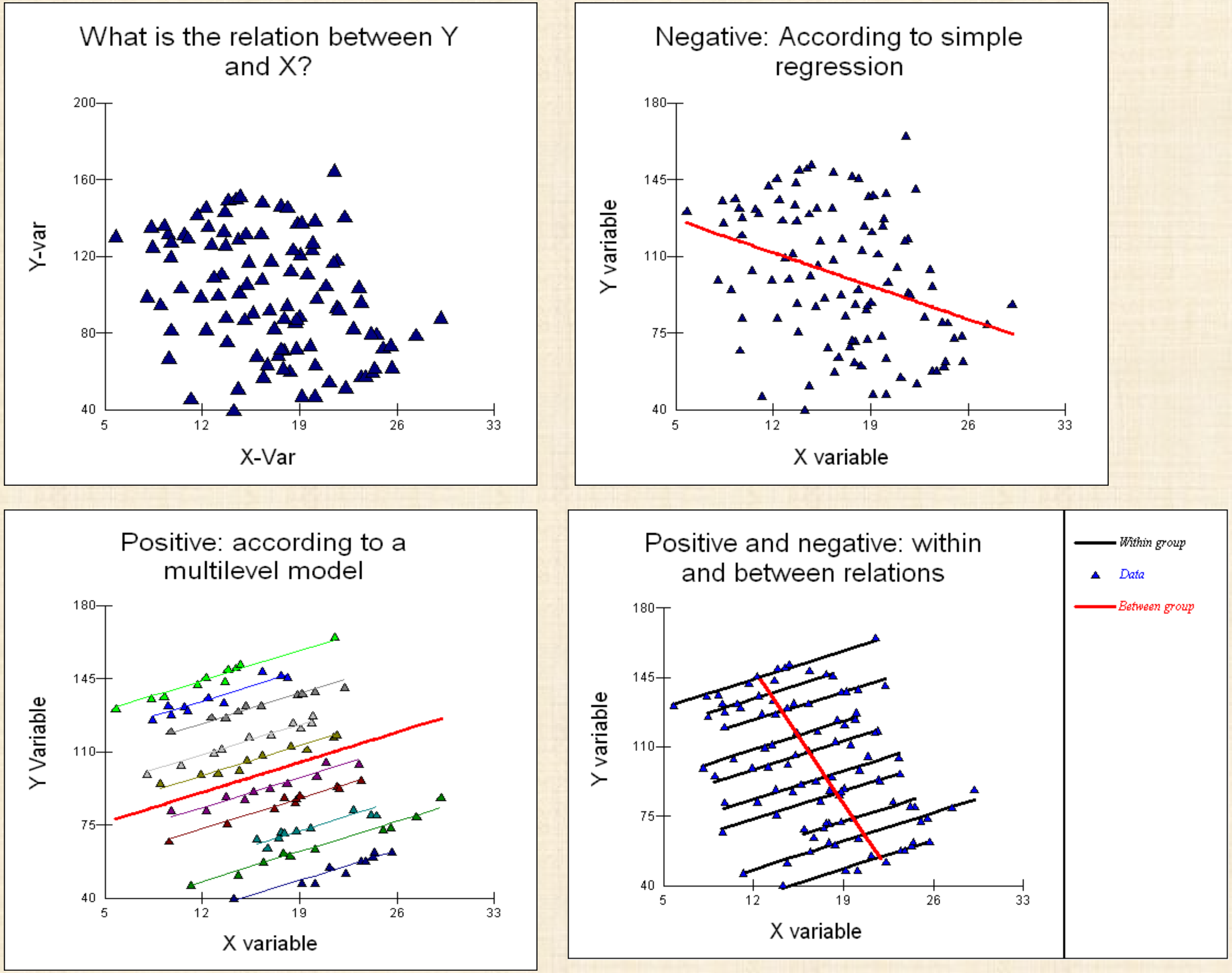
Pourquoi modèles Multi-niveaux?

réponse

parce que les modèles uni-niveaux peuvent conduire à des conclusions erronées

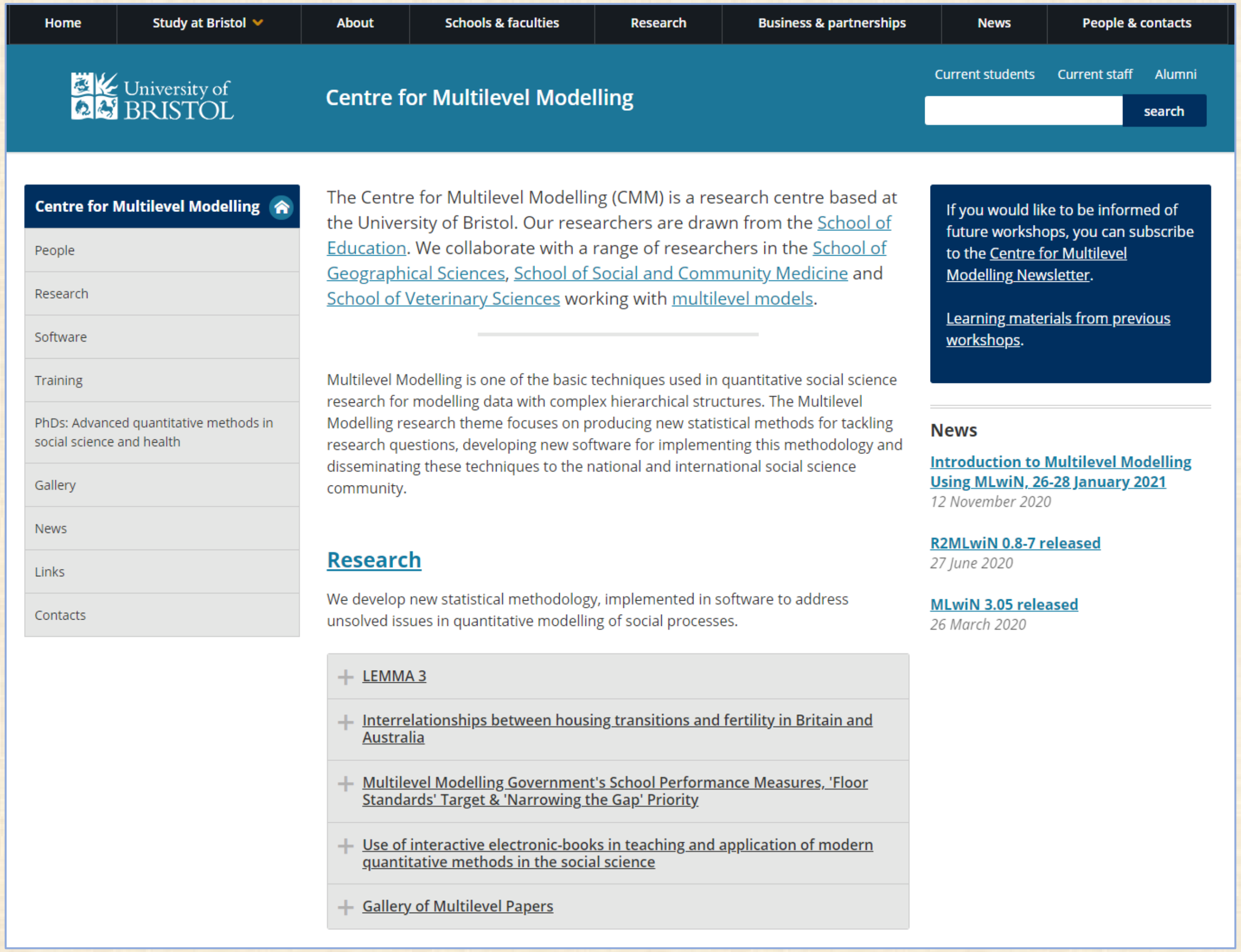
ou

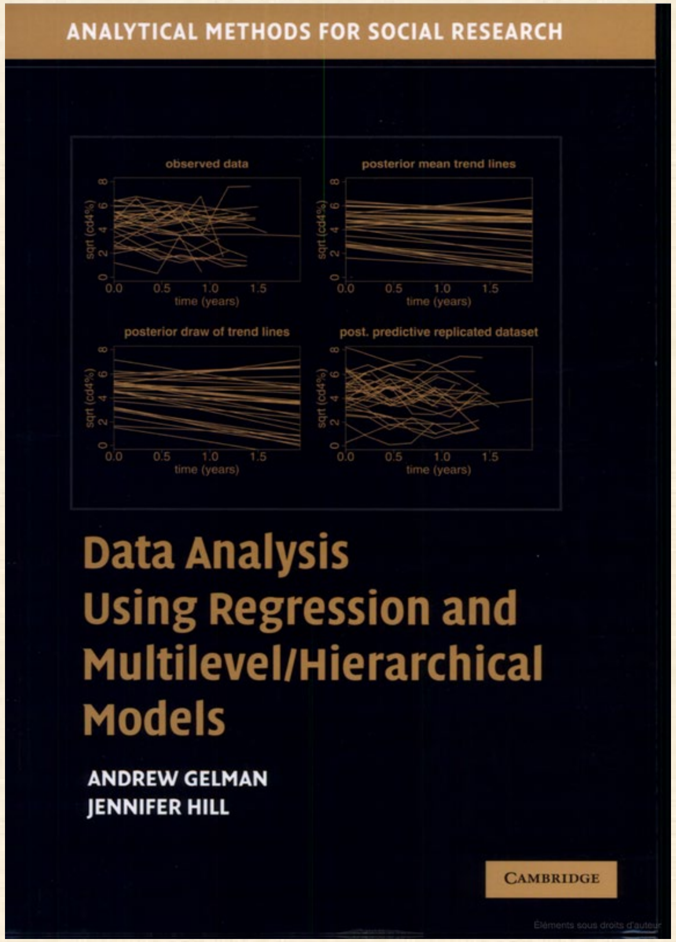
conduire à de la la confusion



RÉFÉRENCE

Centre for Multilevel Modelling | Centre for Multilevel Modelling | University of Bristol





un des nombreux livres …

beaucoup d'applications en sciences sociales, sciences médicales, sciences de la vie mais aussi en …. sciences physiques, ingénierie …

Références

- Data Analysis Using Regression and Multilevel_Hierarchical Models- Andrew Gelman
- LightBulbExperiment
- UnivBristol
- UnivBristol-data
- Bohning-Hierarchical Model.pdf
- Bohning-Random effect and Hierarchical structure.pdf
- Doncaster-Anova&Ancova Examples.pdf
- Gelman-Hierarchical modeling.pdf
- Kaltenback-Hasse diagram.pdf
- Krzywinski-Nested design.pdf
- LectureNotes-chap28-Nested designs.pdf
- Medical-Enginering.pdf
- Meier-Nested_Mixed_Part_I.pdf
- Montgomery-splitplot.pdf
- Oehlert-chap12-nested models.pdf
- Oehlert-Hasse Diagram.pdf
- SAS-GLM procedure.pdf
- SAS-Multilevel Model.pdf
- SAS-Proc GLM and Proc Mixed..pdf
- SAS-Proc Nested.pdf
- Schielzeth-Nested and mixed design.pdf
- Wikipedia-Mixed-design_analysis_of_variance.pdf
- Wikipedia-Multilevel model.pdf
- Anderson&McLean-Medical-Engineering Experiment.docx
- Medical Enginering Experiment.docx
- Medical Enginering Experiment2.docx
- STATISTICA-module VEPAC.docx
- STATISTICA-Nested&Crossed Designs.docx
- 3 Factors Crossed & Nested.jmp
- 3 Factors Nested & Crossed-2.jmp
- Animals.jmp
- Growth Measurements.jmp
- Growth.jmp
- Data Analysis Using Regression and Multilevel_Hierarchical Models-
- ~$U Bristol extraits.pptx
- McDermot-Modèle facteurs croisés-emboités.pptx
- Processus mesure avec STATISTICA.pptx
- U Bristol extraits.pptx
- Nested&Crossed Designs-examples.stw
- Valve.stw
- Wheeler EMP.stw

EXEMPLE 1

	1 ID	2 Temperature	3 Chamber	4 Gender	5 Person	6 Y_Comfort
1	1	65	1	Male	p1	5
2	2	65	1	Male	p2	4
3	3	65	1	Female	p3	1
4	4	65	1	Female	p4	2
5	5	65	2	Male	p5	5
6	6	65	2	Male	p6	4
7	7	65	2	Female	p7	5
8	8	65	2	Female	p8	5
9	9	65	3	Male	p9	4
10	10	65	3	Male	p10	2
11	11	65	3	Female	p11	1
12	12	65	3	Female	p12	3
13	13	70	4	Male	p13	8
14	14	70	4	Male	p14	8
15	15	70	4	Female	p15	10
16	16	70	4	Female	p16	7
17	17	70	5	Male	p17	6
18	18	70	5	Male	p18	3
19	19	70	5	Female	p19	8
20	20	70	5	Female	p20	8
21	21	70	6	Male	p21	5
22	22	70	6	Male	p22	7
23	23	70	6	Female	p23	8
24	24	70	6	Female	p24	8
25	25	75	7	Male	p25	12
26	26	75	7	Male	p26	8
27	27	75	7	Female	p27	11
28	28	75	7	Female	p28	13
29	29	75	8	Male	p29	8
30	30	75	8	Male	p30	7
31	31	75	8	Female	p31	8
32	32	75	8	Female	p32	8
33	33	75	9	Male	p33	6
34	34	75	9	Male	p34	6
35	35	75	9	Female	p35	6
36	36	75	9	Female	p36	7

Milliken and Johnson, Chapter 30, Figure 30.2

Milliken, G. A., & Johnson, D. E. (1984). *Analysis of messy data: Vol. I. Designed experiments.* New York: Van Nostrand Reinhold, Co.

Milliken, G. A., & Johnson, D. E. (1992). *Analysis of messy data: Vol. I. Designed experiments.* New York: Chapman & Hall.

A comfort experiment was conducted to study the effects of temperature and gender on a person's comfort.
Researchers were interested only in three temperature settings: 65, 70, and 75 degrees Fahrenheit.
Temperature is a fixed effect.
Each temperature setting was randomly assigned to three of nine available environmental chambers.
36 subjects : 18 males and 18 females were randomly assigned to chambers so that two males and two females were assigned to each of the nine chambers.
After the people were subjected to the environmental condition for three hours, their comfort Y_confort was measured.
MODEL (mixed)
Y = Comfort = Temperature + Gender + Temperature*Gender + Chamber(Temperature) + Person(Chamber) + error
Temperature and Gender are fixed effects; they are crossed;
Chamber and Person are random effects; they are nested:
Chamber is nested within Temperature / Person is nested within Chamber

EXEMPLE 2

Kutner 5ed. p 1089 - instructeur est emboité dans ville

	1 ID	2 Ville	3 Instructeur	4 groupe	5 Y_test
1	1	Atlanta	A	gr1	25
2	2	Atlanta	A	gr2	29
3	3	Atlanta	B	gr3	14
4	4	Atlanta	B	gr4	11
5	5	Chicago	C	gr5	11
6	6	Chicago	C	gr6	6
7	7	Chicago	D	gr7	22
8	8	Chicago	D	gr8	18
9	9	San Francisco	E	gr9	17
10	10	San Francisco	E	gr10	20
11	11	San Francisco	F	gr11	5
12	12	San Francisco	F	gr12	2

EXEMPLE 3

Source: EMP III Using Imperfect Data - Donald J. Wheeler - SPC Press 2006

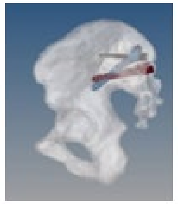
1 new	2 instrument	3 oper	4 batch	5 inXopXba (group)	6 rep	7 test623	8 new	9 instrument	10 oper	11 batch	12 group	13 rep	14 test623
p. 113	inst A	op1	b1	A1b1	1	4,2	p. 118	inst A	op1	b1	A1b1	1	4,2
	inst A	op1	b1	A1b1	2	4,4		inst A	op1	b1	A1b1	2	4,4
two	inst A	op1	b2	A1b2	1	5,3	two	inst A	op1	b2	A1b2	1	5,3
crossed	inst A	op1	b2	A1b2	2	4,8	nested	inst A	op1	b2	A1b2	2	4,8
factor	inst A	op1	b3	A1b3	1	4,0	factor	inst A	op1	b3	A1b3	1	4,0
study	inst A	op1	b3	A1b3	2	4,1	study	inst A	op1	b3	A1b3	2	4,1
	inst A	op2	b1	A2b1	1	4,3		inst A	op2	b1	A2b1	1	4,3
factor 1 =	inst A	op2	b1	A2b1	2	4,4	factor 1 =	inst A	op2	b1	A2b1	2	4,4
instrument	inst A	op2	b2	A2b2	1	5,2	instrument	inst A	op2	b2	A2b2	1	5,2
factor 2 =	inst A	op2	b2	A2b2	2	4,9	factor 2 =	inst A	op2	b2	A2b2	2	4,9
oper	inst A	op2	b3	A2b3	1	3,9	oper	inst A	op2	b3	A2b3	1	3,9
	inst A	op2	b3	A2b3	2	4,2		inst A	op2	b3	A2b3	2	4,2
oper	inst A	op3	b1	A3b1	1	4,5	oper	inst A	op3	b1	A3b1	1	4,5
is	inst A	op3	b1	A3b1	2	4,2	is	inst A	op3	b1	A3b1	2	4,2
crossed	inst A	op3	b2	A3b2	1	5,2	nested	inst A	op3	b2	A3b2	1	5,2
with	inst A	op3	b2	A3b2	2	5,4	in	inst A	op3	b2	A3b2	2	5,4
instrument	inst A	op3	b3	A3b3	1	4,1	instrument	inst A	op3	b3	A3b3	1	4,1
	inst A	op3	b3	A3b3	2	4,1		inst A	op3	b3	A3b3	2	4,1
oper X instrument	inst A	op4	b1	A4b1	1	4,4	oper(instrument)	inst A	op4	b1	A4b1	1	4,4
	inst A	op4	b1	A4b1	2	4,3		inst A	op4	b1	A4b1	2	4,3
	inst A	op4	b2	A4b2	1	5,0		inst A	op4	b2	A4b2	1	5,0
	inst A	op4	b2	A4b2	2	5,1		inst A	op4	b2	A4b2	2	5,1
	inst A	op4	b3	A4b3	1	4,3		inst A	op4	b3	A4b3	1	4,3
	inst A	op4	b3	A4b3	2	4,3		inst A	op4	b3	A4b3	2	4,3
	inst B	op1	b1	B1b1	1	3,5		inst B	op1	b1	B1b1	1	3,5
	inst B	op1	b1	B1b1	2	3,7		inst B	op1	b1	B5B5	2	3,7
	inst B	op1	b2	B1b2	1	4,4		inst B	op1	b2	B5b2	1	4,4
	inst B	op1	b2	B1b2	2	4,3		inst B	op1	b2	B5b2	2	4,3
	inst B	op1	b3	B1b3	1	3,6		inst B	op1	b3	B5b3	1	3,6
	inst B	op1	b3	B1b3	2	3,5		inst B	op1	b3	B5b3	2	3,5
	inst B	op2	b1	B2b1	1	3,8		inst B	op2	b1	B6b1	1	3,8
	inst B	op2	b1	B2b1	2	3,8		inst B	op2	b1	B6b1	2	3,8
	inst B	op2	b2	B2b2	1	4,7		inst B	op2	b2	B6B6	1	4,7
	inst B	op2	b2	B2b2	2	4,3		inst B	op2	b2	B6B6	2	4,3
	inst B	op2	b3	B2b3	1	3,8		inst B	op2	b3	B6b3	1	3,8
	inst B	op2	b3	B2b3	2	3,6		inst B	op2	b3	B6b3	2	3,6
	inst B	op3	b1	B3b1	1	3,5		inst B	op3	b1	B7b1	1	3,5
	inst B	op3	b1	B3b1	2	3,4		inst B	op3	b1	B7b1	2	3,4
	inst B	op3	b2	B3b2	1	4,7		inst B	op3	b2	B7b2	1	4,7
	inst B	op3	b2	B3b2	2	4,5		inst B	op3	b2	B7b2	2	4,5
	inst B	op3	b3	B3b3	1	3,5		inst B	op3	b3	B7b3	1	3,5
	inst B	op3	b3	B3b3	2	3,9		inst B	op3	b3	B7b3	2	3,9
	inst B	op4	b1	B4b1	1	3,7		inst B	op4	b1	B8b1	1	3,7
	inst B	op4	b1	B4b1	2	3,7		inst B	op4	b1	B8b1	2	3,7
	inst B	op4	b2	B4b2	1	4,6		inst B	op4	b2	B8b2	1	4,6
	inst B	op4	b2	B4b2	2	4,2		inst B	op4	b2	B8b2	2	4,2
	inst B	op4	b3	B4b3	1	3,7		inst B	op4	b3	B8b3	1	3,7
	inst B	op4	b3	B4b3	2	3,7		inst B	op4	b3	B8b3	2	3,7

EXEMPLE 4

étude R&R avec test destructif
o = 3 opérateurs
p = 15 paires de pièces (pj pj') j = 1, 2,.., 15
chaque paire les plus semblables possibles
n = 2 pseudo répétitions
Y_mesure

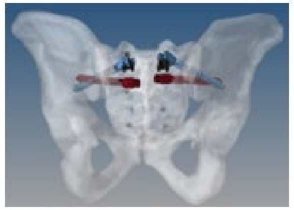
1 ID	2 new	3 Part	4 Operator	5 Y_mesure
1	test	p1	Steve	15,4257
2	destructif	p1'	Steve	16,8677
3		p2	Steve	15,5018
4		p2'	Steve	15,1628
5		p3	Steve	15,7251
6		p3'	Steve	12,8191
7		p4	Steve	15,1429
8		p4'	Steve	13,8563
9		p5	Steve	14,1119
10		p5'	Steve	16,5675
11		p6	Billie	13,1025
12		p5'	Billie	15,5494
13		p7	Billie	13,8316
14		p7'	Billie	14,2388
15		p8	Billie	16,8403
16		p8'	Billie	14,3250
17		p9	Billie	15,1448
18		p9'	Billie	14,5478
19		p10	Billie	16,3736
20		p10'	Billie	17,5779
21		p11	Nathan	14,0156
22		p11'	Nathan	16,0597
23		p12	Nathan	14,7948
24		p12'	Nathan	14,8448
25		p13	Nathan	14,2155
26		p13'	Nathan	13,7057
27		p14	Nathan	16,4566
28		p14'	Nathan	16,2174
29		p15	Nathan	15,0697
30		p15'	Nathan	16,3231

EXEMPLE 5



S1

S1AI



S2AI

plan 1 : complet
plan 2 : ?
moins essais
à déterminer

PROJET PhD - fixations pelviennes - Sophie

===

but : étude de configurations de fixations pelviennes (vis) afin de déterminer
la configuration qui minimise les risques de complications

FACTEURS (X)

===

X1_tra Trajectoire modalités = S1, S1AI (=A1) S2AI (=A2) facteur catégorique
codage de X1_tra avec les variables Z1 et Z2
Z1 = 1 si S1 / = 0 si A2 / = -1 si A2
Z2 = 0 si S1 / = 1 si A2 / = -1 si A2

X2_lon (mm) Longueur modalités = 40 et 50 si S1 / = 80 et 100 si (S1AI et S2AI)
emboité dans trajectoire - facteur continu - effet aléatoire ou fixe ?
(40, 50) si S1 (80, 100) si A1 A2
codage de X2_lon avec Z3 = (X2_ lon - c) / d
c = (40 + 100) / 2 = 70 d = (100 - 40) / 2 = 30

X3_dia (mm) diamètre modalités = 75 85 facteur continu
codage de X3_dia avec z4 = (X3_dia - c) / d
c = (75 + 85) / 2 = 80 d = (85 - 75) / 2 = 5

X4_qua Qualité osseuse normale (NO) ostéoporotique (OS) facteur catégorique
codage de X4_qua avec Z5
Z5 = 1 si X4_qua = NO Z5 = -1 si X4_qua = OS

X5_oss Ossification modalité = oui non facteur catégorique
codage de X5_oss avec Z6
Z6 = 1 si X5_oss = oui Z6 = -1 si X5_oss = non

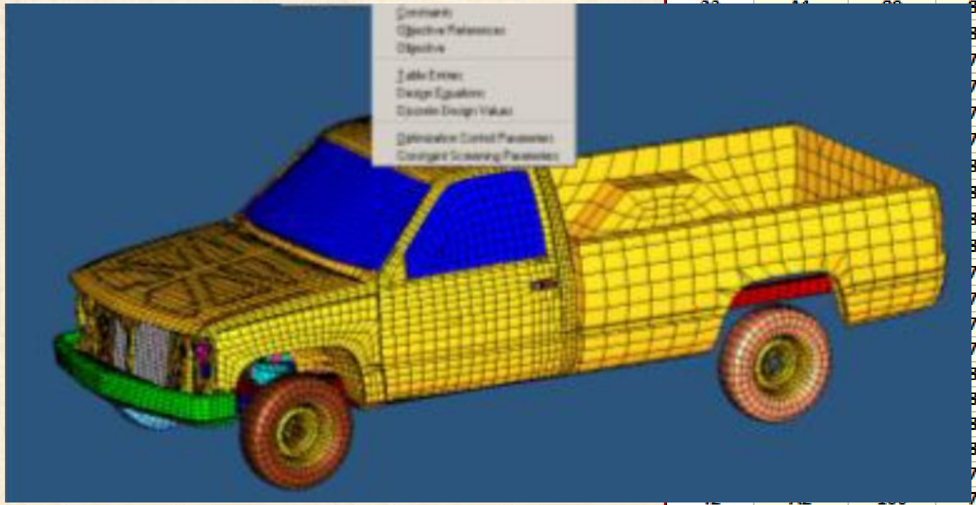
nombre de traitments = 3 x 2 x 2 x 2 x 2 = 48 plan factoriel complet
recherche d'un plan avec moins d'essais (projet à faire)

RÉPONSE (Y) Y_déf déformation (unité = ?) obtenu par programme (calcul) ?
possibilité : création d'un émulateur ?
pas de répétition car réponse ne varie pas si re-calcul

objectif : minimiser Y_def car si Y_def > seuil (seuil = ?) risque de déformation
ici les valeurs sont simulées sans aucune relation aux 5 facteurs
Y_def =100 + Rnd(20) normale avec moyenne = 100 et écart-type = 20
l'analyse ne devrait pas révéler auun effet significatif
plus tard : nous proposerons d'autres modèles en relation avec les facteurs

===

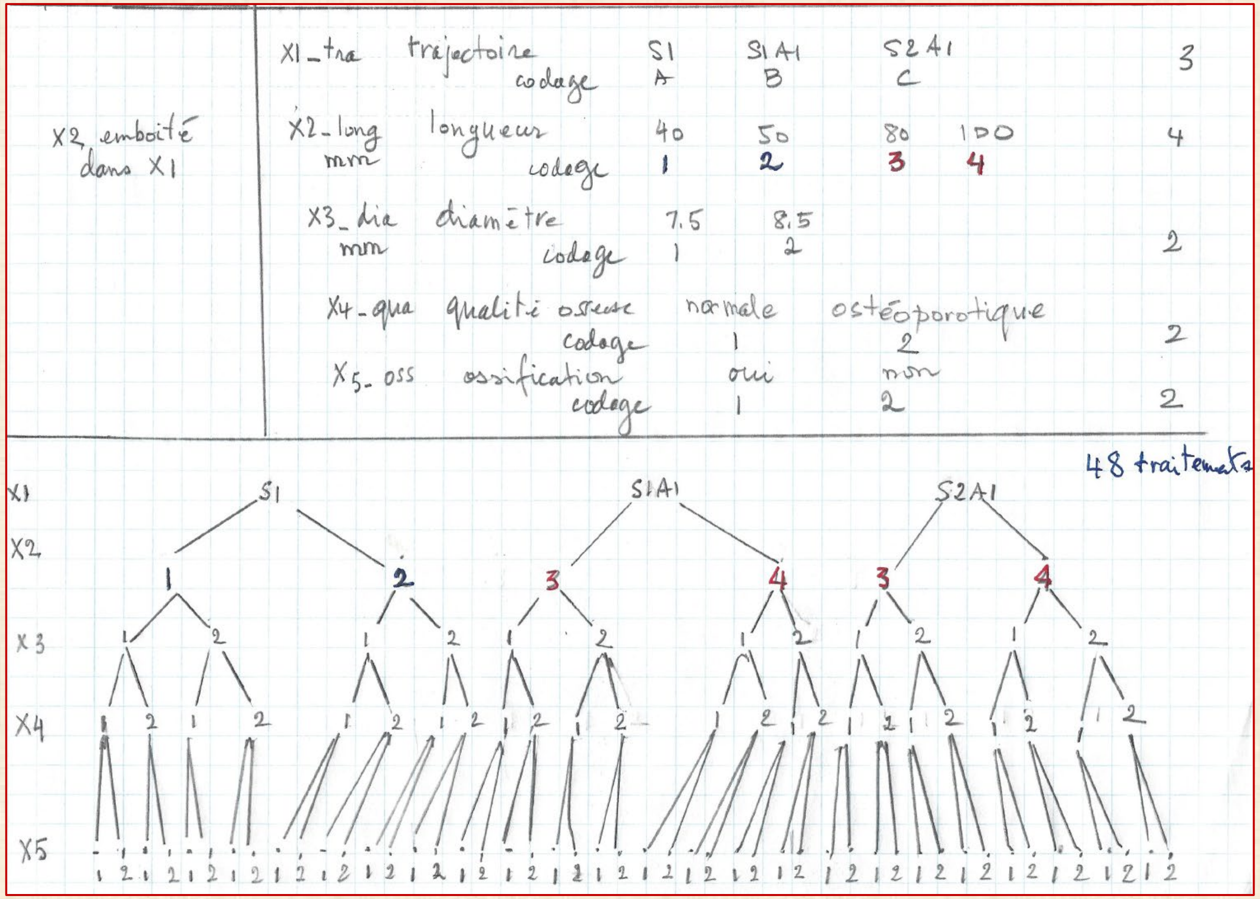
calculs par méthodes d'éléments finis :
prend beaucoup de temps …

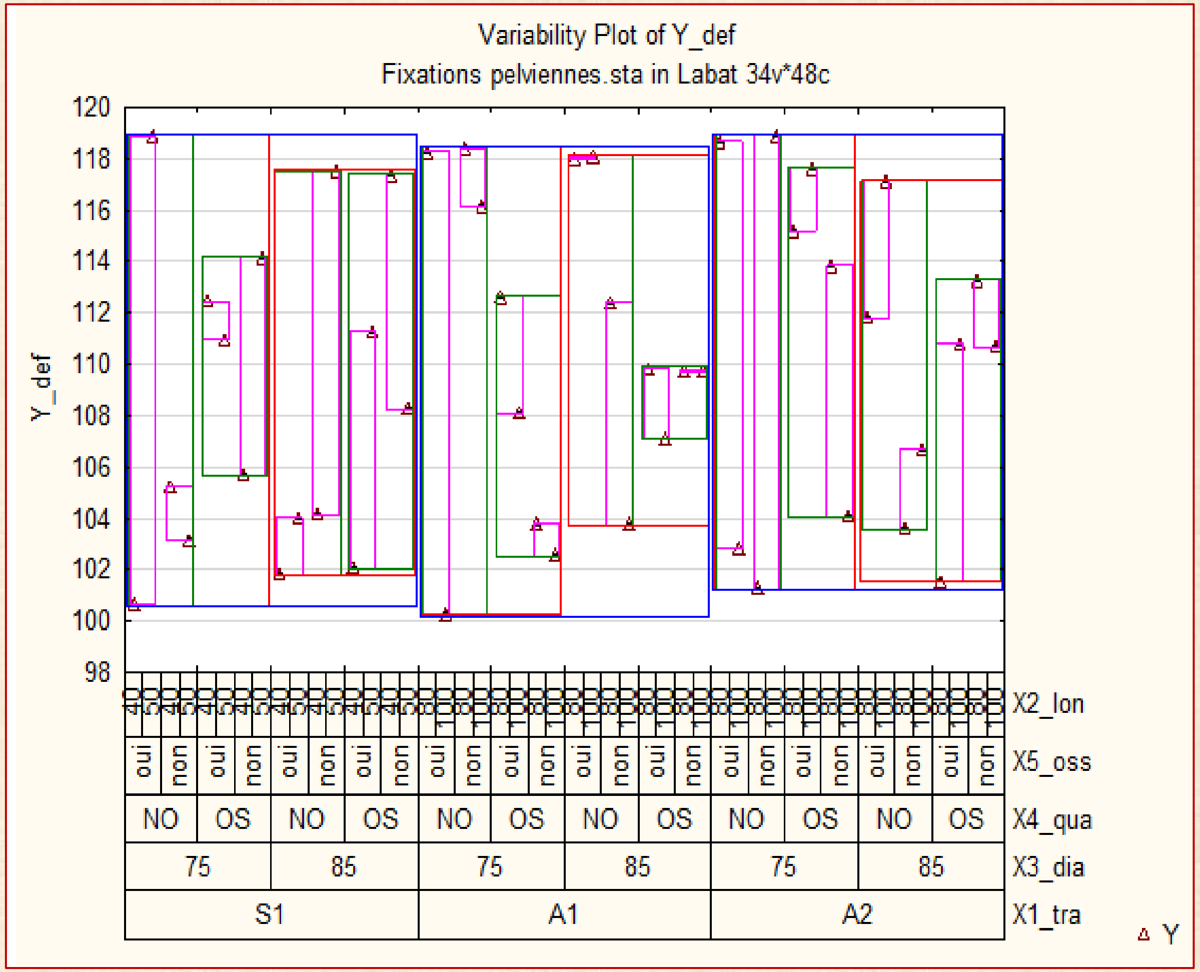


1 ID_trait	2 X1_tra	3 X2_lon	4 X3_dia	5 X4_qua	6 X5_oss	7 Y_def
1	S1	40	75	NO	oui	
2	S1	40	75	NO	non	
3	S1	40	75	OS	oui	
4	S1	40	75	OS	non	
5	S1	40	85	NO	oui	
6	S1	40	85	NO	non	
7	S1	40	85	OS	oui	
8	S1	40	85	OS	non	
9	S1	50	75	NO	oui	
10	S1	50	75	NO	non	
11	S1	50	75	OS	oui	
12	S1	50	75	OS	non	
13	S1	50	85	NO	oui	
14	S1	50	85	NO	non	
15	S1	50	85	OS	oui	
16	S1	50	85	OS	non	
17	A1	80	75	NO	oui	
18	A1	80	75	NO	non	
19	A1	80	75	OS	oui	
20	A1	80	75	OS	non	
21	A1	80	85	NO	oui	
22	A1	80	85	NO	non	
			85	OS	oui	
			85	OS	non	
			75	NO	oui	
			75	NO	non	
			75	OS	oui	
			75	OS	non	
			85	NO	oui	
			85	NO	non	
			85	OS	oui	
			85	OS	non	
			75	NO	oui	
			75	NO	non	
			75	OS	oui	
			75	OS	non	
			85	NO	oui	
			85	NO	non	
			85	OS	oui	
			85	OS	non	
			75	NO	oui	
			75	NO	non	
43	A2	100	75	OS	oui	
44	A2	100	75	OS	non	
45	A2	100	85	NO	oui	
46	A2	100	85	NO	non	
47	A2	100	85	OS	oui	
48	A2	100	85	OS	non	

EXEMPLE 5 : suite

données simulées





Complex modelling : 4 key notions

1. Modelling data with a complex structure

large range of structures that ML can handle routinely; e.g houses nested in neighbourhoods

2. Modelling heterogeneity

standard regression models 'averages', ie the general relationship

ML additionally models variances; e.g individual house prices vary from n'hood to n'hood

3. Modelling dependent data

potentially complex dependencies in the outcome over time, over space, over context;

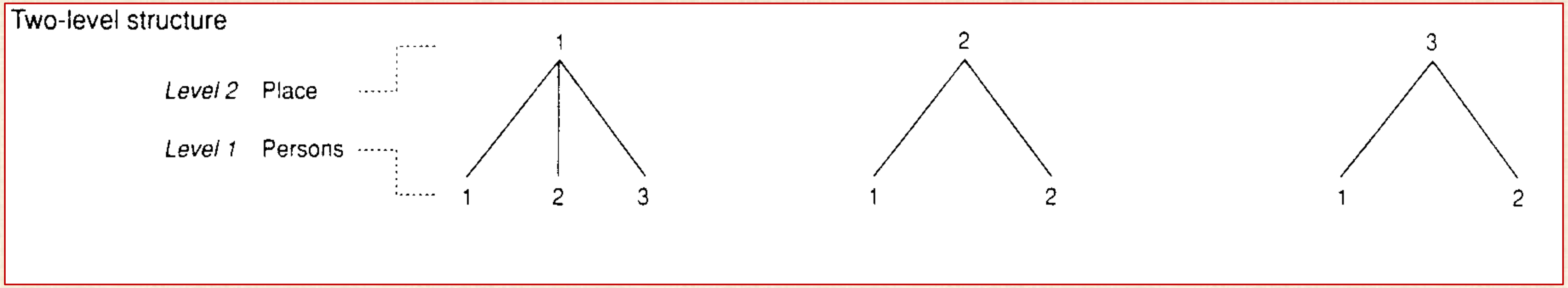
e.g houses within a n'hood tend to have similar prices

4. Modelling contextuality: micro & macro relations

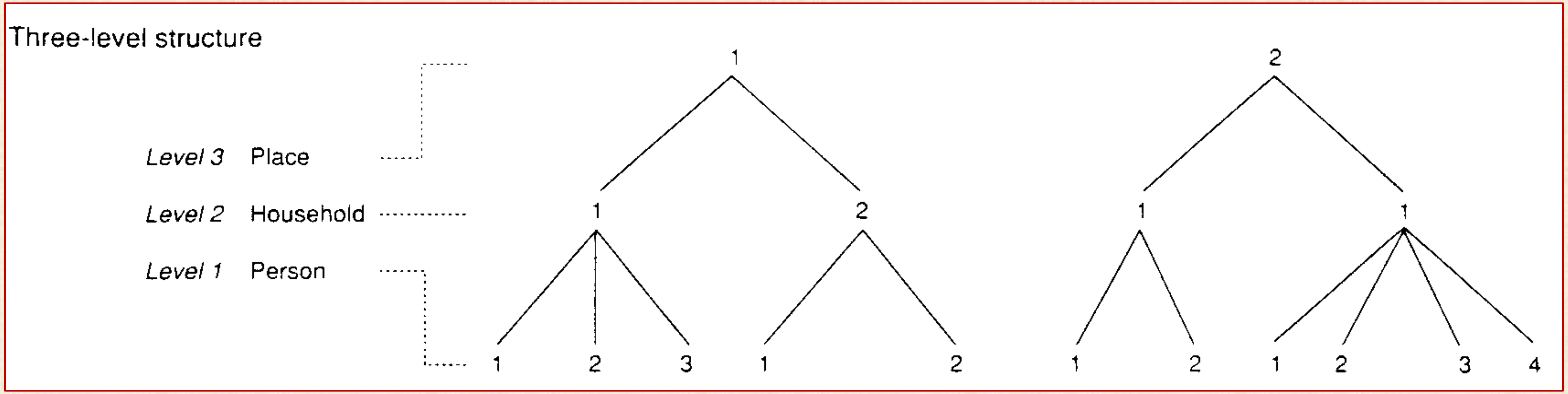
e.g individual house prices depends on individual property characteristics + on neighbourhood characteristics

Modelling data with complex structure : hiérarchique , emboitées (nested)

People nested within places : two-level model

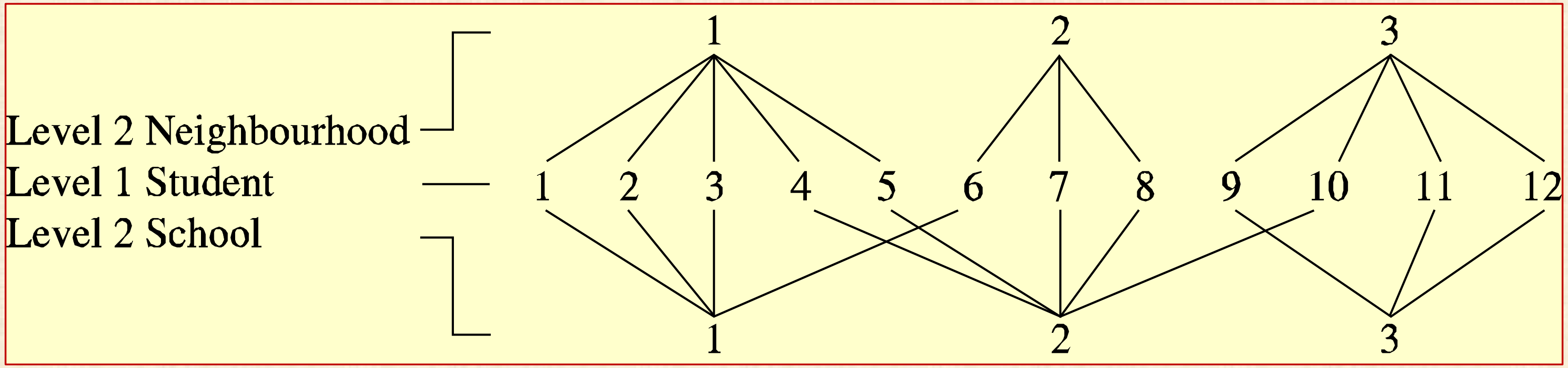


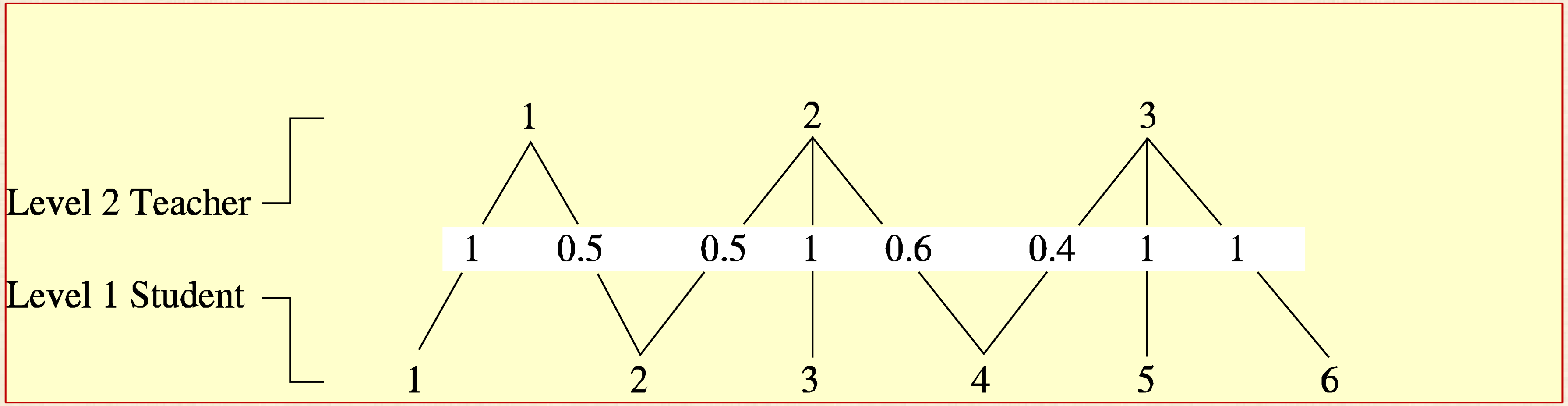
People nested within households within places : three-level model



Non-Hierarchical structures

cross-classified structure





CLASSIFICATION DIAGRAMS

Classification

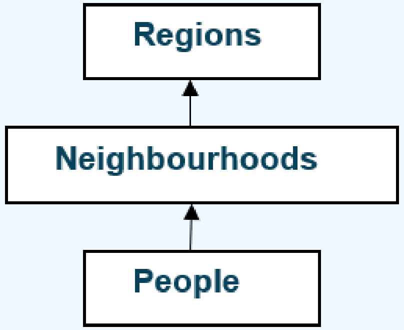
Hierarchy

Multiple membership

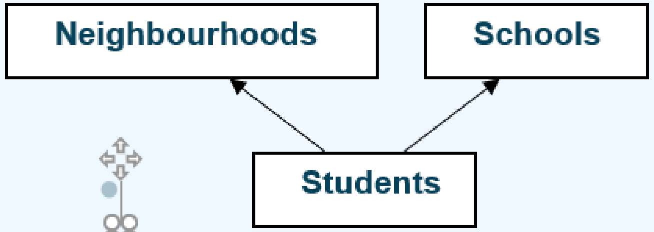
no linkage

Cross-classified

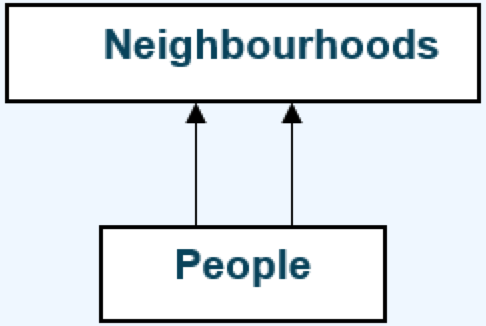
a) 3-level hierarchical structure



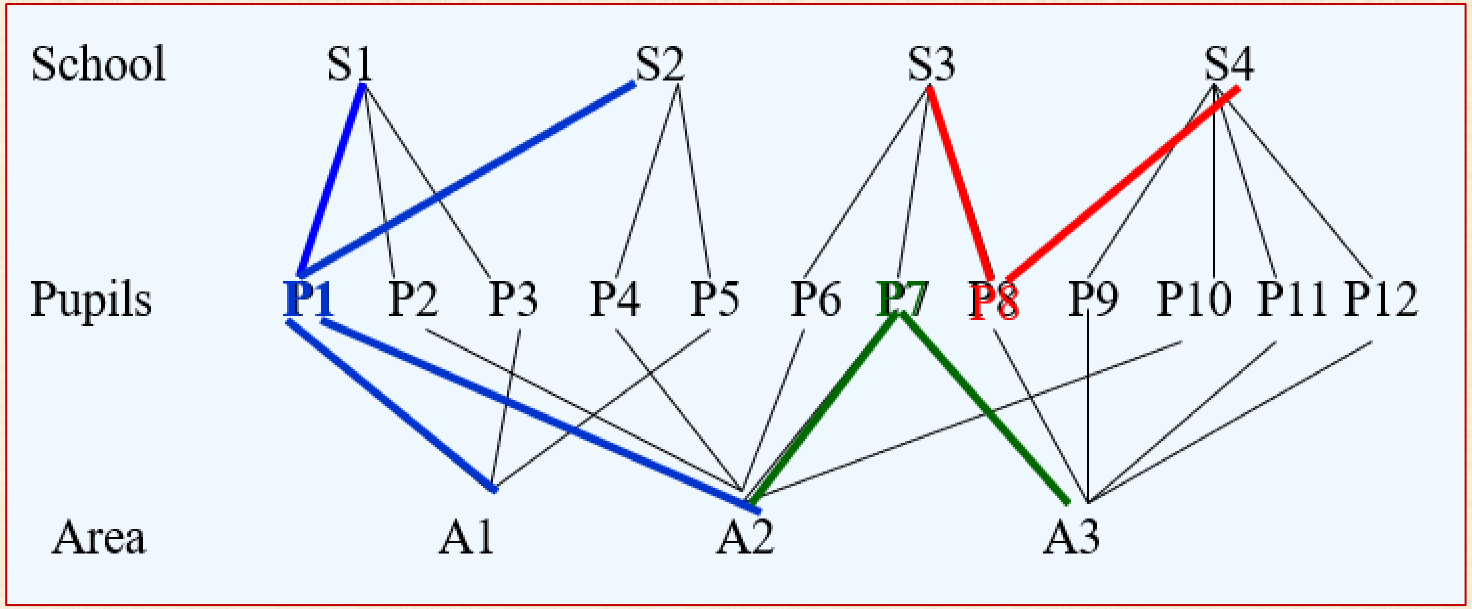
b) cross-classified structure



c) multiple membership structure



Combining structures : crossed-classifications and multiple membership relationships



Pupil 1 moves in the course of the study from residential area 1 to 2 and from school 1 to 2

Pupil 8 has moved schools but still lives in the same area

Pupil 7 has moved areas but still attends the same school

data-frame : examining neighbourhood effects on price of houses

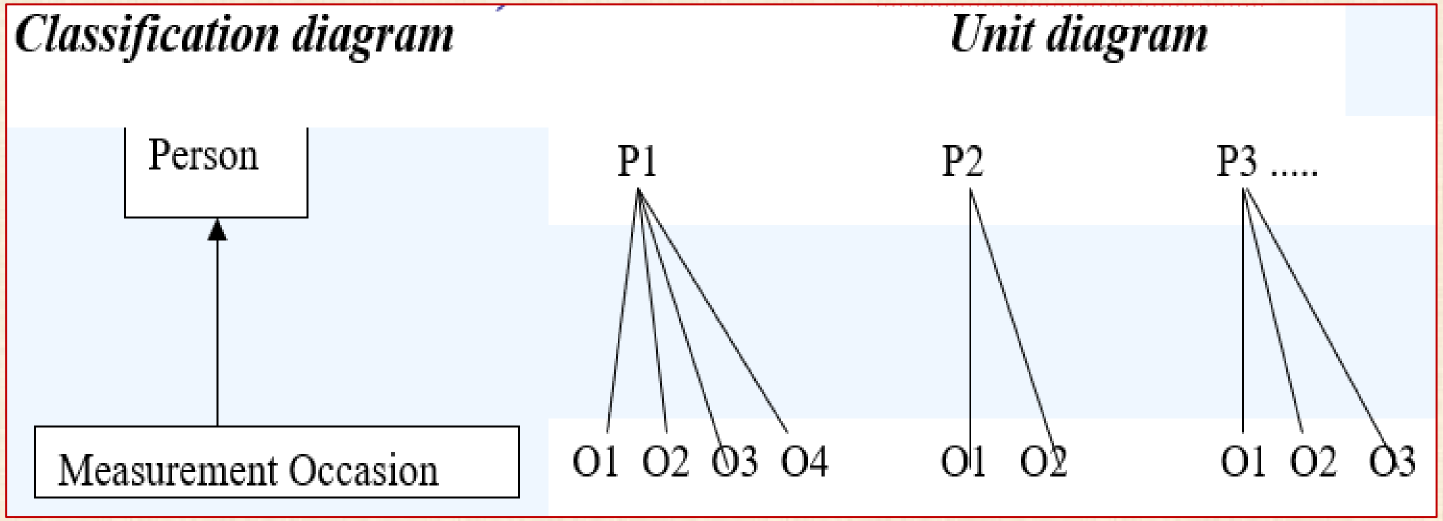
Classifications or levels		Response	Explanatory variables		
House i	N'hood j	House Price I j	No of Rooms ij	House type ij	N'hood Type j
1	1	75	6	Semi	Suburb
2	1	71	8	Semi	Suburb
3	1	91	7	Det	Suburb
1	2	68	4	Ter	Central
2	2	37	6	Det	Central
3	2	67	6	Ter	Central
1	3	82	7	Semi	Suburb
2	3	85	5	Det	Suburb
1	4	54	9	Terr	Central
2	4	91	7	Terr	Central
3	4	43	4	Semi	Central
4	4	66	55	Det	Central

Questions for multilevel (random coefficient) models

- What is the between-neighbourhood variation in price taking account of size of house ?
- Are large houses more expensive in central areas ?
- Are detached houses more variable in price ?

Two level repeated measures design : classifications, units and dataframes

Classifications or levels		Response	Explanatory variables	
Occasion i	Person j	Income Ij	Age Ij	Gender j
1	1	75	25	F
2	1	85	26	F
3	1	95	27	F
1	2	82	32	M
2	2	91	33	M
1	3	88	45	F
2	3	93	46	F
3	3	96	47	F



Person	Inc-Occ1	Inc-Occ2	Inc-Occ3	Age-Occ1	Age-Occ2	Age-Occ3	Gender
1	75	85	95	25	26	27	F
2	82	91	*	32	33	*	M
3	88	93	96	45	46	47	F

Distinguishing classifications: variables or levels ?

Classifications or levels			Response	Explanatory Variables	
House I	Nhood j	Type k	Price ijk	Rooms ijk	House type ijk$_{ijk}$
1	1	Suburb	75	6	Det
2	1	Suburb	71	4	Det
3	1	Suburb	91	7	F
1	2	Central	68	9	F
2	2	Central	37	6	M
…					

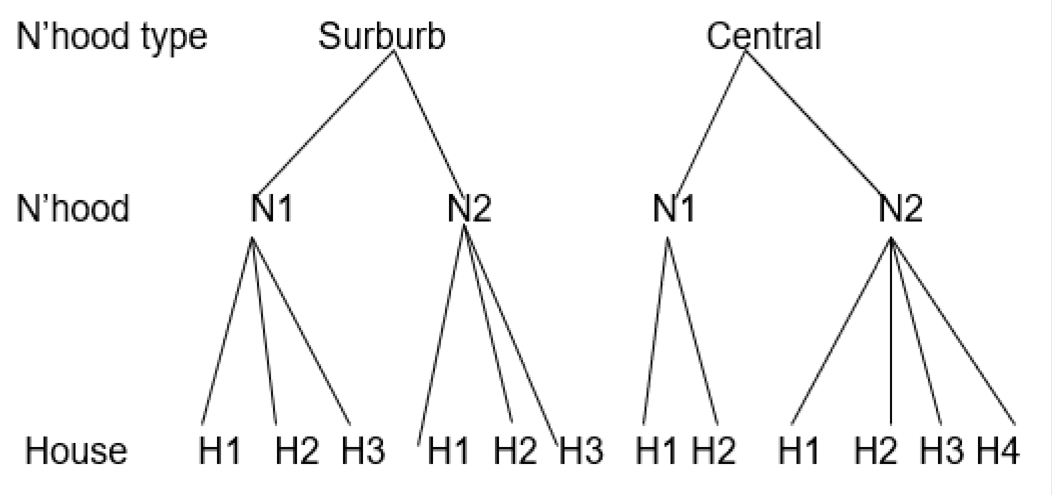
Random classification

if units can be regarded as a *random* sample from a wider population of units. Ex. houses and n'hoods
Beaucoup de modalités : facteur aléatoire

Fixed classification

is a small *fixed* number of categories.

Ex. Suburb and Central are not two types sampled from a large number of types
On the basis of these two we cannot generalise to a wider population of types of n'hoods



N'hood type is NOT a random classification but a fixed classification (fixed effect type)

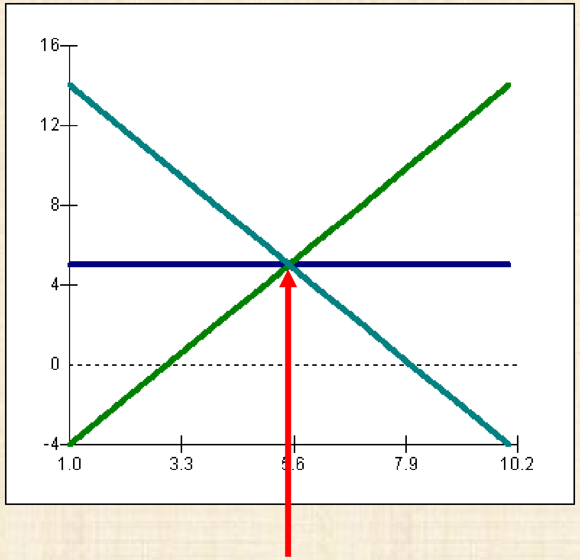
= level of a VARIABLE

Analysis Strategies for Multilevel Data

1. Group-level analysis. Move up the scale: analyse only at the macro level;

Aggregate to level 2 and fit standard regression model.

problem : cannot infer individual-level relationships from group-level relationships (ecological or aggregation fallacy)



same mean $\bar{Y}_j$, $\bar{X}_j$
but three very different within school relations (elitist; egalitarian, bizarre!)

Example: research on school effects

Response: Current score on a test, turned into an average for each of j schools;

Predictor: past score turned into an average for each of j schools

Model: regress means on means

Means on means analysis is meaningless!

Mean does not reflect within group relationship

Analysis Strategies for Multilevel Data

2 Contextual analysis. Analysis individual-level data but include group-level predictors
problem : asumes all group-level variance can be explained by group-level predictors; incorrect SE's (standard errors) for group-level predictors

EXEMPLE

- Do pupils in single-sex school experience higher exam attainment?
- Structure: 4059 pupils in 65 schools
- Response: Normal score across all London pupils aged 16
- Predictor: Girls and Boys School compared to Mixed school

Parameter	Single level	Multilevel
Cons (Mixed school)	-0,098 (0.021)	-0.101 (0.070)
Boy school	0,122 (0.049)	0,064 (0.149)
Girl school	0,245 (0.034)	0,258 (0.117)
Between school variance(σ_u^2)		0,155 (0.030)
Between student variance (σ_e^2)	0,985 (0.022)	0,848 (0.019)

Analysis Strategies for Multilevel Data

3 Analysis of covariance with fixed effects model

Include dummy variables for each and every group

problems

- What if number of groups very large, eg households?
- No single parameter assesses between group differences
- Cannot make inferences beyond groups in sample
- Cannot include group-level predictors as all degrees of freedom at the group-level have been consumed
- Target of inference: individual school versus schools

4 Fit single-level model but adjust standard errors for clustering (GEE approach)

problems Treats groups as a nuisance rather than of substantive interest; no estimate of between-group variance; not extendible to more levels and complex heterogeneity

5 Multilevel (random effects) model. LA solution définitive

Partition residual variance into between- and within-group (level 2 and level 1) components. Allows for un-observables at each level, corrects standard errors, Micro AND macro models analysed simultaneously, avoids ecological fallacy and atomistic fallacy:

richer set of research questions

BUT we need to be careful with the model specification + assumptions are met.

Type of questions tackled by ML : fixed effects (F) + random effects (R)

- Even with only 'simple' hierarchical 2-level structure
- EG 2-level model: current attainment given prior attainment of pupils (1) in schools (2)
- Do Boys make greater progress than Girls (modelling averages : fixed effect (F)
- Are boys more or less variable in their progress than girls? (modelling variances R)
- What is the between-school variation in progress? (modelling random effect R)
- Is School X different from other schools in the sample in its effect? (F)
- Are schools more variable in their progress for pupils with low prior attainment? (R)
- Does the gender gap vary across schools? (R)
- Do pupils make more progress in denominational schools? (F)
- Are pupils in denominational schools less variable in their progress? (R)
- Do girls make greater progress in denominational schools? (F)

  - cross-level interaction (correct SE's computed)

in general : a focus on variances σ^2

segregation, inequality : differences between units