ANALYSES DE TRAJECTOIRES PAR CLASSES LATENTES: PAR OÙ COMMENCER?

Sarah O'Connor, Dt.P, M.Sc, PhD(c)

INSPQ
CENTRE DE RECHERCHE INSTITUT UNIVERSITAIRE DE CARDIOLOGIE ET DE PNEUMOLOGIE DE QUÉBEC
UNIVERSITÉ LAVAL
UQTR Université du Québec à Trois-Rivières

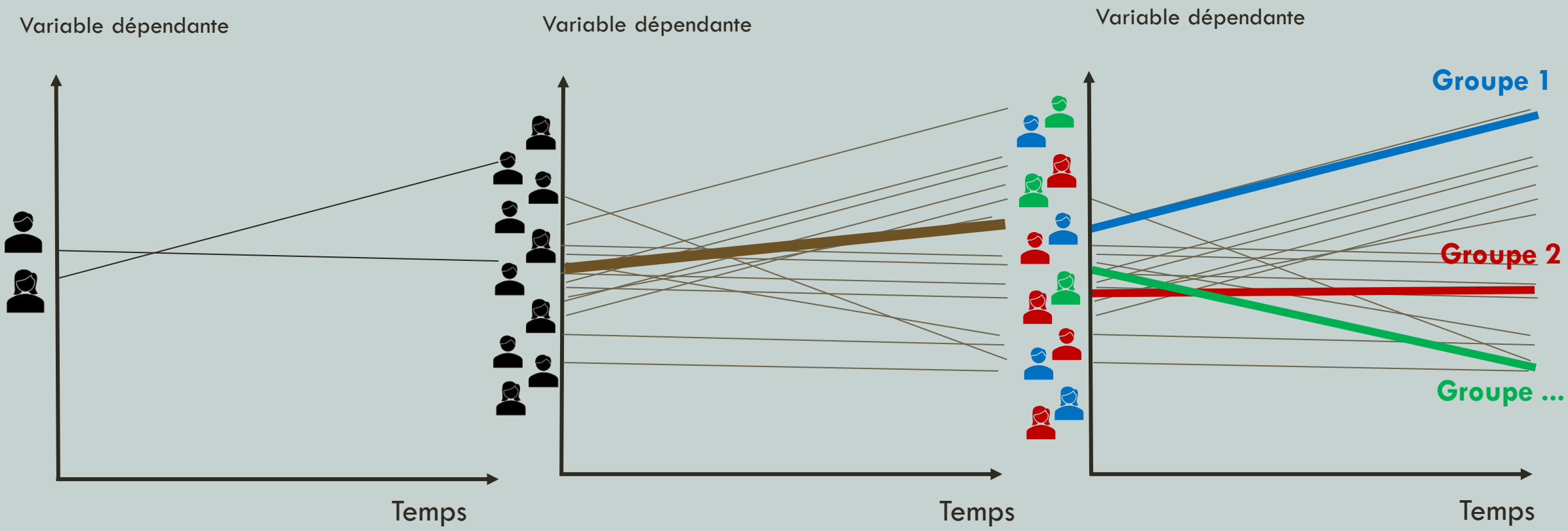
PLAN DE LA PREMIÈRE PARTIE

1) Analyse de trajectoires par classe latente

2) Procédure proc TRAJ de SAS

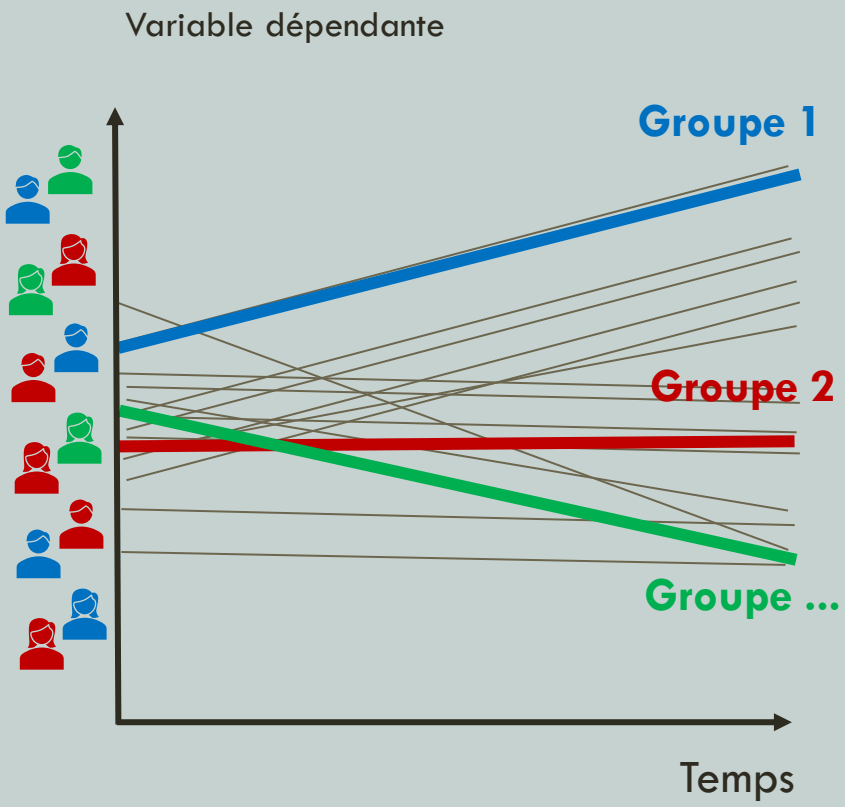
ANALYSE DE TRAJECTOIRES PAR CLASSE LATENTE: PAR OÙ COMMENCER

QU'EST-CE QU'UNE ANALYSE DE TRAJECTOIRES PAR CLASSES LATENTES?



Nagin & Odgers. 2010; Tsai et al. 2019; O'Connor et al. 2022

QUELLES GENRES DE QUESTIONS DE RECHERCHE?



Y-a-til plusieurs patrons de changement pour une variable donnée?
Combien de patrons de changements y a-il?
Quelles sont leur forme de changement dans le temps?
Qu'est-ce qui prédit l'appartenance des trajectoires dans chaque groupe?
Quelles sont les caractéristiques qui diffèrent entre les groupes?

Et autres...

LES GRANDES ÉTAPES

1. Variable dépendante, la métrique du temps et le nombre plausible de groupes de trajectoires, selon la littérature
   1. Distribution
   2. Forme fonctionnelle
2. Choix des classes optimales
3. Spécification du modèle: Formes fonctionnelles
4. Confirmation de l'appartenance aux groupes
5. Présentation en graphique
6. *Calcul des intervalles de confiance*
7. *Interprétation clinique et plausibilité*
8. *Analyses de sensibilité*

Lennon et al. BMJ Open; 2018

ÉTAPE 1: METTRE LA TABLE

Définir la métrique du temps

Âge?
Années? Mois?
Temps à l'échelle? Ex: début d'un diagnostic

ÉTAPE 1: METTRE LA TABLE

Variable dépendante

Variable dépendante et sa distribution

Variable continue ou ordinale (échelles)? → Distribution normale

Variables avec plusieurs valeurs à zéro? → Distribution de Poisson *zero-inflated*

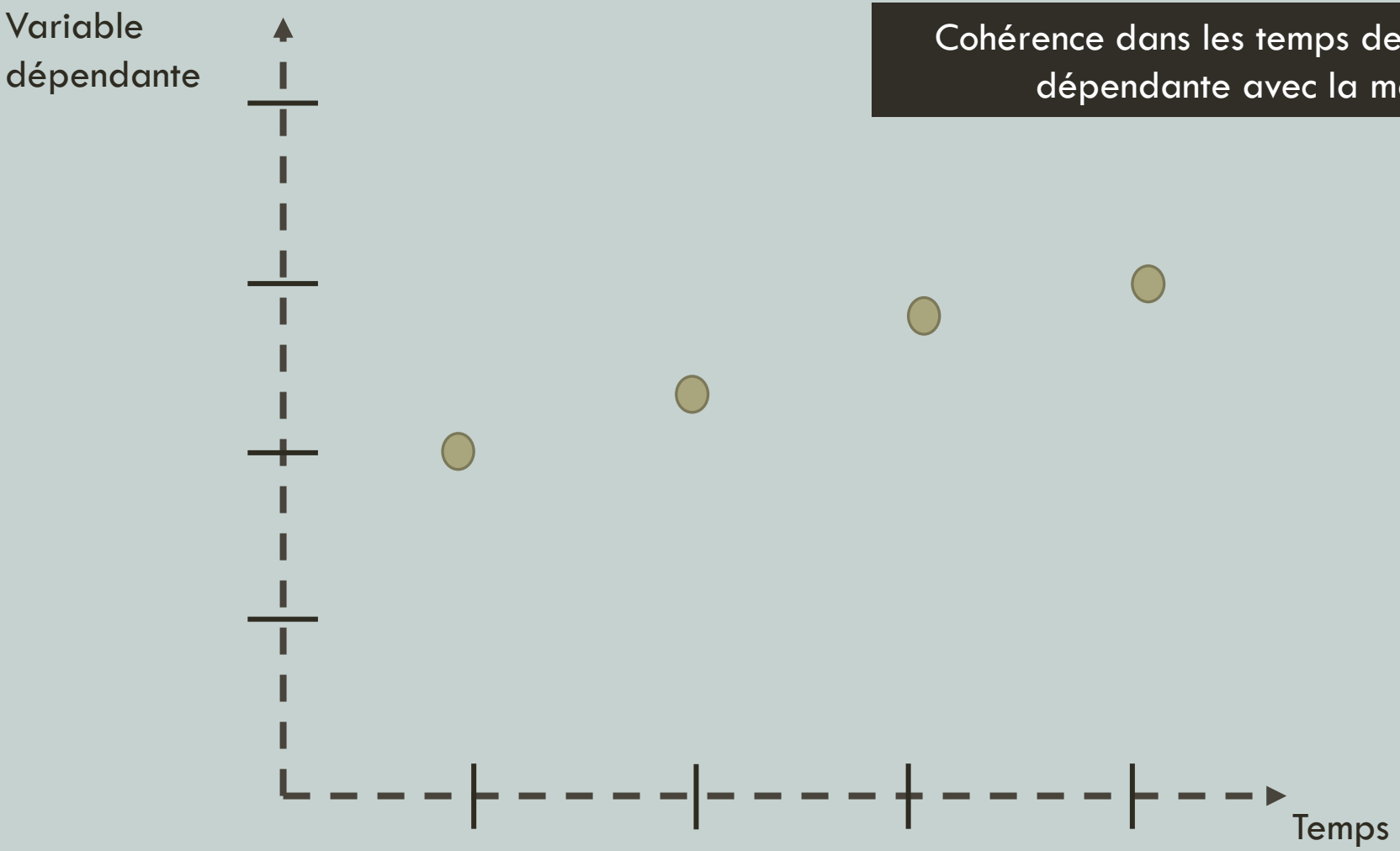
Variable dichotomique? → Distribution logistique

Taux? → Distribution de Poisson

Temps

ÉTAPE 1: METTRE LA TABLE

Cohérence dans les temps de mesure de la variable dépendante avec la métrique du temps



ÉTAPE 1: METTRE LA TABLE

Base de donnée en format large: Une ligne par participant

Identifiant

Variable dépendante

Métrique du temps

ID	Sex	Byr	Csmk01	Csmk02	Csmk03	Wez01	Wez02	Wez03	Yr01	Yr02	Yr03
001	0	1947	0	0	0	0	0	0	1992	1994	1999
002	1	1953	1	.	0	0	.	1	1992	1994	1999
003	0	1951	0	1	1	1	0	1	1992	1994	1999
004	0	1946	0	0	.	1	1	.	1992	1994	1999
005	1	1950	1	0	1	1	1	1	1992	1994	1999

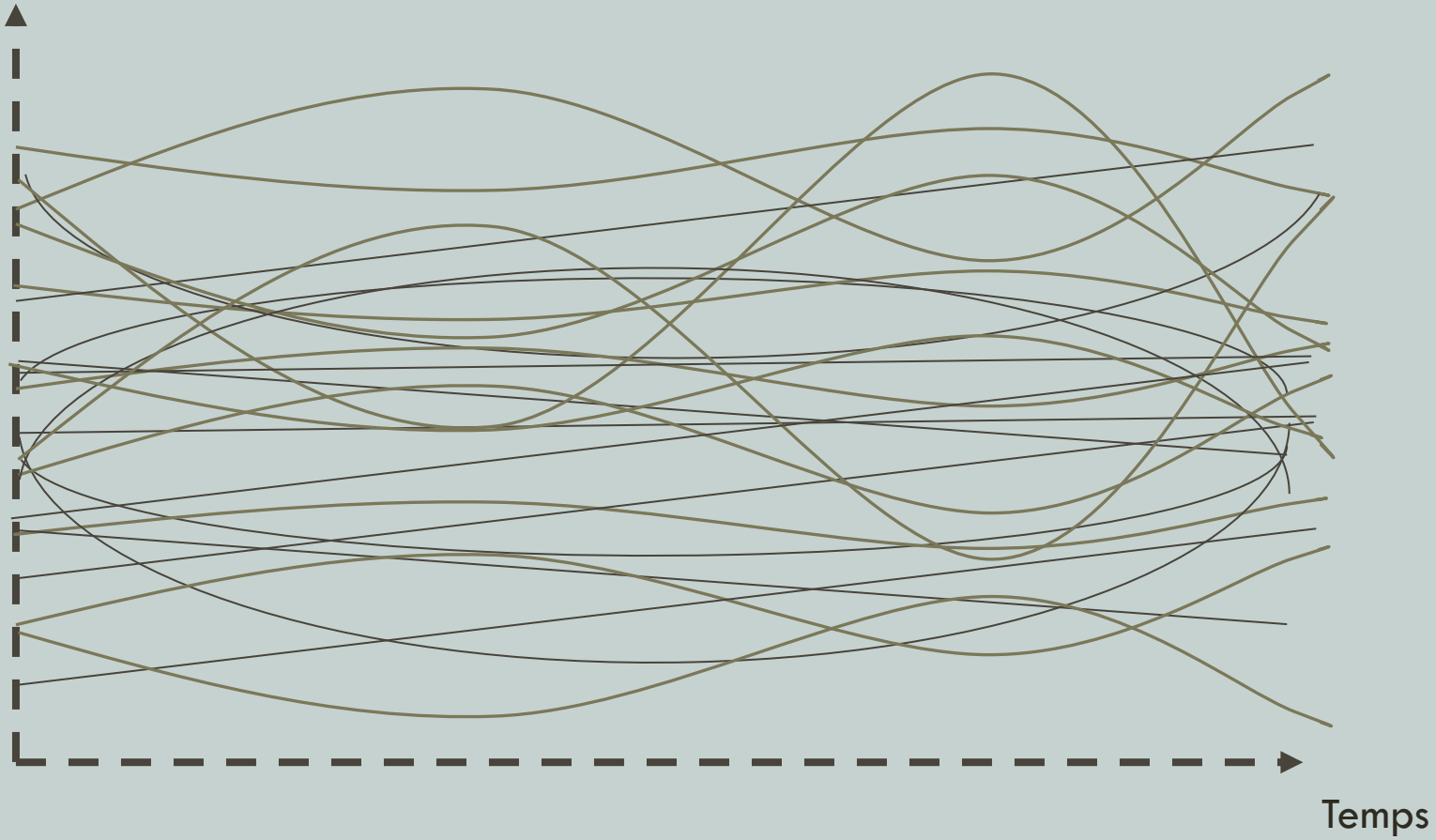
Ajustement de la variable dépendante selon la métrique du temps choisie

Exemple de Arrandale et al. 2006

ÉTAPE 1: METTRE LA TABLE

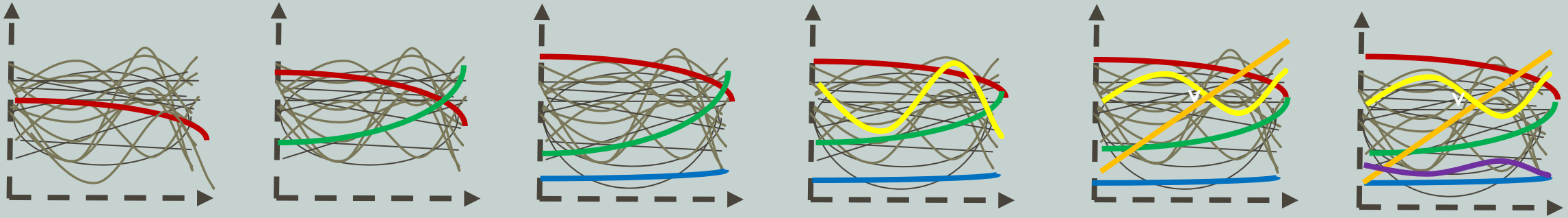
Variable dépendante

Réflexion: Groupes de trajectoires cliniquement plausibles?



ÉTAPE 2: NOMBRE DE CLASSES OPTIMALES

Tester différents scénarios de groupes de trajectoires



1 groupe | 2 groupes | 3 groupes | 4 groupes | 5 groupes | X groupes

Critères d'adéquation: Bayesian Identification Criteria (BIC)

Arrandale et al. 2006; Jones et al. 2001; Nagin 2005;

ÉTAPE 2: NOMBRE DE CLASSES OPTIMALES

Qu'est-ce que le BIC et comment l'utiliser?

Métrique d'adéquation du modèle: Basé sur la fonction de vraisemblance

Log-vraisemblance ajustée pour le nombre de paramètres et la taille d'échantillon.

$$BIC = k*\ln(n) + 2\ln(L)$$

K: Nombre de paramètres

n: Taille d'échantillon

L: Valeur maximale de la fonction de vraisemblance

Deux types de BIC:

-BIC selon le nombre trajectoires

-BIC selon le nombre total d'observations

BIC qui s'ajuste le mieux aux données: La plus petite valeur négative

Arrandale et al. 2006; Jones et al. 2001; Nagin 2005;

ÉTAPE 2: NOMBRE DE CLASSES OPTIMALES

Comment sélectionner le nombre de classes?

Différence entre les BIC de deux modèles (complexe vs modèle nul (1 classe)

2 x ▲BIC	Évidence vs Ho
0 à 2	Insignifiant
2 à 6	Positif
6 à 10	Forte
> 10	Très forte

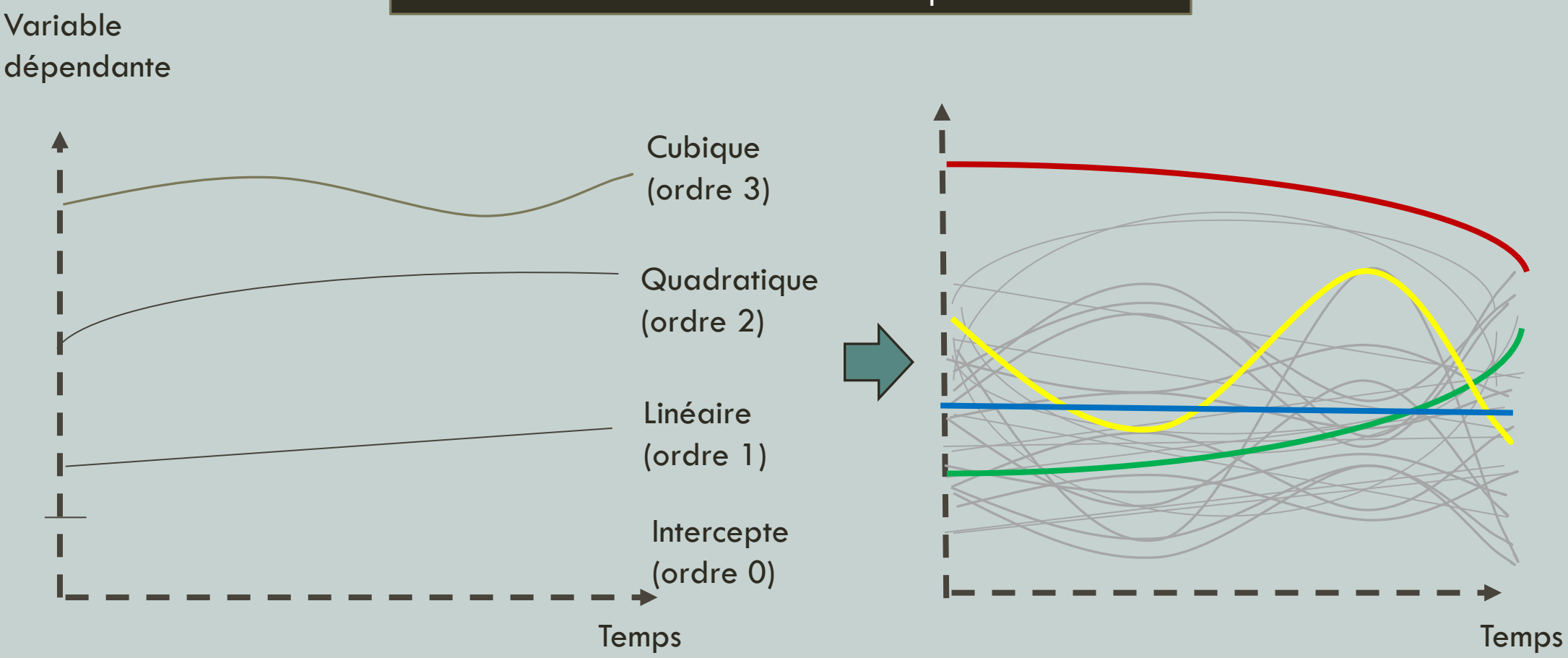
Calcul du facteur Bayes= $e^{BICi-BICj}$

Facteur Bayes	Interprétation
<1/10	Évidence forte pour modèle simple
1/10 à 1/3	Évidence moyenne pour modèle simple
1/3 à 1	Évidence faible pour modèle simple
1 à 3	Évidence faible pour modèle complexe
3 à 10	Évidence moyenne pour modèle complexe
>10	Évidence forte pour modèle complexe

Arrandale et al. 2006; Jones et al. 2001; Nagin 2005;

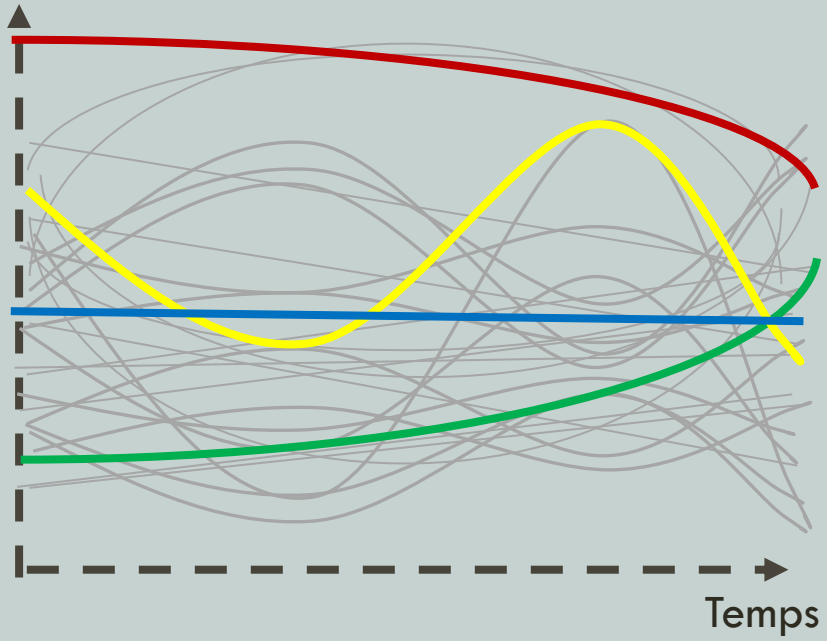
ÉTAPE 3: SPÉCIFICATION DU MODÈLE

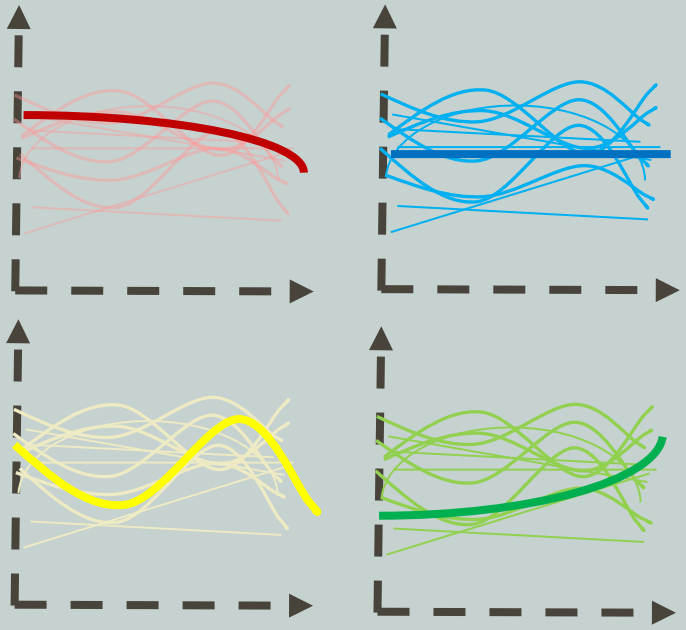
Formes fonctionnelles de chaque classe



ÉTAPE 4: CONFIRMATION DE L'APPARTENANCE

Probabilité postérieure moyenne





PROC TRAJ DANS SAS

PROCÉDURE PROC TRAJ

Développé par Nagin en 1999: Version semi-paramétrique de modèle par mixture discret

Produit de SAS/TOOLKIT: Procédure SAS adaptée pour PC

Doit être téléchargé individuellement, gratuit

Objectif: Identifier des classes d'individus partageant des progressions similaires de variables dépendantes selon le temps / âge

Paramètre des modèles: Maximum de vraisemblance

Jones & Nagin 2007

PROCÉDURE PROC TRAJ

$$f(\mathbf{y}) = \sum_{k=1}^{K} \underbrace{\Pr(C = k)}\underbrace{\Pr(\mathbf{Y} = \mathbf{y} | C = k)} = \sum_{k=1}^{K} p_k f(\mathbf{y}, \lambda_k).$$

Probabilité d'appartenir au groupe K

Estimé avec fonction logisitique multinomiale, puis normalisée à 0.

Probabilité de trajectoire individuelle, conditionnelle à l'appartenance au groupe K

Estimé selon une distribution plausible, selon les trajectoires

Jones & Nagin 2007

PROCÉDURE PROC TRAJ

```
PROC TRAJ statement-options;
    ID;
    VAR;
    INDEP;
    MODEL;
    ORDER;
    START;
    WEIGHT;
    RISK; REFGROUP;
    TCOV; PLOTTCOV;
    MIN; MAX; RORDER;      /* cnorm options */
    IORDER; EXPOS;         /* zip options */
    OUTCOME; OMODEL; OCOV; /* outcome model */
    DROPOUT; DCOV; OBSMAR; /* dropout options */
    MULTGROUPS; MULTRISK; /* multi-trajectory */
```

Jones & Nagin 2007

ÉTAPE 1: METTRE LA TABLE

Base de donnée en format large: Une ligne par participant

Identifiant

Variable dépendante

Métrique du temps

ID	Sex	Byr	Csmk01	Csmk02	Csmk03	Wez01	Wez02	Wez03	Yr01	Yr02	Yr03
001	0	1947	0	0	0	0	0	0	1992	1994	1999
002	1	1953	1	.	0	0	.	1	1992	1994	1999
003	0	1951	0	1	1	1	0	1	1992	1994	1999
004	0	1946	0	0	.	1	1	.	1992	1994	1999
005	1	1950	1	0	1	1	1	1	1992	1994	1999

Ajustement de la variable dépendante selon la métrique du temps choisie

Exemple de Arrandale et al. 2006

ÉTAPE 1. METTRE LA TABLE

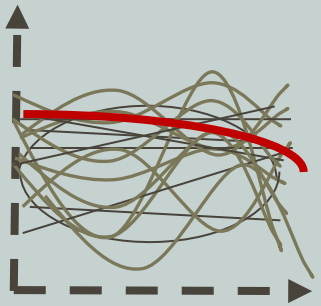
ID	Sex	Byr	Csmk01	Csmk02	Csmk03	Wez01	Wez02	Wez03	Yr01	Yr02	Yr03
001	0	1947	0	0	0	0	0	0	1992	1994	1999
002	1	1953	1	.	0	0	.	1	1992	1994	1999
003	0	1951	0	1	1	1	0	1	1992	1994	1999
004	0	1946	0	0	.	1	1	.	1992	1994	1999
005	1	1950	1	0	1	1	1	1	1992	1994	1999

Variable dichotomique: Logit

```
PROC TRAJ DATA=CLUBSAS2022;
ID ID;
    VAR Wez01-Wez03;
    INDEP Yr01-Yr03;
    MODEL LOGIT;
    NGROUPS X;
    ORDER X X X;
RUN;
```

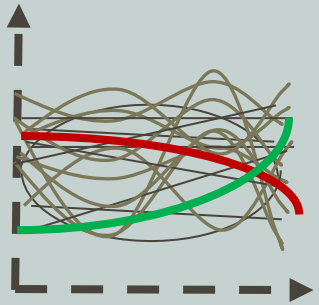
Exemple de Arrandale et al. 2006

ÉTAPE 2: NOMBRE DE CLASSES OPTIMALES



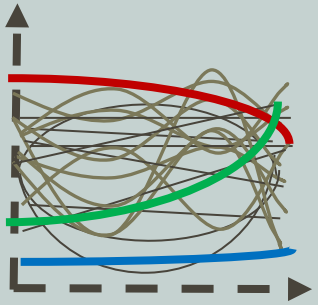
1 groupe (modèle de base)

```
PROC TRAJ DATA=
ID ID;
    VAR Wez01-Wez03;
    INDEP Yr01-Yr03;
    MODEL LOGIT;
    NGROUPS 1;
    ORDER 2;
RUN;
```



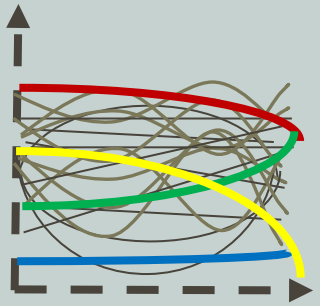
2 groupes

```
PROC TRAJ DATA=
ID ID;
    VAR Wez01-Wez03;
    INDEP Yr01-Yr03;
    MODEL LOGIT;
    NGROUPS 2;
    ORDER 2 2;
RUN;
```



3 groupes

```
PROC TRAJ DATA=
ID ID;
    VAR Wez01-Wez03;
    INDEP Yr01-Yr03;
    MODEL LOGIT;
    NGROUPS 3;
    ORDER 2 2 2;
RUN;
```



4 groupes

```
PROC TRAJ DATA=
ID ID;
    VAR Wez01-Wez03;
    INDEP Yr01-Yr03;
    MODEL LOGIT;
    NGROUPS 4;
    ORDER 2 2 2 2;
RUN;
```

Exemple de Arrandale et al. 2006

ÉTAPE 2: EXEMPLE DE SORTIE AVEC 2 CLASSES

Maximum Likelihood Estimates
Model: Logistic (LOGIT)

Group	Parameter	Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
1	Intercept	-3.43508	0.23871	-14.390	0.0000
	Linear	-0.08482	0.31429	-0.270	0.7873
	Quadratic	-1.97020	0.53759	-3.665	0.0003
2	Intercept	-0.83588	0.16420	-5.091	0.0000
	Linear	-0.61386	0.17273	-3.554	0.0004
	Quadratic	-1.65258	0.28794	-5.739	0.0000
Group membership					
1	(%)	80.10078	4.21882	18.987	0.0000
2	(%)	19.89922	4.21882	4.717	0.0000

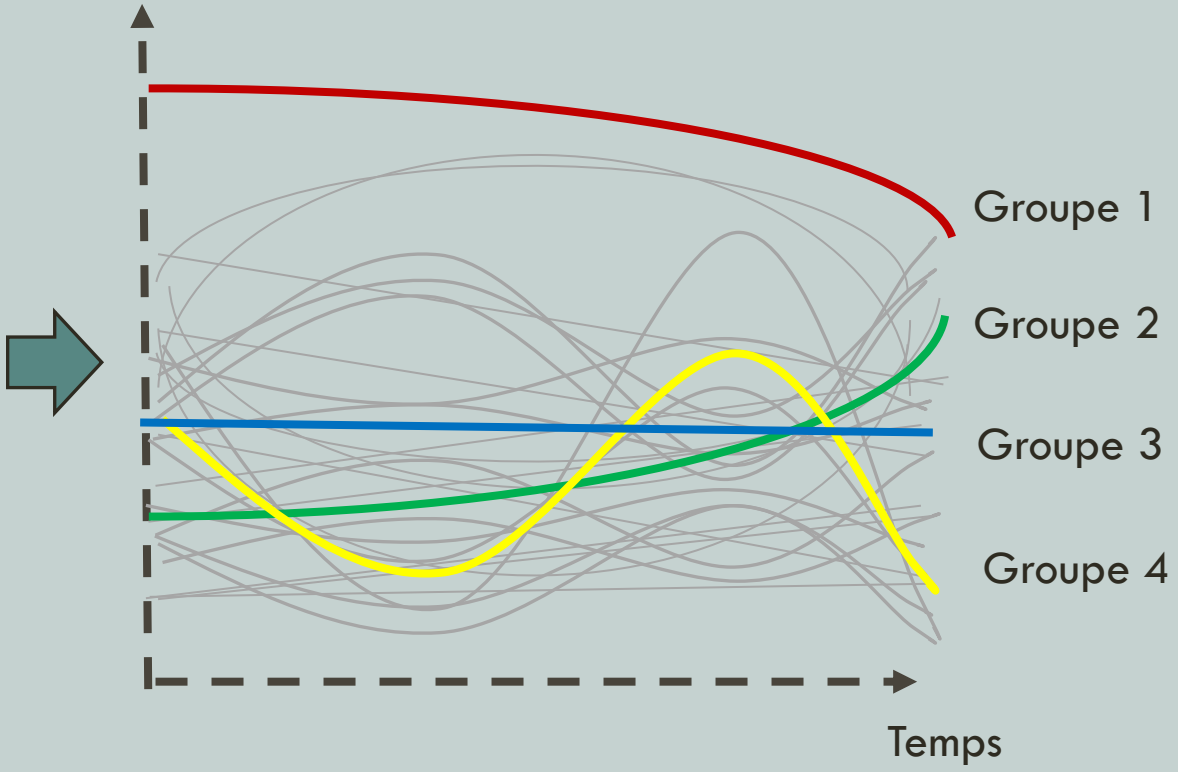
BIC= -818.24 (N= 4485) BIC= -807.27 (N= 195) AIC= -795.81 ll= -788.81

Exemple de Arrandale et al. 2006

ÉTAPE 3: SPÉCIFICATION DU MODÈLE

Formes fonctionnelles de chaque classe

```
PROC TRAJ DATA=CLUBSAS2022;
ID ID;
    VAR Wez01-Wez03;
    INDEP Yr01-Yr03;
    MODEL LOGIT;
    NGROUPS 4;
    ORDER 2 2 1 3;
RUN;
```



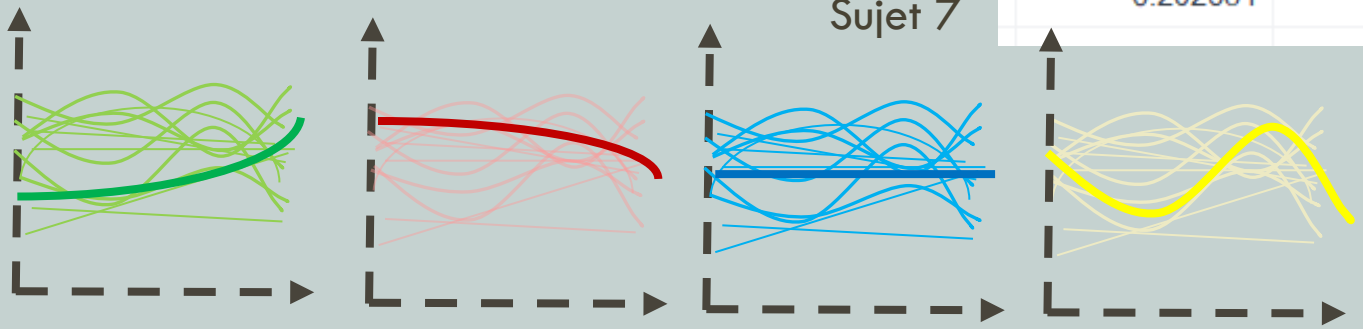
Exemple de Arrandale et al. 2006

ÉTAPE 4: CONFIRMATION DE L'APPARTENANCE

Probabilité postérieure d'appartenance

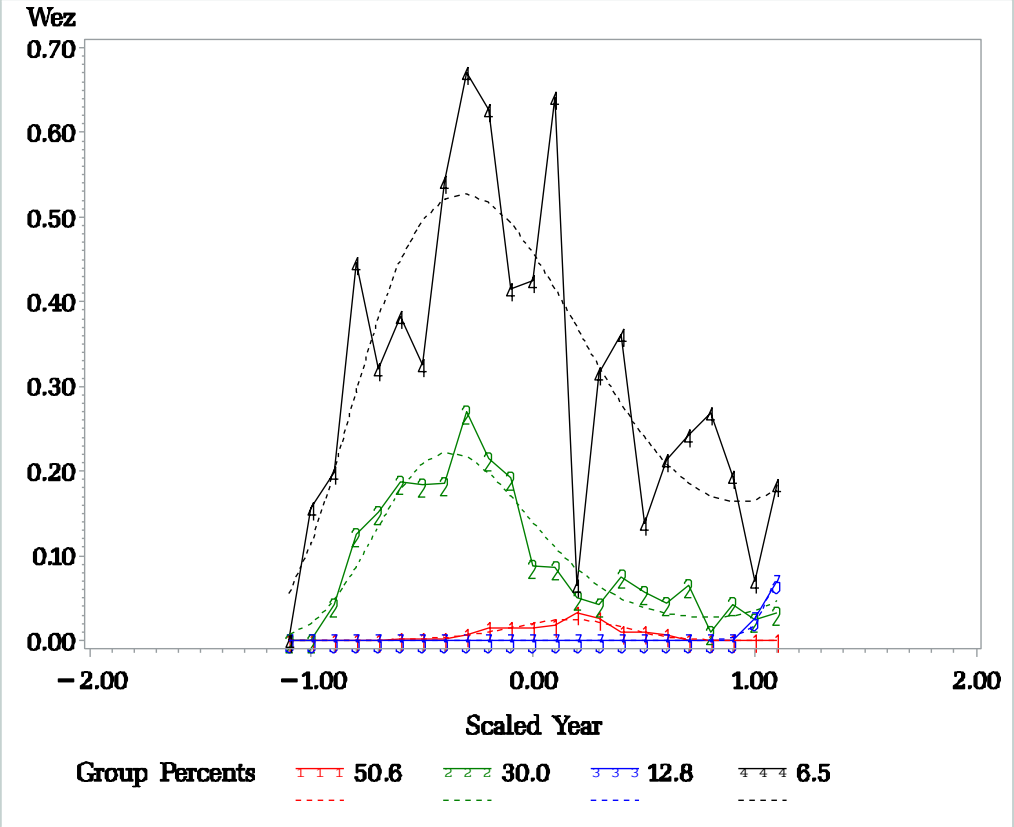
```
PROC TRAJ DATA=CLUBSAS2022 OUT=OF;
ID ID;
    VAR Wez01-Wez03;
    INDEP Yr01-Yr03;
    MODEL LOGIT;
    NGROUPS 4;
    ORDER 2 2 1 3;
RUN;
```

	GRP1PRB	GRP2PRB	GRP3PRB	GRP4PRB	GROUP
Sujet 1	0.860214	0.000000	0.139677	0.000109	1
Sujet 2	0.864054	0.000000	0.135840	0.000106	1
Sujet 3	0.545644	0.000000	0.453067	0.001289	1
Sujet 4	0.071444	0.000000	0.895643	0.032913	3
Sujet 5	0.231886	0.000000	0.760332	0.007782	3
Sujet 6	0.292634	0.696084	0.011280	0.000002	2
Sujet 7	0.292634	0.696084	0.011280	0.000002	2



ÉTAPE 5: PRÉSENTATION EN GRAPHIQUES

```
PROC TRAJ DATA=CLUBSAS2022 OUTPLOT=OP OUTSTAT=OS;
ID ID;
    VAR Wez01-Wez03;
    INDEP Yr01-Yr03;
    MODEL LOGIT;
    NGROUPS 4;
    ORDER 3 3 1 3;
RUN;
%TRAJPLOT(OP,OS,
'Wez vs. years','Logistic Model','Wez',
'Scaled years')
```



Exemple de Arrandale et al. 2006

PROCÉDURE PROC TRAJ: AJUSTEMENT

$$f(\mathbf{y}_i|\mathbf{z}_i,\mathbf{w}_i)=\sum_{k=1}^{K}\underbrace{\Pr(C_i=k|\mathbf{Z}_i=\mathbf{z}_i)}\underbrace{\Pr(\mathbf{Y}_i=\mathbf{y}_i|C_i=k,\mathbf{W}_i=\mathbf{w}_i)}.$$

Probabilité d'appartenir au groupe K

Estimé avec fonction logisitique multinomiale, puis normalisée à 0.

Z: Covariables fixes dans le temps

Probabilité de trajectoire individuelle, conditionnelle à l'appartenance au groupe K

Estimé selon une distribution plausible, selon les trajectoires

W: Covariables dépendantes du temps

Jones & Nagin 2007

AUTRES OPTIONS

```
PROC TRAJ DATA=CLUBSAS2022 OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE;
ID ID;
    VAR Wez01-Wez03;
    INDEP Yr01-Yr03;
    MODEL LOGIT;
    NGROUPS 4;
    ORDER 2 2 1 3;
 RISK SEX;
TCOV Csmk01-Csmk03;
RUN;

%TRAJPLOT(OP,OS, 'Wez vs. years', 'Logistic Model', 'Wez', 'Scaled years')
```

Ajout de covariables fixes dans le temps, conditionne la probabilité d'appartenir au groupe K

Exemple de Arrandale et al. 2006

AUTRES OPTIONS

```
PROC TRAJ DATA=CLUBSAS2022 OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE;
ID ID;
    VAR Wez01-Wez03;
    INDEP Yr01-Yr03;
    MODEL LOGIT;
    NGROUPS 4;
    ORDER 2 2 1 3;
 RISK SEX;
TCOV Csmk01-Csmk03;
RUN;

%TRAJPLOT(OP,OS, 'Wez vs. years', 'Logistic Model', 'Wez', 'Scaled years')
```

Introduire des variables dépendantes du temps pour estimer la probabilité de trajectoire individuelle, conditionnelle à l'appartenance aux groupes

Exemple de Arrandale et al. 2006

AUTRES OPTIONS

```
PROC TRAJ DATA=CLUBSAS2022 OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE;
ID ID;
    VAR Wez01-Wez03;
    INDEP Yr01-Yr03;
    MODEL LOGIT;
    NGROUPS 4;
    ORDER 2 2 1 3;
 RISK SEX;
TCOV Csmk01-Csmk03;
RUN;

%TRAJPLOT(OP,OS, 'Wez vs. years', 'Logistic Model', 'Wez', 'Scaled years')
```

Création d'une base de données avec les BIC et autres critères d'adéquation

Exemple de Arrandale et al. 2006

AUTRES OPTIONS

```
PROC TRAJ statement-options;
    ID;
    VAR;
    INDEP;
    MODEL;
    ORDER;
    START;
    WEIGHT;
    RISK; REFGROUP;
    TCOV; PLOTTCOV;
    MIN; MAX; RORDER;      /* cnorm options */
    IORDER; EXPOS;         /* zip options */
    MULTGROUPS; MULTRISK; /* multi-trajectory */
```

Jones & Nagin 2007

FIN DE LA PARTIE 1….

Pour en savoir plus:

Pour toute la documentation et le téléchargement : https://www.andrew.cmu.edu/user/bjones/index.htm

Jones, B. L. & Nagin, D. S. Advances in Group-Based Trajectory Modeling and an SAS Procedure for Estimating Them. *Sociol. Methods Res.* **35**, 542–571 (2007).

Jones, B. L. & Nagin, D. S & Roeder, K. A SAS Procedure Based on Mixture Models for Estimating Developmental Trajectories. *Sociol. Methods Res.* **29**, 374–393 (2001).

Arrandale, V., Koehoorn, M., MacNab, Y. & Kennedy, S. M. How to use SAS® Proc Traj and SAS® Proc Glimmix in Respiratory Epidemiology. (2006).