

Etude descriptive de données longitudinales

Ndèye Niang
Conservatoire National des Arts
et Métiers

STA112 N.Niang

Plan

- > Semaine1
 - > Introduction - Généralités sur les données
 - > Description simple de données longitudinales
 - Variabilité intra inter
 - Représentations graphiques (notion de trajectoire)
- > Semaine2 et Semaine3
 - > Méthodes d'analyse multidimensionnelle
 - > STATIS TUCKER PARAFAC
 - > Typologies de trajectoires

STA112 N.Niang

Introduction - définition

- Enquêtes **ponctuelles**
 - Elles visent à étudier des caractéristiques de personnes interrogées ou leurs opinions **à un moment donné** -> **un tableau de données**
- Données multivariées **longitudinales**
 - Recueil des valeurs de plusieurs variables sur un échantillon d'individus **à différents moments**

STA112 N.Niang

Introduction - définition

- Mesures répétées dans le temps -> plusieurs tableaux de données tableau à 3 entrées totalement rempli
- Mesures répétées
 - Mesure de variables sur des patients dans différentes conditions: assis, couché, après effort...
 - Cas particulier : mesure de variables sur des patients toutes les semaines -> données longitudinales

STA112 N.Niang

Introduction exemple 1

- Epidémiologie : on s'intéresse aux histoires personnelles c'est à dire à des successions d'événements qui concernent les personnes.
 - suivi médical de femmes volontaires afin de voir quelles sont celles qui développeront un cancer du sein ->
 - traitement rapide si cancer, données sur antécédent, mode de vie ce qui peut permettre de mieux comprendre les causes

STA112 N.Niang

Introduction exemple 2

- Le ministère de l'Éducation nationale enregistre les histoires scolaires d'un grand nombre d'élèves de manière à mieux connaître les circonstances de la réussite et de l'échec scolaire des jeunes. Dans ces projets les personnes sont suivies nominalement.

STA112 N.Niang

Exemple illustratif simple

http://www.ats.ucla.edu/stat/sas/examples/alda/chapter2/aldasasch2_v92.htm

- Etude sur adolescence: mesure de la résistance (tolérance) à l'exposition à un comportement déviant
- Mesures entre 11 et 15 ans des 2 variables et le sexe
- Données longitudinales avec 3 variables

STA112 N.Niang

Exemple illustratif simple

extrait d'un tableau de 16 observations

ID	TOL11	TOL12	TOL13	TOL14	TOL15	MALE	EXPOSURE
9	2.23	1.79	1.9	2.12	2.66	0	1.54
45	1.12	1.45	1.45	1.45	1.99	1	1.16
268	1.45	1.34	1.99	1.79	1.34	1	0.9
314	1.22	1.22	1.55	1.12	1.12	0	0.81
442	1.45	1.99	1.45	1.67	1.9	0	1.13
514	1.34	1.67	2.23	2.12	2.44	1	0.9
569	1.79	1.9	1.9	1.99	1.99	0	1.99
624	1.12	1.12	1.22	1.12	1.22	1	0.98
723	1.22	1.34	1.12	1	1.12	0	0.81
918	1	1	1.22	1.99	1.22	0	1.21
949	1.99	1.55	1.12	1.45	1.55	1	0.93
978	1.22	1.34	2.12	3.46	3.32	1	1.59
1105	1.34	1.9	1.99	1.9	2.12	1	1.38
1542	1.22	1.22	1.99	1.79	2.12	0	1.44
1552	1	1.12	2.23	1.55	1.55	0	1.04
1653	1.11	1.11	1.34	1.55	2.12	0	1.25

STA112 N.Niang

Autre présentation des données

ID	AGE	TOL	MALE	EXPOSURE
9	11	2.23	0	1.54
9	12	1.79	0	1.54
9	13	1.9	0	1.54
9	14	2.12	0	1.54
9	15	2.66	0	1.54
45	11	1.12	1	1.16
45	12	1.45	1	1.16
45	13	1.45	1	1.16
45	14	1.45	1	1.16
45	15	1.99	1	1.16
1653	11	1.11	0	1.25
1653	12	1.11	0	1.25
1653	13	1.34	0	1.25
1653	14	1.55	0	1.25
1653	15	2.12	0	1.25

STA112 N.Niang

Autre présentation des données

■ Un programme sas

```
data tolerance_pp;
set tolerance;
array Atol [11:15] tol11-tol15;
do age=11 to 15;
tol = Atol[age];
time = age - 11;
output;
end;
keep id age time tol male exposure;
run;
```

STA112 N.Niang

Autre exemple illustratif

Kroonenberg 2007 Applied multiway data analysis

■ Données multivariées longitudinales:

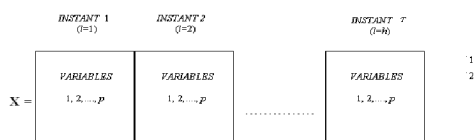
- Croissance corporelle de jeunes filles françaises
- Indice i 1 à I : individus = 30 jeunes filles
- Indice j 1 à J : Variables : Poids, taille, distance taille-coccyx, tour de poitrine, circonférence avant bras, mollet gauche, tête, largeur maximale du pelvis
- Indice k 1 à K : temps = 12 années de 4 à 15 ans

Blocs de $30 \times 8 \times 12$ données longitudinales

STA112 N.Niang

Présentation générale des données

■ FAT DATA: juxtaposition des tableaux par instant



- Le temps n'est pas considéré comme une vraie variable

■ TALL DATA: superposition des tableaux

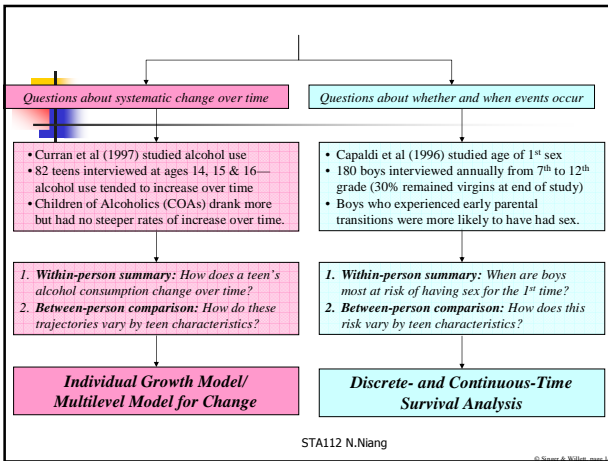
- On prend le temps comme variable

STA112 N.Niang

Méthodes d'analyse de données longitudinales

- Plusieurs objectifs ou questions sur
 - l'évolution au cours du temps
 - suivi individuel, étude de changement dans le temps des caractéristiques des individus
 - Comparaison des évolutions entre individus
 - la survenue et la date d'un événement
 - Variabilité intra et variabilité inter
- Résumé cf Singer J.D:

STA112 N.Niang



STA112 N.Niang

Analyse de données longitudinales

- Méthodes exploratoires
 - Réduction de dimension, synthèse descriptive
 - Typologies
- Méthodes de modélisation
 - Modèles de régression
 - Modèles mixtes
 - ...

STA112 N.Niang

Deux classes de méthodes

- Distinction entre 2 types de méthodes d'analyse
 - Celles qui ne tiennent pas compte de l'information longitudinale : **cross-sectional methods**, chaque tableau de mesures est traité comme s'il provenait d'un ensemble d'individus différents
 - Celles qui tiennent compte de l'information longitudinale, la répétition: **Méthodes à trois voies : (three-way methods)** : suivi de l'individu

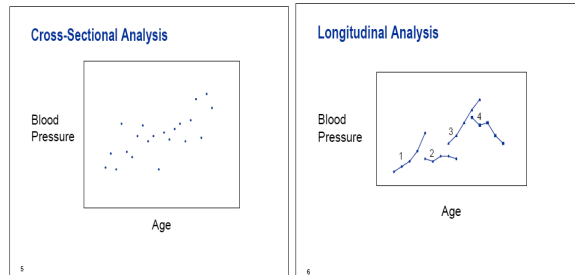
STA112 N.Niang

Prise en compte du temps

- Insuffisance des cross sectional methods
- Étude de l'évolution des phénomènes dans le temps
- Eviter des interprétations fausses
- Mais besoin de méthodes spécifiques : corrélation entre mesures répétées
- Exemple:

STA112 N.Niang

Liaison positive cas 1 et cas 2
mise en évidence différence selon les individus cas 2 seul



STA112 N.Niang

Méthodes à trois voies:

- Plusieurs techniques en fonction du rôle du temps:
 - Modélisation par un processus stochastique
 - Modèles descriptifs faisant intervenir le temps
 - Modèles utilisant le temps pour l'interprétation
 - On se limite aux modèles utilisant le temps pour l'interprétation

STA112 N.Niang

Analyse exploratoire de données longitudinales

- Données multidimensionnelles ->
 - Etape 1
 - Analyses descriptives simples préalables
 - Unidimensionnelles / Bidimensionnelles
 - Résumés numériques et graphiques
 - Etape 2 analyses multidimensionnelles

STA112 N.Niang

Description simple

- Etudes unidimensionnelles
 - Méthodes graphiques :
 - Nuage de points
 - Courbes individuelles, lissage
 - Courbes moyennes par groupes
 - Résumés numériques

STA112 N.Niang



Description simple

- **Recommendations** Diggle, Liang, and Zeger (1994)
- Graph as much of the relevant raw data as possible.
- Highlight aggregate patterns of potential scientific interest.
- Identify both cross-sectional and longitudinal patterns.
- Identify unusual individuals or observations.

STA112 N.Niang



Description simple

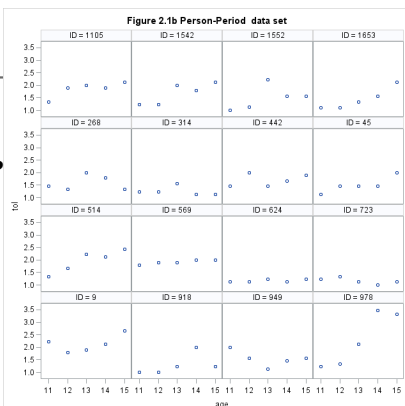
- Représentation de chaque variable en fonction du temps pour chaque individu séparément:
- Exemple adolescence

STA112 N.Niang

Exemple adolescence



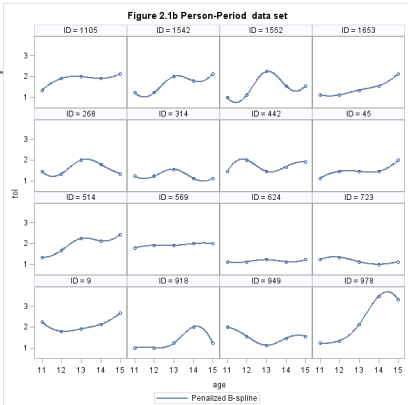
Commentaires?



Exemple adolescence trajectoires



Commentaires?



Description simple

- Représentation simultanée sur un graphique
- Visualisation trajectoire moyenne
- >
- **Profils individuels, graphique très utile**
 - Identification de tendance générale commune
 - Détection de changements non linéaire dans le temps
 - Informations sur la variabilité à des moments donnés
 - Détection de patterns communs à des groupes d'individus
 - Informations sur la variabilité entre les individus

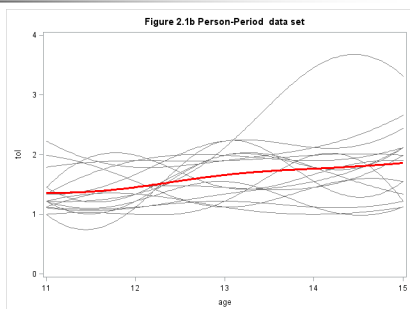
STA112 N.Niang

Exemple de profils individuels



Si beaucoup d'individus
Pas très informatif

Visualisation trajectoire moyenne



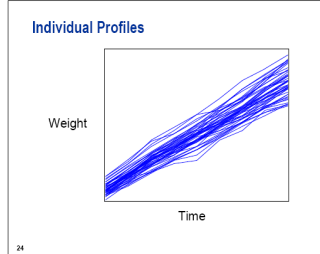
STA112 N.Niang

Autre exemple de profils individuels

Analyse:
Augmentation du poids pour
tous les individus

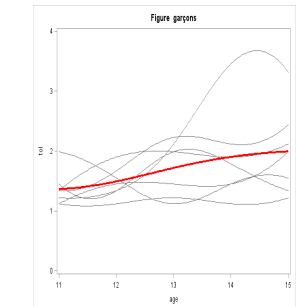
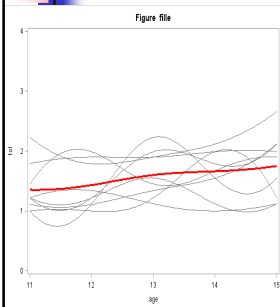
Tendance linéaire: les plus lourds
au début le restent à la fin

Augmentation de la variabilité
Entre individus (plus faible au début)



STA112 N.Niang

profils moyen de groupes



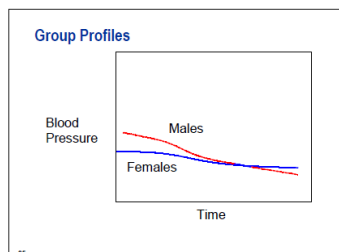
STA112 N.Niang

profils moyen de groupes

Analyse:
Décroissance de la
pression Sanguine
pour les hommes
Comme pour les femmes

Mais pente plus
prononcée pour
les hommes

Présence d'interaction?



STA112 N.Niang

Description simple

■ Etudes bidimensionnelles

- Nuage de points Coefficient de corrélation
- Prise en compte aspect longitudinal, non indépendance des mesures

Table 2.1: Estimated bivariate correlations among tolerance scores assessed on five measurement occasions ($n = 16$)

	<i>TOL11</i>	<i>TOL12</i>	<i>TOL13</i>	<i>TOL14</i>	<i>TOL15</i>
<i>TOL11</i>	1.00				
<i>TOL12</i>	0.66	1.00			
<i>TOL13</i>	0.06	0.25	1.00		
<i>TOL14</i>	0.14	0.21	0.59	1.00	
<i>TOL15</i>	0.26	0.39	0.57	0.83	1.00

STA112 N.Niang

Description simple

■ Etudes bidimensionnelles

- Etude comparative des variables dans le temps
- Variance intra inter

- Etude plus détaillée sur l'exemple de croissance jeunes françaises

STA112 N.Niang

Conclusions partie 1

- Etude préalable indispensable comme dans la démarche classique faire une idée sur les données
- mais attention particulière à la dimension temporelle, suivi des individus, évolution dans le temps, éviter interprétation erronée.
- Insuffisante -> analyse multidimensionnelle

STA112 N.Niang



Programme SAS

```
/* mélange de points*/
proc sgpanel data = tolerance_pp; panelby id /columns=4 rows= 4;
scatter y = tol x = age; run;

/* Lissage trajectoires*/
proc sgpanel data = tolerance_pp; panelby id /columns=4 rows= 4;
pbspline y = tol x = age; run;

proc sgplot data=tolerance_pp noautolegend ;
yaxis min = 0 max = 4;
pbspline x=age y=tol
/group = id nomarkers LINEATTRS = (COLOR= gray PATTERN = 1 THICKNESS = 1) ;
pbspline x=age y=tol
/nomarkers LINEATTRS = (COLOR= red PATTERN = 1 THICKNESS = 3) ;
run;
quit;
```

STA112 N.Niang
