

Introduction à l'Analyse de Données Longitudinales

Pr Roch Giorgi

 roch.giorgi@univ-amu.fr

Introduction

- Etude longitudinale
 - ✓ Le même échantillon (panel) est utilisé de façon répétée
- Données longitudinales
 - ✓ Mesures sur un même sujet effectuées plusieurs fois (au moins 2) au cours du temps
 - ✓ On parle de mesures répétées
- Exemple simple
 - ✓ Mesure de la pression artérielle systolique (PAS) avant et après traitement
- Exemples plus complexes
 - ✓ Mesure de la PAS tous les 3 mois pendant 2 ans chez tous les sujets, à des temps différents selon les sujets,...

Avantages des Etudes Longitudinales

- Economise le nombre de sujets
- Sujets sont leurs propres témoins
- Variation intra-sujet éliminée de l'erreur
- Peut fournir des estimateurs plus efficaces que lors d'études transversales
- Peut séparer l'effet lié à l'âge (évolution individuelle au cours du temps) de l'effet cohorte (différence entre les sujets à la baseline)
- Peut fournir des informations sur les évolutions individuelles

Défis des Etudes Longitudinales

- Observations non indépendantes
 - ✓ Nécessité de le prendre en compte la corrélation entre les mesures
- Mélange de données fixes (sexe, centre,...) et de données évoluant au cours du temps
- Phénomène d'attrition possible
 - ✓ Perte de sujets de l'échantillon au cours du temps
- Données pouvant être incomplètes
 - ✓ Nécessité de gérer ces données manquantes
- Données pouvant avoir une dimension à plusieurs niveaux

Exemple

- Marqueurs principaux de l'infection par le VIH
 - ✓ Charge virale
 - ✓ Lymphocytes T CD4+
- Evolution de ces marqueurs au cours du temps largement étudiée
 - ✓ Sous traitement antirétroviral
 - ✓ Pour étudier les déterminants de cette évolution

Représentation des Données (1)

- Format large

Sujet	Temps 1	Temps 2	...	Temps n
1	y_{11}	y_{12}	...	y_{1n_1}
2	y_{21}	y_{22}	...	y_{2n_2}
...			...	
N	y_{N1}	y_{N2}	...	y_{Nn_N}

Représentation des Données (2)

- Format long

Sujet	Temps	Réponse
1	1	y_{11}
1	2	y_{12}
...
1	n_1	y_{1n_1}
...
...
N	1	y_{N1}
N	2	y_{N2}
...
N	n_N	y_{Nn_N}

Notations

- Sujets : $i = 1, \dots, N$
- Observations : $j = 1, \dots, n_i$ (n pour étude équilibrée)
- Nb total observations = $\sum_i^N n_i$
- Vecteur de réponses : $y_i = n_i \times 1$
- Vecteur de covariable du sujet i au temps j : x_{ij}
 - ✓ Covariables non dépendantes du temps (entre sujets)
 - ✓ Covariables dépendantes du temps (intra sujets)
- Matrice de covariables du sujet i (avec intercept) :
 $\mathbf{X}_i = n_i \times p$

Représentation des Données (3)

- Format long

Sujet	Temps	Réponse	Var 1	...	Var p
1	1	y_{11}	x_{111}	...	x_{11p}
1	2	y_{12}	x_{121}	...	x_{12p}
...
1	n_1	y_{1n_1}	x_{1n_11}	...	x_{1n_1p}
...
...
N	1	y_{N1}	x_{N11}	...	x_{N1p}
N	2	y_{N2}	x_{N21}	...	x_{N2p}
...
N	n_N	y_{Nn_N}	x_{Nn_N1}	...	x_{Nn_Np}

Analyses : Généralités (1)

- Variable réponse
 - ✓ Continue (normale ou non-normale)
 - ✓ Catégorielle (binaire, ordinale, nominale, comptage)
- Nombre de sujets N
- Nombre d'observations par sujet n_i
 - ✓ $n_i = 2$ pour tous : ANCOVA (analyse changement de score)
 - ✓ $n_i = n$ pour tous : étude équilibrée, ANOVA ou MANOVA pour mesures répétées
 - ✓ n_i varie : méthodes plus générales

Analyses : Généralités (2)

- Nombre et type de covariables – $E(y_i)$
 - ✓ Un échantillon
 - ✓ Plusieurs échantillons
 - ✓ Régression (covariables continues ou catégorielles)
 - ✓ Covariables dépendantes du temps
- Nature de la structure de la matrice de variance-covariance – $V(y_i)$
 - ✓ Variances homogènes ou hétérogènes
 - ✓ Covariances homogènes ou hétérogènes

Analyses : Méthodes

- Variable dérivée : réduction des observations observées dans une variable résumée
 - ✓ Moyenne au cours du temps
 - ✓ Evolution d'un score
 - ✓ Tendance linéaire au cours du temps
 - ✓ Dernière observation
- Analyse longitudinale
 - ✓ ANOVA pour mesures répétées
 - ✓ MANOVA pour mesures répétées
 - ✓ Modèles à effets mixtes
 - ✓ Equations d'estimation généralisée