

Cet enseignement de statistique pour sciences sociales de niveau master comprend un cours de 2 heures et un séminaire obligatoire de 2 heures qui forment un tout (6 crédits).

Ce cours est consacré à l'analyse statistique de données longitudinales et plus particulièrement à l'analyse de l'histoire d'événements (Event History Analysis), connue aussi sous le nom d'analyse des biographies ou encore d'analyse de la survie. L'objectif est de déterminer les facteurs influençant le temps qui s'écoule entre deux événements, entre le mariage et la première naissance par exemple, ou de façon équivalente le risque de changer d'état après une durée donnée dans un état. En premier lieu, le cours introduit les notions centrales de fonction de survie et de risque instantané (hazard rate) ainsi que la problématique des données tronquées. Il présente ensuite les modèles descriptifs du type Kaplan-Meier. La suite du cours porte sur la modélisation des risques en fonction de facteurs explicatifs. On discute en particulier le modèle semi-paramétrique en temps continu de Cox qui permet d'introduire des facteurs variant au cours du temps, ainsi que la modélisation en temps discret.

L'accent est mis sur la pratique de la modélisation avec des logiciels statistiques et sur l'interprétation des résultats fournis par les logiciels. Les étudiants pourront s'exercer sur des données réelles mises à leur disposition par l'enseignant.

Organisation de l'enseignement

L'enseignement comprend 4 heures hebdomadaires.

Cours	André Berchtold	vendredi	10h15 - 12h	Mail 5250
Séminaire	Jonathan Zufferey	lundi	12h15 - 14h	Mail 5383

Dans le cadre du séminaire, les étudiants seront appelés à mettre en œuvre les méthodes discutées au cours sur des jeux de données réelles. Les étudiants seront à cette fin initiés aux possibilités d'analyse longitudinale offertes par le logiciel statistique R.

Evaluation

- Réalisation d'un travail de recherche, individuellement ou par groupe de deux.
- Examen oral individuel portant sur le travail de recherche.

L'évaluation vaut pour l'ensemble du cours et du séminaire et permet d'obtenir 6 crédits.

Coordonnées et heures de réception

		bureau	téléphone	e-mail
André Berchtold	vendredi 9h-10h ou sur rdv	5224	022 379 81 72	<i>andre.berchtold@unil.ch</i>
Jonathan Zufferey	lundi 14h15 - 15h ou sur rdv	5237	022 379 92 45	<i>jonathan.zufferey@unige.ch</i>

Site web : ***http://www.andreberchtold.com***

Plan du cours

Données longitudinales et modèles de survie

1 Introduction

2 Données longitudinales et censure

- Données longitudinales : nature et organisation des données
- Événements multiples et à répétition
- Fonctions de densité, de survie et de risque
- Données censurées

3 Distributions empiriques de la durée

- Exemples de données à analyser
- Temps discret : table actuarielle de survie
- Temps continu : méthode de Kaplan-Meier et Nelson-Aalen
- Comparaison de distributions de durées de survie

4 Modèles en temps continu de Cox

- Modèles semi-paramétriques à risques proportionnels (Cox)
- Evaluation et sélection de modèles
- Estimation de la fonction de survie
- Modèles stratifiés
- Prise en compte de facteurs variant dans le temps
- Test de l'hypothèse de risques proportionnels

5 Modèles de régression en temps discret

- Données personnes périodes
- Formulation sous forme de régression logistique : modèle à ratios de cotes (*odd ratios*) constants
- Prise en compte de facteurs variant dans le temps

6 Distributions théoriques

- Distribution exponentielle
- Distribution de Weibull
- Distribution de Gompertz
- Distribution log-logistique

7 Extensions (en fonction du temps à disposition)

- Événements à répétition
- Événements multiples
- Modèles paramétriques de régression en temps continu
- Approches multi-niveau
- ...

Séminaire

Dans le cadre du séminaire, les étudiants seront appelés à faire des exercices et à réaliser un travail de recherche. Les documents et données nécessaires seront mis à disposition des étudiants par l'enseignant sur son site internet : <http://www.andreberchtold.com>.

Références bibliographiques

- ALLISON, P. D. (1982). Discrete-time methods for the analysis of event histories. In S. Lienhardt (Ed.), *Sociological Methodology*, pp. 61–98. San Francisco : Jossey-Bass Publishers.
- ALLISON, P. D. (1984). *Event history analysis, Regression for longitudinal event data*. QASS 46. Beverly Hills and London : Sage.
- BLOSSFELD, H.-P., A. HAMERLE, AND K. U. MAYER (1989). *Event history analysis, Statistical theory and application in the social sciences*. Hillsdale NJ : Lawrence Erlbaum.
- BLOSSFELD, H.-P. AND G. ROWHER (2002). *Techniques of Event History Modeling, New Approaches to Causal Analysis* (2nd ed.). Mahwah NJ : Lawrence Erlbaum.
- COLLETT, D. (2003). *Modelling Survival Data in Medical Research*. London : Chapman and Hall, 2nd ed.
- COURGEAU, D. AND É. LELIÈVRE (1989). *Analyse démographique des biographies*. Paris : Editions de l'INED.
- FLINN, C. J. AND J. J. HECKMAN (1982). New methods for analyzing individual event histories. In S. Lienhardt (Ed.), *Sociological Methodology*, pp. 99–140. San Francisco : Jossey-Bass Publishers.
- HOSMER, D. W. AND S. LEMESHOW (1999). *Applied Survival Analysis, Regression Modeling of Time to Event Data*. New York : Wiley.
- KLEINBAUM, D. G. AND KLEIN, M. (2005). *Survival Analysis, a Self-learning Text*, 2nd edition. New York : Springer-Verlag.
- LELIÈVRE, É. ET A. BRINGÉ (1998). *Manuel pratique pour l'analyse statistique des biographies. Présentation des modèles de durée et utilisation des logiciels SAS, TDA et STATA*. Paris : INED.
- MILLS, M. (2011). *Introducing Survival and Event History Analysis*. London : Sage.
- REHER, D. S. AND R. SCHOFIELD (Eds.) (1993). *Old and New Methods in Historical Demography*. Oxford : Clarendon Press.
- RITSCHARD, G. AND M. ORIS (2005). Life Course Data in Demography and Social Sciences : Statistical and Data Mining Approaches. in R. Lévy, P. Ghisletta, J.-M. Le Goff, D. Spini et E. Widmer (eds) *Towards an Interdisciplinary Perspective on the Life Course*, Advances in Life Course Research, Vol. 10. Amsterdam : Elsevier, pp. 283-314.
- ROHWER, G. AND PÖTTER, U. (2009). *Basic Notions of Event History Analysis*. Disponible gratuitement à l'adresse suivante : <http://www.stat.ruhr-uni-bochum.de/pub/eha/deha1e.pdf>
- RUSPINI, E. (2002). *Introduction to Longitudinal Research*. London : Routledge.
- SINGER, J. D. AND J. B. WILLETT (2003). *Applied longitudinal data analysis : Modeling change and event occurrence*. Oxford : Oxford University Press.
- TUMA, N. B. (1982). Nonparametric and partially parametric approaches to event-history analysis. In S. Lienhardt (Ed.), *Sociological Methodology*, pp. 1–60. San Francisco : Jossey-Bass Publishers.
- YAMAGUCHI, K. (1991). *Event history analysis*. ASRM 28. Newbury Park and London : Sage.

Consignes pour le travail de recherche

Le travail peut être fait seul ou par groupe de deux. Il doit démontrer votre capacité à mettre en œuvre les méthodes descriptives et de modélisation enseignées durant le cours.

Le travail doit être remis au plus tard le dimanche 1er juin 2014. Il doit être envoyé sous format électronique par email à l'enseignant et à l'assistant.

Le corps du texte devrait comprendre entre 8 et 15 pages, plus éventuellement quelques annexes présentant les outputs les plus utiles. Le corps du texte ne doit pas inclure des copies d'outputs, mais uniquement des tableaux synthétiques.

La structure du document pourrait par exemple être la suivante :

1. **Introduction** (1 page). Description de la problématique considérée, des hypothèses que vous souhaitez tester et énumération succincte des données et méthodes qui seront utilisées. L'introduction peut aussi annoncer les principaux enseignements qu'apporte votre étude.
2. **Données** (1-2 pages). Il s'agit de donner toutes les indications utiles sur les données : sources, population concernée, nombre de cas, définition précise des variables retenues, des codages et des éventuels recodages et filtres utilisés.
3. **Analyse exploratoire** (1-2 pages). Tableaux synthétiques et commentaire des distributions univariées des variables explicatives, avec indication notamment du nombre de valeurs manquantes. Analyse préliminaire par table de survie et estimation de fonctions de survie.
4. **Modélisation utilisant le modèle de Cox** (2-5 pages). Spécification initiale du modèle avec justification. Explication de la démarche de modélisation suivie (modèle, variables variant avec le temps, essais de plusieurs spécifications, ...). Présentation des résultats quantitatifs et évaluation statistique.
5. **Interprétation et discussion** (2-4 pages) des résultats par rapport à la problématique.
6. **Conclusion** (1 page). Récapitulation de la démarche suivie et des principaux résultats.
7. **Bibliographie** (1 page). Liste des références consultées (livres, articles, manuels, sites web, ...) classées par ordre alphabétique de leur auteur.

De plus, les étudiants doivent joindre à leur travail l'ensemble des fichiers de syntaxe R utilisés pour le réaliser.

Remarques :

- Cette structure est uniquement donnée à titre indicatif!
- N'oubliez pas, pour les tableaux et figures, de mettre des libellés compréhensibles et de donner, le cas échéant, toutes les indications utiles à leur lecture!
- La recherche effectuée doit être *originale*. Il ne s'agit en aucun cas de recopier des résultats existants. Si un tel cas devait se produire, l'étudiant concerné se verrait refuser les crédits du cours et son nom serait transmis au décanat.
- L'originalité est volontiers récompensée!
- Une des caractéristiques d'un travail de bonne qualité est d'être *reproductible*, ce qui signifie qu'à partir des éléments fournis (données utilisées, explications statistiques, etc ...), une personne lisant le travail doit pouvoir reproduire les analyses effectuées et bien sûr obtenir les mêmes résultats numériques.