

INFORMATIONS GENERALES

Le cours Statistique IV donné durant le semestre d'hiver 2011-2012 est consacré à l'analyse des relations entre variables catégorielles. Nous commencerons par étudier différentes mesures d'association plus performantes que le chi-2 et permettant de mettre en évidence un éventuel lien entre deux variables. Nous introduirons ensuite le concept de modélisation des tables de contingence et nous le déclinerons sous différentes formes (modèles log-linéaires, modèles logits, régression logistique). Finalement, une méthode de présentation graphique des données catégorielles, l'analyse factorielle des correspondances, sera présentée.

Ce cours est destiné à toutes les personnes désirant se perfectionner dans le domaine de l'analyse des données catégorielles multivariées. Il est nécessaire d'avoir suivi au préalable un cours d'introduction à la statistique et de connaître les notions de base de l'inférence statistique. Une part importante du cours sera dédiée aux applications informatiques sur de vraies données catégorielles à l'aide du logiciel statistique R.

Ce cours comprend 4 heures d'enseignement hebdomadaire durant un semestre. Le cours est partagé entre présentation théorique de différentes méthodes, exemples et applications pratiques. Une partie du cours a lieu en salle informatique. Les étudiants sont fortement encouragés à appliquer les méthodes vues durant le cours sur leurs propres ensembles de données.

Divers compléments au cours sont fournis sur le site web suivant : **www.andreberchtold.com**. Ces compléments incluent notamment le travail demandé pour chaque semaine, des livres et articles complétant le cours, des références bibliographiques et des liens vers d'autres sites dédiés aux statistiques.

Horaire

		jour	heure	salle
Cours	André Berchtold	lundi	13h15–17h	Anthropole 3077 & 4120

Heures de réception

	jour	bureau	téléphone	email
André Berchtold	jeudi 12h–14h ou sur rdv	Anthropole 4128	021 692 3123	Andre.Berchtold@unil.ch
Mélanie Glasson-Cicognani	mercredi 10h30–12h30 ou sur rdv	Anthropole 4125	021 692 3124	Melanie.Glasson@unil.ch

Evaluation

Le cours permet d'obtenir 6 crédits. L'évaluation du cours consiste en un travail de recherche et un examen écrit. La note finale est constituée de la moyenne arithmétique simple, arrondie au demi-point, des notes du travail de recherche et de l'examen. Une note finale supérieure ou égale à 4 donne droit aux crédits liés au cours.

L'examen est un travail écrit de 4 heures portant sur l'ensemble de la matière traitée durant le cours. Il est organisé durant les sessions habituelles d'examens de la faculté des SSP. La documentation est libre à l'exception des ordinateurs.

Plan du cours

1. Introduction
2. Mesures d'association
3. Modèles log-linéaires
4. Modèles logit
5. Régression logistique
6. Analyse factorielle des correspondances

Références

- AGRESTI, A. (2002) *Categorical Data Analysis*, 2nd edition. Wiley-Interscience.
- AGRESTI, A. (2007) *An Introduction to Categorical Data Analysis*, 2nd edition. Wiley-Interscience.
- BISHOP, Y.M.M., FIENBERG, J.E., HOLLAND, P.W. (2007) *Discrete Multivariate Analysis*. Springer.
- CLAUSEN, S.-E. (1998) *Applied Correspondence Analysis : An Introduction*. Quantitative Applications in the Social Sciences (QASS), 121, Sage University Papers, London.
- DEMARIS, A. (1993) *Logit Modeling : Practical Applications*. Quantitative Applications in the Social Sciences (QASS), 86, Sage University Papers, London.
- KNOKE, D., BURKE, P.J. (1980) *Log-Linear Models*. Quantitative Applications in the Social Sciences (QASS), 20, Sage University Papers, London.
- LIEBETRAU, A.M. (1983) *Measures of Association*. Quantitative Applications in the Social Sciences (QASS), 32, Sage University Papers, London.
- LONG, J. SCOTT (1997) *Regression Models for Categorical and Limited Dependent Variables*. Advanced Quantitative Techniques in the Social Sciences, SAGE Publications, London.
- MENARD, S.W.. (2001) *Applied Logistic Regression Analysis*. Quantitative Applications in the Social Sciences (QASS), 106, Sage University Papers, London.

Travail de recherche

Principe

Ce travail consiste en la réalisation d'une analyse statistique complète sur un ensemble de données catégorielles au moyen des outils théoriques et informatiques présentés durant le cours. Il peut être réalisé soit individuellement, soit par groupes de 2 personnes. Il sera évalué par une note comprise entre 1 et 6.

Démarche

1. Sélectionner un ensemble de données comportant environ 10 variables catégorielles. Il est possible de créer des variables catégorielles à partir de variables numériques en les catégorisant. Le nombre d'observations doit être suffisant pour permettre des analyses fiables.
2. Formuler une ou plusieurs hypothèses de travail.
3. Analyser les données en utilisant les outils appropriés.
4. Synthétiser les résultats et écrire un rapport comportant *au minimum* les éléments suivants :
 - (a) Une introduction complète mentionnant clairement les hypothèses et les objectifs du travail.
 - (b) Une description des données utilisées avec l'indication précise de leur source.
 - (c) Une synthèse des principaux résultats numériques obtenus. (Les listings informatiques complets peuvent éventuellement être annexés au travail.)
 - (d) Une interprétation pertinente et critique des résultats.
 - (e) Une conclusion synthétisant les principaux résultats obtenus et les mettant en relation avec les hypothèses de départ.

Remarques

1. L'originalité (lorsqu'elle est de qualité) est volontiers encouragée !
2. Une analyse statistique peut être un processus constitué de nombreuses étapes. Si l'on n'obtient pas tout de suite des résultats intéressants, il peut être nécessaire de reformuler des hypothèses, utiliser d'autres méthodes, etc ...
3. L'une des caractéristiques d'un travail de bonne qualité est d'être *reproductible*, ce qui signifie qu'à partir des éléments fournis (données, explications statistiques, etc ...), une personne lisant le travail doit pouvoir reproduire les analyses effectuées et bien sûr obtenir les mêmes résultats numériques.
4. La recherche effectuée doit être *originale*. Il ne s'agit en aucun cas de recopier des résultats trouvés dans un article ou un livre. Toute tentative de tricherie sera automatiquement sanctionnée de la note minimale et dénoncée auprès des autorités universitaires.

Reddition des travaux

La date limite de reddition est fixée au **lundi 12 décembre 2011 à minuit**. A cette date, les étudiants devront avoir envoyé par email à l'enseignant les documents suivants :

- Un fichier pdf contenant le rapport final ainsi que ses éventuelles annexes.
- Le fichier de données utilisé au format R.
- Le ou les fichiers de syntaxe R utilisés pour effectuer les analyses.