

PLAN DE COURS

STT 6415 - Régression - 3 crédits (Automne 2011)

Professeur(e): Pierre Lafaye de Micheaux

Coordonnées: Bureau: 4249

Courriel: lafaye@dms.umontreal.ca

Téléphone: 514-343-6607

Web: <http://www.biostatisticien.eu/moodle>

Disponibilités: Jeudi 13h-15h

1 Description succincte du cours

Rappels sur la régression linéaire multiple. Diagnostics. Transformations, moindres carrés pondérés, méthodes robustes, régression « ridge ». Régression non linéaire. Modèles spécifiques : logistique, probit, de Poisson.

2 Horaire et format du cours

Cours théoriques : 39h

Activité	Jour	De - À	Du - Au	Nb séances	Local	Immeuble
Cours	Jeudi	09:00-12:00	1 septembre - 20 octobre	8	4186	A.-AISENSTADT
			3 novembre - 1 décembre	5		

Chaque séance de cours prendra la forme d'un exposé magistral assuré par le professeur. Ces séances seront parsemées de périodes de résolution d'exercices. Suivant la disponibilité des ressources informatiques, des séances de travaux pratiques sur ordinateur (TP) pourraient être organisées. Ces séances permettent à l'étudiant de mettre en pratique, à l'aide de l'outil informatique (logiciel **R**), les notions théoriques vues en cours.

3 Approche pédagogique

L'orientation pédagogique du cours privilégie les méthodes d'apprentissage et les modes d'évaluation combinés. La première méthode pédagogique repose sur l'exposé magistral (exposé professoral hebdomadaire). Accompagnés de périodes de résolution d'exercices en classe, les exposés magistraux feront acquérir des connaissances théoriques et empiriques et un bagage conceptuel qui constitueront le cœur de notre enseignement. La deuxième méthode est la mise en pratique des apprentissages par les étudiants, d'une part lors des séances de travaux pratiques, et d'autre part dans la réalisation des devoirs maison et du projet. Cette mise en pratique impliquera plusieurs niveaux cognitifs visés par notre enseignement : compréhension, application des connaissances acquises, analyse et synthèse, évaluation (sens critique).

L'étudiant pourra également utiliser certains outils mis à disposition sur le site internet associé à ce cours : forum de discussion, notes de cours, etc.

4 Évaluation des apprentissages

Devoirs à faire à la maison et/ou examens en classe : 60%

Un projet final : 40%

Le détail des épreuves :

- *Devoir maison* : Un devoir maison doit être réalisé **seul**. Tout **retard non justifié**¹ dans la remise du devoir entraînera la non correction de celui-ci et la note 0 sera alors donnée pour ce devoir.
Un devoir maison permet de s'assurer que l'on maîtrise bien les éléments théoriques et pratiques vus en cours et en TP. Par ailleurs, il s'agit également d'une occasion pour l'étudiant de développer sa maîtrise de la rédaction d'un texte. Ainsi, il sera demandé à l'étudiant de soigner la présentation du travail rendu et la qualité de l'argumentation. L'étudiant utilisera les logiciels \LaTeX et **R** pour réaliser son devoir.
- *Examens* : Un premier examen en classe, de nature plutôt théorique, pourrait porter sur la régression linéaire multiple. Un autre examen en classe pourrait comprendre un mélange de théorie et d'applications et porter sur les modèles pour données catégorielles et sur les modèles linéaires généralisés.
- *Projet* : Le projet sera réalisé en équipe de **deux personnes maximum**, cet effectif pouvant toutefois être modulé par le professeur en fonction du nombre d'étudiants dans le groupe. Une recherche sur un sujet relié à la régression sera entreprise et les résultats seront présentés sous la forme d'un article rédigé en \LaTeX et d'une présentation orale effectuée à l'aide de l'outil **Beamer**. La lecture d'articles de revues scientifiques sera nécessaire. Les détails de ce projet seront disponibles sur le site web associé à ce cours.

Critères généraux d'évaluation :

- Contenu : acquisition des connaissances et de leur application adéquate, capacité d'analyse et de synthèse, cohérence argumentaire, capacité de relier les sujets entre eux, capacité d'évaluer, de faire preuve d'un sens critique appuyé par les connaissances.
- Forme : respect des directives afférentes aux diverses épreuves (travail, examen), respect des règles de présentation matérielle (page de couverture, prescriptions de longueur, etc.).
- Qualité rédactionnelle : la qualité du français (syntaxe, lexique, orthographe, grammaire, ponctuation), la clarté d'expression.
- Ponctualité : le respect des délais prescrits.

Remarques importantes :

- La présence en classe est fortement recommandée pour bien réussir ce cours!
- Tout plagiat² entraînera de graves sanctions.

5 Objectifs

La régression permet de modéliser la relation qui peut exister entre une (ou plusieurs) variable(s) que l'on tente d'expliquer et une ou plusieurs autres variables explicatives. Par exemple, quelle est la relation entre la consommation d'essence et différentes autres variables telles que la taxe sur l'essence, la proportion de gens avec un permis de conduire, etc. ou encore, quelle est la relation entre le fait de guérir ou non d'une maladie et l'âge du patient, son sexe et le fait qu'il fume ou non. Dans un premier cours, on s'intéresse généralement à la régression linéaire, c'est-à-dire que la relation entre les variables est linéaire. Dans ce deuxième cours, nous allons voir certains des aspects les plus poussés de la régression linéaire multiple de même que certaines généralisations où, par exemple, la variable à expliquer n'est plus continue mais est plutôt binaire. La régression offre plusieurs défis intéressants. Par exemple, une seule observation peut-elle grandement influencer les résultats obtenus? Comment doit-on choisir les variables qui feront partie du modèle? Doit-on vérifier l'influence d'une observation avant ou après avoir choisi les variables dans le modèle? Les réponses à ces

¹Sur la définition du motif valable de retard d'absence à une évaluation, consulter le guide pédagogique de la FAS. <http://www.etudes.umontreal.ca/reglements/fas.html>

²Voir <http://www.fas.umontreal.ca/plagiat>

questions sont rarement uniques ! Ce cours s'adresse à tout étudiant de maîtrise ou de doctorat en statistique qui veut se familiariser davantage avec certains des outils fondamentaux de la modélisation statistique de même qu'avec certains des développements les plus récents.

Objectifs généraux :

- L'étudiant maîtrisera les aspects essentiels, de même que les limites, de la théorie et de la pratique des méthodes de régression et les appliquera dans des situations réelles.
- L'étudiant se familiarisera avec le processus de recherche.

Objectifs particuliers :

À la fin du cours, l'étudiant devrait être en mesure de :

- résoudre les problèmes reliés à la théorie des estimateurs de régression ;
- comprendre et détecter les dangers associés à certaines méthodes de régression ;
- juger du bien-fondé des hypothèses émises et modifier le modèle en conséquence ;
- choisir un modèle de régression ;
- interpréter correctement les résultats d'une analyse effectuée sur un jeu de données réelles ;
- faire une simulation en **R** afin d'évaluer une méthode ou encore comparer deux méthodes ;
- écrire un mini-article dans le cadre d'un projet ;
- donner une présentation orale.

6 Plan du cours

- Régression linéaire multiple :
 - Estimation et inférence dans les modèles de régression linéaire multiple (RLM) ;
 - Étude des résidus et influence des observations en RLM ;
 - Transformations et colinéarité en RLM ;
 - Sélection de modèle.
- Modèles pour données catégorielles et modèles linéaires généralisés.

7 Bibliographie : ouvrages de référence *recommandés*

Certains de ces ouvrages sont disponibles à la bibliothèque Maths-Info : <http://www.bib.umontreal.ca/MI>

Ouvrages principaux :

- Lafaye de Micheaux, P., Drouilhet, R. et Liquet, B. (2010). *Le logiciel R - Maîtriser le langage - Effectuer des analyses statistiques*, Springer. <http://www.springerlink.com/content/978-2-8178-0114-8>
- Cornillon, P.-A. et Matzner-Lober, E. (2007). *Régression. Théorie et applications*, Springer. <http://www.springerlink.com/content/978-2-287-39692-2>
- Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Advanced Quantitative Techniques in the Social Science Series. Sage.
- McCullagh, P. et Nelder, J.A. (1989). *Generalized Linear Models*. Second Edition. Monographs on Statistics and Applied Probability 37. Chapman and Hall.

Ouvrage secondaires :

- Greene, W. (2005). *Économétrie*. 5^{ème} édition. Pearson Education.
- Draper, N.R. et Smith, H. (1998). *Applied Regression Analysis*. 3rd édition. Wiley Series in Probability and Statistics.
- Hastie, T. et Tibshirani, R. et Friedman, J. (2001). *The elements of Statistical Learning. Data Mining, Inference, and Prediction*. Springer Series in Statistics.
- Maronna, R.A. et Martin, R.D. et Yohai, V.J. (2006). *Robust Statistics. Theory and Methods*. Wiley Series in Probability and Statistics.