Université du Québec à Montréal
Faculté des sciences
Département de mathématiques

PLAN DE COURS

STT5100 **Modèles linéaires appliqués**

Automne 2019

(3 crédits)

Professeur : Arthur Charpentier
Professeur, Département de mathématiques
Bureau : PK-5330
Courriel : charpentier.arthur@uqam.ca
Url : https://github.com/freakonometrics/STT5100
Blog : https://freakonometrics.hypotheses.org

Cours préalables : ACT3035 – Laboratoire d'Actuariat
ACT4400 – Modèles de survie

Horaire du cours : Vendredi 09h00-12h00 (SH-2140) - cours
Vendredi 12h00-14h00 (SH-2140) - exercices

Matériel obligatoire : cf https://github.com/freakonometrics/STT5100

Disponibilité: Prendre rendez-vous par courriel (charpentier.arthur@uqam.ca)

Description du cours

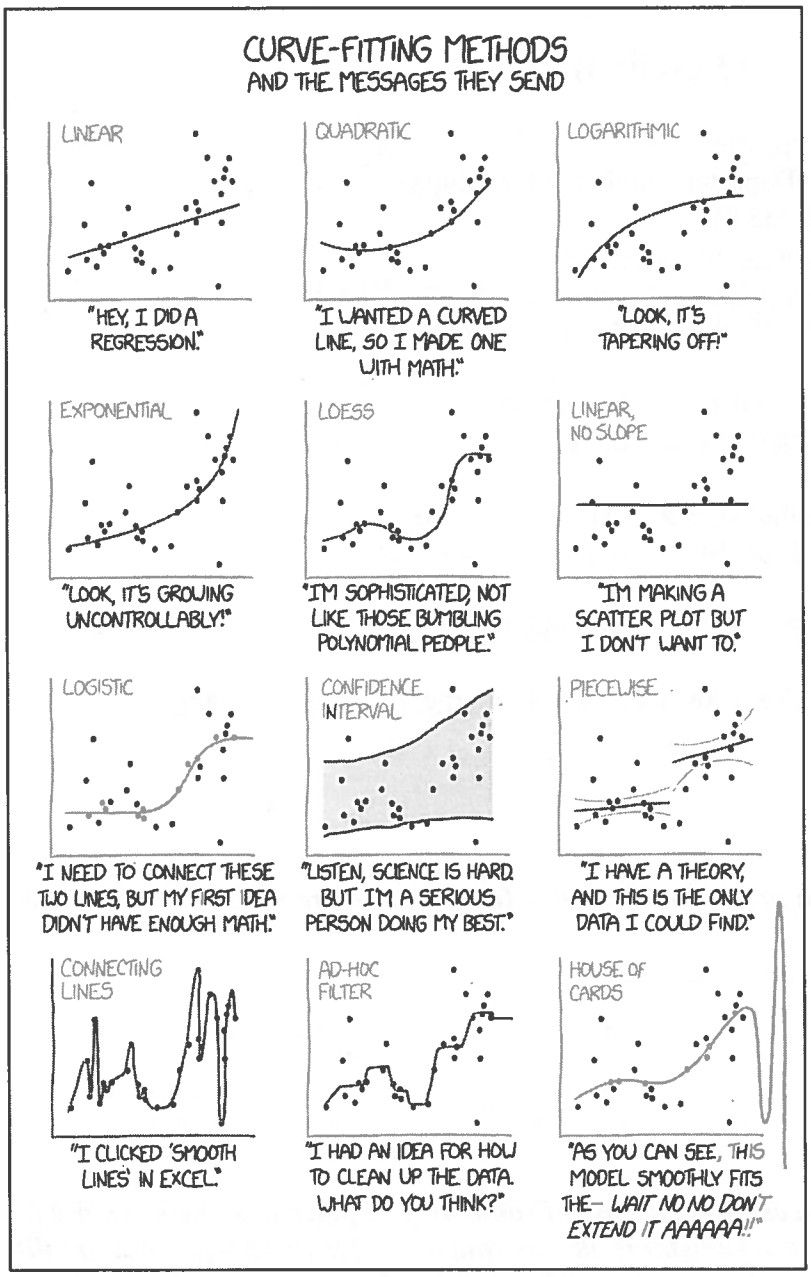
"***In theory there is no difference between theory and practice. In practice there is***" disait Yogi Berra

APPLIQUER, Verbe [via http://www.cnrtl.fr/definition/appliquer]
I.– [*Appliquer* st un verbe à double constr.]
2. Utiliser des principes, des règles abstraites dans un cas particulier, pour régler un problème concret. Faire passer du domaine de la théorie à celui de la pratique :
12. En supposant même que ces observations puissent être négligées dans la recherche des vérités spéculatives, doivent-elles l'être, lorsqu'il s'agit d'**appliquer** ces vérités à la pratique, et de déduire de la science l'art qui en doit être le résultat utile? Condorcet, *Esquisse d'un tableau hist. des progrès de l'esprit hum.*, 1794, p. 12.

« ***Modeling, in the broadest sense, is the cost-effective use of something in place of something else for some cognitive purpose. It allows us to use something that is simpler, safer or cheaper than reality instead of reality for some purpose. A model represents reality for the given purpose; the model is an abstraction of reality in the sense that it cannot represent all aspects of reality*** »
Rothenberg *et al.* (1989). The nature of modeling.

Un modèle est une abstraction, une « ***simplification de la réalité*** ». Construire des modèles appliqués, sur des vraies données, est donc forcément compliqué. Cela suppose de maitriser la théorie sous-jacente, afin d'avoir une théorie à appliquer. Et forcément, on le verra, les modèles simplistes donneront des résultats souvent médiocres. Le cours est ambitieux, et demandera un investissement personnel important de la part de tous les étudiants.

Sommaire du contenu : Méthode d'estimation par moindre carrés, Régression linéaire simple et multiple, Tests d'hypothèses et intervalles de confiance, tests d'ajustement, introduction aux modèles linéaires généralisés (GLM): régression logistique, régression Poisson, régression Gamma. Applications numériques à l'aide de R (via https://etudier.uqam.ca/cours?sigle=STT5100)



Via. https://xkcd.com/2048/

Contenu :

1. **Rappels de probabilités, de statistiques, d'optimisation** (~5% du temps)
   Espérance, Variance, Espérance Conditionnelle, Variance Conditionnelle, Loi Gaussienne, loi de Poisson, loi Gamma, famille exponentielle, vecteur Gaussien, introduction aux méthodes de simulation (Monte Carlo), maximum de vraisemblance et théorie asymptotique, tests basés sur la vraisemblance, optimisation numérique, optimisation avec contraintes (linéaires), rappels d'algèbre linéaire
   (support de cours en ligne sur github)
2. **Modèle linéaire Gaussien** (~45% du temps)
   Variables continues et variables catégorielles, Méthode des moindres carrés, Maximum de vraisemblance, Propriétés sous hypothèse de normalité, théorème de Gauss-Markov, Significativité, Prévision et Intervalle de confiance, Simulation dans le modèle linéaire et incertitude, Choix de modèle et choix de variables (stepwise), Moindres carrées pondérés, Introduction aux méthodes non-linéaire : approche locale et décomposition dans une base (splines), modèle additif vs. multiplicatif et transformation de Box-Cox, diagnostique et analyse des résidus, tests de normalité
   Bonus : méthode de pénalisation et lasso (sélection de variable)
3. **Régression logistique et méthodes de classification** (~20% du temps)
   Maximum de vraisemblance, méthodes numériques d'estimation, logistique vs. Probit, Incertitude dans le modèle logistique et prévision
   Bonus : introduction aux arbres de classification
4. **Régression de Poisson** (~15% du temps)
   Maximum de vraisemblance, méthodes numériques d'estimation, équi-dispersion, offset, méthode des marges et tableaux de contingences
5. **Modèles GLM** (~15% du temps)
   Régression dans la famille exponentielle, maximum de vraisemblance, déviance et analyse des résidus, choix de modèles
   Bonus : quasi-vraisemblance et sur-dispersion

Évaluation

Petit rappel : comme le disait Charles Goodhart, membre du Comité de politique monétaire de la Banque d'Angleterre, "***when a measure becomes a target, it ceases to be a good measure***". Autrement dit, les examens ne sont pas l'objectif du cours. Les examens servent juste à vérifier que la matière présentée a été travaillée, et comprise.

Autre point important : le but du cours est de vous donner quelques outils pour votre boite à outils de *modèles prédictifs*, d'expliquer comment ils fonctionnent, dans quelles conditions, et de vous apprendre a les utiliser. Comme il s'agit du premier cours, on va rester modestes ! Une partie de l'évaluation reposera sur des projets : je vous donne des vraies données, et vous demande de construire un modèle pour prédire une variable d'intérêt. Le but est de montre que vous savez trouver le bon outil parmi ceux que nous aurons vu. Utiliser un outil qui n'a pas été présenté en cours méritera une explication…

Description des évaluations

- Examen #1 **le 1er novembre** (3 heures : 33.33% de la note finale) : modèle linéaire
- Examen #2 **le 13 décembre** (3 heures : 33.33% de la note finale) : GLM

- Travail de session #1, à remettre **le 4 novembre à midi** (16.67% de la note finale)
- Travail de session #2, à remettre **le 20 décembre à minuit (23:59)** (16.67% de la note finale)
    - Les instructions au sujet du travail seront en ligne sur la page https://github.com/freakonometrics/STT5100

Afin de se conformer aux règles du programme d'agrément universitaire de l'Institut Canadien des actuaires, ***les travaux de session seront à faire individuellement.*** Je demanderais

- un .rmd (markdown) qui compile, et qui génère un .html : tout fichier qui ne compile pas entrainera un 0/100
- un .html, produit par le .rmd

(voir plus loin pour des références sur les fichiers markdown et RStudio)

A titre d'exemple,

- les ***examens*** pour la session d'automne 2018 sont en ligne, intra et correction, final et correction, et les énoncés pour les ***devoirs*** de la même session sont en ligne devoir1 et devoir2.
- les ***examens*** pour la session de l'hiver 2019 sont en ligne, intra et correction, final et correction, et les énoncés pour les ***devoirs*** de la même session sont en ligne devoir1 et devoir2.

Politique de reprise d'examen

Aucune reprise d'examen ne sera permise, à moins de conditions exceptionnelles. La reprise d'examen n'est pas un droit mais un privilège. Dans de très rares occasions où une reprise aura été autorisée par le professeur et par le département, l'examen de reprise sera beaucoup plus difficile que l'original pour tenir compte du temps d'étude supplémentaire disponible, par souci d'équité pour les autres étudiants.

Remarques supplémentaires

La note finale (en lettre, A+, A, etc.) pour le trimestre sera attribuée en fonction de l'atteinte des objectifs spécifiques à travers les trois évaluations. La distribution des résultats dans le groupe pourrait aussi être utilisée. Aucune autre opportunité (travail supplémentaire, etc.) d'augmenter le nombre de points ne sera allouée. La grille suivante est donnée ***à titre indicatif.***

Note	Min Inclus	Max exclus	Note	Min inclus	Max exclus
A+	90	100	B+	76	80
A	85	90	B	73	76
A-	80	85	B-	70	73

Note	Min inclus	Max exclus	Note	Min inclus	Max exclu
C+	66	70	D+	55	60
C	63	66	D	50	55
C-	60	63	E	0	50

Seules les calculatrices acceptées par la *Society of Actuaries* seront admises pour les examens i.e. les modèles BA-35, BA II Plus, TI-30X, TI-30Xa, TI-30XIIS et TI-30XIIB.

Bibliographie

Des « *notes de cours* » seront en ligne sur https://github.com/freakonometrics/STT5100

Pour les étudiants intéressés, les livres suivants pourraient être pertinents (pour ce cours et d'autres cours du baccalauréat).

- William H. Greene, 2011, **Econometric Analysis**, Prentice Hall
- James H. Stock & Mark W. Watson, 2007, **Introduction to Econometrics**, Addison Wesley
- Arthur Charpentier & Michel Denuit, 2005, **Mathématiques de l'Assurance-Non Vie, tome 2**, Economica. *chapitre 9:tarification a priori (en ligne)*
- Colin Cameron & Pravin K. Trivedi, 2013, **Regression Analysis of Count Data**, Cambridge University Press
- Piet de Jong, & Gillian Heller, 2008, **Generalized Linear Models for Insurance Data**, Cambridge University Press
- Esbjörn Ohlsson & Björn Johansson, 2010, **Non-Life Insurance Pricing with Generalized Linear Models**, Springer Verlag

Ces livres et sites contiennent des exercices pour se pratiquer…

- Julian, J. Faraway, 2002, **Practical Regression and Anova using R**, CRAN
- *Study tools for CAS exam S* https://www.casact.org/admissions/studytools/examS/

Sur l'utilisation de R (*le premier livre était en commande fin 2018*)

- Pierre-André Cornillon, Arnaud Guyader, François Husson, Nicolas Jégou, Julie Josse, Nicolas Klutchnikoff, Erwan Le Pennec, Eric Matzner-Lober, Laurent Rouvière, Benoît Thieurmel 2018, **R pour la statistique et la science des données**, https://r-stat-sc-donnees.github.io/
- Achim Zeileis, 2008, **Applied Econometrics with R**, Springer Verlag
- Arthur Charpentier, 2014, **Computational Actuarial Science with R**, CRC Press

Pour se familiariser avec RMarkdown

- https://toltex.u-ga.fr/cours/RMarkdown/rmd
- https://daringfireball.net/projects/markdown/syntax
- https://rmarkdown.rstudio.com/lesson-1.html

Calendrier

septembre 2019

dimanche	lundi	mardi	mercredi	jeudi	vendredi	samedi
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	1	2	3	4	5

octobre 2019

dimanche	lundi	mardi	mercredi	jeudi	vendredi	samedi
29	30	1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31	1	2

novembre 2019

dimanche	lundi	mardi	mercredi	jeudi	vendredi	samedi
27	28	29	30	31	1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30

décembre 2019

dimanche	lundi	mardi	mercredi	jeudi	vendredi	samedi
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31	1	2	3	4

Légende

Examen
Travaux
Cours
Conges

Règlements

J'autorise ponctuellement la prise de photos du tableau ou de l'écran projeté. Tout enregistrement vidéo ou audio est en revanche interdit.

Politique 16 sur le harcèlement sexuel
Le harcèlement sexuel se définit comme étant un comportement à connotation sexuelle unilatéral et non désiré ayant pour effet de compromettre le droit à des conditions de travail et d'études justes et raisonnables ou le droit à la dignité.
La Politique 16 identifie les comportements suivants comme du harcèlement sexuel : **1. Manifestations persistantes ou abusives d'un intérêt sexuel non désirées.** **2. Remarques, commentaires, allusions, plaisanteries ou insultes persistants à caractère sexuel portant atteinte à un environnement propice au travail ou à l'étude.** **3. Avances verbales ou propositions insistantes à caractère sexuel non désirées.** **4. Avances physiques, attouchements, frôlements, pincements, baisers non désirés.** **5. Promesses de récompense ou menaces de représailles, implicites ou explicites, représailles liées à l'acceptation ou au refus d'une demande d'ordre sexuel.** **6. Actes de voyeurisme ou d'exhibitionnisme.** **7. Manifestations de violence physique à caractère sexuel ou imposition d'une intimité sexuelle non voulue.** **8. Toute autre manifestation à caractère sexuel offensante ou non désirée.** **Pour plus d'information :** **http://www.instances.uqam.ca/ReglementsPolitiquesDocuments/Documents/Politique_no_16.pdf** **Pour rencontrer une personne ou faire un signalement :** **Bureau d'intervention et de prévention en matière de harcèlement :** **514-987-3000, poste 0886** **http://www.harcelement.uqam.ca**

Tout acte de plagiat, fraude, copiage, tricherie ou falsification de document commis par une étudiante, un étudiant, de même que toute participation à ces actes ou tentative de les commettre, à l'occasion d'un examen ou d'un travail faisant l'objet d'une évaluation ou dans toute autre circonstance, constituent une infraction au sens de ce règlement.

La liste non limitative des infractions est définie comme suit :

- la substitution de personnes ;
- l'utilisation totale ou partielle du texte d'autrui en le faisant passer pour sien ou sans indication de référence ;
- la transmission d'un travail aux fins d'évaluation alors qu'il constitue essentiellement un travail qui a déjà été transmis aux fins d'évaluation académique à l'Université ou dans une autre institution d'enseignement, sauf avec l'accord préalable de l'enseignante, l'enseignant ;
- l'obtention par vol, manœuvre ou corruption de questions ou de réponses d'examen ou de tout autre document ou matériel non autorisés, ou encore d'une évaluation non méritée ;
- la possession ou l'utilisation, avant ou pendant un examen, de tout document non autorisé ;
- l'utilisation pendant un examen de la copie d'examen d'une autre personne ;
- l'obtention de toute aide non autorisée, qu'elle soit collective ou individuelle ;
- la falsification d'un document, notamment d'un document transmis par l'Université ou d'un document de l'Université transmis ou non à une tierce personne, quelles que soient les circonstances ;
- la falsification de données de recherche dans un travail, notamment une thèse, un mémoire, un mémoire-création, un rapport de stage ou un rapport de recherche.

Les sanctions reliées à ces infractions sont précisées à l'article 3 du Règlement no 18

RESPECT DE L'INTÉGRITÉ ACADÉMIQUE

Face à l'importance et à l'ampleur du phénomène de la tricherie et du plagiat dans les universités, ici et à l'étranger, l'UQAM a amorcé, en janvier 2007, une démarche visant à promouvoir le respect de l'intégrité académique. Dans ce contexte et inspirée d'une philosophie de « tolérance zéro », la Commission des études de l'UQAM a modifié son Règlement sur les infractions de nature académique (R. 18) à sa réunion du 2 décembre 2008.

Endossant cette philosophie de « tolérance zéro » relativement aux actes de plagiat, de fraude et de tricherie, la Faculté des sciences de l'UQAM souhaite sensibiliser ses étudiants à l'importance du respect de l'intégrité académique. Puisqu'en sollicitant son admission à l'UQAM, toute candidate, tout candidat s'engage à suivre les politiques et règlements de l'Université, la Faculté souhaite informer ses étudiants des différents articles de ce règlement, des actes répréhensibles et des sanctions applicables. Un extrait de ces articles se trouve ci-dessous. Le Règlement complet et son application à la Faculté des sciences sont disponibles à l'adresse Web suivante : http://www.sciences.uqam.ca/etudiants/integrite-academique.html

Tous ces efforts visent à assurer la validité de la formation dispensée par la Faculté, ainsi qu'un traitement équitable de tous afin de maintenir la qualité de ses diplômes.

Article 2 - Infractions de nature académique	Article 3 - Sanctions
2.1 Infraction Tout acte de plagiat, fraude, copiage, tricherie, falsification de document ou création d'un faux document commis par une candidate, un candidat, une étudiante, un étudiant de même que toute participation à ces actes ou tentative de les commettre, à l'occasion d'un examen, d'un travail ou d'un stage faisant l'objet d'une évaluation ou dans toute autre circonstance, constitue une infraction au sens de ce règlement. *2.2 Liste non limitative des infractions* Sans limiter la généralité de ce qui précède, constitue notamment une infraction le fait de poser ou tenter de poser l'un des actes suivants ou le fait d'y participer : a) la substitution de personnes ou l'usurpation d'identité; b) le plagiat : l'utilisation totale ou partielle du texte ou de la production d'autrui en le faisant passer pour sien ou sans indication de référence; c) l'autoplagiat : le dépôt d'un travail pour fins d'évaluation alors que ce travail constitue essentiellement un travail qui a déjà automne soumis pour fins d'évaluation académique à l'Université ou dans une autre institution d'enseignement, sauf avec l'accord préalable de l'enseignante, l'enseignant; d) la possession ou l'obtention par vol, manœuvre ou corruption de questions ou de réponses d'examen; e) la possession ou l'utilisation de tout document ou matériel non autorisé préalablement, pendant un examen ou lors de la réalisation de travaux, incluant le recours aux outils informatiques ou moyens technologiques; f) l'utilisation pendant un examen de la copie d'examen ou de tout autre matériel provenant d'une autre personne; g) l'obtention de toute aide non autorisée, qu'elle soit collective ou individuelle; h) l'obtention d'une évaluation non méritée notamment par corruption, chantage, intimidation ou toute forme de harcèlement ou la tentative d'obtenir une telle évaluation; i) la falsification d'un document ou la création d'un faux document, notamment d'un document transmis à l'Université ou d'un document de l'Université transmis ou non à une tierce personne, quelles que soient les circonstances; j) la falsification de données de recherche dans un travail, notamment une thèse, un mémoire, un mémoire-création, un rapport de stage ou un rapport de recherche.	*3.1 L'attribution de la mention «P»* L'étudiante, l'étudiant qui commet une infraction est mis en probation et peut se voir imposer une ou plusieurs sanctions. La mise en probation génère l'attribution de la mention «P» au dossier informatisé de l'étudiant, de l'étudiante. La mention «P» n'apparaît pas au relevé de notes de l'étudiante, l'étudiant mais figure en tout temps à son dossier. Lorsque la sanction est la suspension, une mention à cet effet apparaîtra au relevé de notes pour la durée de la suspension. Dans le cas d'une expulsion définitive de l'Université, une mention à cet effet apparaîtra de manière permanente au relevé de notes. *3.2 La mise en probation et autres sanctions* a) la *mise en probation;* La mise en probation constitue la reconnaissance que l'étudiante, l'étudiant a commis une infraction au présent règlement. La mise en probation peut être imposée sans autre sanction, auquel cas, l'enseignant, l'enseignante se voit inviter à attribuer une notation à l'étudiante, l'étudiant pour le cours conformément au résultat obtenu pour les prestations complautomnees. La mise en probation sans autre sanction signifie que la mention «P» est inscrite au dossier de l'étudiante, l'étudiant et que celle, celui qui en est l'objet ne doit commettre aucune autre infraction au présent règlement, à défaut de quoi, l'une ou l'autre des sanctions suivantes lui sera imposée. Outre la mise en probation, l'étudiante, l'étudiant peut se voir imposer une ou plusieurs des sanctions suivantes : b) l'échec au cours ou à l'activité créditée; c) l'obligation de réussir de trois à six crédits additionnels, hors programme, afin d'obtenir son grade, diplôme, certificat ou attestation; les cours doivent être identifiés; d) la suspension de toute activité à l'Université, pour une période maximale de neuf trimestres consécutifs; e) son expulsion définitive de l'Université.