

## STT 2400: Régression linéaire - Hiver 2016

URL: <http://www.dms.umontreal.ca/~murua/cours/stt2400Hiver16/index.html>

Ce cours est une introduction à la théorie, méthodes et applications de l'ajustement et de l'interprétation de modèles de régression multiple. Les sujets traités comprennent l'étude et analyse de résidus, la transformation des données, des stratégies et des critères pour la sélection d'une équation de régression, le biais associé aux variables exclues et aux erreurs de mesure, et l'utilisation des logiciels de régression. La base théorique des méthodes, la relation avec l'algèbre linéaire, et les effets et les violations des hypothèses sont aussi analysées. Les techniques étudiées sont illustrées par des exemples concrets impliquant des données de sciences physiques, sociales ou biologiques.

### Horaire:

Lundi 9h30–11h30	B-3285 Pav. J-Brillant
Vendredi 10h30–11h30	Z-205 Pav. Claire-McNicoll
Vendredi 14h30–16h30	4191 A-Aisenstadt (Travaux pratiques)

### Professeur:

Alejandro Murua 4221 André-Aisenstadt (514) 343-6987 murua@dms.umontreal.ca  
*Disponibilités:* Lundi et Mardi 15h00 16h00

### Démonstrateur:

Nicolas Grenon-Godbut nicolas.grenon-godbout@umontreal.ca  
*Disponibilités:* Jeudi 14h00 16h00

### Livre à utiliser

[Weisberg] Sanford Weisberg (2014). *Applied Linear Regression*. 4<sup>e</sup> édition, John Wiley & Sons, Inc. (*Obligatoire et fortement recommandé.*)

### Des autres livres d'intérêt

[DS] Draper, N. et H. Smith (1998). *Applied Regression Analysis*. John Wiley & Sons (*lecture supplémentaire. Non requis.*)

[Hocking] Hocking, R. R. (2003). *Methods and Applications of Linear Models*. John Wiley & Sons. (*lecture supplémentaire. Non requis.*)

[CML] Cornillon, P.-A. et Matzner-Lober, E. (2007). *Rgression. Thorie et applications*, Springer. (*lecture supplémentaire. Non requis.*)

[LDL] Lafaye de Micheaux, P., Drouilhet, R. et Liquet, B. (2010). *Le logiciel R - Maitriser le langage - Effectuer des analyses statistiques*, Springer. <http://www.springerlink.com/content/978-2-8178-0114-8>. (*lecture supplémentaire. Non requis.*)

**Contenu du Cours** En principe, nous couvrirons les dix premiers chapitres du livre de Weisberg.

1. Régression linéaire simple  
la méthode des moindres carrés  
Analyse de la variance

- Le coefficient de détermination  $R^2$
- Test d'hypothèses
- Graphique de résidus
- 2. Régression linéaire multiple
  - Analyse de la variance
- 3. Interprétation des résultats
- 4. Test d'ajustement
- 5. Modélisation
  - Régression par polynômes
  - Des variables catégorielles et indicatrices
  - Composantes principales
- 6. Colinéarité
- 7. Moindres carrés pondérés
- 8. Transformation de variables
- 9. Des statistiques de diagnostic
  - Des observations aberrantes
  - Des observations influentes
  - Vérification des hypothèses:
    - Non-linéarité, des erreurs qui ne suivent pas une distribution normale
- 10. Sélection de variables et de modèle

## Évaluation

La note final a trois composantes:

**1. Les devoirs ( 20%)** seront assignés, rassemblés, évalués, et retournés. Tout le travail sera dû au début de la classe de la date assignée. Le travail en retard ne sera pas accepté pour aucune raison.

Les devoirs seront distribués selon le programme suivant:

Dvr#1	Vendredi 15 janvier	échéance: le vendredi 29 janvier
Dvr#2	Vendredi 29 janvier	échéance: le vendredi 12 février
Dvr#3	Vendredi 26 février	échéance: le vendredi 11 mars
Dvr#4	Vendredi 18 mars	échéance: le vendredi 1 avril

Chaque devoir aura le même poids dans l'évaluation finale et sera évalué sur une échelle de 0 à 100 points.

**2. Examen partiel [vendredi 19 février (25%)]** de 14h30 à 16h20 dans la salle B-3295 Pavillon 3200 J-Brillant. Un examen à livre fermé (pourtant vous pourrez y apporter et utiliser deux pages (une seule feuille) avec des formules).

**3. Le projet (20%)**. Le projet consistera à des applications particulières de la statistique aux sciences sociales, physiques, biologiques, au génie, ou des autres sciences. Il exigera l'exploration de données et une compréhension claire des matières présentées dans la classe et des tâches de lecture. Les projets sont des outils importantes non seulement pour vous aider à comprendre les

sujets traités dans ce cours, mais aussi pour vous aider également à développer une vue critique de l'analyse de données.

Vous devez former des groupes de deux à trois étudiant(e)s pour travailler dans le projet. Le projet sera annoncé le vendredi 1 avril pendant le cours. **Soumettez votre rapport avec l'analyse des données avant 17h00 lundi 11 avril.**

**3. L'examen final ( 35%)** L'examen (à livre fermé) est programmé pour **le lundi 25 avril de 9h00 à 11h59** dans la salle Z-310 du pavillon Claire McNicoll.

### Présentation des devoirs

Les conditions suivantes simplifieront considérablement l'évaluation des devoirs et sont obligatoires.

1. Vous devez former des groupes de deux à trois étudiant(e)s pour résoudre chaque devoir. Les groupes peuvent être les mêmes pour tous les devoirs, mais **il est fortement recommandé de créer des groupes différents pour chaque devoir.** Il n'y aura aucune compensation dans l'évaluation pour ceux qui décident de travailler tout seuls dans les devoirs. Les devoirs sont des travaux de groupe. Il faut apprendre travailler en groupe.

Ajouter à chaque devoir une page où les membres du groupe déclarent honnêtement le pourcentage de travail effectué par chaque membre du groupe: par exemple,

Phillipe	40\%
Andrea	60\%
-----	
Total	100\%

Cette page doit être signée par au moins un des membres du groupe.

2. En soumettant chaque devoir, mettez vos noms, le numéro du cours, *et le numéro du devoir* sur la première page, comme suit:

Vos Noms  
STT2400 - Hiver 2016  
Devoir #

3. Seul les devoirs lisibles seront acceptés et évalués.
4. Soumettez chaque devoir sur de papier de 8x11 pouces. Collez aussi les sorties d'ordinateur sur de papier de 8x11 pouces.
5. **Les sorties d'un logiciel sans aucune annotation ne sont pas acceptable. Vous devez clarifier quels aspects des sorties d'ordinateur sont appropriées et vous devez montrer comment ils répondent aux questions posées dans le devoir. Des parties non pertinentes ou incorrectes des sorties d'ordinateur devraient être éliminées ou bien clairement biffées.**
6. **Agrafez les pages ensemble.** Nous ne serons pas responsables des pages perdues. N'utilisez pas des trombones, ni pliez et puis déchirez les coins des pages.
7. **Soumettez les problèmes dans l'ordre donné.**

8. Organisez chaque devoir de sorte que les graphiques et leur discussion soient ensemble. **NE METTEZ PAS tous le graphiques à la fin du devoir.** Marquez quels graphiques sont assortis à quels problèmes.

Ces conditions aident à s'assurer que votre devoir est évalué efficacement et dans le meilleur délai. Les règles qui ne sont pas suivies peuvent vous faire perdre des points.

### Dates importantes

21 janvier	Date limite pour modifier le choix de cours
21 janvier	Date limite pour annuler un cours sans frais
Vendredi 19 février	Examen intra.
Du lundi 29 février au dimanche 6 mars	Périod d'activités libres
11 mars	Date limite pour abandonner un cours (avec frais)
25 mars au 28 mars	Congé de Pâques
Vendredi 1 avril	Projet. Échéance: lundi 11 avril.
Lundi 25 avril	Examen final.
Vendredi 15 avril	Dernier jour de cours
Vendredi 29 avril	Fin du trimestre

### Veillez lire ces messages importants

1. Les devoirs ne sont pas facultatives. Si vous manquez la date-limite pour soumettre le devoir, votre devoir recevra zero (0) points.
2. Le plagiat: attention, c'est sérieux! Vous êtes invité à consulter le site [www.integrite.umontreal.ca](http://www.integrite.umontreal.ca)
3. Bien que la discussion des problèmes des devoirs soit autorisée, chaque groupe de deux ou trois étudiant(e)s est requis de préparer et soumettre ses propres solutions (travail d'ordinateur y compris) aux devoirs. Des solutions préparées "en comité" de plus de deux groupes ne sont pas acceptables. **La duplication des solutions des devoirs et des sorties d'ordinateur préparé entièrement ou partiellement par quelqu'un d'autre ne sont pas acceptables et sont considérées plagiats.** Si vous recevez l'aide de n'importe qui, vous devez dûment lui (leur) rendre reconnaissance dans votre rapport (exemple: "puisque les données sont toutes positives et leur distribution est asymétrique, une transformation logarithmique est clairement approprié dans la prochaine étape. Je remercie David Cox de m'indiquer ceci."). **La collaboration de n'importe quelle sorte sur des examens est interdite.**
4. Vous avez l'obligation de motiver une absence prévisible à une évaluation dès que vous êtes en mesure de constater que vous ne pourrez pas être présent. Il appartiendra à l'autorité compétente de déterminer si le motif est acceptable (article 9.9).
5. Nous faisons bon accueil à des commentaires ou à des suggestions au sujet du cours à tout moment, soit en personne, par courriel, ou par téléphone.
6. **Ce programme est prévu pour fournir une vue d'ensemble de STT2400. Vous ne pouvez revendiquer aucun droit de lui. En particulier, les dates d'examen peuvent changer. Tandis que le programme devrait être un guide assez fiable pour la session présente, les annonces officielles sont toujours ceux qu'on fait dans la classe.**