STT 2400: Régression linéaire - Hiver 2023

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 8h30–10h30	Z-220 Pav. Claire-McNicoll
Mardi 14h30–15h30	Y-115 Pav. Roger-Gaudry
Mardi 8h30–10h30	Y-115 Roger-Gaudry (Travaux pratiques)

Professeur:

Alexandro Murua	4221 André-Aisenstadt	(514) 343–6987	alejandro.murua@umontreal.ca
Disponibilités:	Lundi et Mardi	15h00 à 16:00	

Démonstrateur:

TBA	TBA
Disponibilités:	TBA

Livre à utiliser

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

Des autres livres d'interêt

[CML] Cornillon, P.-A. et Matzner-Lober, E. (2007). *Régression. Théorie et applications*, Springer-Verlag France. *(lecture supplémentaire. Non requis.)*

[CML-R] Cornillon, P.-A. et Matzner-Lober, E. (2019). *Régression avec R*, Springer. *(lecture supplémentaire. Non requis.)* QA 278.2 C67 2019

[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.)*

[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.

Les devoirs seront distribués selon le programme suivant:

Dvr#1	mardi 17 janvier	échéance: le mardi 31 janvier
Dvr#2	mardi 31 janvier	échéance: le mardi 7 février
Dvr#3	mardi 14 mars	échéance: le mardi 28 mars

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

2. Examen intra [mardi 14 février (25%)] de 8:30 - 10:29 dans la salle Y-115 Pav. Roger-Gaudry. Un examen à livre fermé, pourtant vous pourrez y apporter et utiliser une feuille recto-verso (deux pages dans une seule feuille) avec des formules.

3. Le projet (25%). 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 presentées 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 mardi 28 mars pendant le cours. **Soumettez votre rapport avec l'analyse des données avant 23h59 vendredi 14 avril.**

3. L'examen final (30%). L'examen (à livre fermé) est programmé pour **lundi 24 avril de 9:00 - 11:59** dans la salle Z-220 Pav. Claire-McNicoll.

Preśentation 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 à travailler en groupe.

   Ajouter à chaque devoir une page ou les membres du groupe declare hônnetement 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 2023
   Devoir #

3. Seul les devoirs lisibles seront acceptés et évalués.

4. **Soumettez chaque devoir dans un document PDF au format lettre directement sur le site StudiUM du cours.**

5. **Les sorties d'un logiciel sans aucune annotation ne sont pas acceptable. Vous devez clarifier quels aspects des sorties d'ordinateur sont appropiées 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. **Soumettez les problèmes dans l'ordre donné.**

7. 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 suivis peuvent vous faire perdre des points.

Dates importantes

mardi 24 janvier	Date limite pour modifier le choix de cours
mardi 24 janvier	Date limite pour annuler un cours sans frais
mardi 14 février	Examen Intra, à livre fermé 9:00 - 11:59 (Z-220 Pav. Claire-McNicoll)
Du lundi 27 février au dimanche 5 mars	Période d'activités libres
vendredi 17 mars	Date limite pour abandonner un cours (avec frais)
mardi 28 mars	Projet pratique (à emporter)
Du vendredi 7 avril au lundi 10 avril	Congé de Pâques
vendredi 14 avril	Écheance du rapport associé au projet pratique
vendredi 14 avril	dernier jour de cours
lundi 24 avril	Examen final, à livre fermé, 9:00 - 11:59 (Z-220 Pav. Claire-McNicoll)
vendredi 28 avril	Fin du trimestre

Veuillez 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

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 considerées plagiats.** Si vous recevez l'aide de n'importe qui, vous devez dûment lui (leur) rendre reconnaissance dans votre rapport (example: "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 ou par courriel.

6. **Ce programme est prévu pour fournir une vue d'ensemble de STT2400. Vous ne pouvez revindiquer aucun droit de lui. En particulier, les dates d'examen ou d'échéance des devoirs peuvent changer. Tandis que le programme devrait**

être un guide assez fiable pour la présente session, les annonces officielles sont toujours ceux qu'on fait dans le cours.

Ressources d'aide au DMS et à l'UdeM

N'hésitez pas à aller chercher de l'aide au besoin. Voici des ressources disponibles à l'Université de Montréal.

1. Le centre de santé et de consultation psychologique (CSCP) de l'Université de Montréal (`http://www.cscp.umontreal.ca/`). La prise de rendez-vous et l'inscription à un premier rendez-vous se font entièrement en ligne à l'adresse suivante : `https://monudem.umontreal.ca/.../Consultation`

2. Le Programme Mieux-être de l'ASEQ.
   Ligne téléphonique ouverte 24 heures / 7 jours : 1 833 851-1363
   Pour plus d'informations: `http://www.aseq.ca/.../FA%C3%89CUM_Programmedaide`

3. N'hésitez pas à contacter votre TGDE (tgdesup@dms.umontreal.ca) ou votre association étudiante (aemsum@dms.umontreal.ca) qui pourront vous guider.