

Modèles Linéaires Appliqués

Arthur Charpentier

Hiver 2022

Intro #1 (plan de cours)

Modèles Linéaires Appliqués (STT5100)

“Méthode d'estimation par moindres 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”

<https://etudier.uqam.ca/cours?sigle=STT5100>

Instructeur: **Arthur Charpentier**

✉ charpentier.arthur@uqam.ca

🐦 [@freakonometrics](https://twitter.com/freakonometrics)

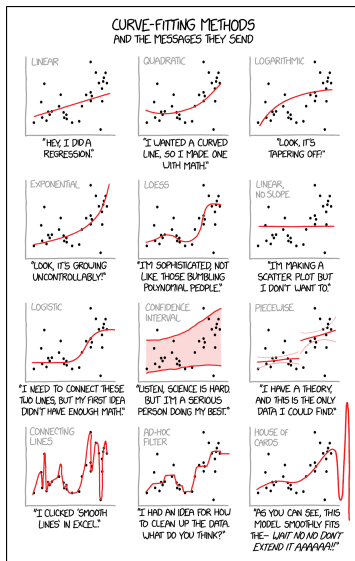
🐱 [freakonometrics](https://github.com/freakonometrics)

🌹 freakonometrics.hypotheses.org

→ [plan de cours.pdf](#)



Modèles Linéaires Appliqués (STT5100)



- R & RStudio
- probas $Y \sim F$, $\mathbb{E}[Y]$, etc
- stats $\{y_1, \dots, y_n\} \sim F_\theta$,
estimateur $\hat{\theta}$, $\mathbb{E}[\hat{\theta}]$, etc
- algèbre linéaire $\langle \vec{x}_1, \vec{\varepsilon} \rangle = 0$
matrices, projection
orthogonales, etc

Source: xkcd.com/2048



Modèles Linéaires Appliqués (STT5100)

① **Rappels**

- R & Markdown
- Probabilités & Algèbre Linéaire
- Statistiques, inférence & tests
- Simulations & Optimisation

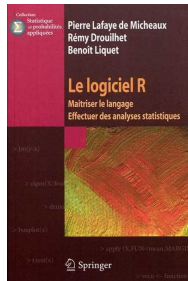
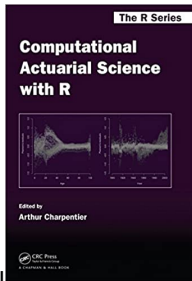
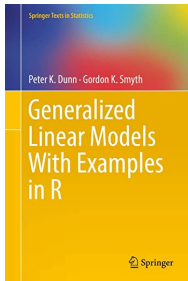
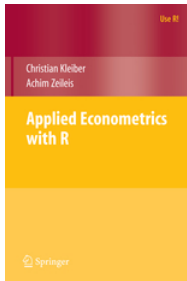
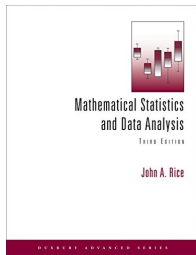
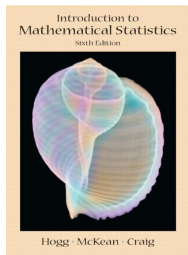
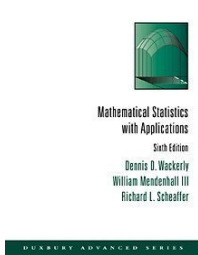
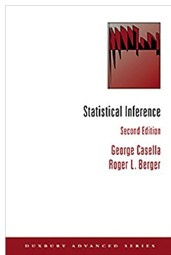
② **Modèles Linéaires (OLS)**

- Moindres Carrés & Modèle Gaussien
- Inférence & Tests
- Incertitude & Prévision

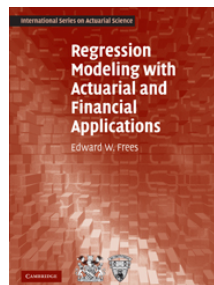
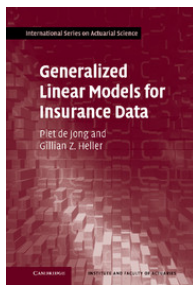
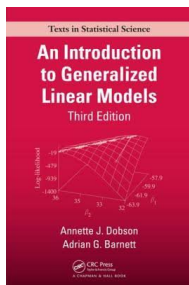
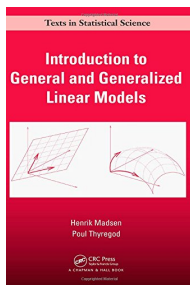
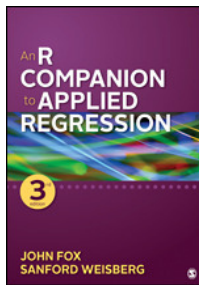
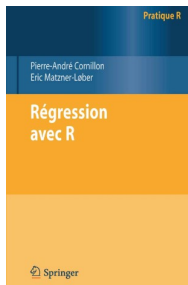
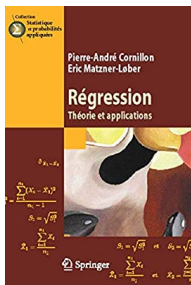
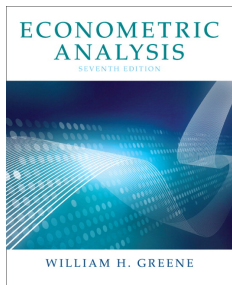
③ **Modèles Linéaires Généralisés (GLM)**

- Logistique, binomiale ($y \in \{0, 1\}$) & multinomiale
- Comptage, Poisson ($y \in \{0, 1, 2, \dots\}$)
- Famille exponentielle & GLM

Statistique Inférentielle & R



Modèles Linéaires (OLS & GLM)



Modèles Linéaires Appliqués (STT5100)

via <https://etudier.uqam.ca/lexique>

Crédit

Un crédit équivaut à 45 heures de travail universitaire, soit 15 heures de **cours** magistraux ou d'ateliers, et 30 heures d'études et de travail personnel. Chaque **cours** comporte normalement 3 crédits. Toutefois, il peut arriver que certains **programmes** incluent des **cours** qui comptent pour plus de 3 crédits chacun.

Un cours de 3 crédit, c'est en moyenne 90 heures de travail personnel, en plus du temps de cours magistral.

Ce cours est ambitieux, la matière est riche, et demandera beaucoup de travail personnel.

Modèles Linéaires Appliqués (STT5100)

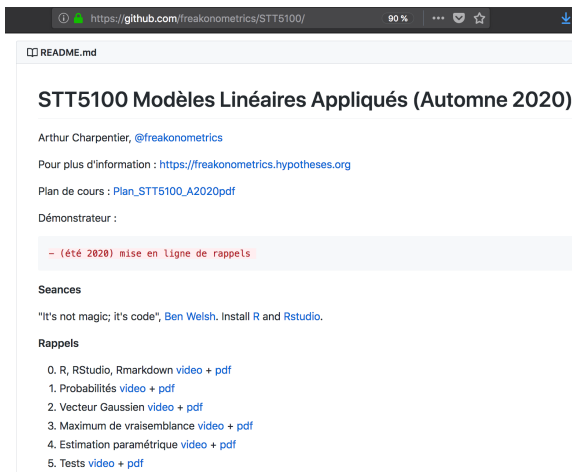
“when a measure becomes a target, it ceases to be a good measure”, Charles Goodhart, membre du Comité de politique monétaire de la Banque d'Angleterre

Autrement dit, les examens ne sont pas l'objectif du cours

- Quizz hebdomadaire ($\times 10$)
- Examen 1 = 30% (OLS)
- Examen 2 = 30% (GLM)
- Projet 1 = 20% (OLS)
- Projet 2 = 20% (GLM)

Modèles Linéaires Appliqués (STT5100)

Dépôt des documents (slides + devoirs): [github](#) + moodle



https://github.com/freakonometrics/STT5100/ 90 %

README.md

STT5100 Modèles Linéaires Appliqués (Automne 2020)

Arthur Charpentier, [@freakonometrics](#)

Pour plus d'information : <https://freakonometrics.hypotheses.org>

Plan de cours : [Plan_STT5100_A2020pdf](#)

Démonstrateur :

- (été 2020) mise en ligne de rappels

Seances

"It's not magic; it's code", [Ben Welsh](#). Install [R](#) and [Rstudio](#).

Rappels

0. R, RStudio, Rmarkdown [video](#) + [pdf](#)
1. Probabilités [video](#) + [pdf](#)
2. Vecteur Gaussien [video](#) + [pdf](#)
3. Maximum de vraisemblance [video](#) + [pdf](#)
4. Estimation paramétrique [video](#) + [pdf](#)
5. Tests [video](#) + [pdf](#)

Modèles Linéaires Appliqués (STT5100)

Dépôt des documents (slides + devoirs): github + moodle



Université du Québec à Montréal
STT5100 - Modèles linéaires appliqués | Groupe 050 Automne 2020

Arthur
Charpentier



Moodle

Soutien

Français (fr) ▼

Rechercher des cours



Les travaux seront à rendre sur moodle !



Devoir 1

Ajouter un travail

Remises de fichiers

Taille maximale des nouveaux fichiers : 2Go, Nombre maximal d'annexes : 20

Fichiers

Vous pouvez glisser des fichiers ici pour les ajouter.

Modèles Linéaires Appliqués (STT5100)

Vidéos du cours: [youtube](#) (+ github + moodle)

The screenshot shows a video player interface. The main video area displays a slide titled "Fonctions Liens : Canonique". The slide content includes:

- Cas particulier : lien canonique $g = b'^{-1}$ i.e. $\eta_i = \theta_i$
- $$\mu(\mathbf{x}) = b'(\mathbf{x}^\top \beta)$$
- Exemple** Pour la loi normale, $\mu(\mathbf{x}) = \mathbf{x}^\top \beta$ (identité)
- Exemple** Pour la loi de Poisson, $\mu(\mathbf{x}) = \exp(\mathbf{x}^\top \beta)$ (logarithmique)
- Exemple** Pour la loi de Bernoulli, $\mu(\mathbf{x}) = \frac{\exp(\mathbf{x}^\top \beta)}{1 + \exp(\mathbf{x}^\top \beta)}$ (logit)
- Exemple** Pour la loi Tweedie $V(\mu) = \mu^\gamma$, $\mathbf{x}^\top \beta = \frac{\mu(\mathbf{x})^{1-\gamma}}{1-\gamma}$ (power)

The video player controls at the bottom show the video is at 8:17 / 41:01. The URL in the address bar is [freakonometrics.hypotheses.org](https://www.youtube.com/watch?v=...).

To the right of the video player is a table of contents for the video series:

STT5100	
arthur charpentier - 14/22	
12	arthur charpentier
13	arthur charpentier
15	arthur charpentier
16	arthur charpentier
17	arthur charpentier

Modèles Linéaires Appliqués (STT5100)

Quizz: moodle



Université du Québec à Montréal
STT5100 - Modèles linéaires appliqués | Groupe 050 Automne 2020

Arthur
Charpentier



Moodle

Soutien

Français (fr) ▼

Rechercher des cours



par exemple

On dispose de 119 observations, dans le fichier [echantillon.csv](#).

Supposons que les observations sont des réalisations d'une loi normale, au sens où les x sont supposés tirés des X , de densité

$$x \mapsto f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

pour $x \geq 0$. On veut l'intervalle de confiance à 95% du paramètre μ , centré sur \bar{x} . Quelle est la borne inférieure de l'intervalle de confiance (symétrique), au seuil de 95%? (une tolérance de 0,02 est accordée)

Réponse :

on sauve le fichier [echantillon.csv](#) puis on ouvre R

Modèles Linéaires Appliqués (STT5100)

```
1 > b = read.csv("echantillon.csv")
2 > n = nrow(b)
3 > m = mean(b$x)
4 > s = sd(b$x)
5 > interval = m + 1.96*c(-1,+1)*s/sqrt(n)
6 > interval
7 [1] 370.0794 407.7525
```

puisque l'intervalle de confiance pour μ , avec un niveau de confiance $1 - \alpha$, est de la forme

$$\left[\bar{x}_n + \Phi^{-1}(\alpha/2) \frac{\hat{s}_n}{\sqrt{n}}; \bar{x}_n + \Phi^{-1}(1 - \alpha/2) \frac{\hat{s}_n}{\sqrt{n}} \right]$$

avec les constantes usuelles ± 1.96 puisque

```
1 > qnorm(c(.025,.975))
2 [1] -1.959964 1.959964
```

Modèles Linéaires Appliqués (STT5100)

Une fois la tentative envoyée, vous n'aurez plus la possibilité de modifier vos réponses pour cette tentative.

Tout envoyer et terminer

Annuler

On peut alors soumettre la réponse (définitivement)

On dispose de 119 observations, dans le fichier [echantillon.csv](#).

Supposons que les observations sont des réalisations d'une loi normale, au sens où les x sont supposés tirés des X , de densité

$$x \mapsto f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

pour $x \geq 0$. On veut l'intervalle de confiance à 95% du paramètre μ , centré sur \bar{x} . Quelle est la borne inférieure de l'intervalle de confiance (symétrique), au seuil de 95%? (une tolérance de 0,02 est accordée)

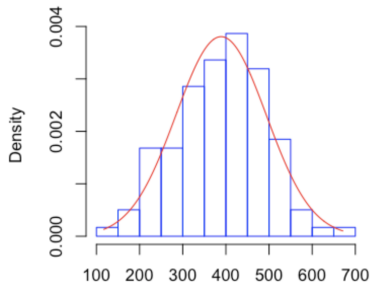
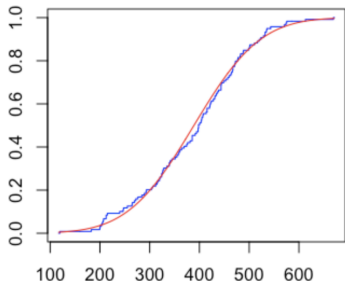
Réponse :

370.0794



Modèles Linéaires Appliqués (STT5100)

Des éléments de réponse (succints) sont alors proposés



Modèles Linéaires Appliqués (STT5100)



Tricherie et intégrité académique

<http://r18.uqam.ca/>

par exemple Comment citer ses sources?

Politique 16, <https://harcelement.uqam.ca/>

Modèles Linéaires Appliqués (STT5100)



Lors d'un cours ou d'une activité d'enseignement en ligne, le personnel enseignant peut décider, selon le cas, de procéder à l'enregistrement audio ou audiovisuel du cours ou de l'activité d'enseignement. Le personnel enseignant peut partager l'enregistrement uniquement à son groupe-cours.



En cas d'enregistrement, l'étudiante, l'étudiant sera informé au début de la séance.



Il est de la responsabilité de l'étudiante, de l'étudiant de désactiver son microphone et/ou sa caméra s'il ne souhaite pas être enregistré.



À défaut de désactiver son microphone et/ou sa caméra, l'étudiante, l'étudiant, consent à l'enregistrement audio ou audiovisuel, à la conservation, à la rediffusion et à l'utilisation de l'enregistrement de son nom, de sa voix et de son image dans le cadre du cours ou de l'activité en ligne. L'étudiante, l'étudiant reconnaît ne détenir aucun droit dans l'enregistrement.



Sauf avec l'autorisation expresse écrite du personnel enseignant, l'étudiante, l'étudiant reconnaît qu'il est interdit de reproduire, d'enregistrer, de publier, de diffuser, de communiquer ou de partager, par quelque moyen que ce soit, tout ou partie de l'enregistrement d'un cours ou d'une activité d'enseignement en ligne de même que tout matériel pédagogique s'y rattachant.

Une étudiante, un étudiant qui contrevient à ce qui précède s'expose aux sanctions prévues dans les règlements et politiques de l'UQAM ou à tout recours légal, notamment en vertu de la Loi sur le droit d'auteur.