

# Modèles Linéaires Appliqués

Arthur Charpentier

Automne 2020

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](mailto:charpentier.arthur@uqam.ca)

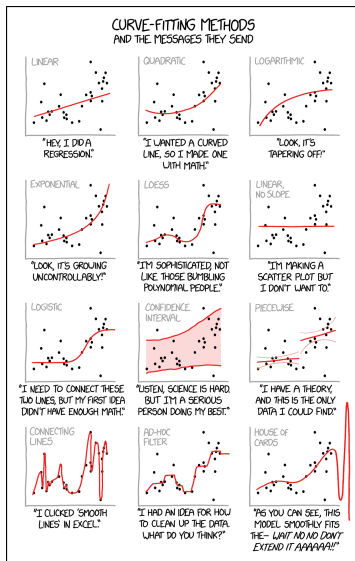
🐦 [@freakonometrics](https://twitter.com/freakonometrics)

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

🔗 [freakonometrics.hypotheses.org](https://freakonometrics.hypotheses.org)



# 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  $\langle \vec{x}, \vec{e} \rangle = 0$   
matrices, projection  
orthogonales, etc

[xkcd.com/2048](https://xkcd.com/2048)



# Modèles Linéaires Appliqués (STT5100)

## ① **Rappels**

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

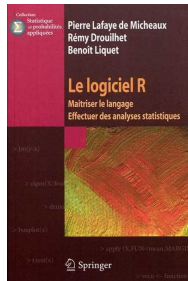
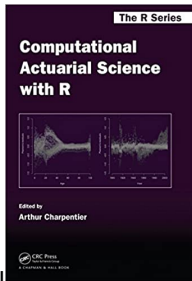
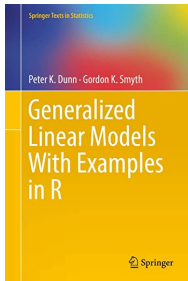
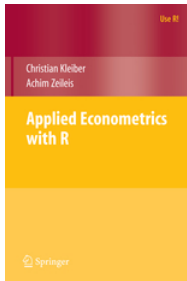
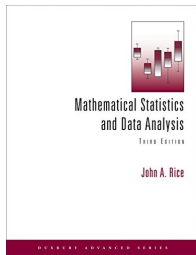
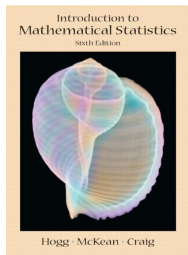
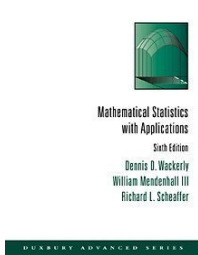
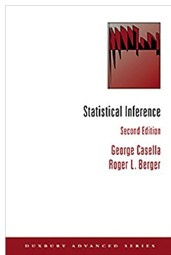
## ② **Modèles Linéaires (OLS)**

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

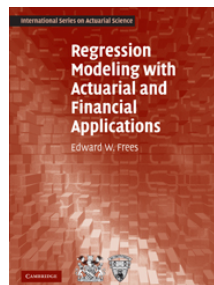
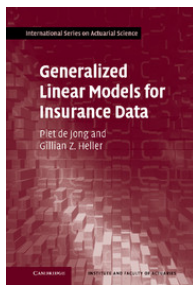
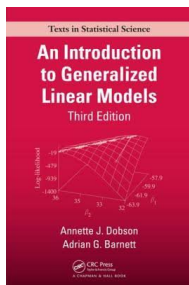
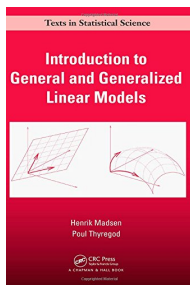
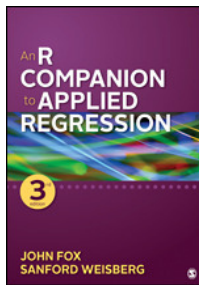
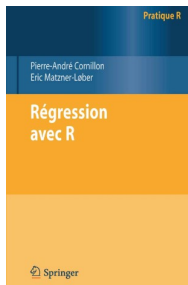
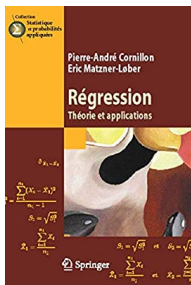
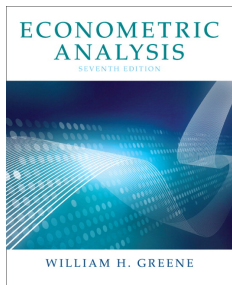
## ③ **Modèles Linéaires Généralisés (GLM)**

- 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)

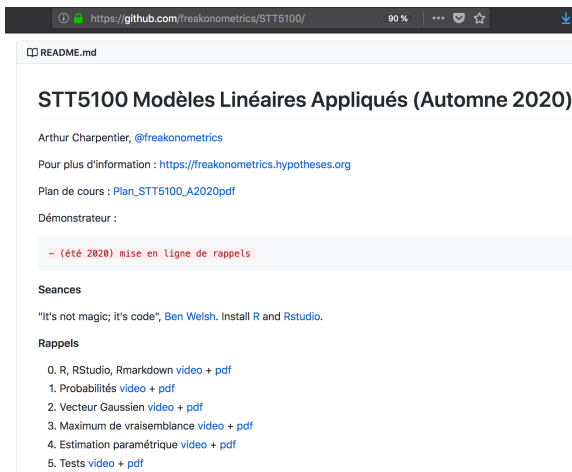
“when a measure becomes a target, it ceases to be a good measure”, Charles Goodhart, membre du Comité de politique monétaire de la Banque d'Angleterre

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

- Quizz hebdomadaire ( $\times 10$ ) = 20%
- Examen 1 = 20% (OLS)
- Examen 2 = 20% (GLM)
- Projet 1 = 18% (OLS)
- Projet 2 = 18% (GLM)
- Participation = 4%

# 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](#).

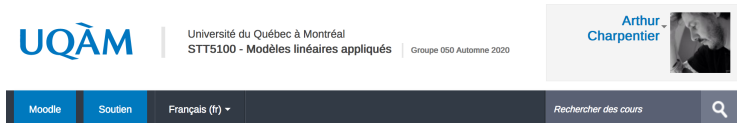
### Rappels

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



# Modèles Linéaires Appliqués (STT5100)

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

The screenshot shows a video player interface. The main content is a slide titled "Fonctions Liens : Canonique". It discusses the canonical link function  $g = b'^{-1}$  where  $\eta_i = \theta_i$ , and the general form  $\mu(\mathbf{x}) = b'(\mathbf{x}^\top \beta)$ . It provides three examples: Normal distribution, Poisson distribution, and Bernoulli distribution, each with its corresponding link function. The video player controls at the bottom show a progress bar at 8:17 / 41:01 and the URL [nonometrics\\_hypotheses.org](https://www.youtube.com/watch?v=nonometrics_hypotheses.org).

**Fonctions Liens : Canonique**

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)

**STT5100**  
arthur charpentier - 14/22

Slide	Duration	Content
12	36:23	arthur charpentier
13	30:01	STT5100 Hiver 2020, GLM, slides 12 arthur charpentier
14	41:02	STT5100, Hiver 2020, GLM, slides 13 arthur charpentier
15	36:05	STT5100 Hiver 2020, GLM slides 14 arthur charpentier
16	15:10	STT5100 hiver 2020 - GLM - slides 15 arthur charpentier
17	29:33	STT5100 Hiver 2020, GLM, slides 16 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  $\mu$ , 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  $\mu$ , 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  $\pm 1.96$  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  $\mu$ , 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)

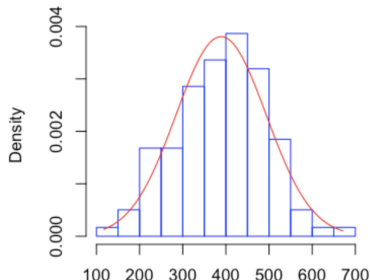
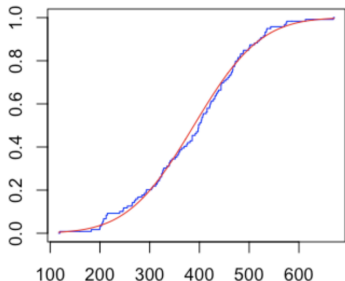
Réponse :

370.0794



# Modèles Linéaires Appliqués (STT5100)

Des éléments de réponse (succints) sont alors proposés



# Modèles Linéaires Appliqués (STT5100)

## Discussion: forum teams

< Toutes les équipes



Groupe UQAM STT5100

Général

**Général** Publications Fichiers Bloc-notes de classe Devoirs Notes +

- Charpentier, Arthur a ajouté vanier, Ceirick à l'équipe.
- Charpentier, Arthur a ajouté Teres, William à l'équipe.
- Charpentier, Arthur a ajouté Tauadros, Edward à l'équipe.
- Charpentier, Arthur a ajouté Tatonetti, Claudia à l'équipe.
- Charpentier, Arthur a ajouté Sylvain, Roxanne à l'équipe.
- Charpentier, Arthur a ajouté Silue, Katieebie Fred à l'équipe.
- Charpentier, Arthur a ajouté Sharma, Amanjot à l'équipe.
- Charpentier, Arthur a ajouté Samaha, Christina à l'équipe.
- Charpentier, Arthur a ajouté Sabourin, Michael à l'équipe.
- Charpentier, Arthur a ajouté Portelance, Vincent à l'équipe.
- Charpentier, Arthur a ajouté Pelletier, Matthieu à l'équipe.
- Charpentier, Arthur a ajouté Paul, Christopher-Junior à l'équipe.
- Charpentier, Arthur a ajouté Paquet, Anne à l'équipe.
- Charpentier, Arthur a ajouté Paiement, William à l'équipe.
- Charpentier, Arthur a ajouté Oueini, Christina à l'équipe.
- Charpentier, Arthur a ajouté Moussa, Amira à l'équipe.
- Charpentier, Arthur a ajouté Morasse, Rachel à l'équipe.
- Charpentier, Arthur a ajouté Moradi-Djanatdoust, Maral à l'équipe.
- Charpentier, Arthur a ajouté Moreau, Loïc à l'équipe.
- Charpentier, Arthur a ajouté Mayaki, Alexandra à l'équipe.
- Charpentier, Arthur a ajouté Mahoro, Liliane à l'équipe.

Démarrer une conversation. Utilisez @ pour mentionner un contact.



# 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.