

Modèles Statistiques (MS) et Programmation Lettrée (PL)

ou comment réaliser une étude en incluant de l'analyse de données

Danilo.Carastan-dos-Santos@univ-grenoble-alpes.fr

Laboratoire LIG
Équipe-Projet Inria DATAMOVE

Grenoble 2026

UE MODÈLES STATISTIQUES

1 ORGANISATION DE L'UE : Modèles Statistiques

2 OBJECTIF DE L'UE

3 RÉFÉRENCES BIBLIOGRAPHIQUES

4 PROBLÉMATIQUE : QUELQUES EXEMPLES

5 REPRODUCTIBITÉ : MOTIVATION Thanks to GitHub SMPE

ORGANISATION

Équipe pédagogique

- ▶ **Danilo Carastan dos Santos (Corrdination) :**
danilo.carastan-dos-santos@univ-grenoble-alpes.fr
- ▶ **Marina Gradvohl :** marina-gradvohl@univ-grenoble-ales.fr
- ▶ **Djoser Simeu :** djoser-simeu@inria.fr
- ▶ **Jean-Marc Vincent :** jean-marc.vincent@imag.fr

COMMUNICATION AVEC L'ÉQUIPE PÉDAGOGIQUE

Mail et adresses électroniques

Adresse Mail enseignant : Prénom.Nom@etu.univ-grenoble-alpes.fr

SUJET : [MIAGE3 :MS] sujet explicite toute adresse de provenance différente risque d'être "grey/black-listée" et d'atterrir dans une poubelle

Page web <https://im2ag-gitlab.univ-grenoble-alpes.fr/l3-mspl/mspl-2025-2026>

UE MODÈLES STATISTIQUES

1 ORGANISATION DE L'UE : Modèles Statistiques

2 OBJECTIF DE L'UE

3 RÉFÉRENCES BIBLIOGRAPHIQUES

4 PROBLÉMATIQUE : QUELQUES EXEMPLES

5 REPRODUCTIBITÉ : MOTIVATION Thanks to GitHub SMPE

OBJECTIF PÉDAGOGIQUE DE L'UE MODÈLES STATISTIQUES

Connaissances

Savoir réaliser une étude d'un objet informatique (ou autre) à partir de données observées :

(répondre à une question, formuler une hypothèse et la confirmer)

- ▶ savoir bâtir une expérimentation simple et produire des données d'observation (non atteint cette année)
- ▶ savoir analyser les résultats obtenus (processus d'analyse)
- ▶ savoir restituer les résultats sous forme synthétique (processus de visualisation, commentaires, analyse et synthèse)

En pratique, savoir réaliser une étude argumentée et correctement présentée.

Savoir utiliser un/des environnement(s) adapté(s) :

- ▶ suivi des développements logiciels (historique, versionning, collaboration) : git, github (non atteint cette année)
- ▶ processus d'analyse (analyse statistique, synthèse, visualisation) : R (R-studio, ggplot2)
- ▶ mise en forme et présentation : LaTeX (via un markdown)

ORGANISATION DE LA SEMAINE

Cours : **guidelines** pour une étude rigoureuse et reproductive

Les cours seront organisés à partir d'études de cas :

- ▶ une partie synthétique sur les **concepts**
- ▶ une partie sur des **exemples** illustrant les concepts
- ▶ une partie sur votre étude de cas

Forme du travail

- ▶ travail individuel
- ▶ travail en groupe
- ▶ travail public (partageable par toute la promotion (et même plus))
- ▶ synthèse en commun (production de fiches)

Travail personnel :

- ▶ prévoir 1 à 2h de travail en moyenne à la maison pour 1 séance de cours,
- ▶ exercices à la maison (pour préparer le matériel des séances suivantes)

Évaluation : une note d'UE

- ▶ partiel et mini-projet avec une présentation (1/3,2/3)

CONTENU INDICATIF

Environnement

- ❶ 13/01 Introduction, Organisation Git
- ❷ 20/01 Analyse de figures, checklist for good graphics
- ❸ 27/01 Installation de l'environnement Git/RStudio exemples de Programmation Lettrée

Traitement de données

- ❹ 03/02 Data management tidy, dplyr
- ❺ 24/02 Processus d'analyse et visualisation ggplot2
- ❻ 03/03 Tests statistiques intervalles de confiance

Mini-projet

- ❻ 24/03 Mini-projet : spécification/études préliminaires / compléments statistiques
- ❼ 31/03 Mini-projet : mise en place du processus d'analyse / compléments statistiques

Évaluation à mi-parcours

- ❾ 15/02 Partiel

Mini-projet (suite)

- ❻ 21/04 Mini-projet : analyse / compléments statistiques
- ❼ 28/04 Mini-projet : rédaction du rapport

Présentation

- ❿ 19/05 Présentation orale (tous les étudiants)



ORGANISATION DU TRAVAIL

Support : Git

- ▶ Git de référence : transparents, checklist, aides, ...
- ▶ Git personnel : fork du git de références, conventions de nommage
- ▶ Git personnel partagé pour les rendus (indication du binôme/quadrinôme)

Cours/TD

Pour chaque TD, un rendu est obligatoire pour assurer le suivi et la progression

- ▶ travail en binôme
- ▶ Rendu au format Rmd et pdf (taille limitée)
- ▶ pas d'évaluation

Mini-projet

Le travail du projet se fera en équipe,

- ▶ travail en quadrinôme
- ▶ Rendu au format Rmd et pdf (taille limitée)
- ▶ Présentation en 5 mn des résultats obtenus

UE MODÈLES STATISTIQUES

1 ORGANISATION DE L'UE : Modèles Statistiques

2 OBJECTIF DE L'UE

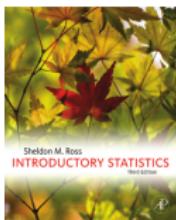
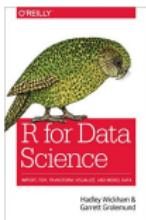
3 RÉFÉRENCES BIBLIOGRAPHIQUES

4 PROBLÉMATIQUE : QUELQUES EXEMPLES

5 REPRODUCTIBITÉ : MOTIVATION Thanks to GitHub SMPE

BIBLIOGRAPHIE : OUVRAGES DE RÉFÉRENCE DU COURS

- ▶ **R** Garrett Grolemund and Hadley Wickham, R for Data Science, O'Reilly 2016
<http://r4ds.had.co.nz/>
- ▶ **Statistiques** Sheldon Ross Introductory Statistics. Academic Press 2010
Également les polycopiés de Frédérique Leblanc
<http://www-ljk.imag.fr/membres/Frederique.Leblanc/>
- ▶ **Historique** Donald E. Knuth Literate Programming. Academic Press 1983



et évidemment de nombreuses ressources sur le web ...

UE MODÈLES STATISTIQUES

1 ORGANISATION DE L'UE : Modèles Statistiques

2 OBJECTIF DE L'UE

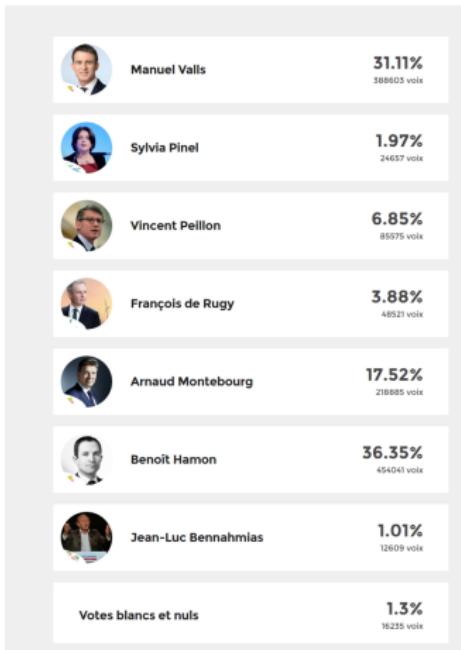
3 RÉFÉRENCES BIBLIOGRAPHIQUES

4 PROBLÉMATIQUE : QUELQUES EXEMPLES

5 REPRODUCTIBITÉ : MOTIVATION Thanks to GitHub SMPE

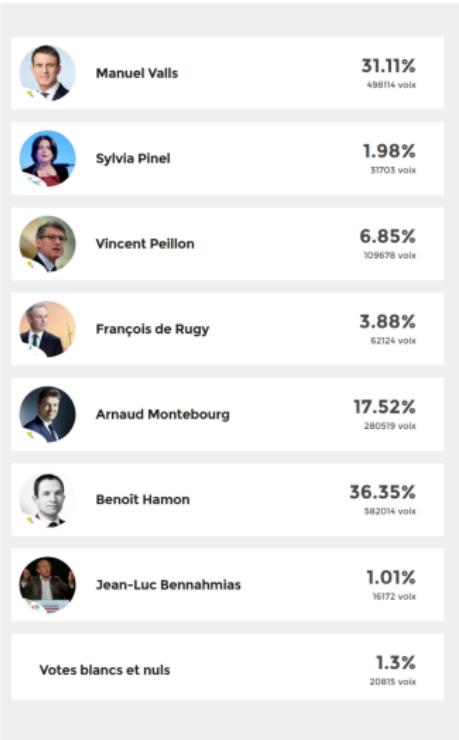
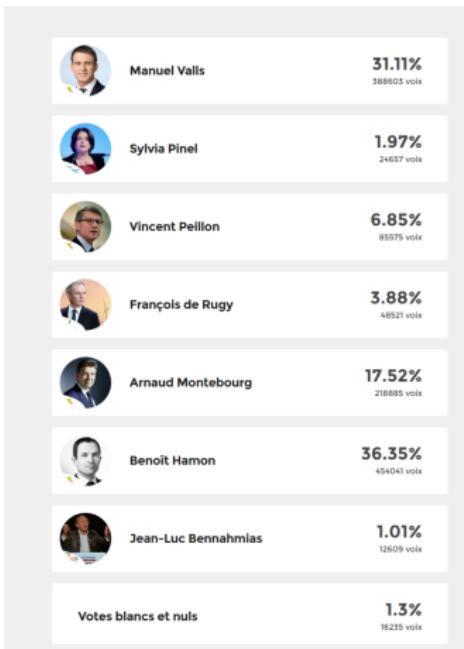
UN Oeil CRITIQUE

Dernière mise à jour le dimanche 22 janvier 2017 à 00h45



UN Oeil CRITIQUE

Dernière mise à jour le dimanche 22 janvier 2017 à 00h45



et lundi matin

UE MODÈLES STATISTIQUES

1 ORGANISATION DE L'UE : Modèles Statistiques

2 OBJECTIF DE L'UE

3 RÉFÉRENCES BIBLIOGRAPHIQUES

4 PROBLÉMATIQUE : QUELQUES EXEMPLES

5 REPRODUCTIBITÉ : MOTIVATION Thanks to GitHub SMPE

FRUSTRATION EN TANT QU'AUTEUR

- ▶ Je pensais avoir utilisé les mêmes paramètres mais **j'obtiens des résultats différents !**
- ▶ Le nouvel étudiant veut comparer avec **la méthode que j'ai proposée l'année dernière**
- ▶ Mon directeur de thèse m'a demandé si j'avais pris soin de configurer ceci ou cela mais je ne me souviens plus
- ▶ Le maudit quatrième relecteur a demandé une révision majeure et veut que je **modifie la figure 3** :(
- ▶ **Quel code et quel jeu de données** ai-je utilisé pour générer cette figure ?
- ▶ Ça **marchait hier** !
- ▶ 6 mois plus tard : **pourquoi** ai-je fait ça ?

FRUSTRATION EN TANT QUE RELECTEUR

Cela peut être une contribution intéressante mais :

- ▶ Cette **valeur moyenne** doit cacher quelque chose
- ▶ Comme d'habitude, il n'y a pas d'**intervalle de confiance**, je m'interroge sur la variabilité et si la différence est **significative** ou non
- ▶ Cela ne peut pas être vrai, je suis sûr qu'ils ont **supprimé certains points**
- ▶ Pourquoi ce graphique est-il en **échelle logarithmique** ? À quoi ressemblerait-il autrement ?
- ▶ Les auteurs ont décidé de ne montrer qu'un **sous-ensemble des données**. Je me demande à quoi ressemble le reste
- ▶ Il n'y a pas d'étiquette/légende/... Quelle est la **signification de ce graphique** ? Si seulement je pouvais accéder au script de génération

LE CHIEN A MANGÉ MES DEVOIRS !!!

► Problèmes de Versionnage

Merci pour votre intérêt dans l'implémentation de notre article. La bonne nouvelle est que j'ai pu trouver du code. Je espère simplement que c'est une version stable et fonctionnelle du code, et qu'elle correspond à l'implémentation que nous avons finalement utilisée pour l'article. Malheureusement, j'ai perdu des données quand mon ordinateur portable a été volé l'année dernière. La mauvaise nouvelle est que le code n'est pas commenté et/ou propre.

Ci-joint le code source <système> de notre algorithme. Je ne suis pas très sûr qu'il s'agisse de la version finale du code utilisée dans notre article, mais il devrait être au moins à 99% proche. J'espère que cela vous aidera.

LE CHIEN A MANGÉ MES DEVOIRS !!!

- ▶ Problèmes de Versionnage
- ▶ Mauvaises Pratiques de Sauvegarde

*Malheureusement, le serveur sur lequel mon implémentation était stockée a subi un **crash disque en avril et trois disques ont planté simultanément**. Bien que le service d'assistance ait fait des efforts importants pour sauvegarder les données, mon implémentation complète pour cet article n'a pas été retrouvée.*

LE CHIEN A MANGÉ MES DEVOIRS !!!

- ▶ Problèmes de Versionnage
- ▶ Mauvaises Pratiques de Sauvegarde
- ▶ Le Code Sera Bientôt Disponible

Malheureusement, le système actuel n'est pas assez mature pour le moment, donc il n'est pas encore disponible publiquement. Nous travaillons activement sur un certain nombre d'extensions et les choses sont quelque peu volatiles. Cependant, une fois que les choses se seront stabilisées, nous prévoyons de le publier pour les utilisateurs externes. À ce moment-là, nous serions heureux de vous envoyer une copie.

LE CHIEN A MANGÉ MES DEVOIRS !!!

- ▶ Problèmes de Versionnage
- ▶ Mauvaises Pratiques de Sauvegarde
- ▶ Le Code Sera Bientôt Disponible
- ▶ Aucune Intention de Publier

Je crains que le code source n'ait jamais été publié. Le code n'était jamais destiné à être publié et n'est donc pas en état d'être utilisé de manière générale.

LE CHIEN A MANGÉ MES DEVOIRS !!!

- ▶ Problèmes de Versionnage
- ▶ Mauvaises Pratiques de Sauvegarde
- ▶ Le Code Sera Bientôt Disponible
- ▶ Aucune Intention de Publier
- ▶ **Le Programmeur est Parti**

(ÉTUDIANT) était un étudiant diplômé dans notre programme mais il est parti il y a quelque temps donc je réponds à sa place. Pour l'article, nous avons utilisé un prototype qui incluait de nombreux éléments mobiles que seul (ÉTUDIANT) savait utiliser et nous n'avons pas eu le temps de les intégrer dans une implémentation prête à partager avant son départ. Néanmoins, j'espère que vous pourrez vous appuyer sur les idées/techniques de l'article.

Malheureusement, l'auteur qui a fait la plupart du codage pour cet article est décédé et le code n'est plus maintenu.

LE CHIEN A MANGÉ MES DEVOIRS !!!

- ▶ Problèmes de Versionnage
- ▶ Mauvaises Pratiques de Sauvegarde
- ▶ Le Code Sera Bientôt Disponible
- ▶ Aucune Intention de Publier
- ▶ Le Programmeur est Parti
- ▶ **Code Commercial**

Puisque ce travail a été réalisé chez <ENTREPRISE> nous ne rendons pas le code open-source à moins qu'il y ait une raison commerciale impérieuse de le faire. Donc malheureusement, je ne pense pas que nous pourrons le partager avec vous.

Le code appartient à <ENTREPRISE>, et pour autant que je sache le code n'est pas open-source. Votre meilleure option est de le réimplémenter :(Désolé.

LE CHIEN A MANGÉ MES DEVOIRS !!!

- ▶ Problèmes de Versionnage
- ▶ Mauvaises Pratiques de Sauvegarde
- ▶ Le Code Sera Bientôt Disponible
- ▶ Aucune Intention de Publier
- ▶ Le Programmeur est Parti
- ▶ Code Commercial
- ▶ **Code Académique Propriétaire**

Malheureusement, les sources du <SYSTÈME> ne sont pas destinées à être open-source (le code est partiellement la propriété de (UNIVERSITÉ 1), (UNIVERSITÉ 2) et (UNIVERSITÉ 3).)

Si cela change, je vous tiendrai au courant, bien que je ne pense pas qu'il y ait l'intention de rendre les sources du <SYSTÈME> open-source dans un avenir proche.

Si vous êtes intéressé par l'obtention du code, nous demandons seulement une description du projet de recherche dans lequel le code sera utilisé (ce qui pourrait mener à une recherche conjointe), et nous avons également un accord de licence logicielle que l'Université devrait signer.

LE CHIEN A MANGÉ MES DEVOIRS !!!

- ▶ Problèmes de Versionnage
- ▶ Mauvaises Pratiques de Sauvegarde
- ▶ Le Code Sera Bientôt Disponible
- ▶ Aucune Intention de Publier
- ▶ Le Programmeur est Parti
- ▶ Code Commercial
- ▶ Code Académique Propriétaire
- ▶ **Recherche vs. Partage**
- ▶ ...
- ▶ ...

Par le passé, lorsque nous avons tenté de le partager, nous nous sommes retrouvés à passer plus de temps à mettre les personnes extérieures à jour qu'à faire notre propre recherche. Alors j'ai finalement dû établir la politique selon laquelle nous ne fournirons pas le code source en dehors du groupe.