


Programmation dans 
Ch 0. Introduction
M2 CEE

Pr. Philippe Polomé, Université Lumière Lyon 2

2018 – 2019

Plan

- ▶ Présentation
- ▶ Motivation
- ▶ Organisation
- ▶ Introduction à **R**

NOS TUTELLES



LABEX



Accueil > Membres > Les membres > Annuaire

Polomé Philippe

Full Professor / Professeur des universités, Université Lumière Lyon 2

- CV
- Directeur du [Master Risques & Environnement](#) [1]
- Coordination of [Resilience, Hazard, Safety, Health, Vulnerability](#) research group in [Labex IMU](#)



Page [Enseignements / Teaching page](#)

Research Interests

- Environmental Economics
 - Environmental risk
 - Forest policy
 - Prosocial Behaviors
 - Compatibility between stated and revealed preferences
- Applied Econometrics
- Pluridisciplinarity



► Toutes les diapo via cette page

Master Risques et Environnement



Risques et Environnement

Mention de master



[Accueil](#) [Parcours](#) [Actualités](#) [Publications](#) [Partenariats](#)

Rechercher :

[>>](#)

[Accueil](#) > [Parcours](#) > [Gouvernance des risques environnementaux](#) > [Présentation](#)

Actualités

Le master "Risques et Environnement" accrédité

Le Master "Risques et Environnement" a été officiellement accrédité ce mois de février 2016 par la Direction générale de (...)

Rapport 2015 de l'Agence Européenne de l'Environnement sur la qualité de l'air en Europe

La pollution atmosphérique est le premier risque sanitaire d'origine environnementale en Europe : plus 430 000 décès prématurés (...)

Organisation de la formation

Toutes les versions de cet article : [\[English \[en\]\]](#) [\[français\]](#)

Le **parcours RISE** se passe en deux temps. En M1, les étudiants s'inscrivent dans un parcours obligatoire, les M1 droit public de Lyon 3 et Analyse des politiques économiques de Lyon 3. En M2, le parcours RISE en M2.

Le **M2 RISE** s'adresse à des étudiants détenteurs du niveau M1 (ou équivalent) en ingénierie généraliste. Il a pour objectif de former au dialogue pluridisciplinaire autour de l'environnement dans un contexte professionnel très diversifié, tant sur le plan des domaines académiques que professionnels.

Qu'est-ce que le risque environnemental ? Lisez [ici](#).

Débouchés — Lisez [ici](#)

S3 : Enseignements : 3 modules pluridisciplinaires de sept. à février (cliquer pour le détail)

risques-environnement.universite-lyon.fr

Table des matières

Organisation

Motivation

Introduction à R

Objectifs

- ▶ Usage de R
 - ▶ Rédaction de fichier de commandes (.R)
 - ▶ qu'on appelle "programme"
 - ▶ Réalisation de régressions & graphiques
 - ▶ = usage de packages existants
 - ▶ Mais aussi création d'outils propres
 - ▶ Via simulation
- ▶ Outils économétriques
 - ▶ Nouveaux & anciens

Plan

1. Commandes de base
 - 1.1 Quelques commandes hors SWIRL
 - 1.2 Graphiques & analyse préliminaire des données
2. Extensions du Modèle de Régression Linéaire
 - 2.1 Régressions linéaires, facteurs, interactions, splines
 - 2.2 Diagnostics
 - 2.3 Régressions quantiles
3. Simulation
 - 3.1 Simulation
 - 3.2 Bootstrap
4. Maximum de Vraisemblance
 - 4.1 Maximiser une vraisemblance
 - 4.2 Maximum de vraisemblance simulée
 - 4.3 nlogit

Évaluation

- ▶ Certaines sections sont suivies de devoirs
 - ▶ Ceux-ci sont évalués
- ▶ Un examen final
 - ▶ L'évaluation des devoirs est ajustée selon l'examen
 - ▶ devoirs \leq examen
- ▶ Idéalement, le devoir devrait prendre la forme d'un rapport
 - ▶ Fonctionnalité **Markdown** (+loin) pour écrire des rapports professionnels dans \LaTeX
 - ▶ J'utilise LyX, interface graphique de \LaTeX
- ▶ Dans tous les cas, seulement fichiers pdf
 - ▶ + le fichier de commande au format .R
 - ▶ + le fichier de données quand il n'est pas dans R

Pédagogie

- ▶ Ce cours, ainsi que celui sur stat non-param
 - ▶ aura 2 types d'enseignements
 1. du CM pour des concepts d'économétrie,
 2. de l'auto-apprentissage sur machine
- ▶ Il faut que vous suiviez le cours avec vos portables
 - ▶ Connectés à Internet
 - ▶ Et que vous ayez installé ce qu'il y a dans ce doc (+bas)
- ▶ L'emploi du temps avec Stat non-param
 - ▶ Ne sera pas forcément respecté
 - ▶ L'an dernier, 6 sessions Programmation
 - ▶ Puis 6 sessions Non-paramétrique

Références

- ▶ Bibliographie
 - ▶ Kleiber & Zeilis, *Applied Econometrics with R*, Springer, 2008
 - ▶ Cameron & Trivedi, *Microeconometrics*, Cambridge, 2005, 2006
- ▶ Adresses
 - ▶ R @ www.r-project.org/
 - ▶ The R-manuals [R-intro.pdf](#), [R-data.pdf](#), full reference
 - ▶ Auto-apprentissage
 - ▶ SWIRL, voir ci-dessous
 - ▶ Quick-R www.statmethods.net/index.html
 - ▶ www.ats.ucla.edu/stat/R/
 - ▶ <http://varianceexplained.org/RData/>
 - ▶ wiki wiki.r-project.org/rwiki/doku.php
 - ▶ R for economists
 - ▶ www.mayin.org/ajayshah/KB/R/R_for_economists.html
 - ▶ En français
 - ▶ eric.univ-lyon2.fr/~ricco/cours/cours_programmation_R.html

Table des matières

Organisation

Motivation

Introduction à R

R est un logiciel d'économétrie

- ▶ Open-source
 - ▶ Tout le monde peut contribuer
 - ▶ Gratuit
 - ▶ Multi-plateforme
 - ▶ Multi-langue (pas tellement)
- ▶ Langage matriciel programmable
- ▶ La plus grande variété de techniques stat/économétrie
 - ▶ et graphiques
- ▶ Autres usages que l'économétrie
 - ▶ Données disponibles + gestion SQL
 - ▶ SIG, math/optimisation

Architecture de R

- ▶ Coeur
 - ▶ comprend des commandes très utilisées
- ▶ Étendu par des *packages*
 - ▶ Que vous ajoutez selon vos besoins
 - ▶ Un package est une collection de commandes
 - ▶ Pour un usage spécifique
 - ▶ p.e. logit multinomial dans le package **mlogit**
- ▶ Il y a des milliers de packages (Oct. 2016 : 9385)
 - ▶ Certains redondants ou anciens
 - ▶ Le “CRAN task view” sur <http://cran.r-project.org/> aide à se localiser
 - ▶ Prochaine diapo
 - ▶ Avec “Econometrics task view” on fait presque tout
- ▶ Un package doit être installé dans R
 - ▶ Pas là par défaut

Classes (non-exclusives) de packages

Bayesian Inference	Statistical Genetics	Survival Analysis
Chemometrics and Computational Physics	Graphic Displays & Visualization	Psychometric Models & Methods
Cluster Analysis & Finite Mixture Models	gRaphical Models in R	Robust Statistical Methods
Optimization & Math. Programming	High Performance and Parallel Computing	Statistics for the Social Sciences
Econometrics	Machine Learning & Statistical Learning	Analysis of Spatial Data
Analysis of Ecological & Environmental Data	Analysis of Pharmacokinetic Data	Multivariate Statistics
Design of Experiments (DoE) & Analysis of Experimental Data	Natural Language Processing	Time Series Analysis
Empirical Finance	Probability Distributions	

Pourquoi utiliser R ?

- ▶ Notion de communauté
 - ▶ Vs. logiciels “propriétaires” (Stata, Gauss, SAS, SPSS...)
 - ▶ Licences vendues
 - ▶ Sur le fond, 99% de tout logiciel est développé dans le secteur public
 - ▶ Au lieu d'un petit groupe de travailleurs
 - ▶ La communauté scientifique contribue
 - ▶ Rend les codes accessibles à tous, gratuitement
 - ▶ Fournissent un support à leurs pairs
- ▶ Évolution rapide
 - ▶ Les chercheurs participent au codage de leur recherche
- ▶ Packages intérêts + variés
 - ▶ Pas besoin de rentabilité commerciale

Table des matières

Organisation

Motivation

Introduction à R

Installer R

- ▶ Coeur <https://www.r-project.org/>
 - ▶ Miroir à Lyon 1 <https://pbil.univ-lyon1.fr/CRAN/>
 - ▶ Usage sur serveur au GATE sur demande
- ▶ R-Studio <https://www.rstudio.com/>
 - ▶ IDE (integrated development environment)
 - ▶ Pas un GUI (graphical user interface)
 - ▶ Mais le + proche qu'on ait
- ▶ Microsoft R Open <https://mran.microsoft.com/open>
 - ▶ Améliore R de diverses façons, transparent
- ▶ Packages
 - ▶ La plupart du temps, dans R-studio
 - ▶ On en installera dans un moment
- ▶ Démarrer R-Studio
 - ▶ R-Studio appelle R

Présentation de R-studio : 4 fenêtres

The screenshot displays the RStudio interface with four main windows:

- Source:** Contains R code for a logistic regression model. The code includes comments explaining the use of `summary()` for coefficient extraction, the `glm()` function with `alpha=0`, and the calculation of mean lead and WTP2. The code is as follows:

```
summary(SB.NP.DC.probit)$medianWTP

# In glm, extract the coefficients & apply the formula from the table in the class note
# alpha/beta, but alpha is individual
# sex is categorical, the regression has managed it, but if we do calculation we have to do it ourselves; it turns out that female=2 and male =1
WTP.probit <- -(SB.NP.glm.probit$coefficients[1]
+SB.NP.glm.probit$coefficients[3]*NPSage
+SB.NP.glm.probit$coefficients[4]*(as.numeric(NPSsex)-1)
+SB.NP.glm.probit$coefficients[5]*NPSincome)/%SB.NP.glm.probit$coefficients[2]
mean(WTP.probit)

# Computing at the sample mean lead (unusually) to a similar, but not identical, number
WTP2.probit <- -(SB.NP.glm.probit$coefficients[1]
+SB.NP.glm.probit$coefficients[3]*mean(NPSage)
+SB.NP.glm.probit$coefficients[4]*mean(as.numeric(NPSsex)-1)
+SB.NP.glm.probit$coefficients[5]*mean(NPSincome)/%SB.NP.glm.probit$coefficients[2]

WTP2.probit
```
- Environment:** Lists objects in the global environment, including `sbchoice`, `sbchoice15`, `sbchoice895`, `tab`, `WTP.probit`, and `WTP2.probit`.
- Console:** Shows the output of the code, including the number of Fisher Scoring iterations (4) and the median WTP value (33.71668).

```
Number of Fisher Scoring iterations: 4
+
+
+
+
+SB.NP.glm.probit$coefficients[5]*mean(NPSincome)/%SB.NP.glm.probit$coefficients[2]
> WTP2.probit
[Intercept]
      33
> mean(WTP.probit)
[1] 33.25962
> median(WTP.probit)
[1] 33
> SB.NP.DC.probit$medianWTP
NULL
> summary(SB.NP.DC.probit)$medianWTP
[1] 33.71668
>
```
- Help:** Displays the documentation for `as.numeric()`, explaining that it returns `TRUE` if the argument is of mode "numeric" and `FALSE` otherwise. It also includes a warning about factor objects and S4 methods.

R-Studio fenêtre en haut à gauche : *éditeur*

- ▶ Invoqué avec n'importe quel des 2 boutons HG (New or Load)
 - ▶ Code-couleur, avec aide en ligne & reconnaissance de commande
- ▶ Programmation : dans l'éditeur
 - ▶ Programmation = séquences de commandes dans un fichier texte "script"
 - ▶ Qui prend une extension .R
 - ▶ Qui est sauvé pour usage futur
 - ▶ Les **commandes** sont transmises par p.e. `plot(x)`
 - ▶ L'éditeur reconnaît les commandes et les colore en **bleu**
 - ▶ Les commandes sont exécutées avec `cmd↔` ligne par ligne
 - ▶ Les résultats de commandes peut être stockés dans des **objets** avec `<-`
 - ▶ `y_lm <- lm(y~x1+x2)`
 - ▶ Pls fichiers de commande peuvent être ouverts simultanément
 - ▶ onglets

R-studio fenêtre en bas à droite : 5 onglets

- ▶ **Files** fichiers dans votre projet
- ▶ Visualisations des **Plots**
- ▶ **Packages** qui sont présents
 - ▶ Chargé si le carré est coché
 - ▶ Bouton Install
 - ▶ Cliquez-le (il faut être connecté)
 - ▶ Écrivez swirl & suivez les instructions
- ▶ **Help**
- ▶ **Viewer**
 - ▶ pour voir un contenu web local (si vous éditez des pages web)
- ▶ Ces 5 onglets ont en commun l'outil de recherche en haut à droite

R-studio 2 autres fenêtres

- ▶ Bas Gauche : *console*
 - ▶ y écrire des commandes pour exécution immédiate
 - ▶ Ne reste pas en mémoire
 - ▶ Imprime les résultats des commandes
 - ▶ données par éditeur ou directes
 - ▶ Programmation : on écrit une ligne, on la teste
- ▶ Haut à Droite
 - ▶ Environnement : Liste des données chargées
 - ▶ Données, résultats, fonction
 - ▶ Pour un projet (+ bas)
 - ▶ Historique des commandes
 - ▶ Peuvent être récupérées

Premières commandes : Projet

- ▶ Un projet est un fichier qui réfère à une collection de fichiers
 - ▶ fichier de commandes .R, fichiers de données, de résultats
- ▶ Icône en Haut à Droite de R-Studio
 - ▶ Cliquez et créez un projet “Programmation R”
 - ▶ Là où vous le créez est votre répertoire de travail
 - ▶ Les fichiers vont là
 - ▶ Ne pas le mettre sur le bureau, la racine ou à un endroit obscur
 - ▶ Déchargez PR2017.R de ma page de cours
 - ▶ Dans le même dossier que votre projet
 - ▶ Ouvrez-le avec l'éditeur : icône en H à G
 - ▶ R-Studio retient les projets
 - ▶ On peut passer d'un projet à l'autre
 - ▶ Tous les fichiers écrits sur le disque restent disponibles

Premières commandes

- ▶ Quelques manipulations dans la Console
 - ▶ Taper `install.views("Econometrics")`
 - ▶ Pour à peu près tous les packages dont on aura besoin
 - ▶ C'est long ! ne faites pas ça en cours !
 - ▶ Pour le futur `update.views("Econometrics")`
 - ▶ Taper `Sys.setenv(LANG = "fr")`
 - ▶ Met un partie de R en français (pas tout)
 - ▶ R-Studio n'est qu'en anglais ainsi que la plupart des packages
- ▶ Éditeur
 - ▶ Écrire ici ce qu'on pense réutiliser
 - ▶ **Éviter** d'utiliser les symboles français é, è, ê, ë, à, ù, ç, ...
 - ▶ **Éviter** les symboles type #, \$, &, -... si vous n'êtes pas sûr
 - ▶ **Uniquement** les caractères latins non accentués
 - ▶ La CAPITALISATION est importante
- ▶ Commencer une ligne par un # pour indiquer un **commentaire**
 - ▶ Pas exécuté, l'éditeur le colore en vert

SWIRL

- ▶ Ensemble de modules basiques d'entraînement
 - ▶ C'est ce qu'on va faire pour commencer
 - ▶ Installer swirl
 - ▶ Taper
 - ▶ `install_course("R Programming")`
 - ▶ `install_course("Regression_Models")`
 - ▶ Autres cours https://github.com/swirldev/swirl_courses
 - ▶ Sur SWIRL : <http://swirlstats.com/students.html>
 - ▶ Diapo <https://github.com/DataScienceSpecialization/courses>
- ▶ Auto-apprentissage : Démarrer `swirl()`
 - ▶ Faites le cours R programming, Lessons 1-9 + 14
 - ▶ Vous-même, de chez vous, avant le 1^o cours ou plus tard
 - ▶ On refera Lesson 1 en cours

Quelques ressources sur R sur le web

- ▶ Utilisez Google!
 - ▶ Posez des questions basées sur des mots-clefs English
 - ▶ p.e. “R read Stata data”
- ▶ de R home page www.r-project.org
 - ▶ Getting help, Manuals, FAQs...
- ▶ Quelques liens
 - ▶ Quick-R www.statmethods.net/index.html
 - ▶ <http://stats.idre.ucla.edu/r/>
 - ▶ <http://varianceexplained.org/RData/>
 - ▶ www.r-bloggers.com
 - ▶ R for economists
 - ▶ www.mayin.org/ajayshah/KB/R/R_for_economists.html
- ▶ En français : oubliez le français avec R

En résumé, pour le 1^o cours, vous devez :

- ▶ Avoir installé R & R-Studio sur vos machines
- ▶ À partir de R-Studio
 - ▶ `install.views("Econometrics")`
 - ▶ avoir installé swirl
 - ▶ avoir installé les 2 modules (R programming & regressions)
- ▶ Dans swirl :
 - ▶ vous pouvez commencer les leçons 1-9 & 14 du course R programming