

Chapitre 0

Introduction à la statistique avec R

R

# Introduction au MOOC

CC BY NC SA

1

Pr. Bruno Falissard

UNIVERSITÉ PARIS SUD

[0:01] Avant de nous lancer dans les premières vidéos de statistique, quelques mots d'introduction.

Plan

Introduction à la statistique avec R > Introduction au MOOC

- Le MOOC (planning, fonctionnement)
- Les modalités d'évaluation
- À quoi servent les statistiques ?
- Présentation d'un « fichier de données »
- L'étude « santé mentale en prison »
- Premiers aspects techniques
  - Exportation d'un fichier de données
  - Introduction à R
  - Importation d'un fichier de données
  - Rcmdr et R Studio

2

Pr. Bruno Falissard

[00:05] Nous allons voir successivement,

- le fonctionnement du MOOC, à partir de la plateforme FUN ;
- les modalités d'évaluation, pour avoir la certification ;
- une petite introduction aux statistiques ;
- la notion de fichier de données, de jeu de données,
- en insistant sur l'étude qui va nous servir tout le long de ce MOOC, l'étude santé mentale en prison ;
- et puis quelques aspects techniques :
  - Comment importer un fichier de données que vous avez à partir d'un tableur comme OpenOffice ou comme Excel
  - Comment une fois dans R importer ce fichier
  - Et puis on parlera des surcouches logicielles de R qui permettent de faciliter un peu le travail : Rcmdr et puis surtout R Studio

Plan

Introduction à la statistique avec R > Introduction au MOOC



- Le MOOC (planning, fonctionnement)
- Les modalités d'évaluation
- À quoi servent les statistiques ?
- Présentation d'un « fichier de données »
- L'étude « santé mentale en prison »
- Premiers aspects techniques
  - Exportation d'un fichier de données
  - Introduction à R
  - Importation d'un fichier de données
  - Rcmdr et R Studio

 3 Pr. Bruno Falissard 

[00:46] Alors, à propos du fonctionnement du MOOC, nous disposons de la plateforme FUN,

The screenshot shows the interface of a MOOC course titled "Introduction à la statistique avec R > Introduction au MOOC". The top navigation bar includes links for "Courseware", "Info Cours", "Discussion", "Wiki", "Progression", "Fonctionnement du cours", "Ressources pédagogiques", and "Student view". A large, empty rectangular box is positioned on the left side of the main content area. Below it, a scatter plot displays several data points. The right side of the interface contains sections for "Course Updates & News" and "Course Handouts".

**Course Updates & News**

OCTOBER 20, 2014

Bonjour à tous,

La cinquième et dernière semaine du cours vient de s'achever. Nous vous remercions pour votre participation au MOOC et vos contributions sur le forum.

Vous disposez d'une semaine supplémentaire pour compléter les évaluations (quiz 'mémoire' et quiz 'exercice') et rendre les devoirs (date limite: le 26 octobre). Compte-rendu des problèmes techniques rencontrés avec le quiz exercice de la semaine 5, celui-ci ne compte pas dans l'évaluation finale.

**Course Handouts**

téléchargement du logiciel :  
<http://www.r-project.org/>

**Pr. Bruno Falissard**

[00:50] qui est globalement assez agréable à utiliser.

The screenshot shows a web-based learning environment for a MOOC titled "Introduction à la statistique avec R > Introduction au MOOC".

- Header:** "FUN" and "Introduction à la statistique avec R > Introduction au MOOC".
- Navigation Bar:** "Courseware" (highlighted in red), "Info Cours", "Discussion", "Wiki", "Progression", "Fonctionnement du cours", "Ressources pédagogiques", "Instructor", and "Student view".
- Sidebar:** "Semaine 1 : Introduction aux statistiques et à R, description d'une variable". Below it is a list of chapters:
  - Introduction Vidéo
  - Introduction Documents
  - Enquête
  - Chapitre 1 : Définitions Quiz 'mémorie'
  - Chapitre 2 : Représentations graphiques Quiz 'mémorie'
  - Chapitre 3 : Mesures de position et de
- Content Area:** A video player titled "VIDEO D'INTRODUCTION AU MOOC" showing a thumbnail of a person speaking. The video player has a play button and a progress bar.
- Bottom Right:** "Pr. Bruno Falissard" and the logo of Université Paris-Sud.
- Bottom Left:** CC BY NC SA license icons.
- Page Number:** "5"

[00:53] On peut visionner les cours comme en ce moment en cliquant sur le premier onglet "courseware". La plupart des vidéos sont sous-titrées ou en cours de sous-titrage. N'oubliez pas qu'au même endroit vous disposez de versions pdf des retranscriptions écrites et des retranscriptions des diapositives, ce qui peut être utile quand on a du mal à suivre le débit verbal des enseignants.

The screenshot shows a Moodle-based MOOC interface for the course "Introduction à la statistique avec R > Introduction au MOOC". The top navigation bar includes links for "Courseware", "Info Cours", "Discussion", "Wiki", "Progression", "Fonctionnement du cours", and "Ressources pédagogiques". The "Discussion" tab is highlighted with a red circle. A large, empty rectangular box on the left side of the page is likely a placeholder for a video or a large image.

The main content area displays a discussion post titled "ANOVA" by user "Kambireollo". The post is visible to everyone and was posted 2 months ago. It contains the text: "Dans le cas d'une ANOVA, lorsque nous avons plus de deux classes et que la distribution ne suit pas une loi normale. Que pouvons nous faire?". Below the post, it says "(this post is about Week 1 / Topic-Level Student-Visible Label)". There are 0 responses to the post.

On the left sidebar, there is a list of pinned discussions under the heading "Attestations", including topics like "début seconde session", "Information", "Statistiques de fréquentation", "Attestation sans l'heure de cours", and "ACP: quels types de variables ?".

At the bottom of the page, there are icons for Creative Commons Attribution-NonCommercial-ShareAlike (CC BY NC SA), a page number "6", the name "Pr. Bruno Falissard", and the logo for Université Paris-Sud.

[01:17] Il existe un forum de discussion. N'hésitez pas à poser des questions en temps réel. Nous répondrons au moins une fois par jour.

The screenshot shows a Moodle-based course interface for the FUN MOOC. At the top, there's a navigation bar with tabs: Courseware, Info Cours, Discussion, **Wiki**, Progression, Fonctionnement du cours, and Ressources pédagogiques. Below the navigation is a large white area containing a scatter plot of gray dots. A sidebar on the right contains links for View, Edit, Changes, and Settings, along with a note about the last modification date (Oct. 24, 2014, 7 p.m.). At the bottom left is a Creative Commons BY-NC-SA license logo, and at the bottom right is a logo for Université Paris-Sud.

FUN  
Introduction à la statistique avec R > Introduction au MOOC

Courseware Info Cours Discussion **Wiki** Progression Fonctionnement du cours Ressources pédagogiques

Instructor

Wiki > Introduction à la statistique a... > + Add article

**Introduction à la statistique avec R - Wiki**

Bienvenue dans ce wiki du cours Introduction à la statistique avec R.

**ATTENTION :** Ne posez pas vos questions sur le Wiki. Nous vous donnons rendez-vous sur le forum de discussion pour répondre à vos questions. Merci.

- Questions fréquentes
- Transcription des vidéos

View Edit Changes Settings

LAST MODIFIED:  
Oct. 24, 2014, 7 p.m.

See all children

7 Pr. Bruno Falissard

UNIVERSITÉ PARIS SUD

[01:25] Il existe un Wiki, que les étudiants ont créé au fil des sessions du MOOC. Il existe des réponses aux questions les plus fréquentes et un grand nombre de ressources sur R que l'on peut trouver sur internet.

The screenshot shows a web page for a MOOC titled "Introduction à la statistique avec R > Introduction au MOOC". The page features a large empty box on the left, likely for a video or presentation. At the top right is the logo for Université Paris-Sud. Below the title, there's a navigation bar with links: "Courseware", "Info Cours", "Discussion", "Wiki", "Progression", "Fonctionnement du cours", and "Ressources pédagogiques", with the latter being underlined in red. The main content area contains two sections: "INSTALLATION DES LOGICIELS R ET RSTUDIO" and "INTRODUCTION À R". The "INTRODUCTION À R" section includes a note about installing R and RStudio, a list of two steps for download, and a link to a document titled "Les bases du langage R". At the bottom left is a Creative Commons license logo (CC BY NC SA), and at the bottom right is the name "Pr. Bruno Falissard" next to the Université Paris-Sud logo.

[01:36] Enfin, quelque chose de très important, l'onglet "Ressources pédagogiques" sur lequel il y a de nombreux documents écrits correspondant au MOOC.

- Il y a les scripts des labs.
- Il y a aussi tous les fichiers de données que vous êtes susceptibles d'analyser pendant ces 5 semaines.
- Et puis des fichiers pdf correspondant à chaque diapositive.

Planning du MOOC

Introduction à la statistique avec R > Introduction au MOOC

R

1. Définitions, représentations graphiques, mesures de position et de dispersion

2. Coefficient de corrélation, risque relatif et odds-ratio

3. Tests statistiques

4. Régression linéaire, Régression logistique

5. Données de survie, méthodes exploratoires multidimensionnelles

- Les « labs »

[01:57] À propos du planning des 5 semaines, nous verrons, successivement,

- les bases des statistiques, avec définitions, représentations graphiques, mesures de position et de dispersion.
- Nous verrons ensuite les statistiques d'association, dont le coefficient de corrélation, risque relatif et odds-ratio ;
- enfin un cours central, le cours relatif aux tests statistiques.
- Nous verrons ensuite les modèles, plus sophistiqués : régression linéaire, régression logistique.
- Et nous terminerons avec la cinquième semaine avec les données de survie et les méthodes exploratoires multidimensionnelles.

Au cours de chacune de ces semaines vous aurez des labs, des cours plus techniques, un peu plus délicats d'approche. Vous n'êtes pas obligés de les suivre, du moins dans une première étape. Vous pouvez y revenir quand vous voulez, au moment où vous vous sentez prêts.

Les modalités d'évaluation

Introduction à la statistique avec R > Introduction au MOOC

- Quiz « mémorisation »
- Quiz « exercices »
- Devoir avec correction par les pairs

10

Pr. Bruno Falissard

UNIVERSITÉ PARIS SUD

[02:45] Dans un MOOC, les modalités d'évaluation font partie intégrante du processus pédagogique. Vous disposerez de trois modalités d'évaluation des connaissances :

- les quizz mémorisation,
- les quizz exercice,
- et enfin un devoir avec correction par les pairs.

Les modalités d'évaluation

Introduction à la statistique avec R > Introduction au MOOC



## Quiz « mémorisation »

- Pour une distribution symétrique:
  - A : La moyenne est égale à la médiane.
  - B : La moyenne est plus petite que la médiane.
  - C : La moyenne est plus grande que la médiane.
  - D : On ne peut pas savoir.

11

Pr. Bruno Falissard





[03:01] Les quizz mémorisation, vous les ferez après avoir visionné chacune des vidéos.

Ils sont faciles, ils sont destinés à fixer ce que vous avez appris, ce que vous devez retenir après avoir vu la vidéo.

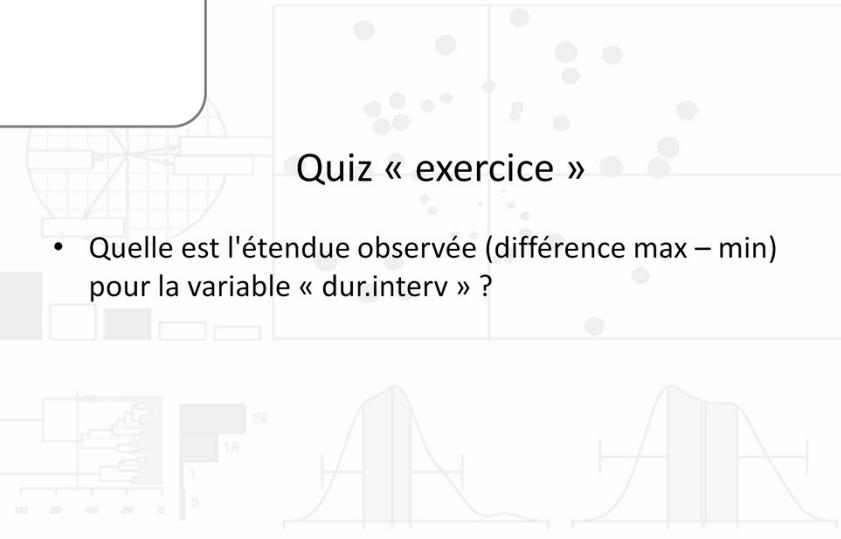
Les modalités d'évaluation

Introduction à la statistique avec R > Introduction au MOOC



Quiz « exercice »

- Quelle est l'étendue observée (différence max – min) pour la variable « dur.interv » ?



CC BY NC SA

12

Pr. Bruno Falissard

UNIVERSITÉ PARIS SUD

[03:13] Les quizz exercices sont un peu plus difficiles. Parfois ce sont des questionnaires à choix multiples, parfois vous devrez faire un calcul et donner le résultat du calcul dans une petite fenêtre.

Les modalités d'évaluation

Introduction à la statistique avec R > Introduction au MOOC

**Devoir**

- Écrire un script R permettant de répondre à une liste de questions, copier/coller dans une fenêtre
- Vous allez tester, valider et noter des scripts réalisés par
- Votre script sera testé, validé et noté par un petit nombre d'étudiants tirés au sort.

CC BY NC SA

13

Pr. Bruno Falissard

UNIVERSITÉ PARIS SUD

[03:23] Enfin, il y a des devoirs, il y en aura 3 : un à la semaine 1, un à la semaine 3 et un à la semaine 5.

Ces devoirs reposeront sur un nouveau jeu de données.

- Vous devrez répondre à des questions, écrire des scripts. Ces scripts vous en ferez des copier/coller, vous les mettrez dans une fenêtre, sur FUN.
- En échange vous allez recevoir des scripts d'autres étudiants que vous devrez tester, valider et noter
- En échange de quoi, votre propre script sera à son tour testé, validé et noté par un certain nombre d'étudiants. C'est ce qu'on appelle une validation par les pairs.

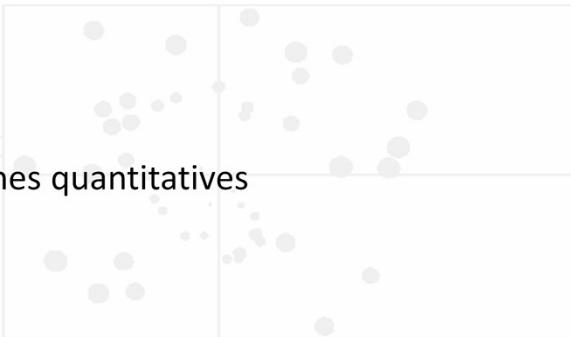
Les recherches quantitatives

Introduction à la statistique avec R > Introduction au MOOC



• Les recherches quantitatives

- Biologie
- Psychologie
- Sociologie
- Économie
- Santé publique
- ...




14

Pr. Bruno Falissard

[03:56] Maintenant, voyons quelques principes fondamentaux qui régissent les statistiques.

Dans un grand nombre de disciplines comme la biologie, la psychologie, la sociologie, l'économie ou la santé publique, on peut réaliser ce que l'on appelle des recherches quantitatives, c'est-à-dire que l'on pratique des mesures sur des individus, des animaux ou des cultures cellulaires, et on essaie de mettre en relation ces différentes variables.

Un exemple d'étude

Introduction à la statistique avec R > Introduction au MOOC

- Santé publique : santé mentale en prison
  - Âge
  - (sexe)
  - Profession
  - Nombre d'incarcérations antérieures
  - Nombre d'enfants
  - Personnalité pathologique
  - Maltraitance infantile
  - Déprimé
  - ...

CC BY NC SA

15

Pr. Bruno Falissard

UNIVERSITÉ PARIS SUD

[04:23] Par exemple, si on fait de la santé publique, on peut s'intéresser aux troubles psychiatriques que présentent les détenus incarcérés. Pour ce faire, on va réaliser une étude à l'aide de questionnaires, où l'on va renseigner

- l'âge des détenus,
- leur sexe,
- leur profession,
- leur nombre d'incarcérations antérieures,
- leur nombre d'enfants,
- l'existence d'une personnalité pathologique,
- d'une maltraitance infantile,
- d'une dépression,
- d'une schizophrénie,
- d'un trouble anxieux,
- etc.

On peut avoir de la sorte plusieurs dizaines, voire plusieurs centaines de variables mesurées sur plusieurs centaines, voire plusieurs milliers d'individus. Une fois l'étude réalisée, quand on essaie de lire l'ensemble des questionnaires, il est quasiment impossible d'avoir une vue d'ensemble de ces données. On a besoin d'un outil et cet outil c'est la statistique. La statistique va permettre de faire une synthèse des données, de chacune des variables prises indépendamment les unes des autres. La statistique va aussi permettre de mettre en relation ces variables les unes avec les autres et de savoir si le hasard peut expliquer à lui tout seul les relations que l'on a mises en évidence. On peut dire de la sorte que la statistique, c'est l'interface entre les nombres mesurés dans l'étude et les hommes qui essaient de donner du sens à ces nombres. La statistique est une interface entre les nombres et les hommes.

Un fichier de données

Introduction à la statistique avec R > Introduction au MOOC

| A                       | B    | C        | D        | E         | F        |
|-------------------------|------|----------|----------|-----------|----------|
| 1 age                   | prof | dep.cons | scz.cons | grav.cons | n.enfant |
| 2 31 autre              |      | 0        | 0        |           | 1        |
| 3 49                    |      | 0        | 0        |           | 2        |
| 4 50 prof intermédiaire |      | 0        | 0        |           | 2        |
| 5 47 ouvrier            |      | 0        | 0        |           | 1        |
| 6 23 sans emploi        |      | 1        | 0        |           | 2        |
| 7 34 ouvrier            |      | 0        | 0        |           | 1        |
| 8 24 autre              |      | 1        | 0        |           | 5        |
| 9 52 artisan            |      | 0        | 0        |           | 1        |
| 10 42 ouvrier           |      | 1        | 0        |           | 5        |

Pr. Bruno Falissard

UNIVERSITÉ PARIS SUD

[05:44] Quand on réalise une étude de ce type, les données sont saisies sur ordinateur et elles sont finalement disponibles sous forme d'une feuille de calcul que l'on peut ouvrir avec Excel ou avec OpenOffice comme ici. Nous avons donc en pratique une feuille de calcul avec une variable par colonne, et un individu évalué par ligne.

Dans le cas de l'étude santé mentale en prison, nous avons donc sur la première ligne

le nom des variables mesurées, avec

- l'âge,
- la profession,
- l'existence d'une dépression,
- d'une schizophrénie,
- etc.

Et puis par ligne chaque détenu, avec par exemple à la ligne 5, le quatrième détenu qui a 47 ans, qui est ouvrier, et qui n'a pas de dépression, puisqu'on a zéro.

L'étude « santé mentale en prison »

Introduction à la statistique avec R > Introduction au MOOC



- Entre 2003 et 2004
- Maisons d'arrêt, centres de détention, maisons centrales
- 799 hommes détenus dans les prisons françaises
- Deux investigateurs : junior / senior
- Données
  - Variables sociodémographiques et biographiques
  - Troubles mentaux
  - Éléments de personnalité
- Fichiers : smp1, smp2

 17 Pr. Bruno Falissard 

[06:27] Voyons donc maintenant avec un peu plus de détails l'objet de cette étude santé mentale en prison.

Elle a été réalisée pendant les années 2003 et 2004.

Un certain nombre d'établissements carcéraux ont été tirés au sort parmi les maisons d'arrêts, les centres de détention, et les maisons centrales de France Métropolitaine.

Dans un second temps, 799 hommes détenus ont été tirés au sort parmi ces établissements eux-mêmes tirés au sort.

Chacun des détenu a été interrogé par deux enquêteurs ; l'un, ayant une expérience clinique d'au moins 5 ans, on l'appelle senior ; l'autre, moins expérimenté, est appelé junior ; les deux étaient soit psychologues, soit psychiatres. Le junior commençait l'entretien par un questionnaire standardisé balayant l'ensemble de la psychopathologie. Le senior, dans un second temps, pratiquait un entretien ouvert, plus à même de rechercher des troubles qui pouvaient ne pas apparaître lors de l'entretien structuré.

Un grand nombre de données ont été mesurées, plusieurs centaines, parmi lesquelles :

- des variables sociodémographiques et biographiques, notamment une histoire de l'enfance du détenu, est-ce qu'il a été maltraité, est-ce qu'il a été placé par un juge pour enfants, etc.
- l'existence ou non d'une grande liste de troubles mentaux
- et enfin des éléments de personnalité.

Nous utiliserons deux fichiers relatifs à cette étude santé mentale en prison. Un petit fichier plus facile à utiliser, smp1, et un fichier plus lourd, utilisé dans la fin du MOOC, qui sera smp2.

The screenshot shows a window titled "Exportation d'un fichier de données" with the subtitle "Introduction à la statistique avec R > Introduction au MOOC". In the top right corner is the logo of Université Paris-Sud. The main area displays a scatter plot with several grey dots. Below the plot is a table titled "smp1.ods - OpenOffice.org Calc". The table has four columns labeled C, D, E, and F. The data is as follows:

|          | C | D | E         | F        |
|----------|---|---|-----------|----------|
| dep.cons | 0 | 0 | grav.cons | n.enfant |
|          | 0 | 0 |           | 1        |
|          | 0 | 0 |           | 2        |
|          | 0 | 0 |           | 2        |
|          | 0 | 0 |           | 1        |
|          | 1 | 0 |           | 2        |
|          | 0 | 0 |           | 1        |
|          | 1 | 0 |           | 5        |
|          | 0 | 0 |           | 1        |
|          | 1 | 0 |           | 5        |

At the bottom left is a Creative Commons BY-NC-SA license logo. At the bottom center is the number "18". At the bottom right are the names "Pr. Bruno Falissard" and "UNIVERSITÉ PARIS SUD".

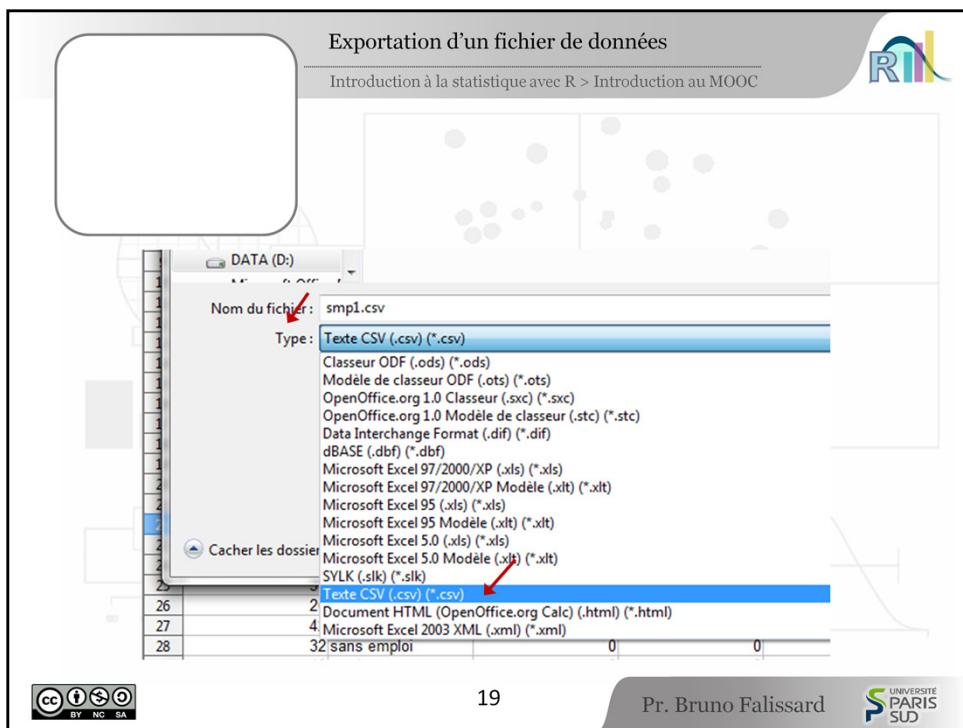
[08:04] Alors pour terminer, quelques aspects techniques, peut-être un petit peu rébarbatif, mais c'est un passage obligé avant de se mettre véritablement à analyser des données.

Premier point essentiel, l'exportation d'un fichier de données.

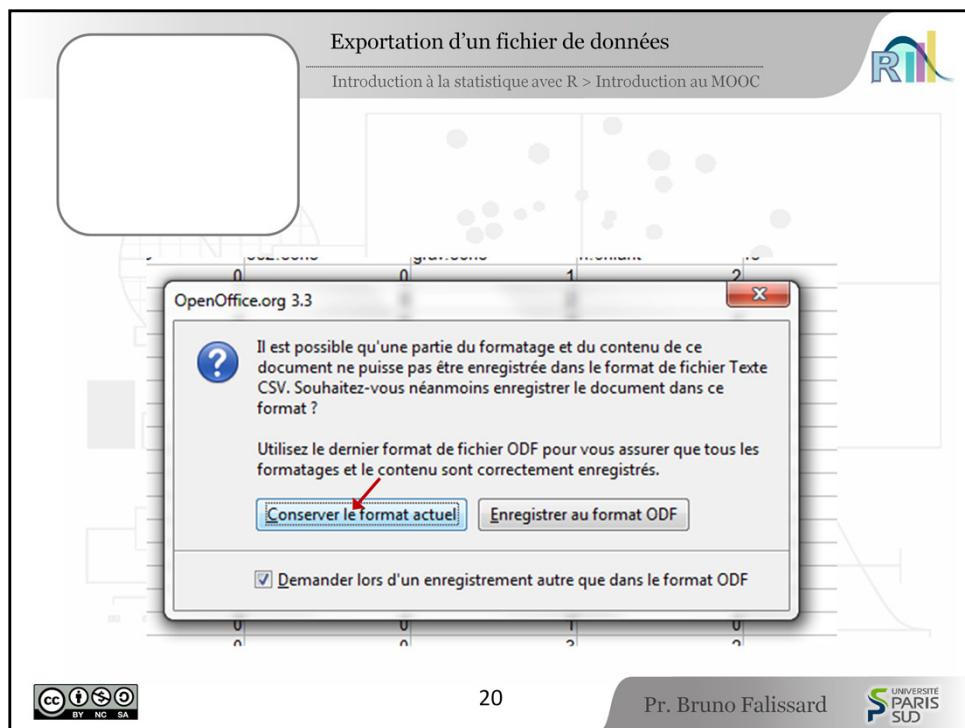
Nous avons donc une feuille de calcul Excel ou ici OpenOffice avec des variables en colonnes et des sujets en lignes.

R ne permet pas d'analyser un fichier Excel ou un fichier OpenOffice. Il faut exporter ce fichier dans un format adéquat, en général un format CSV pour "Comma Separated Value".

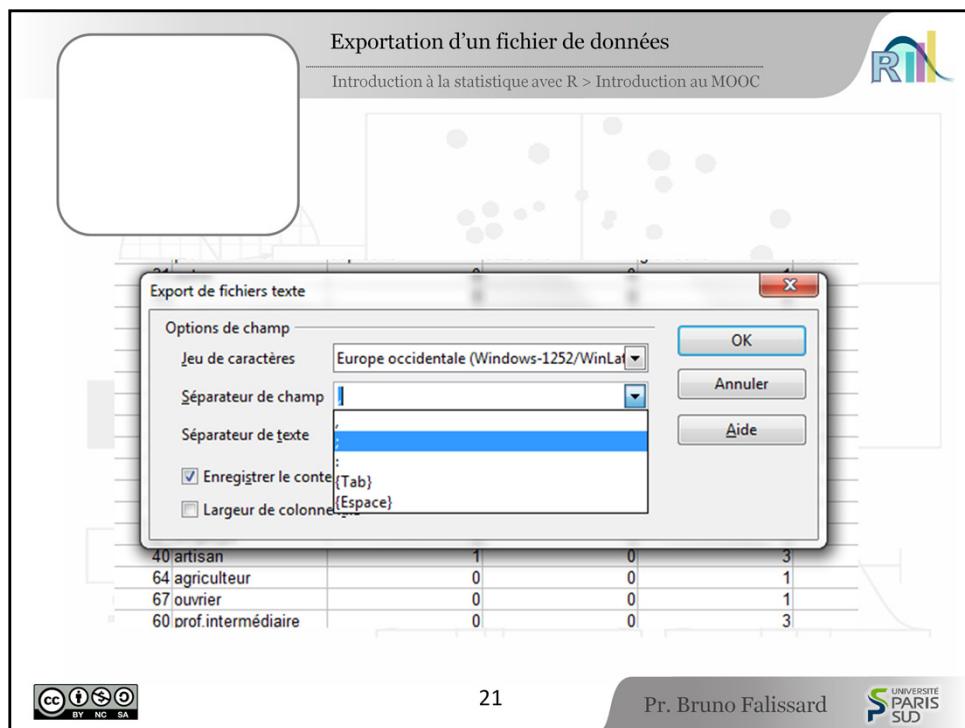
Pour sauvegarder un fichier au format CSV, vous cliquez dans le menu "Fichier" / ensuite, "Enregistrer sous..."



[08:49] et là vous sélectionnez comme type de fichier, le format CSV.



[08:55] Alors, d'un logiciel à l'autre, il va vous être demandé de conserver le format, etc. Vous répondez en général par oui, ça ne pose pas de problèmes.



[09:04] Pour OpenOffice, en particulier mais pas pour Excel, vous devez spécifier explicitement que entre les valeurs il doit y avoir un point-virgule, dans le but d'obtenir un bon format de fichier.

A ce stade, vous cliquez "OK" et vous avez obtenu votre fichier au format CSV que vous allez pouvoir importer dans R.



[09:25] Faisons maintenant nos premiers pas avec R.

Dans un premier temps, il faut le télécharger. Pour cela, vous allez sur le site de R.

Vous pouvez taper "CRAN" sur n'importe quel moteur de recherche, vous tomberez directement sur ce lien.

Vous avez le choix entre télécharger la version de R pour Linux, pour Mac, ou pour Windows. Ici, nous allons installer la version Windows. Nous cliquons donc sur le lien correspondant.

Le logiciel R

Introduction à la statistique avec R > Introduction au MOOC

Boîte de réception - brun... Google Actualités The Comprehensive R Arc...

cran.r-project.org

Applications Gmail Google Ad. Search Actu Scholar Groups Drive Trends GTranslate Dico

**R**

Subdirectories:

- [base](#) ↗
- [contrib](#)
- [Rtools](#)

Binaries for base distribution (mainly Windows).

Binaries of contributed packages (not CRAN Windows services and contributed tools).

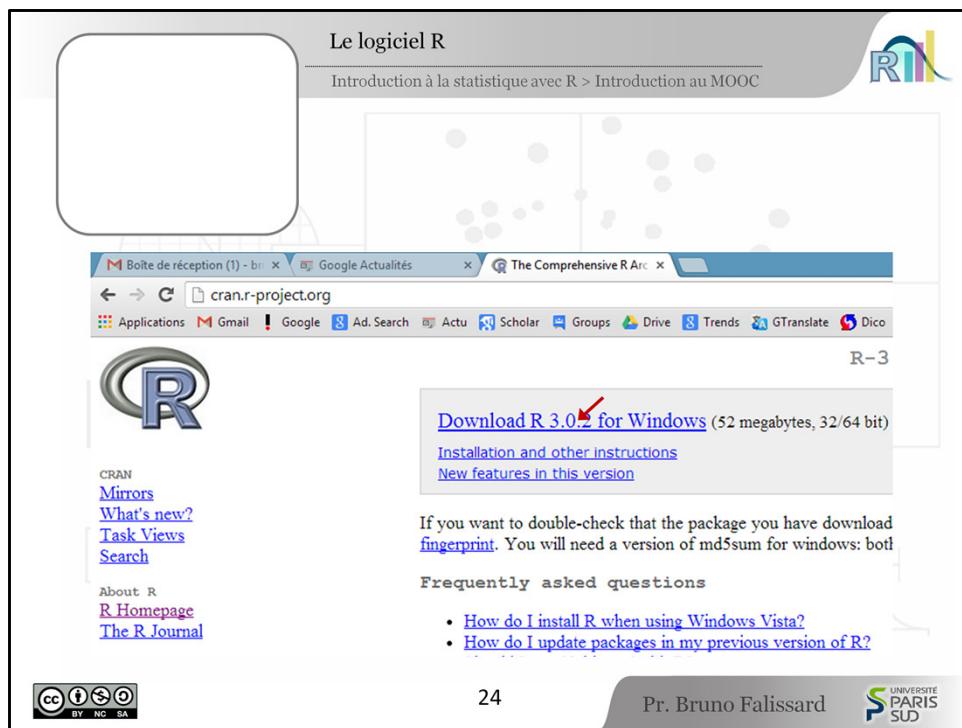
Tools to build R and R packages (rtools) for Windows, or to build R itself.

Please do not submit binaries to CRAN. Package developers might upload Windows binaries.

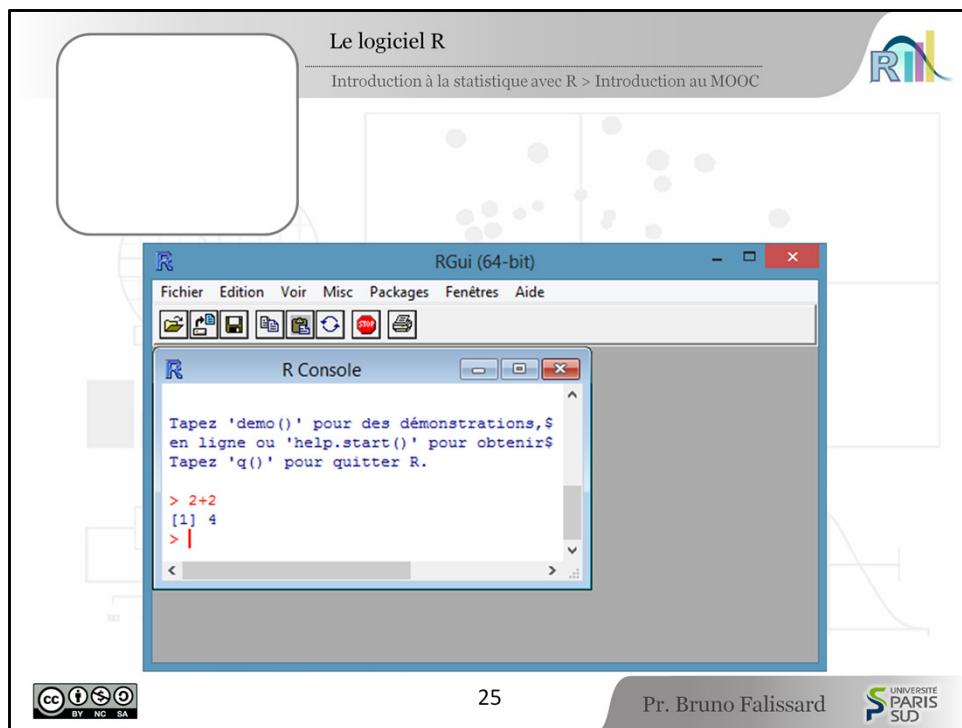
You may also want to read the [R FAQ](#) and [R for Windows FAQ](#).

23 Pr. Bruno Falissard

[09:50] Nous cliquons ensuite sur le lien base.



[09:52] et nous téléchargeons le logiciel que l'on installe comme à peu près tous les logiciels.



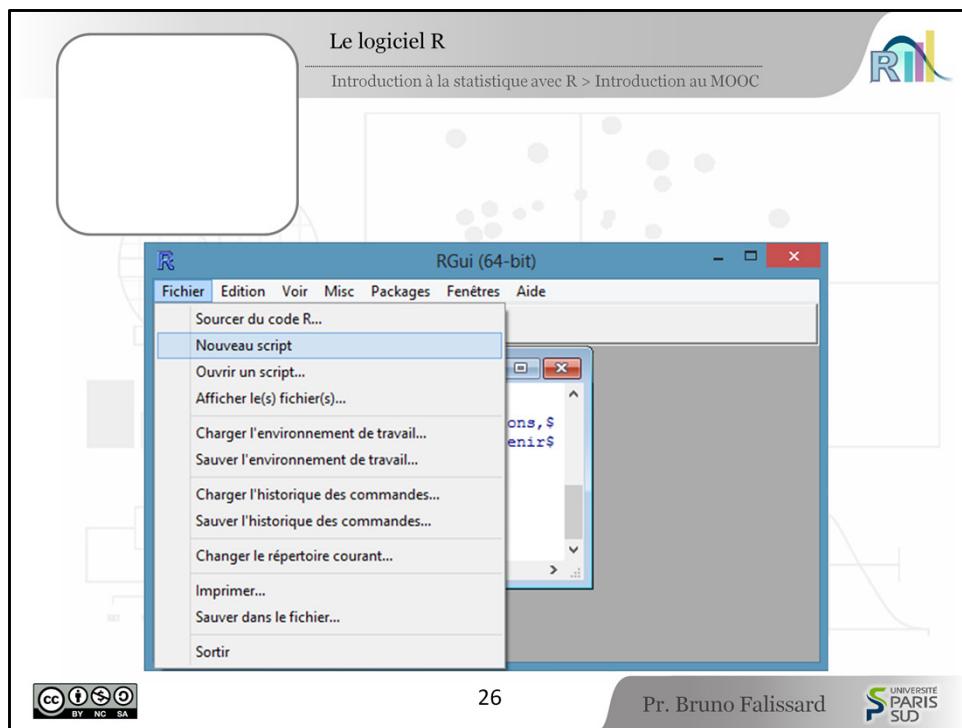
[09:58] Une fois que vous cliquez sur l'icône R présente sur le bureau de votre ordinateur, le logiciel s'ouvre sur une fenêtre que l'on appelle la console.

Vous pouvez vous servir de cette console comme d'une calculatrice.

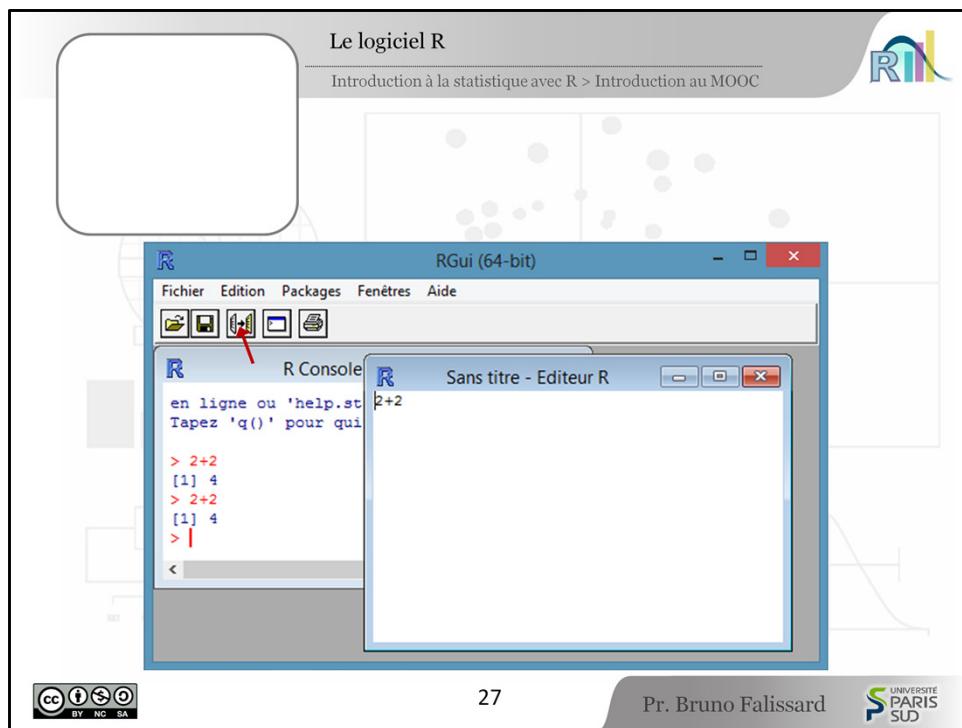
Par exemple,

- si vous tapez "> 2+2"
- vous appuyez sur la touche "Entrée" ou "Return"
- vous obtenez "> 4."

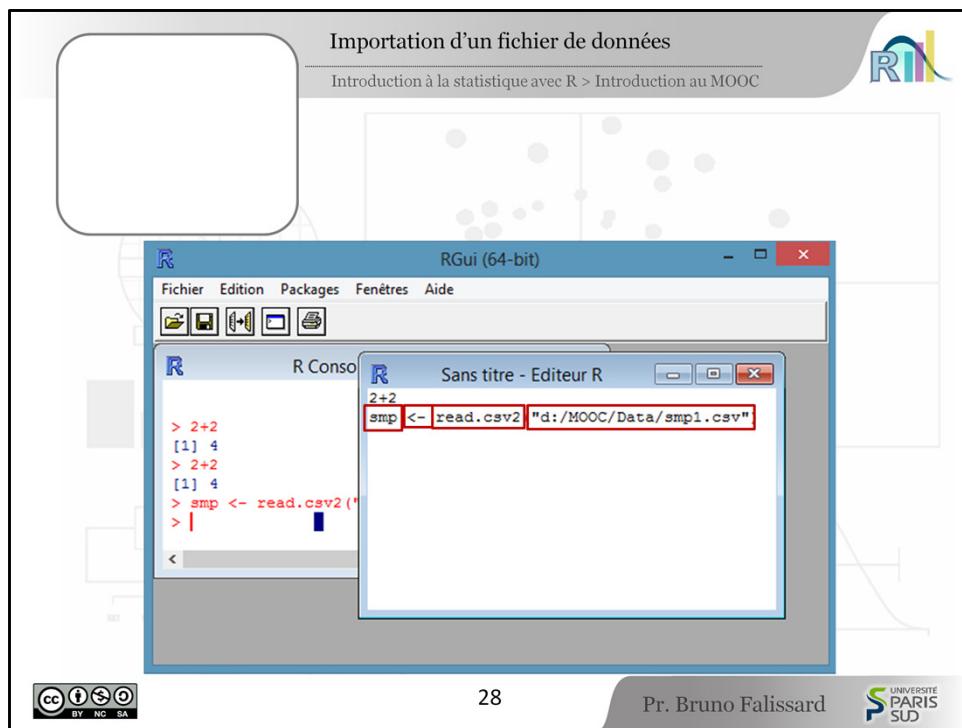
En général, on ne tape pas les instructions directement dans la console de R. Il faut ouvrir une fenêtre appelée "Script".



[10:26] Pour cela, on va dans le menu "Fichier", on clique sur l'option "Nouveau script",



[10:32] une fenêtre s'ouvre dans laquelle vous pouvez taper également "> 2+2"  
vous cliquez sur l'icône "Run", "Action",  
et vous obtenez de nouveau "> 4" dans la fenêtre console.



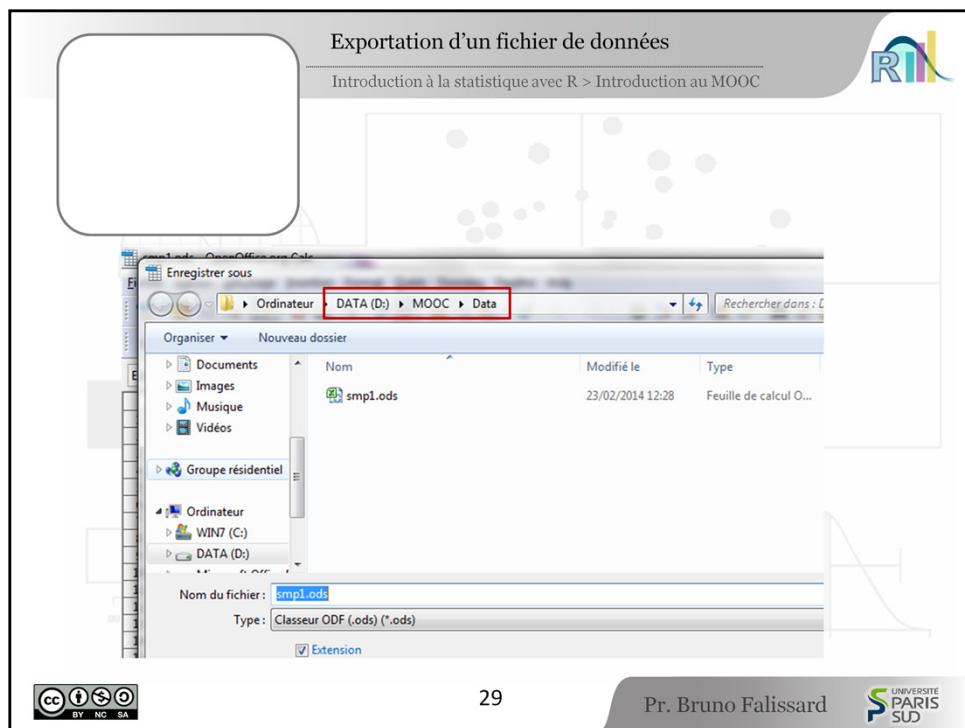
[10:44] Alors, voyons voir maintenant comment importer le fichier de données "Santé mentale en prison" dans R.

Pour cela, dans la fenêtre de script, nous allons utiliser l'instruction suivante :

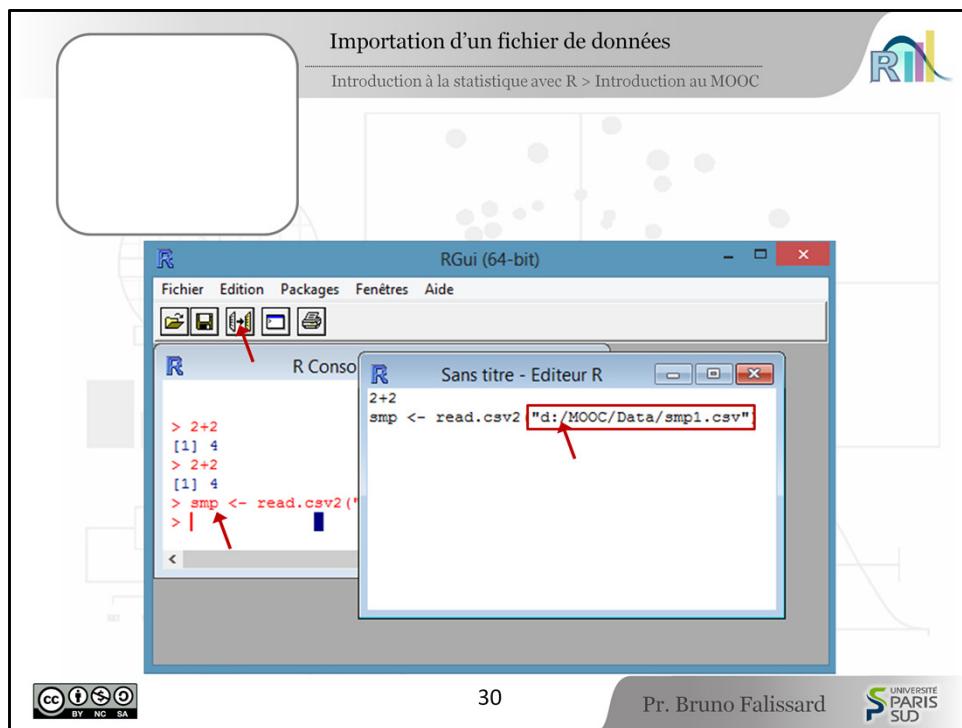
elle est constituée

- à l'extrême gauche du nom de l'objet R qui va contenir le fichier, ici `smp`
- ensuite d'une flèche, une flèche d'attribution, constituée en pratique du signe inférieur et du signe trait d'union (`<-`),
- enfin de la fonction `read.csv2()` qui est la fonction R qui est capable de lire le fichier CSV que nous avons sauvégarde tout à l'heure.

Alors ce fichier CSV, il faut indiquer très précisément à R où il est stocké et en fonction des ordinateurs, des systèmes d'exploitation, ce n'est pas toujours évident d'avoir ce chemin.

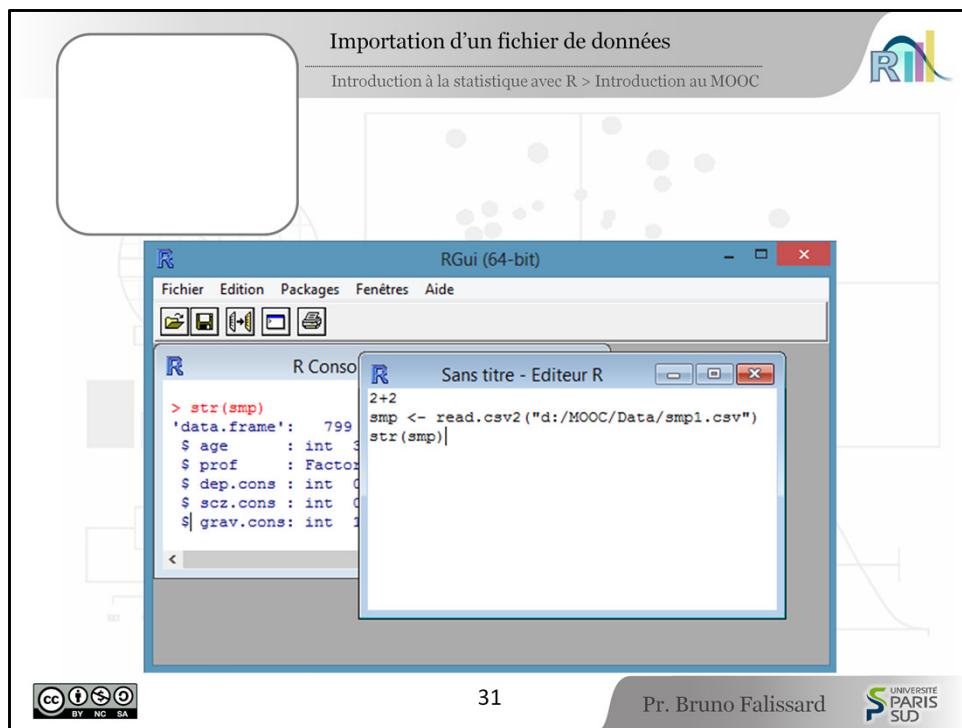


[11:29] Une astuce est de se souvenir, tout à l'heure, quand nous avons stocké au format CSV le fichier, dans la ligne du haut du tableur, il y avait indiqué "D:/MOOC/DATA", qui correspond au chemin précis dans lequel a été stocké le fichier.

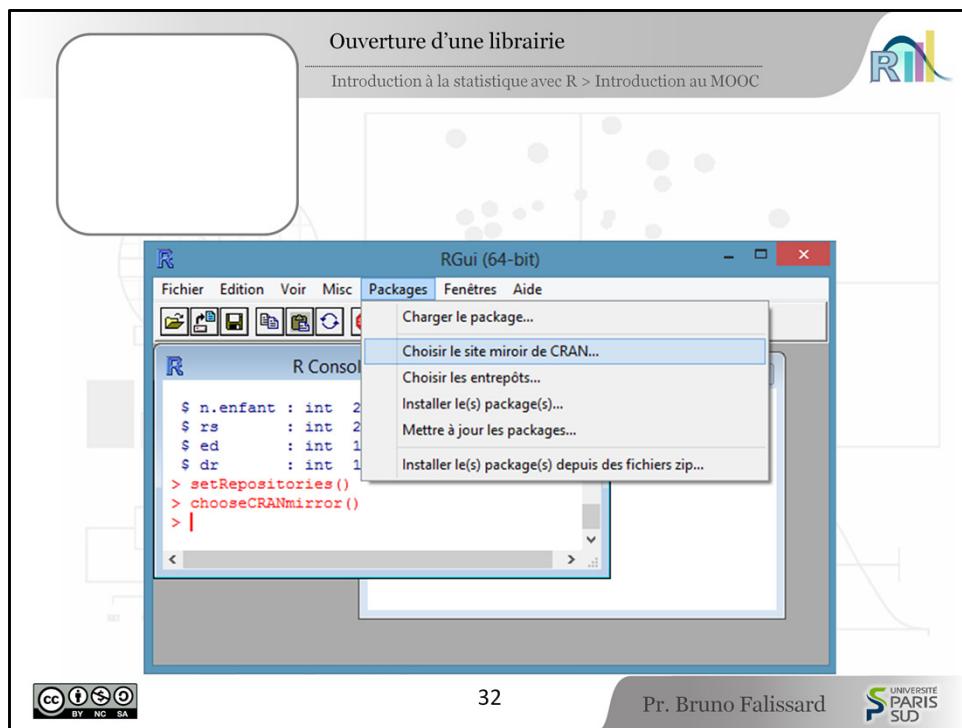


[11:46] Alors bien attendu, faites attention aux lignes de fraction, au sens des lignes de fraction qui sont utilisées, aux guillemets qu'il y a à gauche et à droite du chemin etc., etc.

Alors une fois que cette instruction est précisément écrite, on clique sur l'icône Run et notre ligne de commande est acceptée. R, par défaut, n'indique aucun résultat.



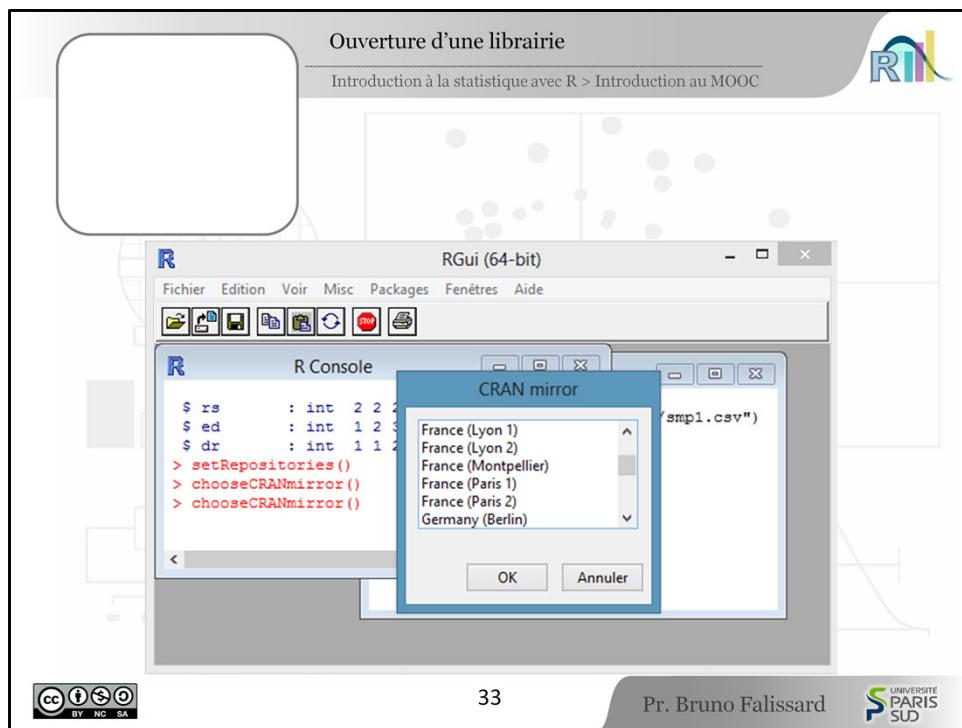
[12:09] Si vous voulez avoir le contenu du fichier `smp`, vous tapez `str(smp)`, et dans la fenêtre, dans la console, nous avons des éléments du contenu du fichier, mais nous verrons ça dans la première vidéo de stat avec beaucoup plus de détails.



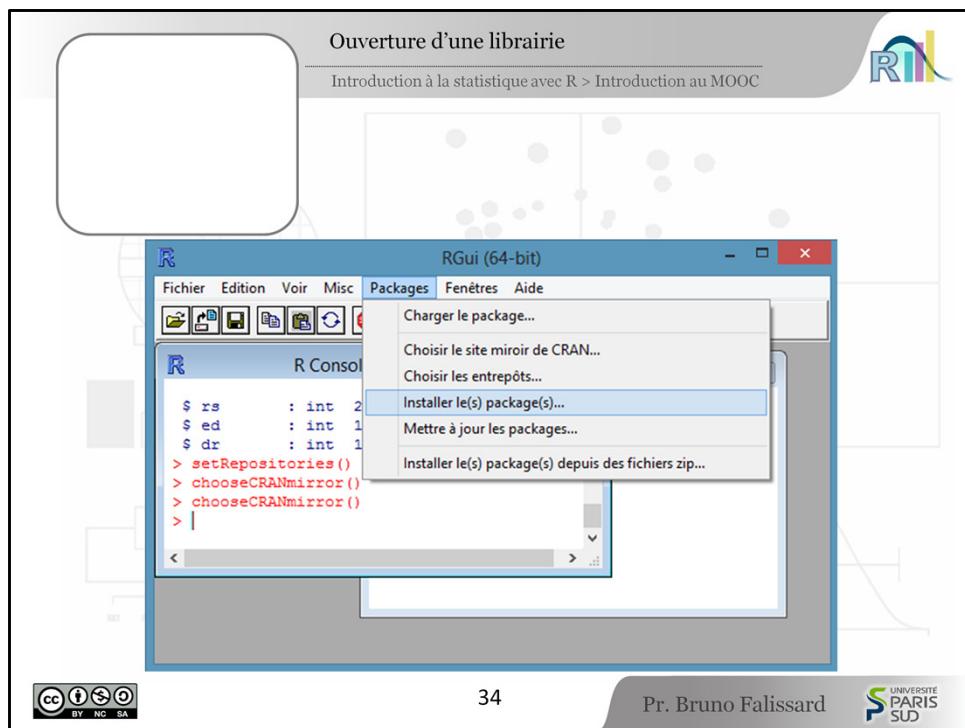
[12:27] Les fonctions R sont regroupées par librairies et toutes les librairies ne sont pas disponibles par défaut quand vous téléchargez la version de base de R.

Aussi, en fonction des méthodes statistiques que vous allez utiliser, vous serez vraisemblablement amenés à importer de nouvelles librairies.

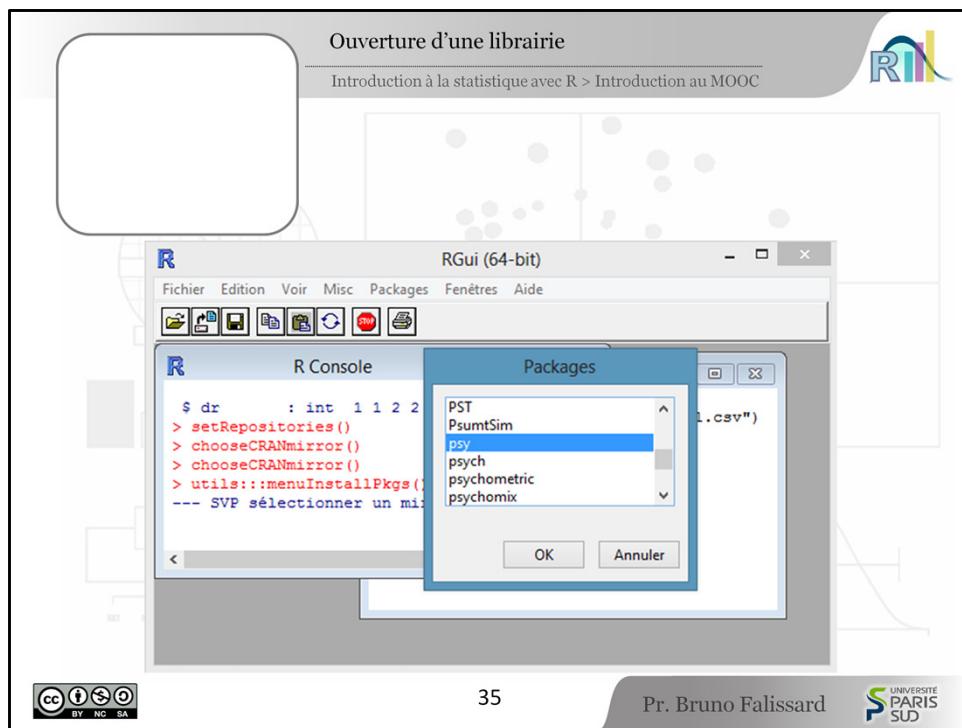
Pour ce faire, vous devez cliquer dans le menu "Packages", tout d'abord, le site miroir de CRAN sur lequel vous allez télécharger votre librairie.



[12:54] Vous allez choisir une des librairies disponibles en France.

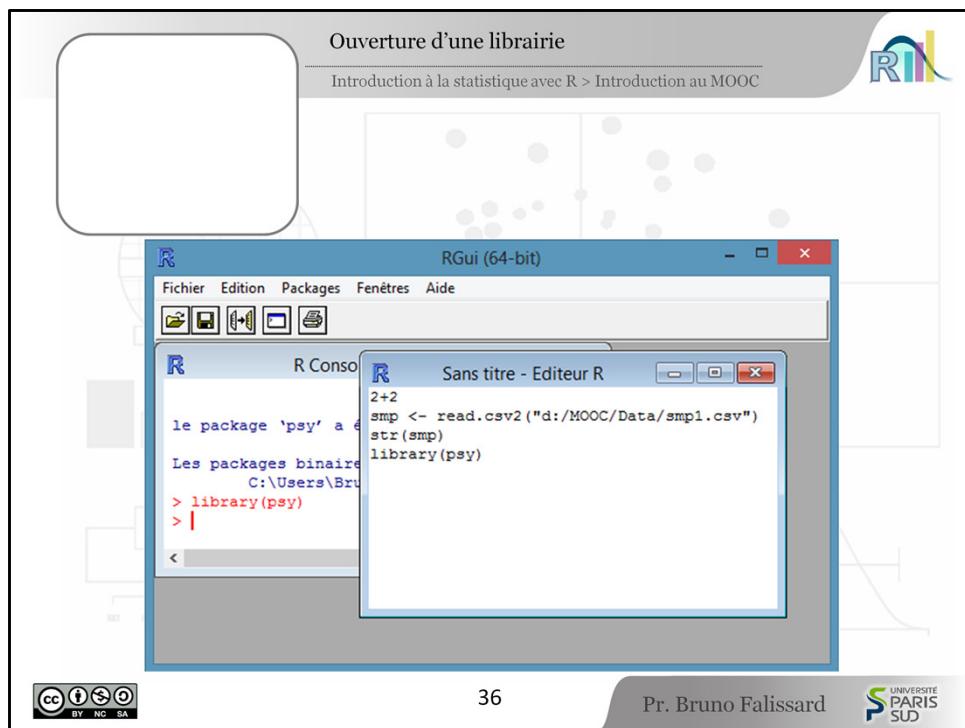


[12:58] Une fois cela fait, vous allez de nouveau dans le menu "Packages", vous installez un package,



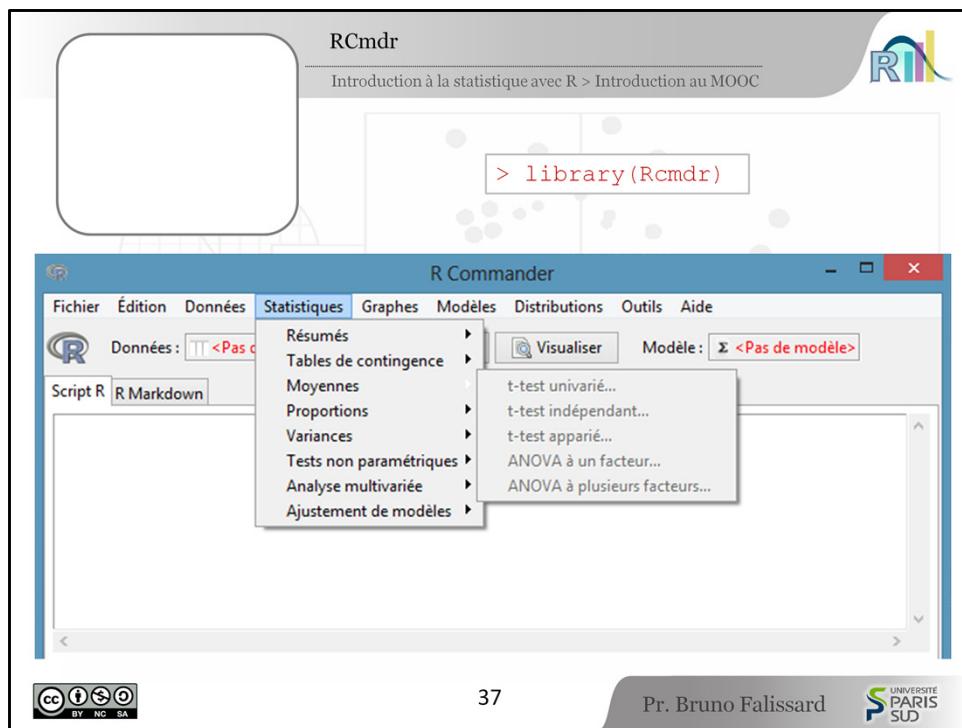
[13:04] vous choisissez le nom du package voulu, par exemple ici "psy", vous cliquez "OK" et le package est installé sur votre ordinateur.

Vous n'avez à faire ça qu'une seule fois pour chaque ordinateur. La librairie est stockée définitivement sur votre disque dur.



[13:20] Par contre, quand vous avez besoin d'appeler une librairie donnée, vous devez écrire dans la fenêtre de script `library` de la librairie correspondante ici par exemple `psy` : `library(psy)`.

Et ça, vous devez le faire à chaque session de travail de R mais nous reviendrons aussi là-dessus dans les vidéos de stat à venir.



[13:41] Vous pouvez parfaitement utiliser R à partir d'une fenêtre de script et de la fenêtre console sur laquelle les calculs sont affichés.

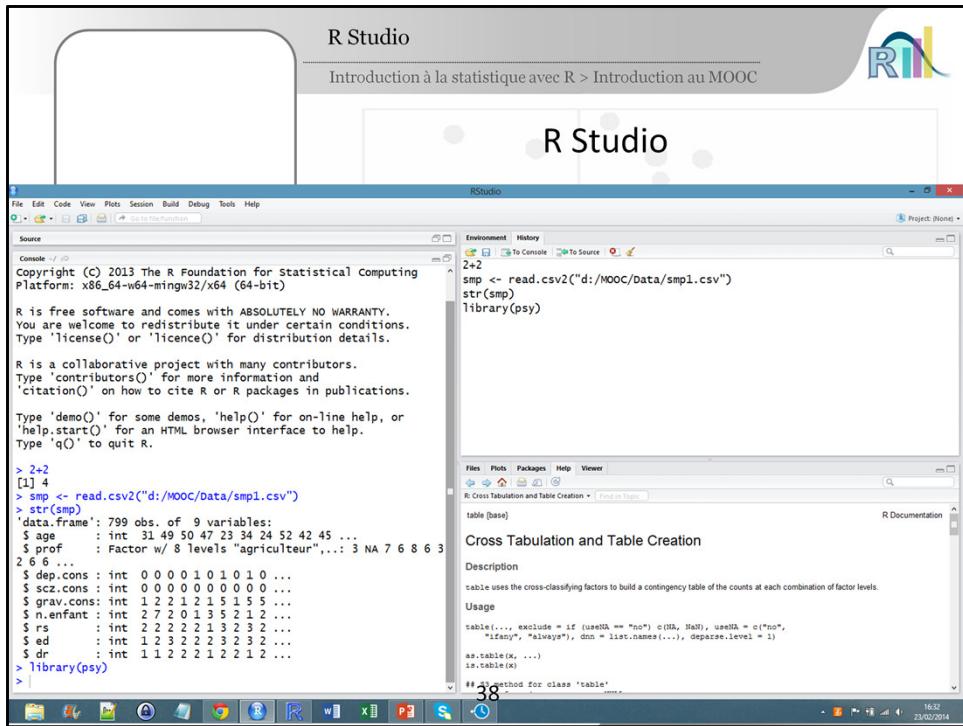
Il existe cependant des surcouches logicielles de R qui permettent de rendre l'écriture des scripts et les calculs plus faciles.

Une première de ces surcouches s'appelle RCommander.

C'est une librairie R que vous pouvez appeler en tapant `library(Rcmdr)` après avoir téléchargé la librairie correspondante naturellement.

Cette surcouche est particulièrement appréciée parce qu'elle permet d'utiliser des menus déroulants au lieu d'utiliser les fonctions R à taper sous forme de script dans la fenêtre de script.

Nous vous déconseillons d'utiliser RCommander. C'est une solution de facilité qui ne permet pas d'avoir une approche véritablement rigoureuse de R et d'ailleurs, nous n'utiliserons pas RCommander dans la suite du MOOC.



[14:38] Une autre surcouche logicielle, elle, très intéressante, est "R Studio".

Ça n'est pas une librairie R, c'est un logiciel à part entière que vous pouvez télécharger gratuitement sur le web.

R Studio permet sous une même interface, de jongler entre une fenêtre de script, une fenêtre de console, une fenêtre d'aide, une fenêtre de représentation graphique...

Cette interfaçage est particulièrement utile, en particulier quand vous avez de gros projets statistiques à écrire.

R Studio sera notamment utilisé pendant tous les labs du MOOC "Introduction à la statistique avec R".

En tout état de cause, vous pouvez bien entendu vous satisfaire de l'interface classique de R avec sa console et ses fenêtres de script.

Les modalités d'évaluation

Introduction à la statistique avec R > Introduction au MOOC

RMOOC

Bon courage !

CC BY NC SA

39

Pr. Bruno Falissard

UNIVERSITÉ PARIS SUD

[15:25] Maintenant, il me reste à vous souhaiter un bon courage, pour ces quatre semaines à venir, à l'issue desquelles vous serez de vrais professionnels de R.