

BEGIN'R

Un projet collaboratif pour la production de supports innovants pour l'enseignement en autonomie et à distance de R

Jean-Pierre Da Costa ¹, Florent Arnal ², Lionel Bombrun ¹, Cécile Dantzer ³, Robin Genuer ⁴, Marthe-Aline Jutand ⁴, Marie Lebreton ⁵, Sébastien Moutault ², Jérôme Saracco ⁶, Raphaëlle Savoie ⁷, Perrine Soret ⁴

¹ *Bordeaux Sciences Agro, prenom.nom@agro-bordeaux.fr*

² *IUT de Bordeaux, Univ. Bordeaux, prenom.nom@u-bordeaux.fr*

³ *Collège Sciences de l'Homme, Univ. Bordeaux, prenom.nom@u-bordeaux.fr*

⁴ *ISPED, Univ. Bordeaux, prenom.nom@u-bordeaux.fr*

⁵ *Collège DSPEG, Univ. Bordeaux, prenom.nom@u-bordeaux.fr*

⁶ *ENSC, Bordeaux INP, prenom.nom@ensc.fr*

⁷ *ENSCBP, Bordeaux INP, prenom.nom@enscbp.fr*

Résumé. La formation à l'utilisation du logiciel de statistique R et de l'environnement de développement intégré R Studio est appréhendé de manière très variable selon les utilisateurs. Afin de faciliter cet apprentissage aux personnes non familières avec la programmation et la statistique, le projet *Begin'R*, lancé par des enseignants du site universitaire bordelais, propose un site internet (<http://beginr.u-bordeaux.fr>) regroupant des supports médiatisés (exercices, vidéos, ...). L'objectif de ce papier est de présenter l'organisation du groupe, d'expliquer le choix des outils pour la création de ces supports innovants, et de présenter les premières réalisations.

Mots-clés. Logiciel R, supports de cours innovants, apprentissage à distance.

Abstract. The learning in the use of the R statistical software and the R Studio integrated development environment is apprehended very differently by the user, depending on their competences. In order to facilitate this learning for people unfamiliar with programming and statistics, the *Begin'R* project, composed by professors from Bordeaux, offers a website (<http://beginr.u-bordeaux.fr>) that includes media (exercises, videos, ...). The aim of this paper is to present the organization of the group, to explain the choice of the tools used for the creation of these innovative supports, and to present the first achievements.

Keywords. R software, innovative course materials, e-learning.

Introduction

R est à la fois un langage de programmation et un environnement logiciel dédiés au traitement statistique de données [1]. Il se présente sous la forme d'un environnement de travail, organisé autour d'une console, d'un éditeur de scripts et d'une fenêtre de sortie graphique. Il peut être complété par un EDI i.e. Environnement de Développement Intégré (par exemple RStudio [2]), facilitant la programmation, ou par des IHM i.e. Interface Homme Machine (par exemple RCommander [3]), facilitant la prise en main par des utilisateurs finaux. Puissant et très complet, cet outil est en constante évolution et tend à devenir le logiciel le plus utilisé dans la recherche, l'industrie et l'enseignement.

Toutefois, la prise en main du logiciel R par des praticiens (étudiants, enseignants, chercheurs, professionnels) non familiers de la programmation ne va pas sans poser certaines difficultés. Dans le cas de la formation initiale, par exemple, l'appropriation de R par les étudiants est très variable selon leur goût pour les mathématiques et les statistiques ou leur aisance en programmation et face à l'outil informatique.

Du fait de ces difficultés, la formation à l'utilisation du logiciel R, sous la forme de modules dédiés ou en appui de modules de formation en statistiques, est apparue indispensable [4, 5]. Toutefois, la nature même de l'objectif pédagogique - i.e. formation à un outil logiciel - exige la mise en place de solutions d'apprentissage originales. Des expériences pionnières, menées notamment dans certains établissements partenaires de ce projet, ont montré que les TICE (i.e. Technologies de l'Information et de la Communication pour l'Enseignement) et la formation à distance s'avèrent pertinentes pour ce type d'apprentissage permettant une plus grande souplesse et une meilleure adaptabilité au public ciblé [6, 7].

Le projet *Begin'R* présenté dans cet article vise à mettre en place un dispositif original, convivial et évolutif de formation à l'utilisation du système d'analyse statistique R et de l'environnement de développement intégré R Studio en proposant, sur un site internet, des supports médiatisés (vidéos, supports pdf, exercices, fichiers de données, etc.). Ce dispositif est construit et géré de manière collaborative par les équipes enseignantes de plusieurs établissements. Il cible en premier lieu des cohortes d'étudiants de filières diversifiées (e.g. chimie, agronomie, agroalimentaire, sciences sociales, médecine, etc.) dont la particularité est de n'être familiers ni de programmation ni de statistique .

Dans ce qui suit, nous présenterons en partie 1 l'organisation du groupe *Begin'R* qui a œuvré à ce projet. Nous décrirons ensuite, de façon succincte en partie 2, la charte et les outils éditoriaux choisis pour la conception des modules de formation. En partie 3, nous présenterons quelques unes de nos réalisations et terminerons par un bilan et quelques perspectives.

1 Organisation du groupe Begin'R

Une gouvernance partagée – Le groupe *Begin'R* s'est formé autour de quelques enseignants provenant de différents établissements du site universitaire bordelais (Université de Bordeaux, Bordeaux Sciences Agro et Bordeaux INP). Animé par un chef de projet, il se réunit deux à trois fois par an pour établir la stratégie de publication, échanger sur ses expériences pédagogiques et définir de nouveaux objectifs, en matière de production et mise en ligne de support.

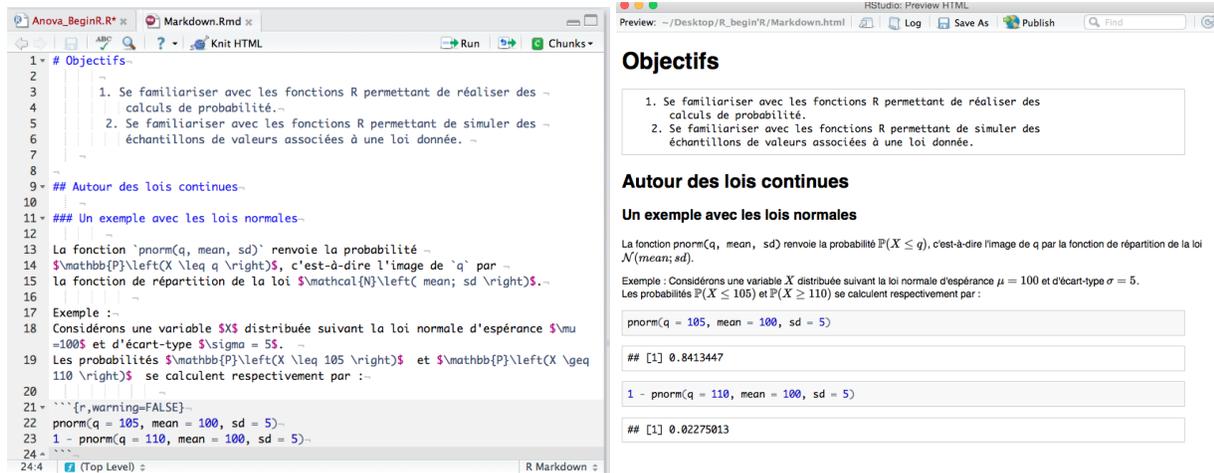
Une cellule technique – Au sein du groupe, une cellule technique, assistée par la cellule MAPI d'innovation pédagogique de l'Université de Bordeaux, se charge de la construction et de la maintenance des outils éditoriaux (voir la partie 3 ci-dessous). Elle gère également les outils de développement collaboratif et la mise en ligne des pages produites par les rédacteurs.

Un groupe de rédacteurs – La rédaction des modules de formation est assurée par les membres du groupes eux-mêmes. Ils se répartissent les chapitres par binômes, en fonction de leur compétences et de leurs priorités pédagogiques.

Une procédure croisée de relecture, harmonisation et validation – Les membres jouent tour à tour le rôle de rédacteur et de relecteur. L'harmonisation des pages produites, en matière de forme et de fond, est réalisée collectivement, de même que leur validation en vue de la publication.

2 Charte et outils éditoriaux

Principes généraux – La stratégie éditoriale choisie est de produire des documents permettant la découverte et l'initiation à l'utilisation de l'environnement R Studio et du langage de programmation R. Compte-tenu du type de public ciblé, relativement large et non spécialiste, une attention particulière a été portée au fait de concevoir des documents simples et attractifs, en évitant tout excès de technicité, à la fois du point de vue algorithmique que du point de vue théorique. Il a été choisi de produire des supports pédagogiques prenant la forme de pages HTML enrichies et dynamiques. Les supports doivent pouvoir incorporer de façon souple des écritures mathématiques, des exemples de codes et leurs résultats. Des redirections vers des vidéos de démonstration, des fichiers de données et des scripts en langage R doivent être possibles. Des exercices et leurs corrections ainsi que des prolongements de cours doivent également pouvoir être proposés de façon dynamique.



(a)

(b)

FIGURE 1 – Exemple de page produite avec R-Markdown. (a) Code R-Markdown incluant des définitions de sections, du texte, des formules \LaTeX et du code R. (b) Page HTML produite avec R-Markdown.

R Markdown, une première piste – Une piste naturelle pour la rédaction de tels supports consiste à combiner les technologies KnitR, Latex et R Markdown [8]. Exécutable depuis l’environnement R studio, ce langage permet d’éditer des documents intégrant des expressions mathématiques au format \LaTeX mais aussi et surtout des lignes de code R. Il permet enfin l’inclusion simplifiée de contenus multimédia (fichiers, images, vidéos). A titre d’exemple, un extrait de code R Markdown est fourni à la figure 1.(a). Son exécution conduit, de façon automatique à l’édition de formules mathématiques et de codes correctement formatés. Lors de la compilation, le code R peut lui même être interprété et le résultat de son exécution peut figurer dans le document final (cf. Figure 1.(b)).

PyWII : une chaîne éditoriale conçue pour les tutoriels – Les documents produits par R Markdown sont simples, efficaces et professionnels. Ils restent cependant statiques et d’apparence relativement austère et n’offrent pas la possibilité de concevoir des documents interactifs. Afin de permettre l’édition de tutoriels satisfaisant nos différents critères, il a été choisi de développer en interne l’outil éditorial PyWII : Python Web Instructions Insertion. La figure 2 montre un exemple de code et de page HTML produite avec PyWII. Le principe est ici d’intégrer dans le code source du document à éditer, en plus des balises R Markdown, des balises spécifiques PyWII gérant la mise en forme du document final. PyWII comble ainsi l’ensemble des fonctionnalités non couvertes par les compilateur Markdown disponibles. Il s’agit, pour l’essentiel, de la structuration en blocs adaptés à la production de tutoriels et de la sélection a posteriori de contenu : exemple, exercice, corrigé, remarque, prolongement, etc.

```

part4-proba.Rmd
1 @document(title = 'calculs de probabilités', \
2 version = 1, \
3 chapterNb = 4)
4 @body()
5
6 @objectifs('chapo')
7 1. se familiariser avec les fonctions R permettant de réaliser des
8 calculs de probabilité.
9 2. se familiariser avec les fonctions R permettant de simuler des
10 échantillons de valeurs associées à une loi donnée.
11 3. réaliser un graphique quantile-quantile pour "vérifier" la normalité
12 d'une distribution.
13 @end objectifs
14
15 ## Autour des lois continues
16
17 ### un exemple avec les lois normales
18
19 La fonction pnorm(q, mean, sd) renvoie la probabilité
20  $P(X \leq q)$ , c'est-à-dire l'image de  $q$  par
21 la fonction de répartition de la loi normale dont l'argument mean
22 est associé à l'espérance et sd à l'écart-type.
23
24 @exemple('exemple_pnorm')
25 Considérons une variable  $X$  distribuée suivant la loi normale
26 d'espérance  $\mu=100$  et d'écart-type  $\sigma=5$ .
27 Les probabilités  $P(X \leq 105)$  et
28  $P(X \leq 110)$  se calculent
29 respectivement par :
30
31 rnorm(q = 105, mean = 100, sd = 5)
32 1 - pnorm(q = 110, mean = 100, sd = 5)
33
34
35 @end
36
37 La fonction qnorm(p, mean, sd) calcule le quantile d'ordre p.
38
39
40 @exemple('exemple_qnorm')
41 La valeur  $x$  telle que  $P(X \leq x) = 0,975$ 
42 est donnée par :
43
44 qnorm(p = 0.975, mean = 100, sd = 5)
45
46
47 @end exemple
48
49
50

```

(a)

Exercice 8 : [\[1\]](#)

Prédire la circonférence du bras pour les valeurs du poids total suivantes : 65, 60, 67, 72 kg.

[\[Aide\]](#)

[\[1\]](#)

```

x_new_Sante <- data.frame(POIDS = c(65, 60, 67, 72))
x_pred_Sante <- predict(reglin_Sante, newdata = x_new_Sante)
x_pred_Sante

```

```

##      1      2      3      4
## 29.50731 28.29702 29.99142 31.20170

```

Ainsi, notre modèle prédit que la circonférence du bras est de 29.5, 28.3, 30, 31.2 cm pour un poids total de 65, 60, 67, 72 kg respectivement.

5.B. INTERVALLE DE CONFIANCE

Si on veut, de plus, avoir un intervalle de confiance pour la prédiction, on complète la commande avec l'option `interval` :

Exemple :

```

x_pred <- predict(reglin, newdata = x_new, interval = "pred")
x_pred

```

```

##      fit      lwr      upr
## 1 1592.449 304.4516 2880.447

```

Nous obtenons l'intervalle de confiance à 95% du prix d'un appartement pour une surface de 65m² :

[304.5, 2880.4]

Remarque :

Par défaut on obtient l'intervalle de confiance à 95%. Si on veut modifier ce niveau, on renseigne le paramètre `level`.

Ce qu'il faut retenir...

Dans cette partie, nous avons vu comment utiliser la fonction `lm()` pour réaliser une régression linéaire simple, ainsi que quelques outils utiles pour la représentation graphique des résultats. Pour finir, voici deux remarques pour vous aider à mettre en place des modifications plus complexes :

- Pour la régression linéaire multiple, on utilise encore la fonction `lm()`. Il faudra juste aller plus loin dans l'utilisation de la formule indiquant le modèle (par exemple si y est expliqué par x et z et qu'on ne veut pas tester l'interaction entre x et z , il faudra utiliser `- x + z`). Consulter l'aide de `formula()` pour plus de détails.
- Pour le modèle linéaire généralisé, on utilise la fonction `glm()`. Il faudra alors consulter l'aide de `family()` (par exemple pour la régression logistique, `family = binomial(link = "logit")` est nécessaire).

(b)

FIGURE 2 – Exemple de page produite avec PyWII. (a) Code mêlant PyWII et R-Markdown. (b) Page HTML produite.

3 Réalisations

Mise en ligne et politique de publication du site – L'ensemble des supports produits a été utilisé dès 2016 dans le cadre d'enseignements conduits dans différents établissements. Il est disponible en libre accès depuis le 13/03/2017 sur le site de l'Université de Bordeaux : <http://beginr.u-bordeaux.fr>.

Chapitres produits – A l'heure actuelle, onze chapitres sont disponibles sur le site. Ces modules pédagogiques concernent la prise en main de R, les statistiques descriptives et inférentielles, l'analyse de variance et les plans d'expériences, la régression linéaire simple et multiple, l'algorithmique et la programmation ainsi que la création de graphiques 3D.

Le matériel pédagogique en appui – De nombreux supports (jeux de données, scripts, mémentos, vidéos, ...) viennent compléter ces ressources pédagogiques. De plus, au delà de ces ressources, des instruments de mesure et d'évaluation des connaissances sont proposés tels que des QCMs sur Moodle. Dans nos pratiques, ces outils revêtent un caractère à la fois formatifs et sommatifs. Ils sont développés en parallèle par des membres

du groupe et seront bientôt disponibles sur le site internet du projet.

4 Conclusion et perspectives

Le projet *Begin'R* propose un site internet (<http://beginr.u-bordeaux.fr>) regroupant des supports médiatisés pour l'apprentissage à distance et en autonomie du logiciel de statistique R et de l'environnement de développement intégré R Studio. Ces supports sont utilisés depuis deux ans dans différents écoles d'ingénieurs et formations de master du site bordelais : Université de Bordeaux, Bordeaux INP et Bordeaux Sciences Agro.

De nouveaux modules pédagogiques ainsi que des QCMs évaluant les objectifs pédagogiques de chaque module sont en cours de réalisation et seront mis en ligne prochainement.

Au delà du projet Begin'R, le groupe de travail a l'ambition de développer une communauté de pratique, au sens de Wenger [9], autour de l'enseignement de la statistique. Celle-ci permettra notamment de proposer des formations concernant la pratique du logiciel R et/ou de la statistique en s'appuyant sur les ressources produites, en direction des enseignants du supérieur, mais aussi du secondaire, au niveau régional voire national.

Remerciements

Les auteurs tiennent à remercier l'IdEx Bordeaux (Initiative d'Excellence) pour son support financier et la cellule MAPI de l'Université de Bordeaux pour son appui technique.

Références

- [1] R Core Team, *R : A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2014.
- [2] RStudio Team, *RStudio : Integrated Development Environment for R*, RStudio, Inc., Boston, MA, 2016.
- [3] John Fox and Milan Bouchet-Valat, *Rcmdr : R Commander*, 2017, R package version 2.3-2.
- [4] E. Paradis, "R pour les débutants, https://cran.r-project.org/doc/contrib/Paradis-rdebuts_fr.pdf," 2005.
- [5] V. Goulet, "Introduction à la programmation en R, https://cran.r-project.org/doc/contrib/Goulet_introduction_programmation_R.pdf," 2016.
- [6] B. Falissard and C. Lalanne, "Introduction à la statistique avec R, <https://www.fun-mooc.fr/courses/course-v1:UPSUD+42001+session07/about>," 2017.
- [7] F. Husson, J. Pagès, and M. Houée-Bigot, "Analyse des données multidimensionnelles, <https://www.fun-mooc.fr/courses/agrocampusouest/40001S03/session03/about>," 2017.
- [8] Yihui Xie, *Dynamic Documents with R and Knitr, Second Edition*, Chapman & Hall/CRC, 2nd edition, 2015.
- [9] E. Wenger, *Communities of practice learning, meaning, and identity*, Cambridge University Press, 1998.