

Frédéric Bertrand • Myriam Maumy-Bertrand

Initiation à la statistique avec R

4^e ÉDITION

DUNOD

Des mêmes auteurs :

Mathématiques pour les sciences de l'ingénieur, Tout le cours en fiches – 3^e édition, Dunod, 2022
PASS Biostatistiques, Mathématiques, Probabilités – 4^e édition, Dunod, 2021
Mathématiques Licence 1 – Exercices et méthodes, Dunod, 2020
Modélisation statistique par la pratique avec R, Cours et exercices corrigés, Dunod, 2019
Mathématiques Licence 2 – Exercices et méthodes, Dunod, 2018
Mathématiques : concours des catégories A et B, Dunod, 2012
Maxi fiches de Statistique. En 80 fiches, Dunod, 2011
Probabilités, Statistiques et Économétrie : concours des catégories A et B, Dunod, 2011
Mathématiques L1/L2 : Statistique et Probabilités en 30 fiches, Dunod, 2009
Mathématiques L1/L2 : Analyse en 30 fiches, Dunod, 2009
Mathématiques L1/L2 : Algèbre/Géométrie en 30 fiches, Dunod, 2009

Illustration de couverture : © Diana913 – Shutterstock.com

© Dunod, 2010, 2014, 2018, 2023

11 rue Paul Bert, Malakoff
www.dunod.com

ISBN 978-2-10-085589-6

Le Code de la propriété intellectuelle n'autorisant, aux termes de l'article L. 122-5, 2° et 3° a), d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective » et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite » (art. L. 122-4).

Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles L. 335-2 et suivants du Code de la propriété intellectuelle.

Table des matières

COMMENT UTILISER CE LIVRE ?	VIII
AVANT-PROPOS	IX
CONVENTIONS TYPOGRAPHIQUES	X
CHAPITRE 1 • INITIATION AU LANGAGE R	1
1.1 Présentation du langage R	1
1.2 Maintenant c'est à vous	4
1.3 Vos premières lignes de commande	8
1.4 Quelques remarques sur la fenêtre R Console	14
1.5 Pour en savoir plus sur le langage R	14
1.6 R sans les mains	15
EN RÉSUMÉ	16
EXERCICES	16
SOLUTION DES EXERCICES	18
CHAPITRE 2 • OBJETS, FONCTIONS ET DONNÉES AVEC LE LANGAGE R	21
2.1 Le fichier de données iris	21
2.2 Quelques définitions	23
2.3 Quelques manipulations élémentaires	24
2.4 Les objets sous R	26
2.5 Les fonctions de R	48
2.6 Les données sous R	51

EN RÉSUMÉ	72
EXERCICES	74
PROBLÈME	76
SOLUTION DES EXERCICES	79
SOLUTION DU PROBLÈME	89
CHAPITRE 3 • STATISTIQUES DESCRIPTIVES UNIVARIÉES	91
3.1 Introduction	91
3.2 Vocabulaire de la statistique	92
3.3 Les fichiers de données : Mesures et Mesures5	95
3.4 Description d'une série statistique	100
3.5 Représentations graphiques d'une série statistique	107
3.6 Représentations numériques d'une série statistique	141
EN RÉSUMÉ	156
EXERCICES	157
PROBLÈME	161
SOLUTION DES EXERCICES	164
SOLUTION DU PROBLÈME	178
CHAPITRE 4 • LOIS DE PROBABILITÉS USUELLES	189
4.1 Variable aléatoire - Variable aléatoire discrète - Variable aléatoire continue	190
4.2 Variables aléatoires discrètes : cas fini	191
4.3 Variables aléatoires discrètes : cas infini	195
4.4 Variables aléatoires continues	197
4.5 Approximations	205
EN RÉSUMÉ	206
EXERCICES	207
SOLUTION DES EXERCICES	211

CHAPITRE 5 • STATISTIQUES DESCRIPTIVES BIVARIÉES	223
5.1 Présentation des données	223
5.2 Distributions conjointe, marginale et conditionnelle	224
5.3 Représentations graphiques	231
5.4 Indépendance statistique	232
5.5 Liaison entre deux variables	232
5.6 Et si vous avez plus que deux variables	236
EN RÉSUMÉ	248
EXERCICES	249
SOLUTION DES EXERCICES	251
CHAPITRE 6 • ÉCHANTILLONNAGE. ESTIMATION D'UN PARAMÈTRE. ESTIMATION PAR INTERVALLE DE CONFIANCE	256
6.1 Échantillonnage	257
6.2 Estimation d'un paramètre	258
6.3 Estimation par intervalle de confiance	264
EN RÉSUMÉ	279
EXERCICES	279
PROBLÈME	281
SOLUTION DES EXERCICES	284
SOLUTION DU PROBLÈME	288
CHAPITRE 7 • TESTS PARAMÉTRIQUES ET NON PARAMÉTRIQUES	293
7.1 Introduction à la notion de tests	293
7.2 Tests de comparaison avec une norme	298
7.3 Test de comparaison entre deux populations indépendantes	311
7.4 Test de comparaison entre deux populations non indépendantes	321
EN RÉSUMÉ	322
EXERCICES	323

PROBLÈMES	325
SOLUTION DES EXERCICES	330
SOLUTION DES PROBLÈMES	335
CHAPITRE 8 • TESTS DU KHI-DEUX	342
8.1 Test du χ^2 d'indépendance	342
8.2 Test du χ^2 d'adéquation d'une loi à une loi donnée	345
8.3 Test exact de Fisher	346
8.4 Test de Mac-Nemar	349
EN RÉSUMÉ	351
EXERCICES	351
SOLUTION DES EXERCICES	357
CHAPITRE 9 • RÉGRESSION LINÉAIRE SIMPLE	365
9.1 Ajustement linéaire en statistique descriptive	365
9.2 Estimation ponctuelle des paramètres de la droite de régression	369
9.3 Intervalle de confiance pour les paramètres β_1 et β_0	371
9.4 Intervalle de confiance de la valeur moyenne	372
9.5 Intervalle de prévision d'une valeur individuelle	373
9.6 Comparaison des paramètres de la droite à des valeurs de référence	373
EN RÉSUMÉ	375
EXERCICES	376
SOLUTION DES EXERCICES	378
CHAPITRE 10 • ANALYSE DE LA VARIANCE À UN FACTEUR	400
10.1 Introduction	400
10.2 Modélisation statistique	402
10.3 Test de comparaison de plusieurs espérances	405
10.4 Vérification des trois conditions fondamentales et conclusion	410

10.5 Comparaisons multiples	417
10.6 Puissance des tests de l'analyse de la variance à un facteur	420
EN RÉSUMÉ	423
EXERCICES	424
SOLUTION DES EXERCICES	426
INDEX	441

Comment utiliser ce livre ?

Les rubriques



Les points clé à retenir sont regroupés en fin de chapitre



Des remarques pour aller plus loin



Les pièges et difficultés sont signalés



Des conseils méthodologiques sont donnés tout au long de l'apprentissage



Pour aller plus loin et mettre toutes les chances de votre côté, des ressources complémentaires sont disponibles sur le site www.dunod.com.

Connectez-vous à la page de l'ouvrage (grâce aux menus déroulants, ou en saisissant le titre, l'auteur ou l'ISBN dans le champ de recherche de la page d'accueil). Sur la page de l'ouvrage, sous la couverture, cliquez sur le lien « LES + EN LIGNE ».

Pour vous, Anaëlle et David, nos deux modèles !

Avant-propos

Chers lecteurs,

Vous tenez entre vos mains le fruit de nos expériences.

Au cours de nos activités d'enseignement, de recherche ou de consultance, nous avons acquis une connaissance et une maîtrise avancées du langage R. Suite à des demandes de plus en plus fréquentes et nombreuses, nous avons partagé celles-ci depuis plusieurs années avec des publics variés : étudiants en Licence ou en Master de biologie, de mathématique ou de statistique, élèves dans chacune des trois années d'écoles d'ingénieurs, utilisateurs de la statistique lors de stages donnés dans le cadre d'écoles doctorales et de la formation continue de chercheurs, enseignants-chercheurs ou de statisticiens.

Ce livre est articulé autour des notions de base des programmes de Licence et Master. Il possède l'avantage d'offrir simultanément une formation progressive à la statistique et au langage R en réduisant au strict nécessaire les notions théoriques abordées. Ce n'est pas un livre-audio mais presque !

Certains d'entre vous se font peut-être une montagne de la statistique mais il suffit d'être bien chaussé, chaudement vêtu et bien guidé pour atteindre le sommet et pouvoir profiter de la vue.

Les chapitres 1 et 2 vous permettent de vous chauffer : installation et prise en main du langage R. Puis vous vous équipez dans les chapitres 3, 4 et 5 en étudiant les statistiques descriptives pour une ou deux variables et les lois de probabilité usuelles. L'escalade commence à partir du chapitre 6 où vous êtes initiés à la théorie de l'échantillonnage et à une notion trop mal connue : les intervalles de confiance. Le chapitre 7 introduit la théorie des tests d'hypothèse qui est complétée au chapitre 8 par les tests du Khi-deux. Le sommet est atteint avec l'exposé de la régression linéaire simple au chapitre 9 et celui du modèle d'analyse de la variance à un facteur au chapitre 10.

Les données utilisées pour la plupart des exercices ont été recueillies par les auteurs avec l'aide du jardinier. Ce sont donc des données réelles sur lesquelles vous découvrirez la statistique et vous exercerez.

Toutes vos remarques, vos commentaires, vos critiques, et mêmes vos encouragements seront accueillis avec plaisir aux adresses électroniques suivantes :
frederic.bertrand@utt.fr et myriam.maumy@utt.fr

Bonne lecture (ascension) à tous !

Frédéric Bertrand
Myriam Maumy-Bertrand

Conventions typographiques

Cet ouvrage suit un certain nombre de conventions typographiques qui ont pour objectif de faciliter la lecture.

Les noms de fichiers sont imprimés avec une police de type machine à écrire.

Exemple : R, SAS, Linux, ade4, etc.

Les noms de fichiers sont également imprimés avec la même police que celle ci-dessus.

Exemple : test.R, donnees.txt, etc.

Il en va de même pour les fonctions de R qui apparaissent dans le corps du texte.

Exemple : summary, mean, <-, etc.

Nous présentons des commandes saisies sous R et leurs résultats de la manière suivante. La commande saisie est indiquée, dans un encadré, avec une police à chasse fixe précédée de l'invite de commande >.

Exemple :

```
> summary(donnees)
```

Lorsque la commande R doit être répartie sur plusieurs lignes, les lignes suivantes sont précédées du symbole + :

Exemple :

```
> filled.contour(volcano, color.palette = terrain.colors, asp = 1,  
+ main = "jeu de données volcano : courbes de niveaux")
```

Le résultat de la commande tel qu'il est affiché par R est indiqué avec une police à chasse fixe dans un cadre avec une trame en gris clair.

Exemple :

```
Min. 1st Qu. Median Mean 3rd Qu. Max.
```

Version utilisée : nous avons utilisé dans ce livre le langage R version 4.2.1 (2022-06-23).

Chapitre 1

Initiation au langage R

« Comme dans toute initiation, c'est dans le fait même de sourire qu'est le triomphe. »

De Paul Auster, *La chambre dérobée*, 1994

Objectifs de ce chapitre

- présenter R
- installer R
- prise en main des premières lignes de commande incontournables



Les informations sur R sont disponibles sur le site internet¹ dédié au projet :
<https://www.r-project.org/>

1.1 PRÉSENTATION DU LANGAGE R

1.1.1 Qu'est-ce-que le langage R ?

- Le langage R est un langage de programmation et un environnement mathématique utilisés pour le traitement de données. Il permet de faire des analyses statistiques aussi bien simples que complexes comme des modèles linéaires ou non-linéaires, des tests d'hypothèse, de la modélisation de séries chronologiques, de la classification, etc. Il dispose également de nombreuses fonctions graphiques très utiles et de qualité professionnelle.

1. En juillet 2018, il s'agit de l'adresse officielle qui peut depuis avoir été éventuellement modifiée.

- R un langage de programmation complet. Cette caractéristique rend par conséquent le langage R différent des autres logiciels de statistique.
- R est disponible pour Microsoft Windows, Macintosh et de nombreux systèmes de type Unix.
- Actuellement, quatre à cinq nouvelles versions de R apparaissent par an.
- R est distribué gratuitement sous les termes de la « GNU », *General Public Licence Version 2*, Juin 1991.
- R est écrit en C, C++, FORTRAN et Java. De plus, R est plus orienté programmation objet que la plupart des autres logiciels ou langage de programmation statistique.
- R est un clone gratuit du langage S-Plus, actuellement commercialisé par Tibco Software Inc., créé autour du langage S qui a été développé par John Chambers des laboratoires Bell.
- R a été créé en 1993 par Ross Ihaka et Robert Gentleman à l'Université d'Auckland, Nouvelle Zélande, et est maintenant développé par la R Development Core Team. L'origine du nom du langage provient, d'une part, des initiales des prénoms des deux auteurs (Ross Ihaka et Robert Gentleman) et, d'autre part, d'un jeu de mots sur le nom du langage S auquel il est apparenté.

1.1.2 Comment se procurer le langage R ?

- L'adresse <https://www.r-project.org/> est le premier résultat pour la recherche de la lettre R avec le moteur de recherche Google™ et la meilleure source d'informations sur le langage R. Vous y trouverez les différentes distributions du langage, de nombreuses bibliothèques de fonctions et des documents d'aide.
- Le langage R est gratuit et se télécharge directement depuis l'internet. Il évolue très rapidement et à peu près tous les six mois une nouvelle version du langage est proposée au public. Elle est accessible via la page officielle consacrée au projet <https://www.r-project.org/>.
- Pour faire face au très grand nombre de téléchargements du langage, un système de miroirs, le Comprehensive R Archive Network <https://cran.r-project.org/>, a été mis en place. Les mêmes fichiers sont ainsi disponibles simultanément sur différents serveurs situés à plusieurs endroits dans le monde². Cette organisation présente au moins deux avantages majeurs pour vous : pouvoir choisir un miroir proche de chez vous où que vous soyez dans le monde ou un miroir de secours lorsque le miroir que vous avez l'habitude d'utiliser est indisponible.

2. La liste des serveurs est disponible à l'adresse <https://cran.r-project.org/mirrors.html> et leur état à l'adresse https://cran.r-project.org/mirmon_report.html.



En août 2022, il existait quatre miroirs en France dont les adresses sont les suivantes :

1. Laboratoire de Biométrie et Biologie Évolutive, CNRS et Université Lyon I
<https://pbil.univ-lyon1.fr/CRAN/>
2. Institut de Biologie et Chimie des Protéines, CNRS et Université Lyon I
<https://mirror.ibcp.fr/pub/CRAN/>
3. Institut de Biologie du Développement de Marseille, CNRS et Université Aix-Marseille
<https://cran.bi.tools.fr/>
4. Institut de Radioprotection et de Sécurité Nucléaire, Paris
<https://cran.irsn.fr/>

Ces miroirs proposent un accès sécurisé avec le protocole https.

1.1.3 Remarques sur le langage R avant l'installation

- R fonctionne avec une grande variété de systèmes d'exploitation et en particulier avec Microsoft Windows, Mac OS X et de nombreux systèmes de type Unix. R est soit disponible sous la forme de fichiers prêts à être installés soit sous la forme de fichiers source à compiler soi-même.
- Il existe une version française du langage R même si le site officiel est rédigé en langue anglaise.
- La très grande majorité des fonctions du langage ne diffère pas d'un système d'exploitation à l'autre bien que les interfaces graphiques ne sont pas similaires.

Il existe plusieurs interfaces graphiques, en anglais « GUI » pour *Graphical User Interface*, qui permettent d'accéder à une partie des fonctions du langage R.

- RGUI, l'interface graphique installée par défaut sous Windows.
- JGR, une interface graphique programmée en Java pour R. Elle fonctionne pour tous les systèmes d'exploitation sur lesquels le langage Java est disponible et donc aussi bien pour Microsoft Windows que pour Mac OS X et de nombreux systèmes de type Unix.
- Rattle, une interface graphique pour le data mining utilisant R.
- R Commander, une interface graphique pour faire des statistiques usuelles avec R.
- Statistical Lab.
- RExcel, pour exécuter les fonctionnalités de R et de R Commander à partir de Microsoft Excel.
- rggobi, une interface pour le logiciel GGobi spécialisé dans la visualisation de données multidimensionnelles.
- RKWard, une interface graphique basée sur les bibliothèques de KDE, une interface graphique disponible sur de nombreux systèmes de type Unix.

Il existe également des programmes qui facilitent l'écriture des instructions et des programmes en langage R.

Ces programmes se regroupent en deux grandes catégories :

- Les éditeurs de texte comme le bloc-note de Microsoft Windows, Microsoft Wordpad ou Microsoft Word.
- Les environnements de programmation, appelés en anglais *Integrated development environments* (IDEs), comme Tinn-R, Emacs (*Emacs Speaks Statistics*), Jedit, Kate, WinEdt (R Package RWinEdt) ou Vim.

Pour les utilisateurs qui en auraient besoin, par exemple pour rédiger un mémoire ou un rapport de stage, il est également possible d'intégrer directement des résultats obtenus avec R dans des documents L^AT_EX à l'aide du langage Sweave ou dans des documents au format OpenDocument (ODF) à l'aide du langage *odfWeave*.

Il existe plusieurs sites internet, principalement en langue anglaise, consacrés au langage R.

- La section sur R à l'Open Directory Project.
- RSeek et R site search qui sont des moteurs de recherche spécialisés dans la recherche de documents consacrés à R.
- Plusieurs listes de diffusion permettent aux utilisateurs novices ou expérimentés de R de poser directement leurs questions aux autres utilisateurs afin qu'ils leur fassent partager leur expérience du langage.
- Le R Journal est un journal à comité de lecture comportant des articles consacrés aux problèmes de calcul statistique et au développement du langage R. Il peut aussi bien intéresser de simples utilisateurs du langage que des programmeurs.
- R Books comprend une liste importante de livres consacrés à R.
- Le R Graphical Manual explique par l'exemple comment construire des graphiques à partir de n'importe quelle bibliothèque du langage. Il comporte également un index exhaustif des fonctions de toutes les bibliothèques d'extensions du langage R.
- Le R Wiki est un site coopératif de documentation sur le langage R.

1.2 MAINTENANT C'EST À VOUS

1.2.1 Installer le langage R

Ces premières instructions sont communes aux trois systèmes d'exploitation.

1. Rendez-vous sur le site <https://www.r-project.org/>.
2. Puis, à gauche sur la page d'accueil, vous trouverez un menu Download. Dans ce menu, cliquez sur CRAN.
3. Choisissez un site miroir proche de chez vous.

4. Un encadré blanc intitulé `Download and Install R` doit apparaître sur votre écran.



1.2 Installation de la version 4.2.1 (2022-06-23) de R sous Windows

1. Cliquez sur Download R for Windows puis sur `base`.
2. Un encadré grisé doit apparaître dans lequel, à la première ligne, est inscrit `Download R-4.2.1 for Windows (79 megabytes, 64 bits)`, la version 4.2.1 étant celle disponible au mois d'août 2022. Cliquez dessus.
3. Procédez au téléchargement.
4. Exécutez le fichier que vous venez de télécharger en choisissant une installation par défaut.

Remarque : le téléchargement de R n'est pas très long. En effet, sa taille est de 79,0 Mo.



1.3 Installation de la version 4.2.1 (2022-06-23) de R sous Mac OS X

1. Cliquez sur Download R for macOS.
2. Une liste de fichiers à télécharger, intitulée `Files` : apparaît. Le premier élément de la liste est `R-4.2.1.pkg`, la version 4.2.1 étant celle disponible au mois d'août 2022. Cliquez dessus. Attention, il vous faudra peut-être, en fonction de votre navigateur internet, appuyer sur la touche « control » en cliquant pour pouvoir télécharger le fichier.
3. Pour installer R, double-cliquez sur l'icône du package d'installation `R-4.2.1.pkg`.



1.4 Installation de la version 4.2.1 (2022-06-23) de R sous Linux

1. Cliquez sur Download R for Linux (Debian, Fedora/Redhat, Ubuntu) puis sur le nom de la distribution Linux installée sur votre ordinateur.
2. Suivez les instructions détaillées sur le site. Celles-ci varient trop d'une distribution à l'autre pour être reproduites ici mais si vous utilisez Linux vous ne devriez pas avoir de mal à installer ou à faire installer le langage par votre administrateur réseau.

Remarque : dans la grande majorité des situations, l'installation de R est très simple et ne nécessite que peu de connaissances techniques.

1.2.2 Le fonctionnement du langage R sous les trois principaux systèmes d'exploitation

L'interface graphique du langage est très similaire d'un système d'exploitation à l'autre. Voici, en détails, la procédure à suivre dans le cas de l'environnement Windows.



1.5 Environnement Windows

- R fonctionne avec plusieurs fenêtres sous Windows. La fenêtre R Console est la fenêtre principale où sont réalisées par défaut les entrées de commandes et les sorties de résultats en mode texte. À celle-ci peuvent s'ajouter des fenêtres facultatives, telles que les fenêtres graphiques, les fenêtres d'informations (historique des commandes, aide, visualisation de fichier, etc), toutes appelées par des commandes spécifiques via la fenêtre R Console.
- Le menu File ou Fichier contient les outils nécessaires à la gestion de l'espace de travail, tels que la sélection du répertoire par défaut, le chargement de fichiers sources externes, la sauvegarde et le chargement d'historiques des commandes exécutées, etc.
- Le menu Edit ou Edition contient les habituelles commandes de copier-coller, ainsi que la boîte de dialogue autorisant la personnalisation de l'apparence de l'interface.
- Le menu View ou Voir permet d'afficher ou de masquer la barre d'outils et la barre de statut.
- Le menu Misc traite de la gestion des objets en mémoire et permet d'arrêter un calcul ou des calculs en cours de traitement.
- Le menu Packages automatise la gestion et le suivi des bibliothèques de fonctions, permettant leur installation et leur mise à jour de manière transparente depuis l'un des miroirs du CRAN (Comprehensive R Archive Network) <https://cran.r-project.org/>.
- Enfin, le menu Windows ou Fenêtres et le menu Help ou Aide assument des fonctions similaires à celles qu'ils occupent dans les autres applications Windows, à savoir la définition spatiale des fenêtres et l'accès à l'aide en ligne et aux manuels de références du langage R.

1.2.3 Installer des packages du langage R

Qu'est ce qu'un package ? Un package est une compilation d'outils. Certains sont déjà présents dans l'installation de base de R. En effet, lors de l'installation de R, un dossier `library` s'est créé par défaut. Il comprend les packages de base de R. Mais d'autres packages qui vous seront utiles pour réaliser vos analyses statistiques seront à télécharger puis à installer.



1.6 Pour les trois environnements

1. Reprenez la procédure de téléchargement de R vue à la section « Installer le langage R ».
2. Cette fois-ci au lieu de cliquer sur Windows, MacOS X ou Linux, cliquez sur `packages` dans la liste intitulée `Source Code for all Platforms`.
3. Une page apparaît sur laquelle est indiqué le nombre de packages actuellement disponibles sur le CRAN (18 489 fin août 2022). Pour obtenir la liste des packages rangés dans l'ordre alphabétique, cliquez sur Table of available packages, sorted by name.
4. La liste des packages apparaît alors.
5. Puis cliquez sur le package dont vous avez besoin.
6. Une brève description du package apparaît suivie d'une liste proposant plusieurs versions du package. La première (`.tar.gz`) est celle contenant le code source du package et ne sert a priori qu'aux utilisateurs de Linux. La deuxième (`.tgz`) est destinée aux utilisateurs de Mac OS X et la troisième (`.zip`) aux utilisateurs de Windows.
7. Sous Mac OS X et Windows, il faut alors démarrer l'interface graphique de R.
8. Sous Mac OS X, allez dans le menu `Packages` et utilisez le `Package Manager`. Indiquez alors à R le fichier `.tgz` que vous venez de télécharger.
9. Sous Windows, allez dans le menu `Packages` et choisissez `Installer depuis un fichier .zip`. Indiquez alors à R le fichier `.zip` que vous venez de télécharger.

Remarque : vous verrez bientôt comment installer très facilement les packages disponibles sur le site du CRAN directement depuis R.



1.7 Environnement Windows

Il existe une procédure alternative pour les utilisateurs d'un environnement Windows.

1. Reprenez la procédure de téléchargement de R vue à la section « Installer le langage R ».
2. Cette fois-ci au lieu de cliquer sur `base`, cliquez sur `contrib`.
3. Cliquez ensuite sur le dossier de la version R que vous avez installée.
4. Puis cliquez sur le package dont vous avez besoin.
5. Un fichier `.zip` est enregistré sur votre disque dur.
6. Démarrez l'interface graphique de R, allez dans le menu `Packages` et choisissez `Installer depuis un fichier .zip`. Indiquez alors à R le fichier `.zip` que vous venez de télécharger.



Lorsque vous aurez besoin de `packages` qui ne sont pas installés par défaut pour réaliser les analyses statistiques qui vous seront demandées dans les exercices, cela vous sera signalé. Il faudra alors installer ces `packages` supplémentaires sur votre ordinateur.



Beaucoup de bibliothèques contiennent un ou plusieurs documents détaillant leurs fonctionnalités et montrant leur application pas à pas à un exemple. Pour obtenir la liste des vignettes présentes sur votre ordinateur, il suffit d'exécuter la commande `vignette()`.

1.3 VOS PREMIÈRES LIGNES DE COMMANDE

1.3.1 Démarrer R

Pour démarrer R, vous pouvez par exemple lancer le langage R en double-cliquant sur l'icône R qui se trouve par exemple sur votre bureau. La fenêtre R console s'ouvre. Elle vous affiche tout ce texte :

```
R version 4.2.1 (2022-06-23) - "Funny-Looking Kid"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform : x86_64-apple-darwin17.0 (64-bit)

R est un logiciel libre livré sans AUCUNE GARANTIE.
Vous pouvez le redistribuer sous certaines conditions.
Tapez 'license()' ou 'licence()' pour plus de détails.


R est un projet collaboratif avec de nombreux contributeurs.
Tapez 'contributors()' pour plus d'information et
'citation()' pour la façon de le citer dans les publications.
Tapez 'demo()' pour des démonstrations, 'help()' pour l'aide
en ligne ou 'help.start()' pour obtenir l'aide au format HTML.
Tapez 'q()' pour quitter R.

[R.app GUI 1.79 (8095) x86_64-apple-darwin17.0]
```

Puis sous ce texte, le symbole rouge `>`, appelé prompt, apparaît également. Il signifie que R est prêt à travailler.



Il ne faut jamais taper ce symbole au clavier car il est déjà présent en début de ligne sur la fenêtre R Console.

C'est à la suite de `>` que vous taperez les lignes de commande de R. Une fois la commande tapée, vous devez toujours la valider en appuyant sur la touche « Entrée » du clavier.



Si votre commande est incomplète, le symbole `>` est remplacé par `+`. Ce symbole `+` signifie que R attend la suite de la commande. Si vous ne savez pas compléter la ligne de commande ou qu'elle présente une erreur vous pouvez appuyer sur la touche « Echap » pour annuler la commande et créer un nouveau symbole `>`.

1.3.2 Quitter R

Pour quitter R, vous utilisez la commande suivante :

```
> q()
```

```
Save workspace image? [y/n/c]
```

R vous propose de sauvegarder le travail effectué. Trois réponses vous sont proposées : y (pour *yes*), n (pour *no*) ou c (pour *cancel*, annuler).

Si vous tapez y, cela permet que les commandes exécutées pendant la session et les objets enregistrés en mémoire soient conservés et soient donc « rappelables » et « réutilisables ».

Si vous tapez n, vous quittez R qui oubliera tout le travail que vous avez réalisé. Attention, vous risquez de tout perdre !

Si vous tapez c, la procédure de fin de session sous R est annulée.

Sous l'environnement Windows ou Mac OS X, c'est une boîte de dialogue en français qui apparaîtra à l'écran et qui comportera les mêmes questions.

1.3.3 Sauvegarder sous R

Si vous quittez R en choisissant la sauvegarde de l'espace de travail, deux fichiers sont créés :

1. le fichier `.Rdata` ou `.Rapp.history` contient des informations sur les variables utilisées,
2. le fichier `.Rhistory` contient l'ensemble des commandes utilisées.

1.3.4 Consulter l'aide de R

Il y a quatre sources principales d'aide :

1. les fichiers d'aide,
2. les manuels,
3. les archives R-help,
4. et enfin R-help lui-même.

Pour une fonction, dont le nom est `fonction`, vous pouvez consulter une fiche de documentation en tapant `?fonction` ou `help("fonction")`. Grâce à cette aide, il suffit que vous reteniez le nom de la fonction, mais pas forcément toute la syntaxe.

Exemple : vous cherchez à obtenir des informations sur la fonction `read.table`, vous taperez alors la commande suivante :

```
> ?read.table
```

ou encore :

```
> help(read.table)
```

Pour une bibliothèque d'extension, dont le nom est `package`, vous pouvez consulter une fiche de documentation en tapant :

```
> help(package="package")
```

Grâce à cette aide, il suffit que vous reteniez le nom de la bibliothèque où se trouve la fonction que vous souhaitez utiliser, mais pas forcément le nom exact de cette fonction.

Les pages d'aide sont généralement très détaillées. Elles contiennent souvent, entre autres :

- une section *See Also* qui donne les pages d'aide sur des sujets apparentés.
- une section *Description* qui précise ce que fait la fonction.
- une section *Examples* avec des lignes de commande illustrant ce que fait la fonction documentée. Ces exemples peuvent être exécutés directement en utilisant la fonction `example`, essayez par exemple :

```
> example(plot)
```

Pour afficher successivement les différents graphiques ainsi créés, vous devez cliquer plusieurs fois de suite sur la fenêtre où sont situés les graphiques.

a) Affichage de l'aide dans la console

Lorsque l'aide s'affiche dans la console, vous pouvez faire défiler le texte ligne par ligne avec la touche « Entrée » ou « Flèche vers la bas » ou page par page en appuyant sur la barre « Espace ». Une fois arrivé à *END*, tapez `q`.

Il s'agit du mode d'affichage par défaut de l'aide dans un terminal R et donc dans un environnement Linux.

b) Affichage de l'aide en-dehors de la console

Lorsque l'aide s'affiche dans la console, elle n'est pas facile à consulter. Il existe des versions au format `.html` de tous les fichiers d'aide dans les trois environnements Linux, MacOS X et Windows. Pour s'en servir à la place des versions texte qui s'affichent dans la console, il faut utiliser l'option `help_type="html"` de la fonction `help`.

Exemple : vous cherchez à obtenir des informations au format `.html` sur la fonction `read.table`, vous tapez alors la commande suivante :

```
> help("read.table", help_type="html")
```

L'option `help_type="text"` de la fonction `help` permet d'afficher l'aide au format texte `.txt`.

Exemple : vous cherchez à obtenir des informations au format `.txt` sur la fonction `read.table`, vous taperez alors la commande suivante :

```
> help("read.table", help_type="text")
```

Le format `.html` est celui utilisé par défaut lorsque vous utilisez l'interface graphique de R dans un environnement Mac OS X ou Windows.

La fonction `help.start` permet d'accéder à la page d'accueil de l'aide au format `.html`.

```
> help.start()
```

c) Changer le mode d'affichage par défaut de l'aide

Il est possible de modifier une des options de R qui gère le mode d'affichage par défaut des fichiers d'aide de R. Pour afficher, pour toute la durée d'une session R, les fichiers au format :

- `.html`, tapez l'instruction `options(help_type="html")` dans la console.
- `.txt`, tapez l'instruction `options(help_type="text")` dans la console.



À la fin de l'aide, il y a presque toujours une ou plusieurs lignes de commande d'exemple. Il est judicieux de les exécuter afin d'avoir une idée de ce que les fonctions ont besoin comme informations et des résultats qu'elles vous renvoient. Pour cela, il suffit de copier ces lignes, de les coller sur la fenêtre R console et d'observer ce qui se passe. Vous pouvez, bien sûr, les modifier suivant vos besoins. De plus, ces lignes de commande sont souvent un exemple d'analyse de données et pointent parfois sur d'autres fonctions utiles à essayer. La notion de « fonctions » sous R sera développée dans le chapitre suivant.

1.3.5 Travailler avec R

Cette section va être courte car elle sera limitée, dans le cadre de ce sous-paragraphe, à quelques fonctionnalités de R.

a) R comme calculatrice

Exemple : tapez la ligne de commande suivante et validez par la touche « Entrée » du clavier :

```
> 2 + 8
```

Le résultat s'affiche sous la forme :

```
[1] 10
```

Le chiffre 1 entre crochets indique l'indice du premier élément de la ligne. Le second chiffre, 10, est le résultat de la ligne de commande.

Exemple : voici un autre exemple. Tapez la ligne de commande suivante :

```
> 120:155
```

```
[1] 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135  
[17] 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151  
[33] 152 153 154 155
```

Dans le résultat ci-dessus, l'indice de l'élément 120 est 1, celui de 136 est 17 et celui de 152 est 33.



Le chiffre qui se trouve entre les crochets indique l'indice du premier élément de la ligne sur laquelle il se trouve.

Exemple : une fonction dont vous pourriez avoir besoin est la fonction `sqrt`, qui n'est rien d'autre que la racine carrée d'un nombre. Tapez la ligne de commande suivante :

```
> sqrt(4)
```

```
[1] 2
```

b) R comme logiciel de statistique

Vous obtiendrez des précisions sur cette fonctionnalité tout au long des chapitres suivants.

c) R comme éditeur graphique

Vous obtiendrez des précisions sur cette fonctionnalité tout au long des chapitres suivants.

1.3.6 Écrire des scripts avec un éditeur de texte et les compiler sous R

Il est souvent plus pratique de composer le code R dans une fenêtre spécifique de l'interface graphique : la fenêtre de script.

Les entrées `Nouveau script` ou `Ouvrir un script` permettent de créer un nouveau script de commandes R ou d'accéder à un ancien script sauvegardé lors d'une session précédente d'utilisation du logiciel.