

Les éléments à télécharger sont disponibles à l'adresse suivante :
<http://www.editions-eni.fr>
Saisissez la référence ENI de l'ouvrage **RIRANADO** dans la zone de recherche et validez. Cliquez sur le titre du livre puis sur le bouton de téléchargement.

Avant-propos

- 1. Introduction 9
- 2. À qui s'adresse cet ouvrage ? 10
- 3. Comment utiliser cet ouvrage ? 11

Chapitre 1
Le langage R

- 1. Histoire du langage R 15
- 2. Installation de R et de RStudio 17
 - 2.1 Comprehensive R Archive Network (CRAN) 17
 - 2.2 Installation de R via le CRAN 17
 - 2.3 Installation de RStudio 27
 - 2.4 Paramétrage de RStudio 32
 - 2.5 Trucs et astuces avec RStudio 34
- 3. Packages 36
 - 3.1 Qu'est-ce qu'un package ? 36
 - 3.2 Installation d'un package 37
 - 3.3 Utilisation d'un package, d'une fonction 39
 - 3.4 Trouver de l'aide à propos d'un package ou d'une fonction . . . 40
- 4. Travail sous forme de projet 42
- 5. Réalisation de rapports avec R Markdown 46

2 ————— Langage R et statistiques

Initiation à l'analyse de données

6.	Initiation au codage en R	50
6.1	Nomenclature du langage R	50
6.2	Importer des données	55
6.3	Manipuler les données	58
6.4	Concaténation de tables	61

Chapitre 2

Les statistiques

1.	Notions clés sur les statistiques	67
1.1	Histoire rapide de la discipline des statistiques	67
1.2	Définitions des mots-clés importants	68
2.	Initiation à la réalisation de graphiques avec {ggplot2} dans R	72
3.	Lois de probabilités statistiques	78
3.1	Introduction aux lois de probabilités	78
3.2	Loi uniforme	82
3.3	Loi binomiale	85
3.4	Loi de Poisson	87
3.5	Loi normale	89
3.6	Loi exponentielle	91
3.7	Résumé des paramètres liés aux lois de probabilités	94
4.	Réalisation de statistiques descriptives dans R	95
4.1	Principe des statistiques descriptives	95
4.2	Statistiques descriptives univariées dans R	95
4.2.1	Type de données	95
4.2.2	Variable qualitative	97
4.2.3	Variable quantitative	104
4.3	Statistiques descriptives bivariées dans R	116
4.3.1	Deux variables qualitatives	117
4.3.2	Une variable qualitative et une variable quantitative	118
4.3.3	Deux variables qualitatives	121
5.	Détection et visualisation des données manquantes dans R	123

Chapitre 3
La régression

- 1. Définitions et limites de la régression 127
- 2. Régression linéaire simple 128
 - 2.1 Objectif de la régression linéaire simple 128
 - 2.2 Réalisation d'une régression linéaire simple dans R 130
 - 2.3 Limites d'application de la régression linéaire simple 132
 - 2.3.1 Visualisation de la corrélation entre les données 132
 - 2.3.2 Indépendances des données 133
 - 2.3.3 Analyse des résidus 133
 - 2.4 Intervalle de confiance de la régression dans R 136
- 3. Généralisation de la régression paramétrique dans R 137
 - 3.1 Régression polynomiale 137
 - 3.2 Régressions exponentielle et logarithmique 143
 - 3.3 Introduction à la régression logistique binomiale 148
 - 3.4 Régression multiple 153
- 4. Régression non paramétrique dans R 163
 - 4.1 Définitions et limites de la régression non paramétrique 163
 - 4.2 Régression non paramétrique univariée 166
 - 4.3 Modèles additifs généralisés 168

Chapitre 4
Les tests d'hypothèses

- 1. Vocabulaire lié aux tests d'hypothèses 173
- 2. Démarche du test d'hypothèses 177
- 3. Test de Shapiro-Wilk 182

4 ————— Langage R et statistiques

Initiation à l'analyse de données

4. Test du Khi-2	186
4.1 Paramétrique ou non paramétrique, le test du Khi-2 ?	186
4.2 Loi du Khi-2 à k degrés de liberté.	187
4.3 Test du Khi-2 d'adéquation	189
4.4 Test du Khi-2 d'indépendance	193
4.5 Test du Khi-2 d'homogénéité.	199

Chapitre 5

La comparaison à une valeur théorique

1. Quand comparer à une valeur théorique ?	205
2. Théorème central limite.	206
2.1 Histoire et utilisation du théorème central limite	206
2.2 Énoncé du théorème central limite	206
3. Comparaison d'une proportion à une valeur théorique	209
3.1 Exemples de comparaison d'une proportion à une valeur théorique	209
3.2 Test de comparaison d'une proportion à une valeur théorique avec R	212
4. Comparaison d'une moyenne à une valeur théorique	218
4.1 Test de Student	218
4.2 Exemple d'utilisation du test de Student dans R	222
4.3 Test de Student apparié	224
4.4 Test de Wilcoxon apparié.	228
5. Comparaison d'une variance à une valeur théorique	231
5.1 Variance d'une variable	231
5.2 Test de comparaison de la variance à une valeur théorique	232
5.3 Exemple de comparaison d'une variance à une valeur dans R	233
6. Test de Kolmogorov-Smirnov	235

Chapitre 6

La comparaison de deux groupes

- 1. Généralités sur la comparaison de deux groupes 245
- 2. Comparaison de proportions de groupes dans R 246
 - 2.1 Utilisation de la comparaison de proportions. 246
 - 2.2 Test z à deux proportions et test du Khi-2. 247
 - 2.3 Test exact de Fisher 251
 - 2.4 Test de McNemar. 254
- 3. Comparaison de variances de deux groupes dans R. 257
 - 3.1 Mise en situation de la comparaison de variances 257
 - 3.2 Test F de Fisher. 259
 - 3.3 Test de Bartlett. 261
 - 3.4 Test de Levene 264
 - 3.5 Test de Fligner-Killeen 267
- 4. Comparaison de moyennes de deux groupes dans R. 270
 - 4.1 Choix du test de comparaison de moyennes à utiliser. 270
 - 4.2 Test t de Student 271
 - 4.3 Test t de Welch. 275
 - 4.4 Test de Wilcoxon-Mann-Whitney 277
- 5. Comparaison de moyennes/médianes de plus de deux groupes ... 279
 - 5.1 Pourquoi ne pas comparer deux à deux ? 279
 - 5.2 ANOVA : analyse de variance dans R. 281
 - 5.3 Test de Kruskal-Wallis dans R. 287

6 ————— Langage R et statistiques

Initiation à l'analyse de données

Chapitre 7

Les modèles linéaires généralisés

1. Modèles linéaires classiques	291
1.1 Rappel	291
1.2 Les tests post-hoc	294
2. Généralisation des modèles linéaires en R	299
2.1 Utilisation des modèles linéaires généralisés.	299
2.2 Régression logistique binomiale	302
2.3 Régression de Poisson.	308
3. Modèles linéaires mixtes	315
3.1 Généralités et utilisation des modèles linéaires mixtes	315
3.2 Réalisation de modèles linéaires mixtes dans R	317
4. Traitement des données manquantes dans les GLM	329
5. Modèles non linéaires à effets fixes ou mixtes	330

Chapitre 8

L'analyse en composantes principales

1. Utilisation de l'ACP	331
2. Bases mathématiques de l'ACP	336
2.1 Approche par les individus.	336
2.2 Approche par les variables	343
3. Mise en place de l'ACP	346
3.1 Déroulement de l'ACP	346
3.2 Sélection des variables actives	347
3.3 Réduction des variables	348
3.4 Réalisation de l'ACP dans R.	349
3.4.1 Lignes de code et sorties automatiques.	349
3.4.2 Représentativité des axes.	352
3.4.3 Graphique des individus.	355
3.4.4 Graphique des variables	356

- 3.5 Sélection du nombre de dimensions à interpréter dans R 358
 - 3.5.1 Significativité des dimensions obtenues
par l'ACP dans R 358
 - 3.5.2 Nombre de dimensions à interpréter. 359
- 3.6 Interprétation des sorties de l'ACP
et indicateurs supplémentaires dans R 361
- 3.7 Ajout de variables supplémentaires dans R 364
- 4. Limites et ouverture des analyses factorielles. 374
 - 4.1 Données manquantes en ACP 374
 - 4.2 Analyse factorielle des correspondances 383
 - 4.3 Pour conclure sur les analyses factorielles. 388

- Index 389

Chapitre 1

Les data, pour quoi faire ?

A. Introduction	11
B. Data mining, big data, smart data, open data : comprendre l'univers des data	12
1. L'arrivée des data dans la gestion des entreprises	12
a. L'exemple de la distribution	13
b. Le développement des entreprises d'études de marché	14
2. Comprendre les termes data	17
a. Le data mining	17
b. Le big data	18
c. Le smart data	22
d. Les open data	22
3. L'intelligence artificielle	23
a. Qu'est-ce que l'IA est et n'est pas ?	24
b. Application de l'IA : le marketing automation	25
c. Application de l'IA : les chatbots	26
d. Application de l'IA : la reconnaissance faciale	26
4. La réalité virtuelle	26
C. Les acteurs du data marketing	27
1. Le chief data officer (CDO)	27
2. Le data scientist/analyst	27
3. Le data protection officer (DPO)	28
4. Le data marketer	28
D. Data et marketing : quelles finalités ?	28
1. Stratégie d'entreprise et data marketing	29
a. L'itération entre stratégie d'entreprise, stratégie marketing et stratégie data	29
b. Pas de stratégie data sans objectifs SMART	30
2. Les data au sein du plan marketing	31
3. Analyser la demande : comprendre ses clients actuels et potentiels	32
a. Les études de marché traditionnelles	32
b. Les échanges spontanés entre consommateurs	32
c. Les mesures comportementales	33
4. Analyser la concurrence	34
5. Mettre en place une stratégie : segmentation et ciblage	35

6. Optimiser ses actions marketing : le bon produit, au bon prix, au bon endroit avec la bonne communication et les bonnes actions commerciales	37
a. Rappels sur le marketing mix	37
b. Le produit	39
c. Le prix	40
d. La distribution	41
e. La communication	44

Chapitre 2

Les outils du data marketing

A. Introduction	49
B. Structurer et analyser ses données	50
1. Excel : usages et fonctionnalités	50
a. Visualiser rapidement la moyenne, le nombre et la somme d'une variable	50
b. Filtrer les données	51
c. Trier les données	52
d. Figurer les volets	53
e. La mise en forme conditionnelle	55
f. Le tableau de contingence : le tableau croisé	55
g. Convertir	57
2. R : Prise en mains	61
a. L'installation de R	61
b. RStudio	64
c. Définir le répertoire de travail	65
d. Télécharger un fichier de données dans R	66
e. Ouvrir un onglet de script	69
f. Installer et charger des packages	70
g. Le package Rcmdr	71
C. Gérer les big data	74
1. La problématique de l'analyse de logs pour comprendre les comportements des consommateurs en ligne	75
2. L'univers Hadoop	76
a. Naissance et enjeux d'Hadoop	76
b. Le principe MapReduce	77
c. Hadoop, un framework puissant, mais demandant un niveau de technicité élevé	78
d. Les briques Hadoop	78

3. Les big data dans R	79
a. Hadoop dans R	79
b. Traitements statistiques de big data dans R	80

Chapitre 3

Auditer son système marketing

A. Collecter des données	85
1. Le type de données et ce que l'on peut en faire	85
a. Les données structurées	85
b. Les données non structurées	87
2. La richesse de vos données	90
a. Les sources d'information en interne	90
b. Type d'informations disponibles en interne	91
3. Les données externes	91
a. Open data	91
b. Données de panel	92
c. Données de tracking	92
B. Stocker ses données	93
1. Faire un état des lieux de ses données	93
2. Les solutions techniques	93
3. Vision à 360° du client : DMP, data lakes	94
a. Les data management platform (DMP)	94
b. Les data lakes	94
c. Est-ce pour mon entreprise ?	95
C. Nettoyer ses bases de données	95
1. Qu'est-ce qu'une base de données propre ?	96
2. Pourquoi « nettoyer » sa base de données ?	96
a. Éviter la perte de temps	96
b. Ne pas renforcer ou susciter de la résistance à une stratégie data-driven	96
c. Faciliter la communication entre les différents outils de data	97
d. Ne pas nuire à la relation client	97
3. Comment procéder (les données manquantes, la clé d'identification, les données anormales) ?	97
a. Définir l'unité statistique d'analyse	98
b. Adopter une logique de table	99
c. Supprimer les variables inutiles	100
d. Donner une clé unique à vos sujets	100

e. Visualiser ses données dans R	100
f. Effectuer quelques tests statistiques basiques pour s'assurer que les variables soient bien définies.	101
g. Supprimer les doublons.	101
h. Nommer les variables de manière efficace	105
i. Coder les variables de manière efficace.	105
j. Traiter les données manquantes	107
k. Traiter les données anormales ou aberrantes (les outliers)	110

Chapitre 4

Que s'est-il passé ? L'analyse descriptive des données et le diagnostic

A. Introduction	115
B. Choisir ses KPI	115
1. Plan : planifier les actions marketing et les objectifs	116
2. Do : la phase d'action.	117
3. Check : la phase d'analyse et de diagnostic.	117
4. Act : ajuster les actions ou poursuivre sur la même voie	117
C. Analyse univariée : étudier les variables une à une	118
1. La tendance centrale	119
a. La moyenne.	119
b. La médiane	120
c. Le mode	122
2. La dispersion.	126
a. La variance	126
b. L'écart-type	128
c. La fréquence	129
d. L'étendue, le minimum et le maximum	131
e. Le fractile (quartile).	132
f. Un outil de visualisation des principaux indices : la boîte à moustaches.	133
3. L'intervalle de confiance	137
D. Analyse bivariée : faire des rapprochements entre deux variables	138
1. Deux variables qualitatives	138
a. Créer un tableau de contingence dans R.	138
b. Test du Khi-deux dans R.	140
2. Deux variables numériques : la corrélation	142
a. La covariance	143
b. Le coefficient de corrélation linéaire	146

c. La significativité du coefficient de corrélation linéaire	148
3. Deux variables ordinales	149
a. Le coefficient de corrélation de Spearman	150
b. La significativité du coefficient de corrélation de Spearman	151
4. Test de comparaison de moyennes	152
a. Préalable statistique à la comparaison de moyenne	153
b. Test de comparaison de moyennes : t de Student	154
E. Conclusion	157

Chapitre 5

Que peut-il se passer dans le futur ? Le marketing prédictif

A. Introduction	161
B. La classification : segmenter ses clients ou ses prospects	161
1. La classification, une des méthodes d'analyse multivariée	162
2. Méthodologie	162
3. Étape 1 : Analyse en composantes principales	163
a. La matrice des coefficients de corrélation	164
b. Réalisation de l'ACP sous R	164
c. Interprétation des dimensions	168
4. Étape 2 : la classification ascendante hiérarchique	169
a. Regroupement des individus par classes	169
b. Réalisation de la Classification Ascendante Hiérarchique (CAH) sous R	169
c. Lecture des résultats de la CAH	173
C. Les modèles de régression	178
1. Méthodologie	178
a. Prévoir le prix de vente d'un bien immobilier : valeur continue et régression multiple	178
b. Prévoir l'achat de glaces : valeur catégorielle et régression logistique	181
2. Applications pratiques	184
D. Le scoring	184
1. Data mining et scoring	185
2. Applications des méthodes de scoring	186
a. Score de risque	186
b. Score d'appétence	186
c. Score d'attrition	187
3. Les différents types d'analyses	187

4. Le scoring par régression logistique	188
a. La régression logistique binaire	188
b. Mise en œuvre du scoring par la régression logistique binaire dans R	189
E. Conclusion	198

Chapitre 6

Que devrions-nous faire ? Le marketing prescriptif

A. Introduction	201
B. À quelles problématiques répond le marketing prescriptif ?	202
C. Mettre en place une expérimentation	204
1. Conditions réelles ou en laboratoire	204
a. L'expérimentation en laboratoire	204
b. L'expérimentation en conditions réelles	205
2. Les éléments à considérer pour mettre en place une expérimentation	205
a. Les traitements à comparer	206
b. Les critères d'évaluation de ces traitements	206
c. Les variables externes	207
d. Les sujets	207
3. La validité de l'étude	208
a. La validité interne et externe	208
4. Les biais à contrôler	208
a. Les effets des paramètres extérieurs à l'expérimentation	208
b. L'effet de sélection	208
c. L'effet d'histoire	209
d. L'effet de maturation	209
e. L'effet de test	209
f. L'effet d'instrumentation	209
g. L'effet de mortalité expérimentale	210
h. L'effet de halo	210
5. Les plans d'expérience	210
a. Le plan post-traitement	211
b. Le plan pré et post-traitement	212
c. Le plan de Solomon	212
d. Le plan factoriel complet	213
e. Plan factoriel à blocs aléatoires	213
f. Plan en carré latin ou en carré gréco-latin	215

D. Analyser les résultats	215
1. Quelques préalables	215
a. Qu'est-ce qu'un résultat significatif ?	215
b. Mes données suivent-elles une loi normale ?	216
c. Les variances sont-elles égales entre les populations ?	219
2. Comparaison de deux moyennes	220
3. Analyse de variance (ANOVA)	222
a. ANOVA à un facteur	222
b. ANOVA à plusieurs facteurs	225
c. MANOVA	229
E. Conclusion	230

Chapitre 7

Comment communiquer ? Règles de base, data visualisation et data storytelling

A. Introduction	233
B. Quelques règles basiques pour bien présenter sa stratégie data-driven	234
1. Communiquer sur ses données : quels éléments-clé prendre en compte ?	234
a. L'émetteur du message	234
b. Le message	235
c. Les interlocuteurs	235
d. Les canaux	235
e. Quels effets ?	236
2. Le traitement des messages : que se passe-t-il chez mon interlocuteur quand je communique ?	236
C. Les outils de la data visualisation	238
1. Quelle est l'idée-clé de mon graphique ?	240
2. Quelles données sont les plus pertinentes pour démontrer mon idée ?	240
3. Quelle représentation graphique est la plus adaptée ?	240
4. Applications dans Excel	241
D. Le data storytelling	244
1. Écrivez votre fil conducteur	244
2. Accrocher votre public	244
3. Plantez votre décor	245
4. Ne faites pas trop de teasing	245
5. Ne vous perdez pas dans les détails	245
6. Ayez une chute efficace	245
7. Inventez un personnage	245

E. Conclusion.....	246
--------------------	-----

Chapitre 8

Conclusion

A. 6 éléments-clé de la réussite de votre stratégie data-driven	249
1. Construire une culture data-driven.....	249
2. Travailler en équipe	249
3. Le pragmatisme et le bon sens orienté client.....	250
4. La plus-value des data	250
5. Commencer simple mais efficace.....	250
6. Avoir une attitude responsable vis-à-vis des données	251

Bibliographie.....	253
--------------------	-----

Index	257
-------------	-----