

stat_helpR

Guillaume Papuga

2023-10-26

Contents

1	Origine du projet	5
1.1	Qui suis je?	5
1.2	Pourquoi / pour qui écrire un livre sur les statistiques?	5
1.3	Pourquoi R	5
1.4	Plagiat & Sources	5
1.5	Report un bug	6
1.6	ToDoList	6
2	Gestion des données	7
2.1	Les données sur R	7
2.2	Les jeux de données utilisés	7
2.3	Les fonctions de base	7
2.4	Créer une fonction	7
2.5	Data manipulation	7
2.6	Ma structure d'analyse (workflow)	7
3	Utilisation de R	9
4	Data mining	11
5	Les lois statistiques	13
6	Test simples	15
7	Le modèle linéaire généralisé	17

8	Le modèle mixte	19
9	Statistiques multivariées	21
10	Times series	23
11	Analyses spatiales	25
12	SDM	27
13	Analyse de viabilité des populations chez les plantes	29
14	Soutient de cours	31
14.1	L2 – écologie générale	31
14.2	L3 – écologie des communautés	31
14.3	Master	31
15	References	33

Chapter 1

Origine du projet

1.1 Qui suis je?

- Botanist, plant ecologist, intérêt pour l'analyse
- MCU
- Sites internet, github, etc.

1.2 Pourquoi / pour qui écrire un livre sur les statistiques?

- Moi - futur moi (pense bete géant). J'accumule des notes ecrites (conf, etc) + scripts
- Etudiants - support de cours
- N'importe qui
- Français

1.3 Pourquoi R

- Libre, puissant, mis à jour
- Github pour le partage

1.4 Plagiat & Sources

- Philo
- Livres

- Blog & internet
- Collaborateurs sur des chapitres

1.5 Report un bug

- Via github
- Mail

1.6 ToDoList

Chapter 2

Gestion des données

2.1 Les données sur R

2.2 Les jeux de données utilisés

- Via R
- Perso en ligne sur le repo
- JDD packages
- Comment simuler un jdd?
 - Lois
 - Random

2.3 Les fonctions de base

2.4 Créer une fonction

2.5 Data manipulation

- Apply
- For
- Suite Tidy

2.6 Ma structure d'analyse (workflow)

Chapter 3

Utilisation de R

- Couleurs (RColorBrewer)
- Fonctionnement (parallèle sur des cœurs différents)
- GitHub

Chapter 4

Data mining

Décrire les données (pas de stat) représentation graphiques • Package de data mining • Matrix of scatter plot • Scatter plot 3D et interactif • GGplot2

Chapter 5

Les lois statistiques

A faire

Chapter 6

Test simples

- Tester la normalité
- Variances
- Paramétriques

Les différentes anova

- Non paramétriques
- Autre, Chi2, etc.
- Tester la corrélation entre variables
 - Pearson
 - Spearman

Inférences • bootstrap

Chapter 7

Le modèle linéaire généralisé

Introduction ◦ Lien (quasi) • Modèle linéaire simple (lm) ◦ Comment écrire un modèle • Introduction au GLM • GLM sur données de comptage • GLM sur loi binomiale ◦ Données présence absence Fit validation ◦ sur proportion • Transformation des variables

Les contrastes

Chapter 8

Le modèle mixte

- Effet mixtes (random y, random slope)
- comment écrire le modèle
- Exemple sur loi normale
- Exemple sur loi poisson
- Exemple sur loi binomiale

Chapter 9

Statistiques multivariées

- Introduction
- Les matrices de distance
- Analyses Descriptive
 - AFC
 - ACPP
 - Nmds
- CAH
- KMeans
- Discriminantes
- Quelques tests
- Couplage de tableaux

Chapter 10

Times series

- Format
- Analyses de survie

Chapter 11

Analyses spatiales

- Représentations spatiales
 - Données en ligne
 - Intégrer des données
 - Raster
 - vecteurs
- Opérations basiques d'analyse spatiale
 - Density kernel
 - Spatial regression
- Détecter l'autocorrélation spatiale

Chapter 12

SDM

- Les différents algorithmes
- Modèle averaging
- jointSDM

Chapter 13

Analyse de viabilité des populations chez les plantes

- Introduction : la question du type de données
- Matrice classique
- Lme avec structuration temporelle (données comptages)
- Lme avec structuration temporelle (données p/a)
-

Chapter 14

Soutient de cours

14.1 L2 – écologie générale

14.2 L3 – écologie des communautés

14.3 Master

Chapter 15

References