

Devoir de Statistiques descriptives

Statistiques descriptives avec R

Avril 2023

Le barème est donné à titre indicatif. Une attention particulière sera accordée aux commentaires des résultats obtenus.

Vous privilégiez un rendu sous la forme d'un fichier PDF. Le cas échéant, vous pouvez rendre un fichier R, ou Rmarkdown ou le rendu html de votre fichier Rmarkdown.

0.1 Prise en main de Rmarkdown (1+1 points)

Cette partie n'est pas indispensable pour la suite du devoir. Elle vise à créer un document R Markdown avec RStudio pour produire un rapport qui combine texte, code et résultats des évaluations de codes. Si les conditions ne sont pas réunies sur votre session pour créer de documents Rmarkdown ou si la partie vous semble difficile, vous pouvez simplement enregistrer vos réponses dans un fichier script R.

Le package `{rmarkdown}` peut être installé à l'aide de l'instruction suivante.

```
install.packages("rmarkdown")
```

Le document de référence sur lequel s'appuie cette partie est disponible à l'adresse : https://egallic.fr/Enseignement/L3_EcoFi/TD/r-markdown.html. Sentez vous libre de le consulter. Cependant, les notions qui vous seront le plus utile sont résumés ci-dessous.

Dans RStudio :

- Cliquez sur le menu **File**, puis sur **New File...**
- Cliquez sur **R Markdown...**
- Dans la fenêtre qui s'affiche :
 - assurez-vous d'être dans l'onglet "Document"
 - donnez un titre au document que vous allez créer
 - renseignez le champ "Author" avec vos noms et prénom(s)
 - Laissez le bouton radio sur l'option **HTML** pour que le rapport qui sera créé après soit un document **html**. Si vous avez une distribution LaTeX installé sur votre ordinateur, passez le bouton radio sur l'option **PDF** pour avoir un document PDF.
 - créer le document en cliquant sur le bouton **OK**.
 - Sauvegardez le fichier créé en lui donnant le nom de votre choix (par exemple : **nom-prénom.Rmd**)

Dans le document que vous venez de créer se trouve un en-tête YAML qui indique le titre, l'auteur, la date et le type de format du document.

- Pour les documents html

```
---  
title: "Devoir de statistiques descriptives"  
author: "Nom Prénom"  
date: "2023-04-13"  
output: html_document  
---
```

- Pour les documents PDF

```
---  
title: "Devoir de statistiques descriptives"  
author: "Nom Prénom"  
date: "2023-04-13"  
output: pdf_document  
---
```

Le fichier `.Rmd` peut contenir du texte que l'on rédige en markdown. Il vous suffit d'écrire comme dans un bloc note pour que le texte soit affiché dans le rapport final. Terminer une ligne par deux espaces permet d'aller à la ligne. Laisser une ligne vide entre deux textes permet de créer un paragraphe.

Vous écrirez vos codes dans des blocs encore appelés chunks. Le bloc de code peut-être insérer avec le raccourci **Control + ALT + i**. Remplacer la touche **Control** par **Command** sur les systèmes Mac. Sinon, utiliser trois accent graves avant et après le code et faire suivre le troisième accent grave du début par le nom du langage entre crochets `{r}`. Un exemple est donné ci-dessous :

```
```{r}  
seq(1, 10)
```
```

Le code d'un chunk peut être compilé en appuyant sur la flèche verte qui porte la mention **Run current chunk**.

Pour compiler le document R markdown dans son entièreté, il suffit de cliquer sur le bouton **Knit** (on le repère facilement avec son icône d'aiguille à tricoter et sa pelote de laine). Le résultat s'affiche dans une fenêtre qui s'ouvre à l'issue de la compilation.

0.2 Premières analyses (14 points)

Vous et votre binôme êtes sollicité.e.s par un département ¹ de Bretagne pour proposer votre expertise en statistiques descriptives!

Chaque année, le département attribue de nombreuses subventions à des associations, des tiers privés, des structures privées et des collectivités locales afin de les soutenir dans leurs actions. Il souhaite donc disposer d'un panorama sur les

¹circonscription administrative, territoire de compétence de services de l'Etat (Wikipédia)

principales structures qu'il subventionne et sur les dernières tendances observées à l'heure où ses compétences évoluent et où le contexte budgétaire se contraint. Vous aurez pour mission de caractériser les subventions départementales.

Le département est réparti en Etablissements Publics de Coopération Intercommunale (EPCI). Pour chaque subvention accordée à un EPCI, vous avez à disposition son montant, l'année de la subvention, sa thématique, la politique concernée, le type d'aide accordé, le type de bénéficiaire et le nombre d'habitants dans l'EPCI à cette année.

Le jeu de données fourni par le département est disponible dans le fichier `data.txt`.

1. Charger les données saisies dans le fichier `data.txt` en conservant la table comme un objet pour pouvoir la réutiliser directement.
2. Quelle est la population étudiée ?
3. A l'aide d'une commande unique R, afficher simultanément la dimension de la table, les noms des colonnes et leurs types R. Donner la nature de chaque variable statistique.
4. Votre binôme vous propose de transformer les variables `epci`, `thematique`, `politique`, `type_aide` et `type_beneficiaire` en `factor`. Pourquoi ? Ecrire le code R permettant d'effectuer ces transformations.
5. Le département souhaite savoir les noms et le nombre d'EPCIs uniques qui ont été subventionnés. Ecrire une commande R qui répond à leur question.
6. Vous vous proposez de leur donner le nombre de subventions demandé par chaque EPCI à l'aide d'une commande R.
7. Dresser la table des fréquences des thématiques subventionnées par le département. Quelle est la thématique ayant reçu le plus de financement ?
8. Construire à l'aide d'une commande R, un graphique (au choix) qui résume ces thématiques. Vous prendrez le soin de donner un titre au graphe. Commenter le graphe.
9. Construire deux graphes qui résument la distribution des montants octroyés par le département. Interprétez chacun des graphes.
10. Quel EPCI a reçu le plus faible montant de subventions ? Répondez à l'aide d'un code R.
11. Les variables type de bénéficiaire et thématique vous semblent-elles indépendantes? Répondez à l'aide de figure et/ou tableaux calculés sur R.

0.3 Analyses complémentaires (5 points)

Vous et votre binôme avez répondu avec succès aux questions posées par le département. Vous souhaitez apporter votre touche personnelle en calculant d'autres indicateurs statistiques et en tentant d'apporter une information supplémentaire qui pourrait être utile au département. (Il y a par exemple certaines variables dans le jeu de données qui n'ont pas été utilisées.)

1. Enoncer la question à laquelle vous souhaiteriez répondre et les outils de statistiques descriptives que vous utiliserez.
2. Si possible, écrire le code `R` permettant de répondre à la question et interprétez les résultats obtenus.