

Consignes projet : Établissements de santé

Octobre, 2022

But : Le but du projet est d'étudier et de décrire un ensemble de données en appliquant des outils rencontrés en TD et également de vous faire chercher un peu au-delà.

Une partie du projet consiste à exécuter des tâches précises (détaillées ci-dessous) pour vous aider à prendre en main les données. Une deuxième partie (**importante !**) du projet est dédiée à la **description et l'analyse des données de la manière que vous trouverez pertinente**.

Consignes générales Le projet doit être rédigé en utilisant R Markdown ou Quarto. La rédaction, les explications, l'interprétation des résultats et la clarté du code seront prises en compte dans la notation. De plus :

- Le rendu final doit être : un rapport sous forme html ou pdf et son fichier source .Rmd ou .qmd. Les titres des fichiers seront de la forme Nom1_Nom2_Nom3....
- La modification externe (sans l'utilisation de R) des fichiers de données est interdite.
- Dans le cas où des **packages** supplémentaires sont utilisés ils doivent être chargés dans un même chunk au début.
- Pour la deuxième partie, les choix d'étude devront être **justifiés**. En particulier, dans le cas où des méthodes statistiques seraient appliquées, inclure également une description sommaire de la méthode.

Données Vous avez à disposition quelques fichiers provenant de la *Statistique annuelle des établissements de santé (SAE)* (voir ici pour plus d'informations).

- les fichiers `CANCERO_2019.csv`, `Q20_2019`, `URGENCES2_2017.csv`, `URGENCES2_2018.csv`, `URGENCES2_2019.csv`, `URGENCES2_2020.csv` et `URGENCES2_2021.csv` contiennent des informations sur des établissements de santé en France.
- les fichiers `Desc_Q20.csv` et `Desc_Cancero_Urgences.csv` contiennent une description des variables des fichiers précédents.
- le fichier `Desc_Perso.csv` contient les étiquettes des modalités de la variable `Perso` du fichier `Q20_2019.csv`.

Prise en main de données

Cancero

1. Quel est le nombre de patients adulte (âge > ou égal à 18 ans) atteints de cancer ayant eu un traitement par chimiothérapie dans l'année en MCO (variable `CANCERO_A11`) ?
2. Comparer la proportion des adultes atteints de cancer ayant eu un traitement par chimiothérapie (MCO et HAD) en hospitalisation à domicile (HAD, variable `CANCERO_B11`) à la proportion des enfants ayant eu un traitement par chimiothérapie en hospitalisation à domicile (les variables `CANCERO_C11` et `CANCERO_D11` correspondent aux nombre de patients de moins de 18 ans atteints de cancer ayant eu un traitement par chimiothérapie dans l'année en MCO et HAD respectivement).
3. Créer une variable `Patients` contenant pour chaque hôpital le nombre total de personnes ayant eu un traitement par chimiothérapie (enfants et adultes, en MCO ou HAD). Représenter graphiquement

le nombre de médecins oncologues médicaux effectifs équivalent temps plein (variable `CANCERO_C28`) en fonction du nombre de `Patients`. Commenter les résultats. Faire un *zoom* sur le graphique en ne représentant que les hôpitaux ayant moins de 2000 patients.

Urgences

1. Dans les unités des urgences générales (celles pour lesquelles `URG == "GEN"`), comparer le nombre total de passages aux urgences en 2019 pour des patients de plus de 80 ans au nombre total de passages aux urgences dans la même année pour des patients de moins de 18 ans (variables `dt80` et `dt18`). Quelle est l'unité d'urgence générale pour laquelle la proportion de patients de moins de 18 ans est la plus élevée ?
2. Dans combien d'unités générales a-t-on rajouté des lits en 2020 par rapport à 2019 ? La variable `LIT_UHCD` contient le nombre de lits installés au 31/12.
3. Représenter graphiquement l'évolution de la moyenne annuelle du nombre total de passages aux urgences dans les unités générales entre 2017 et 2021. Indication : la variable `PASSU` donne le nombre total de passages aux urgences dans l'année.

Q20

1. Quelle est la proportion de données manquantes pour chaque variable ?
2. Calculer la proportion d'hommes dans l'effectif des salariés temps plein pour chaque spécialité. Les spécialités peuvent être retrouvées en regardant les modalités de la variable `PERSO`. Afficher à chaque fois le nom de la spécialité (en utilisant les informations du fichier `DescPerso.csv`) dont cette proportion est :
 - la plus grande
 - la plus petite
 - la plus proche de 1/2.
3. Ordonner les spécialités en fonction du nombre total d'effectif équivalent temps plein (ETP) des hospitaliers universitaires titulaires (variable `ETP_PU`). Afficher les 10 premières. Représenter graphiquement par un diagramme en barre ce nombre pour les 5 premières.