

MENTIONS LÉGALES

Étude produite par 143210563+CelineBoutinon@users.noreply.github.com sur la base des données en libre-accès sur <https://www.fao.org/faostat/en/#home>. Les analyses présentées dans ce document sont la responsabilité seule de l'auteur.

Information sur la licence pour l'utilisation des données de la FAO : <https://www.fao.org/contact-us/terms/en/>

[SLIDE 3]

Ces notes présentent dans une première partie les méthodes utilisées pour le téléchargement et la vérification des données, puis expliquent comment la qualité et l'exhaustivité de ces données ont été vérifiées, et enfin comment les données ont été exploitées, en particulier avec la création de nouveaux indicateurs. Dans une seconde partie, on présentera les différents indicateurs de sous-nutrition retenus ainsi que les principaux résultats de l'analyse.

[SLIDE 5]

Les données ont été importées à partir des fichiers .csv fournis, à l'exception des données de disponibilité alimentaire qui ont été sourcées directement sur le site de la FAO ; en effet, les données mises à disposition pour le projet dataient de 2013, et l'analyse étant demandée pour 2017, il a donc fallu les mettre à jour. Comme les 3 autres fichiers (aide alimentaire, population et sous-nutrition) comportaient tous des données allant de 2013 à 2018 (et 2012 à 2019 pour certains), on a choisi d'importer les données de disponibilité alimentaire de 2013 à 2017 afin de pouvoir faire des analyses cohérentes dans le temps sur tous ces fichiers, non seulement pour l'année 2017 mais aussi en identifiant les tendances sur les 5 années précédentes. La méthode .info() de Python a été utilisée pour vérifier le contenu des fichiers : nombre de lignes, intitulé des colonnes, format des données essentiellement. Cela a permis d'identifier le besoin d'uniformiser les noms des colonnes entre les différents fichiers (ce qui sera utile plus loin dans l'analyse quand nous aurons à effectuer des jointures entre les tables, mais aussi a permis d'en retirer les caractères spéciaux comme les accents). Certains types de données ont également été modifiés (les valeurs de la table aide_alim par exemple ont été changées de INT à FLOAT pour ne pas perdre en précision sur les opérations mathématiques). La colonne Valeur du fichier de sous-nutrition sera traitée plus loin car cette variable nécessite des manipulations plus complexes qu'un simple type casting.

[SLIDE 6]

Plusieurs méthodes ont été utilisées. Tout d'abord, on a supprimé les doublons dans les tables avec la méthode .drop_duplicates(). Ensuite, on a vérifié l'intégrité de la liste des pays en vérifiant l'absence de doublons dans le découpage géographique. C'est le cas de la Chine par exemple, pour laquelle on a supprimé les valeurs dupliquées : on garde les subdivisions territoriales de la Chine (à savoir 'Chine - RAS de Hong-Kong', 'Chine - RAS de Macao', 'Chine, continentale' et 'Chine, Taiwan Province de' et on supprime 'Chine' (qui représente la totalité de ces 4 provinces) dans le fichier de disponibilité alimentaire. La Chine étant un pays très peuplé (le 2^{ème} plus peuplé au monde), cela aurait pu considérablement affecter certains calculs. On a ensuite supprimé les colonnes vides dans les fichiers afin de ne pas alourdir inutilement l'affichage des résultats et de ne garder que les données utiles à l'analyse grâce à la méthode .dropna(). Les tables ont ensuite été analysées avec l'aide de la librairie MISSINGNO, afin d'identifier les lignes et colonnes présentant des valeurs partielles (méthode .matrix(df)).

Le fichier de disponibilité alimentaire présentait quelques valeurs manquantes en nombre minime que l'on a donc juste choisi de remplacer par des zéros, leur faible nombre étant peu susceptible d'affecter significativement l'analyse. Le fichier de sous-nutrition s'est avéré plus problématique, les données étant manquantes pour un grand nombre de pays, ou exprimées en termes non-numériques (tels la chaîne de caractères de type texte « <0.1 »). Il a donc fallu examiner les données de plus près. Il y a 203 pays dans le fichier de sous-nutrition. Sur ces 203 pays, 5 pays ont des valeurs partielles et 97 ont des valeurs complètes. 101 ont des valeurs totalement manquantes. La liste des pays pour lesquels aucune donnée de population en état d'insécurité alimentaire n'est incluse dans le fichier comprend de nombreux pays développés, comme notamment le Luxembourg, le Royaume-Uni ou Singapour. Il est probable que pour ces pays, le chiffre réel soit très proche de zéro. Les données sont manquantes également pour certains pays comme le Soudan du Sud, le Yémen ou l'Érythrée, où la situation géopolitique et climatique porte à penser que les données n'ont pas été collectées et/ou fournies à la FAO par le pays en question mais où le taux d'insécurité alimentaire est potentiellement important. Il conviendra donc de faire preuve de prudence dans l'interprétation des données, celles-ci étant manifestement incomplètes (en outre, l'absence de données historiques pour ces pays rend indisponible la méthode de remplacement par la moyenne ou par la médiane). L'absence de données économiques telles que le PIB dans le fichier rend également impossible l'imputation d'une valeur moyenne en fonction du niveau de développement du pays. Pour ces pays, on a donc remplacé les valeurs manquantes par des zéros, tout en étant conscients de l'impact de ce choix sur l'analyse. Les données au format non-numérique (<0.1) ont quant à elles été remplacées par le milieu de l'intervalle 0-0.1 (soit 0.05) ; là encore cela n'est pas neutre sur l'analyse, on aura l'occasion de l'illustrer par un exemple dans la 2^{ème} partie. De la même manière, les années données par intervalles de 3 ans ont été remplacées par le milieu de l'intervalle. On a ensuite vérifié la cohérence des données, les le type casting et la complétude des tables modifiées pour chaque table de données importée avec la méthode `.describe()`. La ligne `Count` permet de déceler les valeurs manquantes, les valeurs numériques renvoient des `NaN` pour les lignes `unique`, `top`, `freq`, et les valeurs string ou catégorie renvoient des valeurs `Nan` pour les indicateurs statistiques dans les lignes suivantes (`mean` à `max`). Cette analyse n'ayant pas révélé d'anomalies problématiques, nous pouvons à présent passer à la création de nouveaux indicateurs pour supporter notre analyse.

[SLIDE 7]

Afin de pouvoir analyser nos données, il a fallu procéder à :

- un filtrage des tables, essentiellement avec la méthode `.loc()` de Python, pour isoler une année ou un pays ;
- on a également utilisé la méthode `.groupby()`, pour analyser par exemple des indicateurs sur plusieurs années consécutives ;
- évidemment des opérations arithmétiques (sommations et moyennes essentiellement), ainsi que des multiplications et divisions (essentiellement pour corriger des problèmes d'unités) ; et enfin
- des jointures (ou parfois concaténations) de tables, pour exploiter ensemble des données présentes dans des tables différentes, essentiellement avec les méthodes `.merge()` et `.concat()`.

Dans certains cas limités, pour simplifier les formules de calcul, des variables ont été créées.

Voilà pour les principes généraux d'écriture du code qui nous ont permis de procéder à cette analyse, passons maintenant en revue les principaux résultats.

[SLIDE 8]

La sous-nutrition (ou insécurité alimentaire) concernait 7.11% de la population mondiale en 2017 sur la base des données disponibles (n'oublions pas l'absence de données pour certains pays

comme on l'a dit plus tôt). Après plusieurs années de recul, la sous-nutrition est à nouveau en hausse légère en 2017, comme le montre ce premier graphique.

[SLIDE 9]

Ces tableaux montrent la liste des 10 pays les plus touchés par l'insécurité alimentaire en 2017. À gauche, la liste avec les retraitements effectués pour remplacer les valeurs non-numériques (<0.1) par 0.05, et à droite le tableau ignorant les pays où ce remplacement a été fait. On constate que le classement change de façon marquée, ce qui nous permet de noter que les ajustements manuels en question ont résulté en une surestimation de la sous-alimentation pour les petits états insulaires à faible population (St Vincent et les Grenadines, Kiribati et la Dominique). Cet exemple est parlant car il faut toujours avoir à l'esprit les conséquences potentielles sur l'analyse de ce type de choix pour des données peu claires ou manquantes. Quel que soit le choix qui est fait, il n'est jamais neutre sur le résultat.

[SLIDE 10]

- Énergie mondiale disponible en 2017 suffisante pour nourrir 8.84 milliards de personnes
- En pratique, population de 7.55 milliards et 536.7 millions de mal nourris soit 7.01 milliards de personnes effectivement nourries
- 1.83 milliards de personnes supplémentaires auraient pu être nourries par la capacité calorique totale produite en 2017, ce qui est énorme, cela représente plus de 26% de la population en 2017.
- Ce qui met en valeur un problème de répartition plus que de production. Nous produisons en 2017 plus que ce qui aurait nécessaire pour nourrir toute la population, or il y avait 7.11% de sous nutrition en 2017.

[SLIDE 11]

Le problème de la répartition devient encore plus flagrant si l'on regarde les calories animales et végétales produites. Si on regarde la part des calories végétales dans cette disponibilité totale, on constate que sur les 8.84 milliards de personnes qui auraient pu être nourries théoriquement en 2017, 7.27 milliards auraient pu l'être à partir de calories d'origine végétale (soit 96.32 % de la population en 2017) contre 1.57 milliards seulement à partir de calories d'origine animale. On illustrera cette idée plus tard avec l'utilisation des céréales et le cas du manioc en Thaïlande, mais nous avons ici le premier indice que le problème de répartition n'est pas seulement un problème de répartition entre les humains « riches » et « pauvres », mais peut-être aussi un problème de répartition entre humains et animaux – est-il efficace de nourrir des animaux avec des céréales pour ensuite en consommer la viande ? Rien n'est moins sûr, et la végétalisation de l'alimentation est un grand débat d'actualité.

[SLIDE 12]

Regardons à présent comment cette disponibilité alimentaire est affectée. Tout d'abord, rappelons comment est constituée et utilisée la disponibilité intérieure. Le fichier de disponibilité alimentaire vérifie les égalités suivantes pour chaque élément de la colonne ['Produit'] (hors population) :

Disponibilité intérieure = Production + (Importations - Quantité) - (Exportations - Quantité) - Variation de stock

= Aliments pour animaux + Semences + Pertes + Traitement + Résidus + Nourriture

Les 2 côtés de cette équation représentent les sources et utilisations de la disponibilité intérieure et sont par définition toujours équilibrés. Utilisations principales de la disponibilité intérieure :

- Nourriture majoritairement
- Suivie de "Traitement" (=utilisations chimiques, industrielles etc... comme le bioéthanol)
- Aliments pour animaux en 3^{ème} position

[SLIDE 13]

- Répartition relativement stable sur 5 années glissantes
- Légère hausse depuis 2015 pour la nourriture, les aliments pour animaux et les applications industrielles en parallèle avec la hausse légère de la disponibilité intérieure globale

[SLIDE 14]

- Ration journalière = 2,500 kcal / pers. (Source : FAO)
- Grandes inégalités selon les pays
- Le pays ayant le plus de disponibilité a plus du double du pays qui en a le moins

[SLIDE 15]

- Classement peu modifié si on regarde les valeurs moyennes sur 5 ans

[SLIDE 16]

La FAO ne publie plus les données des Food Aid Shipments (World Food Program) après 2016 - cf. <https://www.fao.org/faostat/en/#data/FA>, donc malheureusement pas de cohérence historique possible ici et comme nous n'avons pas de données pour 2017, l'analyse s'arrête donc à 2016, mais les chiffres sont néanmoins intéressants et révèlent que l'aide ne va pas nécessairement aux pays les plus en difficulté (guerre, situation politique, cf. Corée du Nord)

[SLIDE 17]

Pour ces calculs, nous avons utilisé :

- Un équivalent calorique sur la base de 2,950kCal / kg de céréales en moyenne (source : Feedtables.com, 2023)
- Un équivalent calorique de la viande produite en 2017 (sur la base de 2,300kCal / kg de viande en moyenne (source : FDA, 2021).

La céréale la plus abondante au niveau mondial est le maïs, qui est utilisé pour près de moitié pour l'alimentation animale alors que seuls 15% environ sont utilisés pour l'alimentation humaine, le reste allant aux applications industrielles comme le biofuel par exemple. Ce qui nous donne 3.44 calories de céréales consommées pour chaque calorie de viande produite en 2017. On aurait pu nourrir 3.04 milliards de personnes en 2017 avec la même quantité de céréales, alors que seules 0.88 milliards ont effectivement été nourries à partir des calories animales produites, ce qui illustre une fois de plus l'inefficience de l'élevage intensif à nourrir la population mondiale.

Comparer ce chiffre à la proportion de personnes nourrissable en végétal vs animal - rapport de 1 à 4 environ, 1 à 3.5 ici => souligne une fois de plus l'inefficience de l'élevage. De plus, on retrouve le même rapport de 1 à 4 environ que dans la disponibilité alimentaire végétale vs. animale, donc on pourrait nourrir une population 4 fois plus importante en tout-végétal. En pratique, les valeurs individuelles sont de 9, 7 et 4 respectivement pour le bœuf, le porc et la volaille en moyenne. Cette inefficience poussée à l'absurde de l'élevage industriel est très bien illustrée par l'exemple de la Thaïlande et du manioc.

[SLIDE 18]

L'inefficience de la répartition alimentaire homme/animal est particulièrement illustrée par la situation du manioc en Thaïlande. La Thaïlande produisait en moyenne 30.85 millions de tonnes

de manioc par an sur la période 2013-2017 soit les calories nécessaires pour nourrir 54.1 millions de personnes par an en moyenne, soit 78.61 % de sa population moyenne sur la période, mais en pratique 91.86 % de la production en moyenne sont exportés et le taux de malnutrition s'élevait à 8.96 % en 2017. Cette quantité de manioc serait suffisante à couvrir les besoins caloriques d'environ 8.73 fois la population mal nourrie en Thaïlande. On note également que la Thaïlande importe du manioc malgré sa production importante. On peut donc penser que la Thaïlande s'est spécialisée dans la production d'une qualité de manioc impropre à l'alimentation humaine et donc utilisée pour les animaux et qu'elle exporte, et importe le manioc pour sa propre consommation humaine intérieure.

[SLIDE 19]

En conclusion, et comme on l'a dit, certains choix ont été faits pour l'analyse des données. Il n'y a pas de manière unique de résoudre un problème de données manquantes ou partiellement adéquates, tout dépend des circonstances et des autres informations disponibles, mais dans tous les cas, il faut être conscient que ces choix créent toujours un biais dans l'analyse finale, comme on l'a montré avec l'exemple du classement des 10 pays avec les taux d'insécurité alimentaires les plus forts. En tout cas, et malgré ces biais, une tendance très nette se dégage de l'analyse, qui met en valeur des inégalités de répartition et des choix de production sous-optimaux. Qu'il s'agisse de nourrir la population ou de préserver l'environnement, il est très clair que l'élevage intensif est inadapté, et qu'un recours massif à la végétalisation de l'alimentation est impératif si on veut être capable de nourrir les presque 10 milliards d'habitants que devrait compter la terre en 2050.