



BUSINESS INTELLIGENCE

À l'attention de Madame Charlier Christine et Monsieur Macq Alexis

Création d'un tableau de bord interactif



DAFFE Tristan – 3IG – Groupe B2
DIEUDONNE Nicolas – 3IG – Groupe B2
ROUSSEAU Benjamin – 2IG – Groupe C

Table des matières

Contexte	2
Indicateurs choisis	2
Fouille de données et signification des variables utilisées	3
COVID19BE_HOSP	3
COVID19BE_MORT	3
COVID19BE_VACC	4
COVID19BE_NH	4
Présentation des requêtes utilisées dans Excel	5
COVID19BE_MORT	5
COVID19BE_VACC	5
COVID19BE_HOSP	6
COVID19BE_NH	6
Exemple de data cleaning sur le dataset COVID19BE_NH	7
Test statistique	8
Fixer les variables et les hypothèses	8
Préciser le seuil de signification alpha	8
Préciser la loi de probabilité.....	8
Déterminer le seuil de rejet et la zone de non-refus	8
Calculer la grandeur expérimentale	9
Webographie.....	10

Contexte

Actuellement en décembre 2022, nous nous situons trois ans après le début de la pandémie COVID-19, causée par le virus SRAS-Cov-2. Les premiers cas du virus ont été signalés à Wuhan, en Chine, en décembre 2019 et s'est rapidement propagée à échelle mondiale. L'Organisation mondiale de la santé l'a déclaré pandémie en mars 2020.

Cette situation a provoqué des perturbations généralisées de la vie quotidienne, les gouvernements mettant en œuvre des mesures telles que des lockdowns, des couvre-feux, des restrictions de voyage et des mandats de masquage dans le but de ralentir la propagation du virus. La pandémie a eu un impact considérable sur les systèmes de santé, les économies et les sociétés du monde entier, entraînant des changements sans précédent dans la façon dont nous vivons et interagissons les uns avec les autres.

La pandémie a également donné lieu à d'importantes avancées scientifiques et technologiques, les chercheurs travaillant jour et nuit pour mettre au point des vaccins et des traitements contre le virus, et c'est ce que nous allons nous pencher.

Via une récolte de donnée et la création d'un tableau de bord interactif, nous allons donc essayer d'observer l'effet de la vaccination de la population belge sur la pandémie (en Belgique) en nous basant sur différents indicateurs.

Indicateurs choisis

Voici les quatre indicateurs sur lesquels nous avons décidé de nous baser afin de créer notre tableau de bord :

- **Le nombre de vaccinations** : Il s'agit de la quantité de la population qui a reçu au moins une dose d'un vaccin. Un nombre de vaccinations plus élevé peut contribuer à réduire la propagation de la maladie.
- **Le nombre de morts** : Il s'agit du nombre de décès dus à la maladie dans une population sur une période donnée. Une baisse du nombre de mortalité peut être une indication que la vaccination a un impact.
- **Le nombre d'hospitalisations** : Il s'agit du nombre de personnes qui sont hospitalisées en raison de la maladie dans une population sur une période donnée. Une baisse de celui-ci peut être une indication que la vaccination a un impact. Nous regarderons aussi le nombre de patients dans les unités de soins intensifs, le nombre de patients sous assistance respiratoire et le nombre de patients sous ECMO (oxygénation par membrane extracorporelle).
- **Le nombre de décès en maison de repos** : Les résidents des maisons de retraite sont souvent parmi les plus vulnérables aux maladies infectieuses, dont le COVID-19, et aux décès qui en découlent. En analysant le nombre de décès parmi les résidents des maisons de retraite après l'introduction d'un vaccin, les responsables de la santé publique peuvent se faire une idée de l'efficacité du vaccin pour protéger cette population vulnérable.

Fouille de données et signification des variables utilisées

Nous avons utilisé différents datasets (collections de données) repérés sur le site *Sciensano* et téléchargés le 26 décembre 2022.

Sciensano est l'Institut belge pour la santé publique. Il s'agit d'un institut scientifique qui mène des recherches et fournit des informations et des conseils sur les questions de santé publique en Belgique. Il est chargé de surveiller et d'analyser la santé de la population belge, ainsi que de fournir des informations et des conseils aux décideurs politiques et au grand public sur les questions de santé publique. Sciensano joue également un rôle clé dans la prévention et le contrôle des maladies infectieuses ainsi que dans l'évaluation et le suivi des interventions sanitaires.

COVID19BE_HOSP

Le premier contient des variables nous fournissant des informations sur les hospitalisations liées au COVID-19.

DATE	Date
PROVINCE	Province des hôpitaux (y compris Bruxelles)
REGION	Région des hôpitaux
NR_REPORTING	Nombre d'hôpitaux
TOTAL_IN	Nombre total de patients hospitalisés pour le COVID-19 et confirmés en laboratoire au moment du diagnostic, y compris les soins intensifs.
TOTAL_IN_ICU	Nombre total de patients hospitalisés pour le COVID-19 et confirmés par le laboratoire, en soins intensifs au moment du diagnostic.
TOTAL_IN_RESP	Nombre total de patients hospitalisés pour le COVID-19, confirmés par le laboratoire, sous assistance respiratoire au moment du diagnostic.
TOTAL_IN_ECMO	Nombre total de patients COVID-19 hospitalisés et confirmés en laboratoire sous ECMO (oxygénation par membre extracorporelle) au moment du diagnostic.
NEW_IN	Nombre de nouveaux patients COVID-19 confirmés par le laboratoire au cours des dernières 24 heures.
NEW_OUT	Nombre de nouvelles sorties d'hôpital confirmées par le laboratoire de patients COVID-19

COVID19BE_MORT

Ce dataset nous fournit des informations sur les décès liés au COVID-19 en Belgique.

DATE	Date du décès.
REGION	Région de l'hôpital dans lequel a été constaté de décès
AGEGROUP	Groupe d'âge dans lequel se situait la personne décédée.
SEX	Sexe de la personne décédée.
DEATHS	Nombre de décès.

COVID19BE_VACC

Celui-ci nous fournit des informations sur les vaccinations liées au COVID-19 en Belgique.

DATE	Date du décès.
REGION	Région de l'hôpital dans lequel a été constaté de décès
AGEGROUP	Groupe d'âge dans lequel se situait la personne décédée.
SEX	Sexe de la personne décédée.
BRAND	Marques de vaccins (les vaccins qui ne sont normalement pas utilisés en Belgique mais qui ont été approuvés par une autorité réglementaire nationale en Europe ou qui figurent sur la liste d'utilisation d'urgence de l'Organisation mondiale de la santé sont regroupés sous la marque "Autre"). Pour les marques produisant des vaccins adaptés à partir de différentes souches de virus différentes, la version du vaccin est spécifiée.
DOSE	Identifiant de la dose (pour les vaccins nécessitant 2 doses : A pour la première dose, B pour la seconde dose ; C pour les vaccins ne nécessitant qu'une seule dose ; E pour une dose supplémentaire/de rappel de vaccin administrée depuis le 9 septembre 2021 ; E2 pour une deuxième dose de rappel administrée depuis le 20 janvier 2022 ; E3 pour la troisième dose de rappel).
COUNT	Nombre de personnes ayant reçu le vaccin (les combinaisons avec 0 vaccination ne sont pas indiquées).

COVID19BE_NH

Ce dataset nous fournit des informations sur les décès en maisons de repos liés au COVID-19 en Belgique.

YEAR_WEEK	Semaine (du lundi au dimanche).
REGION	Région/communauté d'implantation de la maison de repos (Bruxelles, Flandre, Wallonie, Communauté germanophone).
CASES_PER_100	Pourcentage de nouveaux cas confirmés de COVID-19 parmi les résidents de la maison de repos par semaine. Le dénominateur est le nombre de résidents de la maison de repos.
NEW_IN_PER_100	Pourcentage de nouvelles hospitalisations dues à COVID-19 parmi les résidents des maisons de repos par semaine. Le dénominateur est le nombre de résidents de la maison de retraite.
DEATHS_PER_100	Pourcentage de morts causés par le COVID-19 (possibles, confirmés et confirmés radiologiquement) parmi les résidents de maisons de repos. De petites corrections supplémentaires ont été nécessaires pour conserver des dénominateurs réalistes sur une base hebdomadaire.
PART_RATE_PERCENT	Nombre de maisons de repos.

Présentation des requêtes utilisées dans Excel

Afin de créer un tableau de bord interactif en Excel, il est impératif de faire du « Data cleaning », du nettoyage de données. Le nettoyage et la préparation des données constituent une étape importante du processus de création d'un tableau de bord car ils permettent de s'assurer que les données sont dans un format utilisable et sans erreurs ou incohérences.

Cela permet également d'utiliser des outils tels que des filtres, des tableaux croisés dynamiques et des formules pour manipuler les données de manière visuelle et interactive, éléments essentiels à la création d'un tableau de bord.

Tout ceci permet de garantir que le tableau de bord sera basé sur des données exactes et fiables, ce qui peut aider les utilisateurs à prendre des décisions plus éclairées et à tirer des conclusions plus précises du tableau de bord.

Voici une présentation de requêtes utilisées dans Excel sur nos différents datasets pour la création de notre tableau de bord :

COVID19BE_MORT

- **Type modifié** : la date du décès était stockée sous forme de chaîne de caractère. Une chaîne de caractères est un type de données qui représente une séquence de caractères, comme un mot ou une phrase. Il était plus intéressant pour nous de la stocker dans un type « date ». Une date, est un type de données qui représente un moment précis dans le temps, (des secondes, une heure, un jour, un mois et une année spécifiques), ce qui nous sera utile pour créer des graphiques prenant une date sur l'axe des x.
- **Date extraite** : Nous avons simplement récupéré les informations de la date qui nous intéressaient. En l'occurrence, le jour, le mois et l'année.
- **Création d'un tableau dynamique** : reprenant la somme nombre de morts en filtrant par tranche d'âge.
- **Segmentation** du tableau par année.
- **Création d'un histogramme** reprenant le nombre de morts par tranche d'âge, le tout segmenté par année.

COVID19BE_VACC

- **Type modifié**
- **Date extraite**
- **Création d'un tableau dynamique** reprenant par mois la somme totale du nombre de vaccinations.
- **Segmentation** du tableau par année
- **Création d'un histogramme** reprenant par mois et par années la somme totale du nombre de vaccinations.

COVID19BE_HOSP

- **Type modifié**
- **Date extraite**
- **Création d'un tableau dynamique** reprenant par mois :
 - o Le nombre de personnes en soins intensifs,
 - o Le nombre de personnes sous assistance respiratoire,
 - o Le nombre de personnes sous ECMO,
 - o Le nombre total de personnes hospitalisées.
- **Segmentation** du tableau par année
- **Création d'un graphique linéaire** montrant l'évolution par mois du nombre de personnes en soins intensifs, sous assistance respiratoire, sous ECMO et du nombre total de personnes hospitalisées, le tout segmenté par année.

COVID19BE_NH

- **Fractionner la colonne par délimiteur / Colonnes renommées / Préfixe ajouté :**
C'est trois étapes on pour but de récupérer le numéro de semaine et l'année. Ces informations sont stockées sous cette forme dans le dataset : « 20W10 » où '20' indique l'année '2020' et 'W10' indique qu'il s'agit de la semaine 10.
Le but a donc été de récupérer ces informations dans deux colonnes :
 - o YEAR : l'année (par exemple '2020 ')
 - o WEEK NUMBER : le numéro de la semaine dans l'année (par exemple '10')
- **Remplace null par 0 :**
Cette étape a été appliquée aux colonnes PART_RATE_PERCENT, CASE_PER_100, NEW_IN_PER_100, DEATHS_PER_100.
Une valeur nulle est une valeur spéciale qui représente l'absence d'une valeur ou une référence nulle. Il est ici utile de remplacer les valeurs nulles par une valeur par défaut (ici 0) afin de faciliter le travail sur les données. Dans notre cas, cela nous permet de prendre en compte les semaines où nous n'avons pas d'informations sur les morts du COVID-19 en maison de repos dans nos calculs de sommes (il est normal de ne pas avoir d'informations sur certaines semaines, car on ne recensait pas de mort dues au COVID-19 en maison de repos début 2020).
- **Création d'un tableau dynamique** reprenant par an : le numéro de la semaine, le nombre de morts pour 100 résidents, le nombre de cas détectés pour 100 résidents et le nombre d'arrivées pour 1^{er} résidents.
- **Segmentation** du tableau par année
- **Création d'un graphique linéaire** représentant l'évolution de ces différentes variables par an, le tout segmenté par semaine.

Exemple de data cleaning sur le dataset COVID19BE NH

Requêtes [5]

TAB_mort

TAB_hospitalisation

TAB_vaccination

TAB_test

TAB_home

fx

= Table.ReplaceValue("#"Remplace null par 0 NEW_IN_PER_100",null,0,Replacer.ReplaceValue,{"DEATHS_PER_100"})

	WEEK NUMBER	REGION	PART_RATE_PERCENT	CASES_PER_100	NEW_IN_PER_100	DEATHS_PER_100
1	10	Belgium	0	0	0	
2	10	Brussels	0	0	0	
3	10	Flanders	0	0	0	
4	10	German speaking community	0	0	0	
5	10	Wallonia	0	0	0	
6	11	Belgium	0	0	0	
7	11	Brussels	0	0	0	0,
8	11	Flanders	0	0	0	
9	11	German speaking community	0	0	0	
10	11	Wallonia	0	0	0	
11	12	Belgium	0	0	0	0,
12	12	Brussels	0	0	0	0,
13	12	Flanders	0	0	0	0,
14	12	German speaking community	0	0	0	
15	12	Wallonia	0	0	0	0,
16	13	Belgium	0	0	0	0,
17	13	Brussels	0	0	0	0,
18	13	Flanders	0	0	0	0,
19	13	German speaking community	0	0	0	0,
20	13	Wallonia	0	0	0	0,
21	14	Belgium	0	0	0	0,
22	14	Brussels	0	0	0	1,
23	14	Flanders	0	0	0	0,
24	14	German speaking community	0	0	0	1,
25	14	Wallonia	0	0	0	0,
26	15	Belgium	0	0	0	1,
27						

Paramètres d'une requête

PROPRIÉTÉS

Nom

TAB_home

Toutes les propriétés

ÉTAPES APPLIQUÉES

Source

Fractionner la colonne par déli...

Colonnes renommées

Préfixe ajouté

Remplace null par 0 PART_RA...

Remplace null par 0 CASES_PE...

Remplace null par 0 NEW_IN_...

Remplace null par 0 DEATH_P...

7 COLONNES, 725 LIGNES

Profilage de la colonne en fonction des 1000 premières lignes

APERÇU TÉLÉCHARGÉ À 14:00

Test statistique

Ce test statistique a pour but d'observer l'évolution du nombre moyen de personnes sous assistance respiratoire entre 2021 et 2022. Pour ce faire, nous allons effectuer un test unilatéral droit en nous basant sur ces informations prélevées dans nos dataset :

Stats descr ASS-RESP 2021	
Moyenne	7053,83333
Erreur-type	1297,14776
Médiane	5404,5
Écart-type	4493,45166
Variance de l'échantillon	20191107,8
Kurtosis (Coefficient d'aplatissement)	-0,68196172
Coefficient d'asymétrie	0,70609869
Plage	13804
Minimum	1834
Maximum	15638
Somme	84646
Nombre d'échantillons	12
Niveau de confiance(95,0%)	2855,00297

Stats descr ASS-RESP 2022	
Moyenne	1761
Erreur-type	568,0675776
Médiane	934
Écart-type	1967,843813
Variance de l'échantillon	3872409,273
Kurtosis (Coefficient d'aplatissement)	3,722598038
Coefficient d'asymétrie	2,015160479
Plage	6446
Minimum	415
Maximum	6861
Somme	21132
Nombre d'échantillons	12
Niveau de confiance(95,0%)	1250,308308

Fixer les variables et les hypothèses

$H_0 = m_1 - m_2 = 0$ (Le nombre moyen de personnes sous assistance respiratoire reste similaire d'une année à l'autre).

$H_1 = m_1 - m_2 > 0$ (Le nombre moyen de personnes sous assistance respiratoire est plus grand en 2022 qu'en 2021).

Préciser le seuil de signification alpha

$\alpha = 0.05$

Préciser la loi de probabilité

$n_1 = 12$, et est donc > 30

$n_2 = 12$, et est donc > 30

⇒ Utilisation de la loi de Student

Déterminer le seuil de rejet et la zone de non-refus

Région de refus : $]1.645 ; +\infty [$

Région de non-refus : $] -\infty ; 1.645]$

Calculer la grandeur expérimentale

$$\frac{d - E(d)}{\sqrt{\text{var } d}} = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} =$$
$$\frac{7053,833 - 1761 - 0}{\sqrt{\frac{20191107,787}{12} + \frac{3872409,272}{12}}} = 3,737$$

Conclusion

3,737 appartient à la zone de refus, on doit rejeter H_0 au profit de H_1 .

D'après les statistiques descriptives, au seuil de signification $\alpha = 0.05$, le nombre moyen de personnes sous assistance respiratoire semble être plus grand en 2022 qu'en 2021.

Ceci peut être un indicateur que le vaccin a eu un impact sur le nombre de personnes sous assistance respiratoire, le COVID-19 pouvant provoquer une maladie respiratoire grave chez certaines personnes, entraînant la nécessité d'une assistance respiratoire. Les vaccins sont conçus pour offrir une protection contre le COVID-19 en aidant l'organisme à développer une immunité contre le virus. Cela peut contribuer à réduire le risque de maladie grave et de complications, y compris la nécessité d'une assistance respiratoire.

Il ne faut cependant pas oublier qu'il y a P-plusieurs facteurs qui peuvent avoir contribué à une diminution du nombre de personnes nécessitant une assistance respiratoire en raison de COVID-19. Certains de ces facteurs comprennent :

- **Amélioration des soins médicaux** : En acquérant plus d'expérience dans le traitement du COVID-19, les professionnels de la santé ont pu devenir plus efficaces dans la gestion de la maladie et dans la réduction du risque de complications graves, y compris la nécessité d'une assistance respiratoire.
- **Mise en œuvre de mesures de santé publique** : La mise en œuvre de mesures telles que le port du masque, la distanciation sociale et l'hygiène des mains peut avoir contribué à réduire la propagation du virus et la gravité de la maladie chez ceux qui ont été infectés.
- **Changements dans le virus lui-même** : Il est possible que des changements dans le virus, comme des mutations, aient contribué à une diminution de la gravité de la maladie chez certaines personnes atteintes du COVID-19.

Il est important de noter que la contribution relative de ces facteurs à la diminution du nombre de personnes nécessitant une assistance respiratoire en raison du COVID-19 n'est pas entièrement comprise, et il est probable qu'une combinaison de ces facteurs ait joué un rôle.

Webographie

Lien d'obtention des dataset : Sciensano. Repéré à :

<https://epistat.sciensano.be/covid/>

(Consulté le 26 décembre 2022). Belgium COVID-19 Epidemiological Situation. Repéré à

<https://datastudio.google.com/embed/u/0/reporting/c14a5cfc-cab7-4812-848c-0369173148ab/page/ZwmOB>

(Consulté le 26 décembre 2022). COVID19BE OPEN DATA CODEBOOK. Repéré à

https://epistat.sciensano.be/COVID19BE_codebook.pdf

(Consulté le 26 décembre 2022). COVID-19 BULLETIN EPIDEMIOLOGIQUE
HEBDOMADAIRE. Repéré à

<https://covid-19.sciensano.be/sites/default/files/Covid19/Derni%C3%A8re%20mise%20%C3%A0%20jour%20de%20la%20situation%20%C3%A9pid%C3%A9miologique.pdf>

(Consulté le 30 décembre 2022) RESPIRATORY FAILURE Causes and Risk Factors. Repéré à

<https://www.nhlbi.nih.gov/health/respiratory-failure/causes>

Oxygénation par membrane extracorporelle (2022, 1 décembre) Dans *Wikipédia*. Repéré le 26 décembre 2022 à :

https://fr.wikipedia.org/wiki/Oxyg%C3%A9nation_par_membrane_extracorporelle

Icône Excel [Image en ligne] Repéré à :

<https://fr.findicons.com/icon/590631/excel>