

Grau en Estadística

Títol: Anàlisi de dades de panell de la *Covid-19* a la Unió Europea

Autor: Ivette Roig Rodriguez

Director: Lluís Bermudez Morata

Departament: Matemàtica Econòmica, Financera i Actuarial

Convocatòria: Setembre 2022



Resum

En aquest treball de fi de grau, s'estudia el comportament tendència de les dades de *Covid-19* de la Unió Europea utilitzant eines econòmiques existents que poden ser útils per entendre el seguiment d'aquesta pandèmia que ha afectat globalment.

Amb l'ús de dades de panell i l'aplicació de models estadístics, l'objectiu principal d'aquest treball final és comprendre l'evolució de les morts degudes a la *Covid-19* i quines han estat les variables amb més influència durant aquests darrers anys.

Resumen

En este trabajo de fin de grado, se estudia el comportamiento tendencia de los datos de *Covid-19* de la Unión Europea utilizando herramientas econométricas existentes que pueden ser útiles para entender el seguimiento de esta pandemia que ha afectado globalmente.

Con el uso de datos de panel y la aplicación de modelos estadísticos, el objetivo principal de este trabajo final es comprender la evolución de las muertes debidas a la *Covid-19* y cuáles han sido las variables con mayor influencia durante estos últimos años.

Summary

This paper studies the trend behaviour of *Covid-19* data at the European Union level using econometric tools that may be useful for understanding the behaviour of this pandemic and its effects globally.

The main objective is to understand the behaviour of deaths by *Covid-19* and which have been the variables with the greatest influence in recent years. The method used combines panel data and the application of statistical models.

Paraules clau: Panel Data, OLS, Fixed Effect Model, Random Effects Model, heterogeneïtat, homocedasticitat, Covid-19, SARS-Cov-2, taxa exponencial i Pooling.

Classificació AMS

El codi principal de la Societat Americana de Matemàtiques (*AMS*) per a aquest treball de fi de grau és "*62-07 Data analytics*" degut a que es tracta principalment d'un anàlisi de dades. També, a causa de l'ús d'eines estadístiques com les regressions lineals, aquest treball de fi de grau es classificaria com a "*62-J05 Linear regression*".

Índex

1.	Introducció	1
2.	Metodologia	2
3.	Marc teòric	3
3.1.	<i>Introducció al Panel Data</i>	3
3.2.	<i>Panel data i heterogeneïtat</i>	3
3.3.	<i>Models estadístics de Panel Data</i>	4
4.	Estudi de la base de dades	7
4.1.	<i>Captura i obtenció de les dades</i>	7
5.	Modelització	24
5.1.	<i>OLS i Pooled OLS</i>	24
5.2.	<i>(One-Way) Fixed Effect Model</i>	28
5.3.	<i>Two-Way Individual Effects Models</i>	31
5.4.	<i>(One-Way) Random Effects Model</i>	36
5.5.	<i>Selecció dels models</i>	38
5.6.	<i>Modelització tenint en compte les vacunes contra la Covid-19</i>	40
6.	Conclusions	47
7.	Bibliografia	49
8.	Annex	52
8.1.	<i>Representacions gràfiques</i>	52
8.2.	<i>Representacions gràfiques en el temps de les variables per setmana i país</i>	56
8.3.	<i>Codi</i>	61

1. Introducció

Paraules com "taxa exponencial", "onada" i "aplanar la corba" han estat àmpliament citades per tot tipus de persones a les xarxes socials des de l'esclat de la pandèmia causada pel *COVID-19*. Aquest esdeveniment històric comença al desembre del 2019 quan nombrosos casos de pneumònia d'origen desconegut van sortir a la ciutat de Wuhan, China. Més tard, al gener del 2020 es va identificar que aquest síndrome respiratori era causat pel nou coronavirus o també conegut com a *SARS-CoV-2* (*COVID-19*, s.f.).

Des de principis del 2020, el que va començar com a un virus llunyà i que només afectava a la gent gran, va passar a convertir-se en una nova ordre mundial on els governs de tot el món han estat actualitzant amb freqüència les seves polítiques per gestionar la propagació d'aquest virus i reduir la taxa de mortalitat amb els recursos mèdics dels que es disposa.

Per tant, entendre el comportament i la tendència de la pandèmia és crucial des de la perspectiva de l'elaboració de polítiques, ja que no només ha anat deixant un llarg camí de víctimes pel camí, si no que també ha comportat un impacte negatiu econòmic molt gran.

La finalitat principal d'aquest treball és investigar el comportament tendència de les dades de *COVID-19* a nivell dels països de la Unió Europea i dibuixar atenció a algunes eines econòmiques existents que poden ser útils per a treballs futurs. Respecte a la modelització de les dades de la *COVID-19*, és tot un repte ja que ens enfrontem a l'anàlisi de dades massiu canviant en el temps.

El principal objectiu d'aplicar i estudiar les dades de panell és capturar l'heterogeneïtat no observable, donat que aquesta, no es pot detectar ni amb estudis de sèries temporals ni tampoc amb estudis de tall transversal. Es busca explicar l'heterogeneïtat de les morts entre països atenent a unes variables de control, és a dir, les variables explicatives, i, segons alguns models, a la tendència temporal.

2. Metodologia

La metodologia d'aquest treball s'ha organitzat, com es pot observar en el diagrama de Gantt, de la següent manera:

Taula 2.1. Representació de la metodologia en Diagrama de Gantt

Tasques	Setmanes											
	1	2	3	4	5	6	7	8	9	10	11	12
Introducció												
Descripció de les dades												
Captura i obtenció												
Preprocessament												
Outils / Missings												
Correccions bbdd												
Anàlisi descriptiu de les dades												
Metdologia de l'anàlisi												
Models estadístics panel data												
Visualització de les dades												
Redacció del cos												
Conclusions												
Revisió + annex												

En primer lloc, es fa el recull de les dades de panell sobre la informació de la Covid-19 als diferents països de la Unió Europea. Es recullen les dades diàriament i setmanalment publicades a través del lloc web oficial *European Centre for Disease Prevention and Control* (COVID-19, s. f.).

D'aquesta àmplia *web*, se'n fa una extracció de les bases de dades que interessin per realitzar aquest treball, i es tria el període de dos anys d'entre la setmana 12 de l'any 2020 fins a la setmana 12 de l'any 2022.

Un cop feta l'extracció d'aquestes, amb l'eina de programació *R*, es procedeix a realitzar un primer anàlisi amb el qual s'obtindrà un primer contacte amb les dades i es procedirà a fer tots els canvis que es necessiten, així com possibles correccions per obtenir la base de dades final amb la qual es treballarà. En el cas de les dades amb les quals es treballa, són dades referents a la *Covid-19* i tots els països de la Unió Europea.

Per últim, s'utilitzen models de regressió de dades de panell per estudiar el comportament d'aquestes dades mencionades. L'objectiu principal és intentar explicar amb aquests models de regressió per *panel data* les morts que s'han produït degut a aquesta pandèmia que ha afectat mundialment.

Com es pot veure en el diagrama de Gantt, l'organització d'aquest treball s'ha realitzat per setmanes, d'aquesta manera seguim un ordre i encara que sempre surten imprevistos, és una bona manera d'organitzar la feina i les tasques a fer durant la realització d'aquest treball de fi de grau.

3. Marc teòric

3.1. Introducció al Panel Data

Les dades de panell, més conegudes com *panel data*, en anglès, són conjunts de dades que contenen observacions sobre diferents seccions transversals al llarg del temps. Els grups que es poden formar són molt amplis, des de països, empreses, grups demogràfics...

Les dades són en format diari, setmanal, semestral i anual ja que estan recopilades de manera cronològica i amb una freqüència regular. Per una variable y_{it} , se suposa que tenim $i = 1, \dots, N$ observacions de tall transversal i $t = 1, \dots, T$ observacions temporals, on i i t fan referència als individus i al període de temps, respectivament. Les observacions de tall transversal són tipus de dades recopilades mitjançant l'observació de molts subjectes alhora, o sense tenir en compte les diferències en el temps.

Els individus poden ser: persones, empreses, regions, països, etc. Amb aquest tipus de dades es poden modelar tant comportaments comuns com individuals del grup que sigui adequat. També poden contenir molta variabilitat i es poden detectar i mesurar efectes estadístics que amb altres sèries temporals pures o dades transversals no es podria (E., 3 marzo, 2020).

Dins del *panel data* les dades poden ser

- Balancejades: cada individu i es observat en T períodes ($T_i = T$ per tot T)
- No balancejades: cada individu i es observat en T_i períodes ($T_i \neq T$ per tot T),

Una altra forma de classificar els panells és depenent del número d'observacions en els talls transversals i temporals, amb els quals es pot diferenciar entre tres tipus de *panel data*:

- *Micropanel*: si es té un nombre reduït d'observacions temporals per cada individu i el nombre d'individus és molt elevat. Normalment els individus són persones o llars.
- *Macropanel*: si hi ha un gran nombre d'observacions temporals i pocs individus.
- Camp aleatori o *random*: si hi ha un gran nombre d'observacions temporals i d'individus.

En conclusió, les *panel data* són un conjunt de dades obtingudes de diferents individus, que es recullen en intervals regulars de temps i s'ordenen cronològicament.

3.2. Panel data i heterogeneïtat

Modelar dades de panell es centra en abordar probablement dependència entre les observacions de dades del mateix grup. La principal diferència entre els models amb dades de panell i els de sèries temporals és que els de dades de panell permeten l'heterogeneïtat entre grups i a més a més, introdueixen efectes individuals, com veurem més endavant. Entenem per heterogeneïtat, que els grups són diferents entre ells, és a dir, que hi ha

variabilitat en les dades. El contrari a l'heterogeneïtat és l'homoscedasticitat, que significa que tots els estudis mostren el mateix efecte (Siebert, M., 2019).

És important destacar que hi ha tres tipus d'heterogeneïtat:

- Clínica: on es diferencia entre participants, intervencions o resultats
- Metodològica: on es diferencia en els dissenys d'estudi
- Estadística: on varien els efectes o els resultats dels anàlisis

Al llarg d'aquest treball, s'observarà l'heterogeneïtat estadística.

3.3. Models estadístics de Panel Data

El modelatge d'aquestes sèries de dades de panell és l'única branca del modelatge de sèries temporals composta per metodologies específiques per la seva estructura. Respecte a la modelització d'aquestes dades, referents a la *Covid-19*, s'exploraran tres models: el model *Pooled OLS*, *Fixed Effect Model* i *Random Effect Model*.

Pooled OLS

El model *Pooled OLS* (*Ordinary Least Squares*) o el model de mínims quadrats ordinaris també es conegut com a model agrupat o com a model d'homogeneïtat total. És un model de regressió lineal simple, el qual ignora l'estructura del panell. Aquest model és eficient quan els termes d'error del model són homocedàstics i no estan autocorrelacionats (E., 2020).

L'homoscedasticitat és una propietat desitjable dels errors d'un model de regressió simple, ja que ens permet fer models més fiables.

És el model més senzill de tots, formulat així:

$$y_{it} = \beta x_{it} + \delta z_i + \epsilon_{it}$$

En aquest model, s'entén per X la representació de les característiques observades i per Z les característiques no observades. A continuació, una breu descripció de les variables que componen el model:

- x_{it} són les característiques observables, poden ser constants per un individu al llarg del temps o poden variar al llarg de les observacions temporals
- z_i són les característiques no observables, responsables de la heterogeneïtat del model
- ϵ_{it} és el terme d'error estocàstic

Les suposicions del model *OLS* són gairebé les mateixes de model clàssic:

- Lineal en paràmetres
- $E[\epsilon_{it}|x_{it}] = 0; \forall i, t$
- $Var(\epsilon_{it}|x_{it}) = \sigma^2; \forall i, t \rightarrow$ Homoscedasticitat

- $Cov(\varepsilon_{it}, \varepsilon_{js} | x_{it}, x_{js}) = 0 ; \forall i = j, t \neq s ; \forall i \neq j, t = s ; \forall i \neq j, t \neq s \rightarrow$ No autocorrelació
- No multicol·linealitat

Aquest model també pot ser conegut com a “mínim quadrats ordinaris agrupats”, pel qual s’entén que no hi ha efectes individuals observables, és a dir, $\gamma\delta_{zi}$ és constant entre els individus. Això implica que les observacions de cada grup són independents entre si. I només en aquests casos, el model es converteix en:

$$y_{iy} = \beta x_{it} + \alpha + \varepsilon_{it}$$

Per tant, les dades de panell es poden tractar com a un gran conjunt de dades agrupades.

(One-Way) Fixed Effect Model

El Fixed Effect Model o el model d’efecte fix obté diverses variacions d’una secció transversal que poden ser degudes a les característiques del que estiguem fixant.

Aquest model inclou efectes no observables específics en el temps o específics de l’individu. S’assumeix que aquests efectes específics de l’individu estan correlacionats amb les característiques observades x_{it} . Aquest model es veu així:

$$y_{it} = \beta x_{it} + \alpha_i + \varepsilon_{it}$$

Podem observar el terme α_i , el qual a diferència del model anterior, aquest varia entre els individus, però és constant al llarg del temps (UPV, 2015).

(One-Way) Random Effects Model

En els anteriors models vistos, se suposa que α_i recull els efectes no observables de cada individu que suposem constants en el temps, però el model d’efectes aleatoris o *Random Effects Model* es caracteritza perquè α_i passa a formar part del terme error del model. És a dir, partint del model escrit en la seva forma escalar:

$$Y_{it} = x_{it}\beta + \alpha_i + \varepsilon_{it}$$

el model d’efectes aleatoris es formularia així,

$$Y_{it} = x_{it}\beta + u_{it}$$

on el terme d’error u_{it} és $u_{it} = \alpha_i + \varepsilon_{it}$.

Suposicions del model:

- El model és lineal en β

$$Y_{it} = x_{it}\beta + \alpha_i + \varepsilon_{it} \text{ on } \alpha_i + \varepsilon_{it} = u_{it}$$

$$x_{it} = [1 \ X_{2it} \ \dots \ X_{kit}]$$

$$\beta' = [\beta_1 \ \beta_2 \ \dots \ \beta_k]$$

- $E[\varepsilon_{it} | X_i, \alpha_i] = 0 ; \forall i, t \rightarrow$ Exogeneïtat estricta, pel que implica que ε_{it} no es correlaciona amb valors presents, passats o futurs dels regressors. Tampoc amb α_i .
- $Var(\varepsilon_{it} | X_i, \alpha_i) = \sigma^2 ; \forall i, t \rightarrow$ Homoscedasticitat
- $Cov(\varepsilon_{it}, \varepsilon_{js} | X_i, X_j, \alpha_i) = 0 ; \forall i = j, t \neq s ; \forall i \neq j, t = s ; \forall i \neq j, t \neq s \rightarrow$ No autocorrelació
- Condicional a X_i i α_i , s’assumeix $\varepsilon_{it} \sim i. i. d. N(0, \sigma^2)$

A més a més, totes les regressions de *panel data* que s'han vist fins ara, han assumit que els coeficients dels regressors són els mateixos en tots els individus. Però el model de coeficients aleatoris o *Random Effects Model* no té en compte aquesta suposició i introdueix efectes individuals específics a través del coeficient, per tant

$$y_{it} = \beta_i x_{it} + \alpha_i + \epsilon_i$$

$$y_{it} = (b_i + \beta) x_{it} + (\alpha_i + \alpha) + \epsilon_i$$

$$b_i \sim N(0, \tau^2_{i1})$$

$$\alpha_i \sim N(0, \tau^2_{i2})$$

Aquest model introdueix efectes de pendent individuals i permet l'heteroscedasticitat a través de la variació específica dels individus τ^2_{i1} i τ^2_{i2} (E., 2021).

Two-Way Individual Effects Models

Aquest últim model és un model d'efectes fixos bidireccional, és a dir, permet tant veure els efectes específics del temps com dels individus. Amb altres paraules, és el primer model que té en compte la dependència temporal.

La dependència temporal és una característica de les dades de sèrie temporal que indica que el passat afecta al futur (Pacheco, J., 2021).

Com els anteriors, a partir del model lineal simple

$$y_{it} = \alpha + \beta_{xit} + \epsilon_{it}$$

aquest model bidireccional es pot representar de la següent manera

$$y_{it} = \alpha + \beta_{xit} + \mu_i + \lambda_t + \epsilon_{it}$$

En aquest model, μ_i captura els efectes específics de l'individu no observables i en canvi, λ_t captura els efectes específics del temps no observables.

4. Estudi de la base de dades

4.1. Captura i obtenció de les dades

Base de dades

Es treballa amb dades sobre la Covid-19 a la Unió Europea, per tant, les dades recollides són de caràcter setmanal, més concretament des de la setmana 12 del 2020 fins a la setmana 12 del 2022, dos anys complerts i els nostres individus són els països de la Unió Europea: Àustria, Bèlgica, Bulgària, Croàcia, Xipre, República Txeca, Dinamarca, Estònia, Finlàndia, França, Alemanya, Grècia, Hongria, Islàndia, Irlanda, Itàlia, Letònia, Liechtenstein, Lituània, Luxemburg, Malta, Països Baixos, Noruega, Polònia, Portugal, Romania, Eslovàquia, Eslovènia, Espanya i Suècia.

Les dades han estat extretes de *European Centre for Disease Prevention and Control*, una pàgina oficial de la Unió Europea on han fet un recull de dades sobre la *Covid-19* a tot Europa (COVID-19, s. f.). Dins d'aquesta extensa pàgina, ens hem centrat en trobar variables de tipus setmanal, i per això ens hem basat en les següents bases de dades:

- **Vaccine:** amb totes les variables relacionades amb les vacunes, nombre de dosis, tipus de vacuna...
- **Weekcase:** amb les variables relacionades amb els nous casos positius setmanals, ja sigui per regió, país...
- **Hospital:** on es recull la informació de l'ocupació diària dels hospitals i de l'UCI, és a dir, la Unitat de Cures Intensives
- **Test:** amb la informació dels tests realitzats setmanalment i la taxa de positius
- **Deaths:** on es recullen les morts setmanals dels pacients positius en la *Covid-19*

Transformació de les dades i càlcul de noves variables

De totes elles, s'han realitzat diversos canvis per poder obtenir les variables adequades pel treball.

En primer lloc, per la base de dades *vaccine*, la variable referent a la setmana i l'any tenia un format tipus '2020-W15', mentre que d'altres, tenien un format tipus '2020-15', per tant, a l'hora de dur a terme el *merge*, podria donar problemes. D'aquesta variable, s'elimina el paràmetre *W*, utilitzant la funció *gsub* de l'*R*. També en aquesta base de dades, hi havia recollida tota la informació per a cada tipus de vacuna, informació la qual no ens interessa, ja que ens interessa explicar el total de vacunes subministrades. Per tant, es fa un filtratge i ens quedem amb el grup anomenat *ALL*, el qual fa el recull total de les vacunes. A més a més, d'aquesta base de dades, se'n crea tres noves variables anomenades *FirstDoseAcc*, *SecondDoseAcc* i *ThirdDoseAcc*, les quals corresponen al nombre de primera, segona i tercera dosi subministrades, respectivament, acumulades al llarg de les setmanes.

En segon lloc, per la base de dades *week_case*, la qual fa referència al nombre de casos per setmana, el canvi realitzat és referent al codi del país, ja que aquest estava format per 3 dígits, mentre que la majoria estaven formats per dos. Per exemple, pel país Àustria, el seu codi de país seria AT. Aquest canvi també s'ha realitzat per la base de dades *hospital*, on s'han extret les ocupacions dels hospitals i les unitats de cures intensives (UCI). A més a més, en aquesta darrera base de dades també s'ha canviat el format de data, essent inicialment del format tipus '2020-W15', i darrerament '2020-15'.

A continuació, per la base de dades *tests*, d'on s'extreu el nombre de tests realitzats, la taxa de tests realitzats i la taxa de població positiva en el virus, s'han filtrat les dades per grup nacional, ja que aquesta àmplia base de dades tenia informació a nivell de país i també a nivell regional de cada país. Per tant, ens quedem només amb aquella informació per setmana i del total del país.

Finalment, per la base de dades *deaths*, es realitzen de nou els dos canvis ja fets en anteriors bases de dades respecte al format de la data i al codi del país.

Un cop fets tots els canvis pertinents, es fa un filtratge per data, ja que s'agafen les dades d'entre les setmanes 12-2020 i 12-2022, es fa un *merge* i s'obté la base de dades final amb la qual es treballarà. En el cas de les dades sobre la *Covid-19*, es tracten de dades no balancejades, ja que es disposa de més dades d'uns països que d'uns altres, és a dir, hi ha dades mancants, les quals més endavant es tractaran.

A continuació, es fa una descripció sobre les variables de les quals es disposa a la base de dades final, amb la qual es treballarà al llarg d'aquest treball.

Descripció de les variables, anàlisi descriptiu de les dades

Taula 4.1. Descripció de les variables de la base de dades

Variable	Descripció
<i>country_code</i>	Codi del país
<i>year_week</i>	Any i setmana de l'any corresponent
<i>country</i>	País corresponent de la Unió Europea
<i>new_cases</i>	Nombre dels nous casos confirmats
<i>tests_done</i>	Nombre de tests fets
<i>testing_rate</i>	Taxa de tests per 100.000 habitants
<i>positivity_rate</i>	Positivitat de la prova setmanal (%): $100 \times \text{Nombre de nous casos confirmats} / \text{nombre de proves fet per setmana}$

<i>deaths</i>	Nombre total de les morts setmanals
<i>daily_hosp_occupancy</i>	Mitjana del nombre de pacients amb COVID-19 a l'hospital de la setmana corresponent
<i>daily_ICU_occupancy</i>	Mitjana del nombre de pacients amb COVID-19 a l'UCI (Unitat de Cures Intensives) de la setmana corresponent
<i>NumberDosesReceived</i>	Nombre de dosis de vacuna distribuïdes pels fabricants al país durant la setmana d'informe
<i>FirstDose</i>	Nombre de primeres dosis administrades per aquella setmana
<i>FirstDoseAcc</i>	Nombre de primeres dosis administrades acumulades
<i>SecondDose</i>	Nombre de segones dosis administrades per aquella setmana
<i>SecondDoseAcc</i>	Nombre de segones dosis administrades acumulades
<i>ThirdDose</i>	Nombre de terceres dosis administrades per aquella setmana
<i>ThirdDoseAcc</i>	Nombre de terceres dosis administrades acumulades
<i>PopData2020</i>	Població del país en l'any 2020

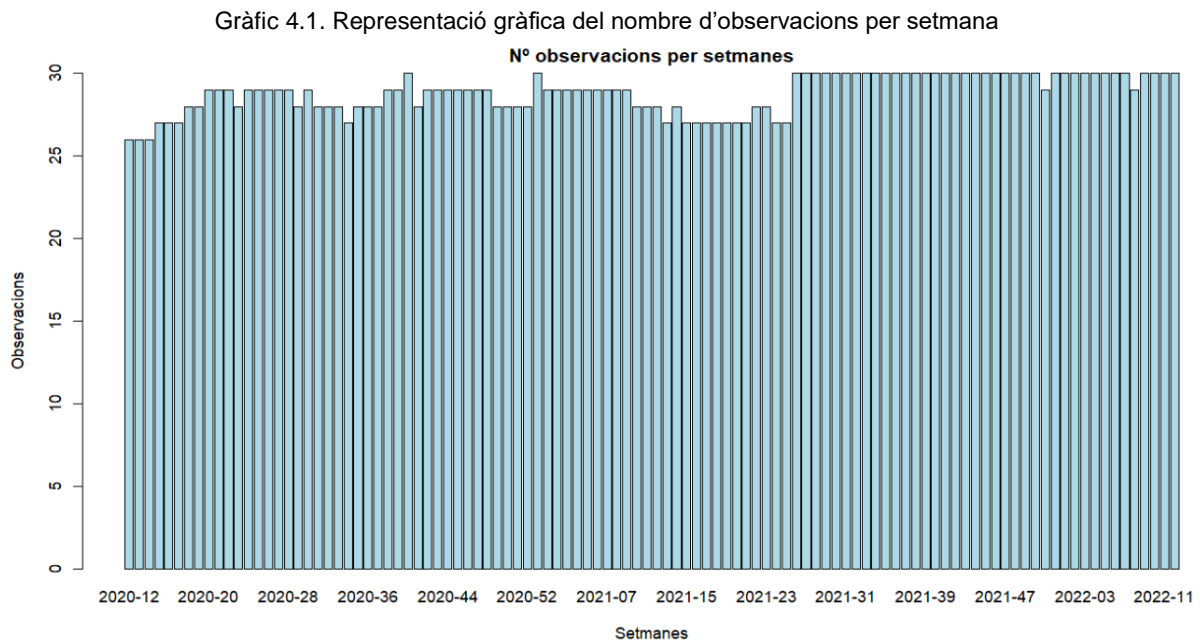
Com s'ha comentat anteriorment, les dades són no balancejades, ja que es disposa de més observacions d'uns països en concret que d'altres, es pot veure a continuació les observacions totals de la base de dades per país.

Taula 4.2. Nombre d'observacions a la base de dades per país

Àustria	93	Bèlgica	106	Bulgària	94	Croàcia	105	Xipre	105
R.Txeca	106	Dinamarca	106	Estònia	105	Finlàndia	106	França	106
Alemanya	106	Grècia	106	Hongria	77	Islàndia	106	Irlanda	106
Itàlia	91	Letònia	81	Liechtenstein	98	Lituània	104	Luxemburg	106
Malta	106	P. Baixos	106	Noruega	106	Polònia	106	Portugal	106
Romania	106	Eslovàquia	105	Eslovènia	106	Espanya	99	Suècia	96

Com s'està observant un total de 2 anys, és a dir, 106 setmanes, veiem que els països amb totes les observacions són Bèlgica, República Txeca, Dinamarca, Finlàndia, França, Alemanya, Grècia, Islàndia, Irlanda, Luxemburg, Malta, els Països Baixos, Noruega, Polònia, Portugal, Romania i Eslovènia. D'altra banda, el país amb menys observacions és Hongria, amb només 77 observacions.

S'observa també, a priori, una creixença del nombre d'observacions al llarg de les setmanes. Com es pot veure a continuació, a les darreres setmanes s'observa que hi ha el total de dades per països, és a dir, trenta observacions, mentre que al principi es veu que en la majoria no estan completes.



La font principal de les dades, extretes de *European Centre for Disease Prevention and Control* (COVID-19, s. f.), han estat enviades per *Members States to the European Surveillance System (TESSy)*, de forma setmanal. D'aquestes, un equip d'epidemiòlegs examinava totes les fonts rellevants dels països per donar lloc a la seva publicació. Per tant, les dades poden estar incompletes perquè certes dades no s'han pogut validar degudament.

Representacions gràfiques de les variables numèriques

En primer lloc, es realitzen diferents representacions gràfiques per totes les variables numèriques, de tal manera que es puguin detectar dades variants d'alguns països i possibles *outliers*. Un *outlier* és una dada anormal o extrema que pot afectar al tractament del conjunt de dades. Aquestes representacions gràfiques, les quals es podran trobar adjuntes a l'Annex, a primera vista no ens permeten identificar quins han estat els països més afectats per la pandèmia, ja que cada país té una població diferent i depenent d'això les dades es poden veure afectades.

És per això, que de la mateixa base de dades de la Covid-19, es fa una duplicació, estandarditzant les variables segons el nombre d'habitants, és a dir, la població de cada país, per tal que es pugui observar gràficament les principals diferències entre països i quins han estat més afectats. Aquesta nova base de dades s'anomenarà *covid_pop*. Per tant, totes les representacions gràfiques es realitzaran amb aquestes noves dades.

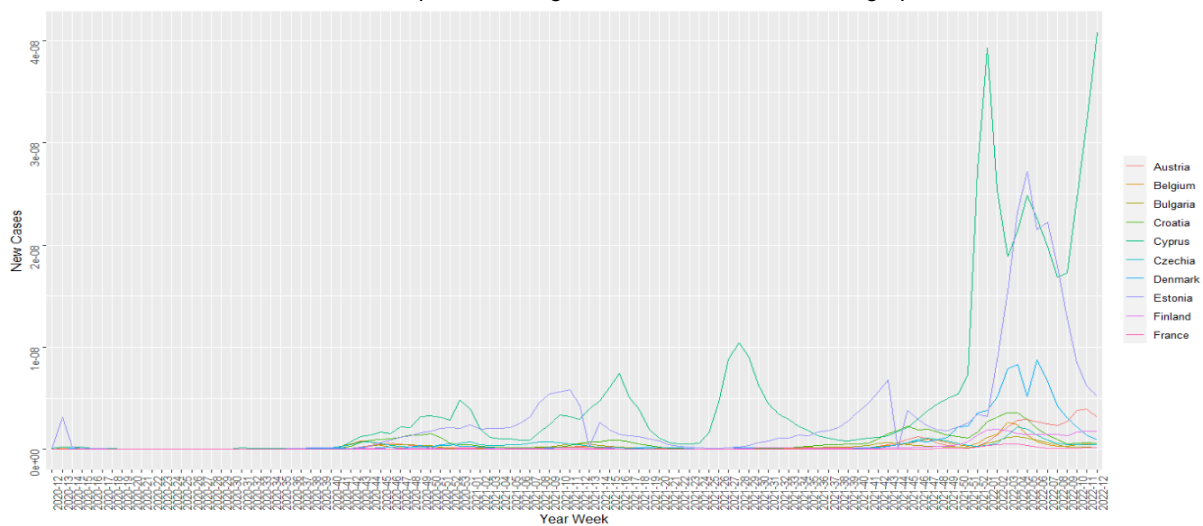
I ara, es realitzen de nou totes les representacions gràfiques corresponents. Aquestes, han estat representades amb grups de 10 països, és a dir, en 3 grups, per poder veure més fàcilment la informació rellevant. Respecte a la selecció dels grups, no s'ha seguit cap criteri, simplement és per fer les representacions gràfiques més visuals i d'aquesta manera, poder veure més clarament els resultats. Els grups han estat distribuïts de la següent manera:

- Grup 1: Àustria, Bèlgica, Bulgària, Croàcia, Xipre, República Txeca, Dinamarca, Estònia, Finlàndia i França
- Grup 2: Alemanya, Grècia, Hongria, Islàndia, Irlanda, Itàlia, Letònia, Liechtenstein, Luxemburg i Lituània.
- Grup 3: Malta, Països Baixos, Noruega, Polònia, Portugal, Romania, Eslovàquia, Eslovènia, Espanya i Suècia.

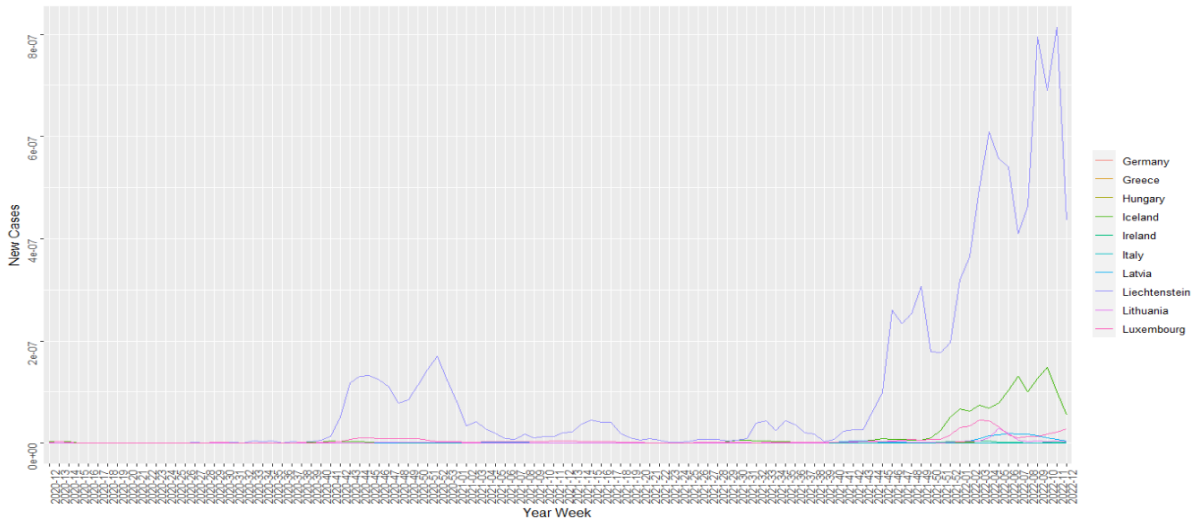
A continuació, s'observen les representacions gràfiques al llarg del temps per variables estandarditzades per nombre d'habitants,

New cases

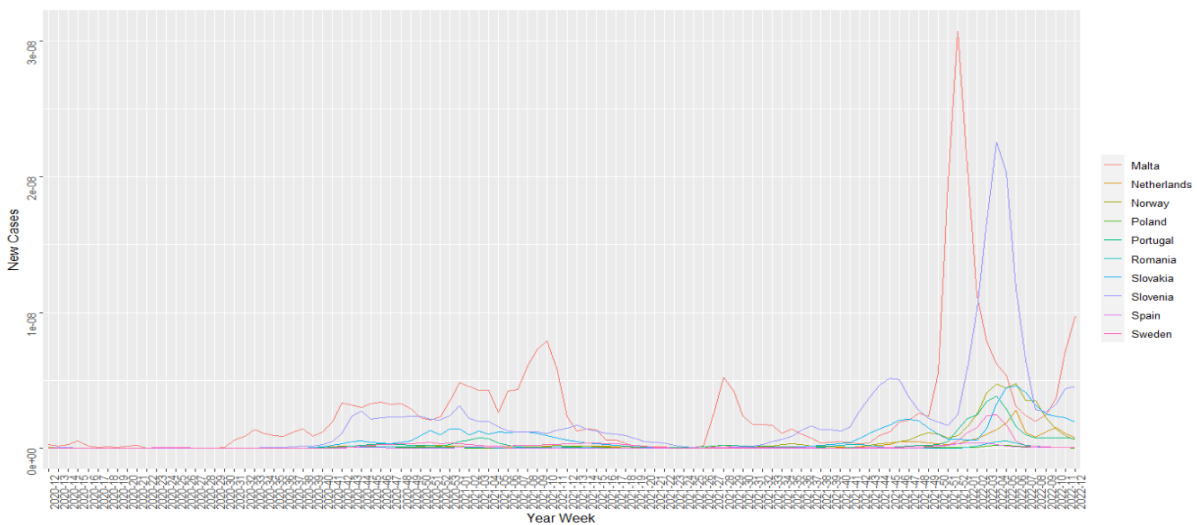
Gràfic 4.2. Representació gràfica dels nous casos en el grup 1



Gràfic 4.3. Representació gràfica dels nous casos en el grup 2



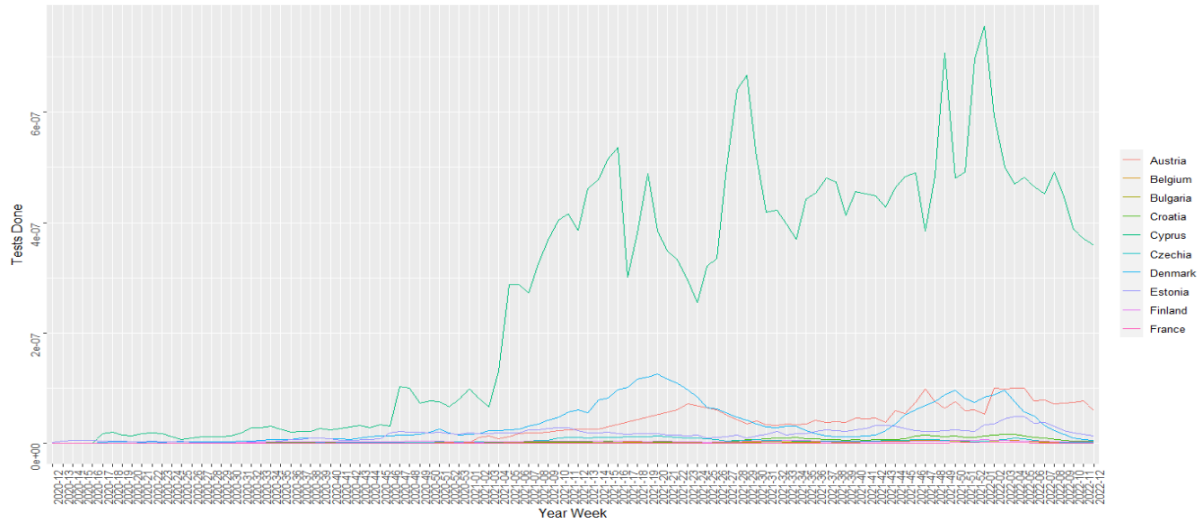
Gràfic 4.4. Representació gràfica dels nous casos en el grup 3



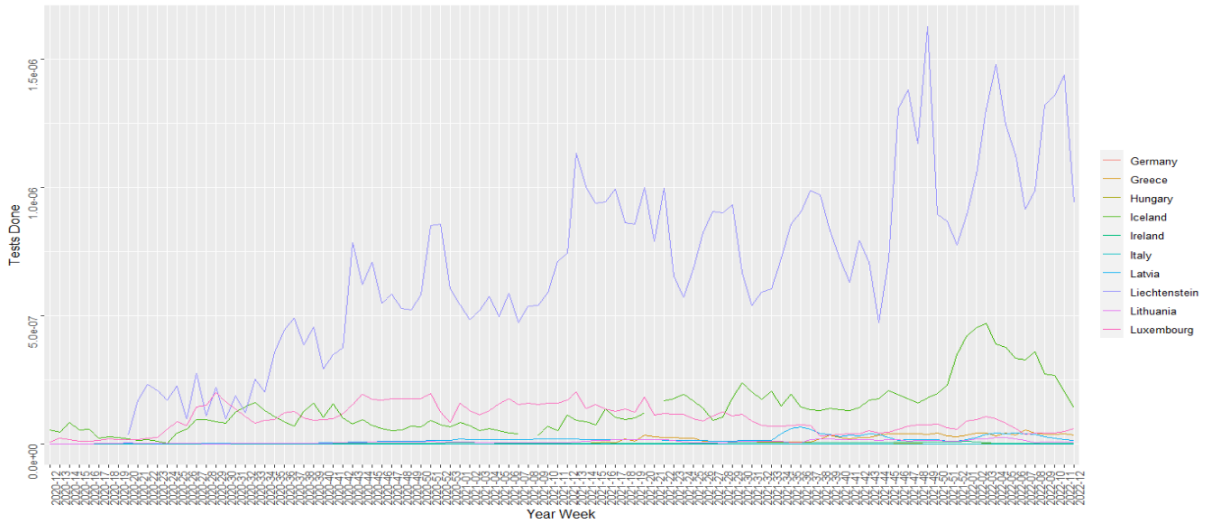
S'observa a primera vista que la variable dels nous casos és variant en el temps, i aquests, augmenten a mesura que passa el temps. Es poden observar diverses onades pels països amb més afectacions, com ara Xipre, Estònia, Liechtenstein, Malta i Eslovènia. D'aquestes, es veu clarament que les puntes on va haver més casos nous són les darreres setmanes.

Tests done

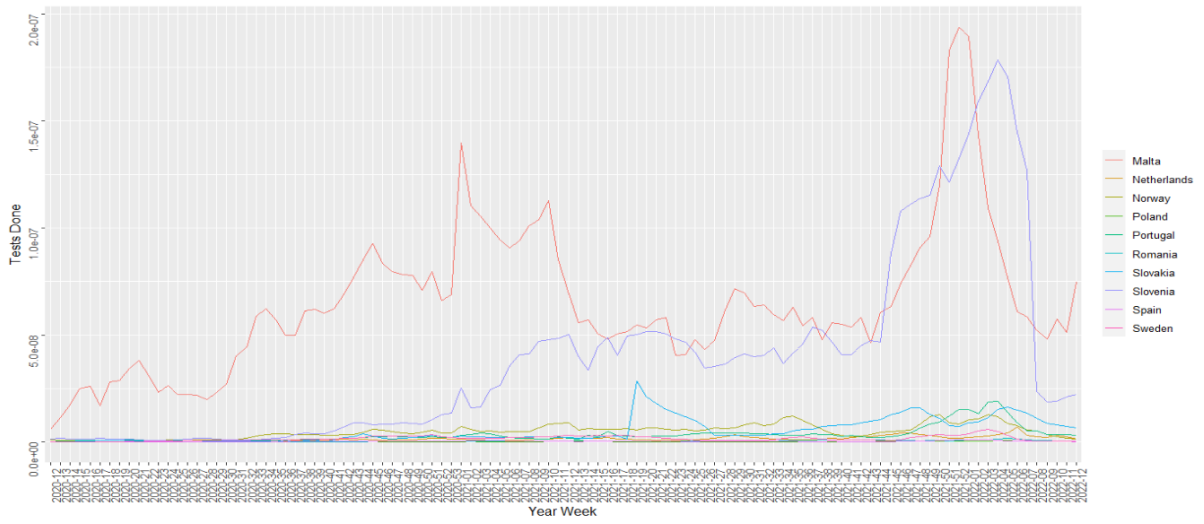
Gràfic 4.5. Representació gràfica dels tests realitzats en el grup 1



Gràfic 4.6. Representació gràfica dels tests realitzats en el grup 2



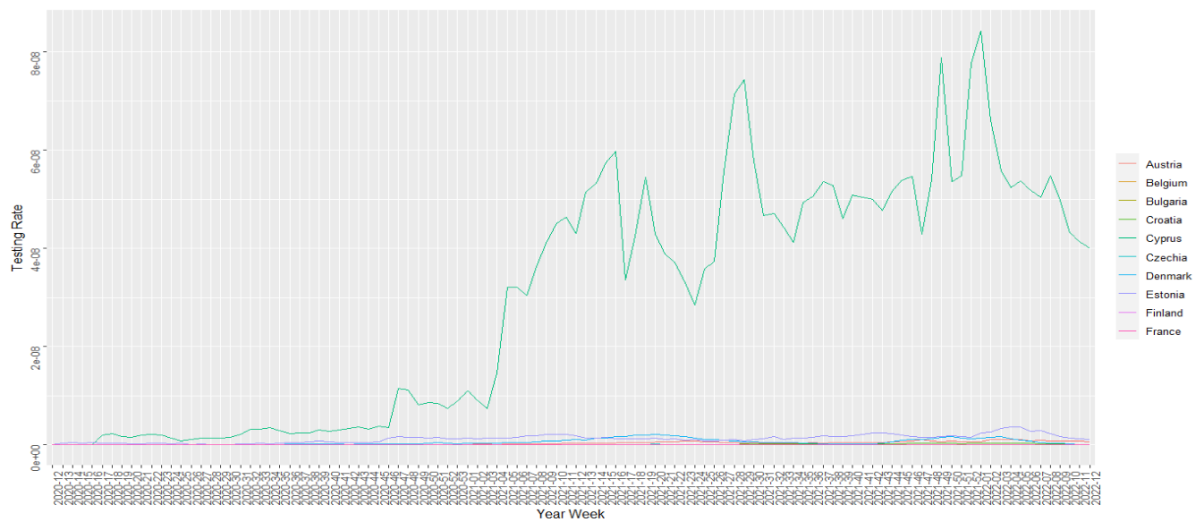
Gràfic 4.7. Representació gràfica dels tests realitzats en el grup 3



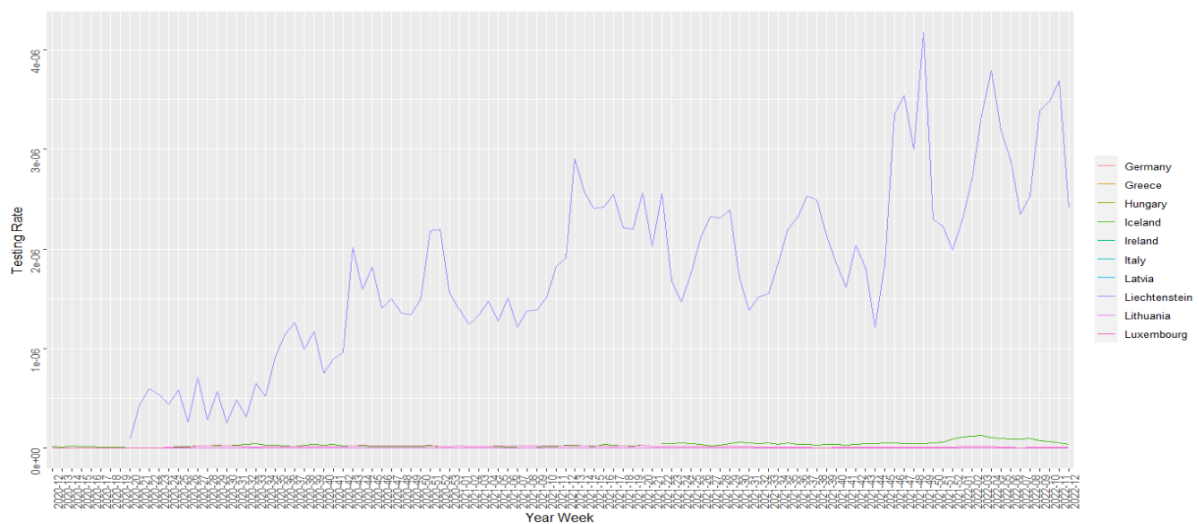
Respecte als tests realitzats per aquests països, es veu que és una variable bastant variant respecte al país. I és que, destaquen amb diferència Xipre, Liechtenstein, Malta i Eslovènia. És a dir, aquells països que més nous casos tenien, per tant, veiem una relació clara entre tests realitzats i nous casos, gràficament.

Testing rate

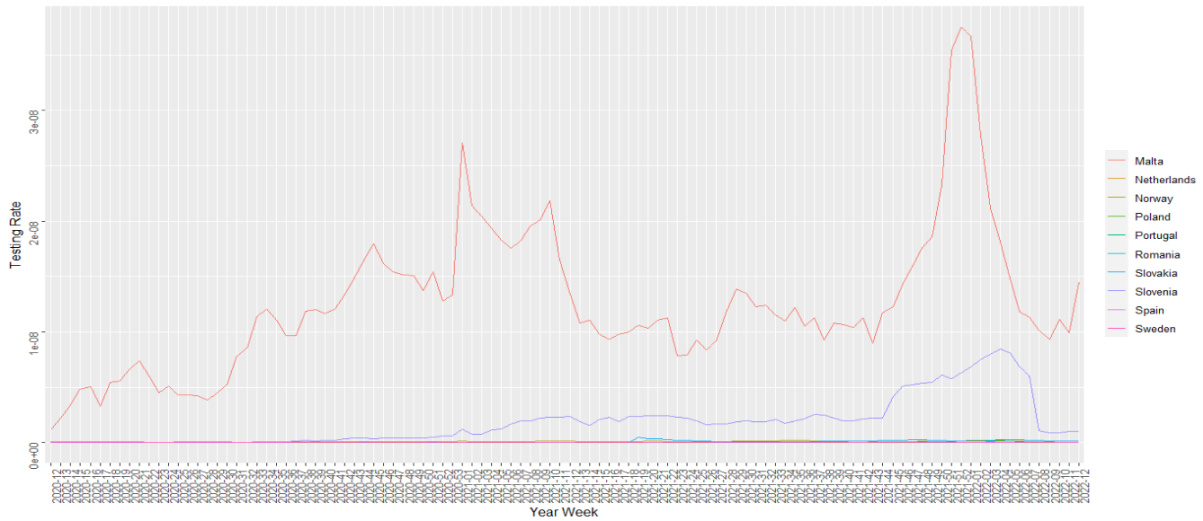
Gràfic 4.8. Representació gràfica del rati de tests realitzats en el grup 1



Gràfic 4.9. Representació gràfica del rati de tests realitzats en el grup 2



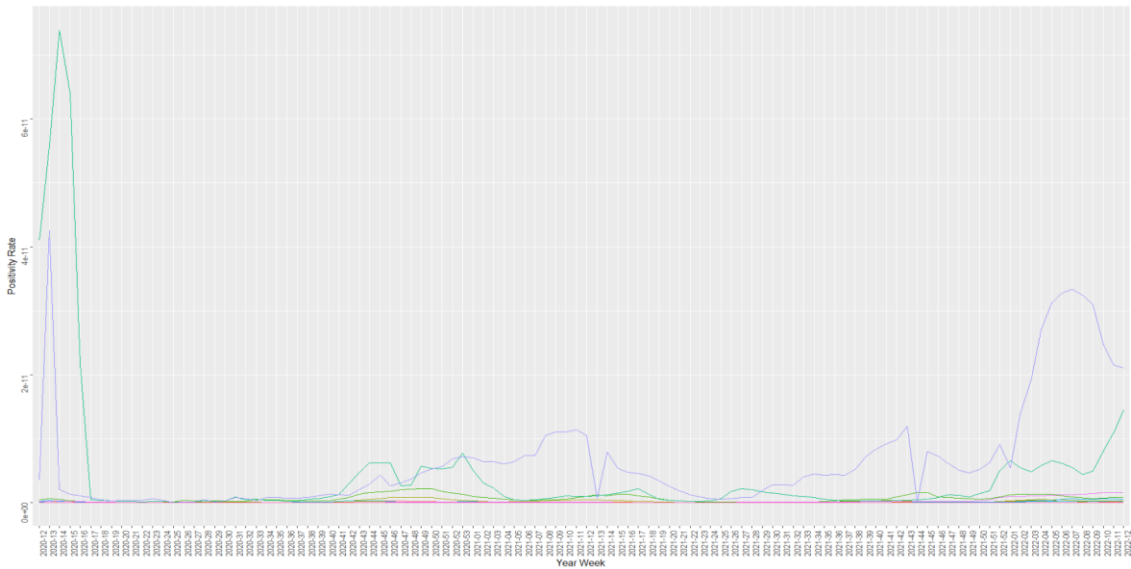
Gràfic 4.10. Representació gràfica del rati de tests realitzats en el grup 3



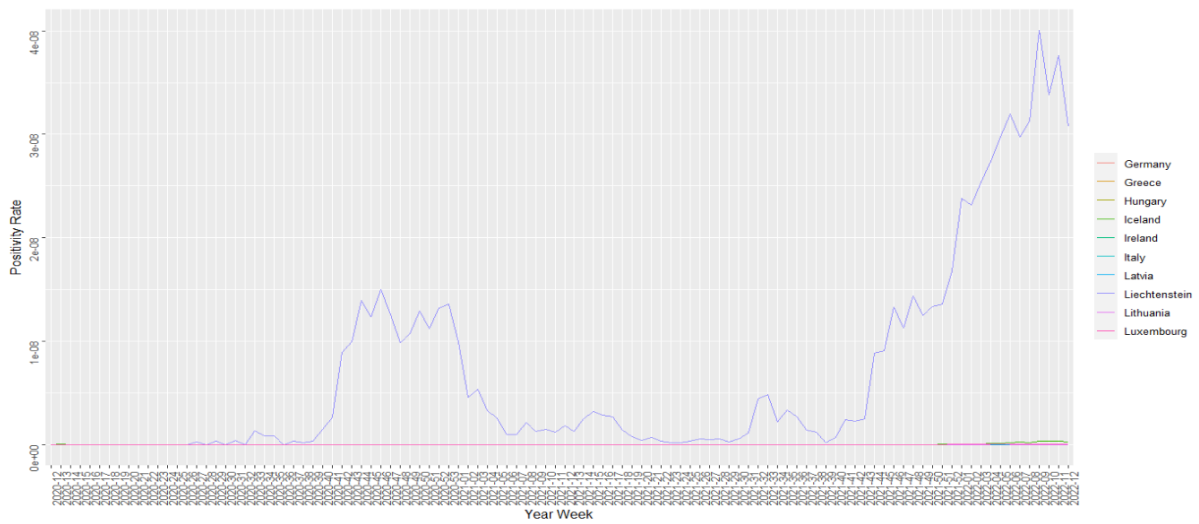
Relacionat amb l'anterior variable, es veu el rati d'aquests tests i destaquen els mateixos països, Xipre, Liechtenstein, Malta i Eslovènia.

Positivity rate

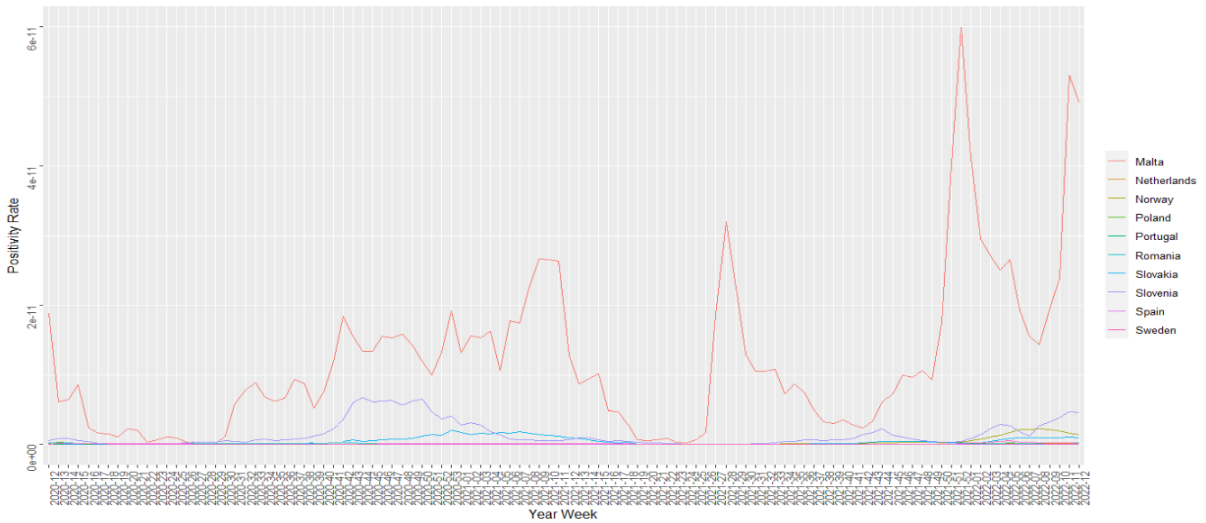
Gràfic 4.11. Representació gràfica del rati de positivitat en el grup 1



Gràfic 4.12. Representació gràfica del rati de positivitat en el grup 2



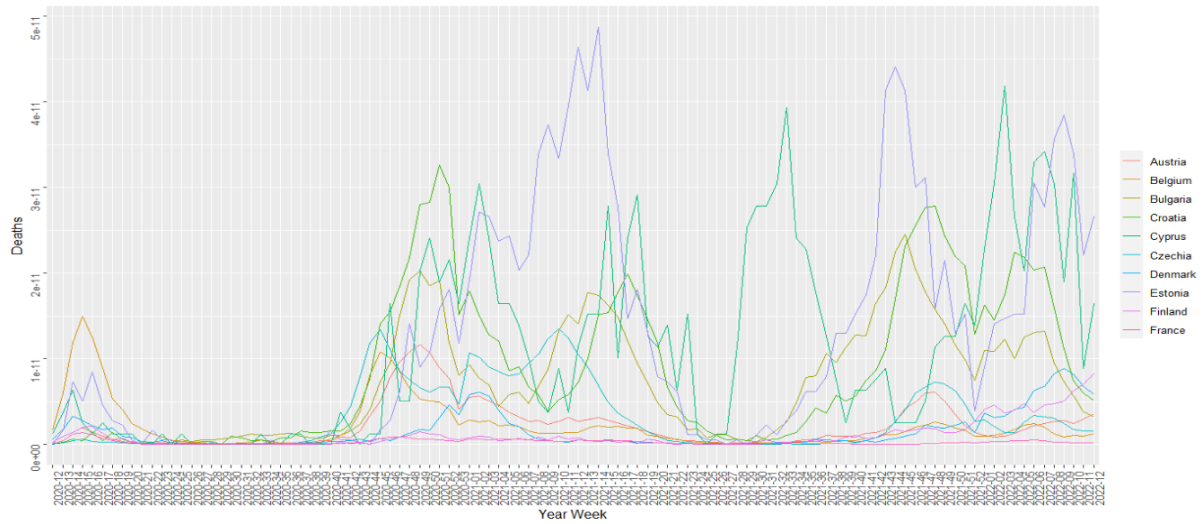
Gràfic 4.13. Representació gràfica del rati de positivitat en el grup 3



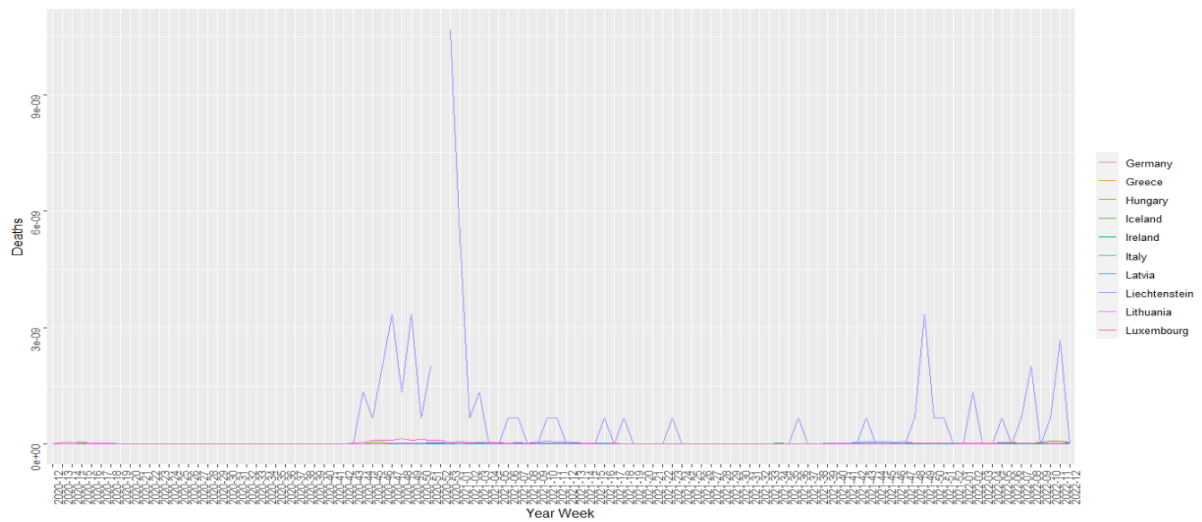
Respecte al rati de positivitat, encara es veu que els mateixos països destaquen, ja que Malta i Liechtenstein continuen destacant per la majoria. Malgrat això, s'observa en les primeres setmanes un gran pic respecte a Xipre, el qual es manté constant després i també s'observa un comportament semblant per Estònia.

Deaths

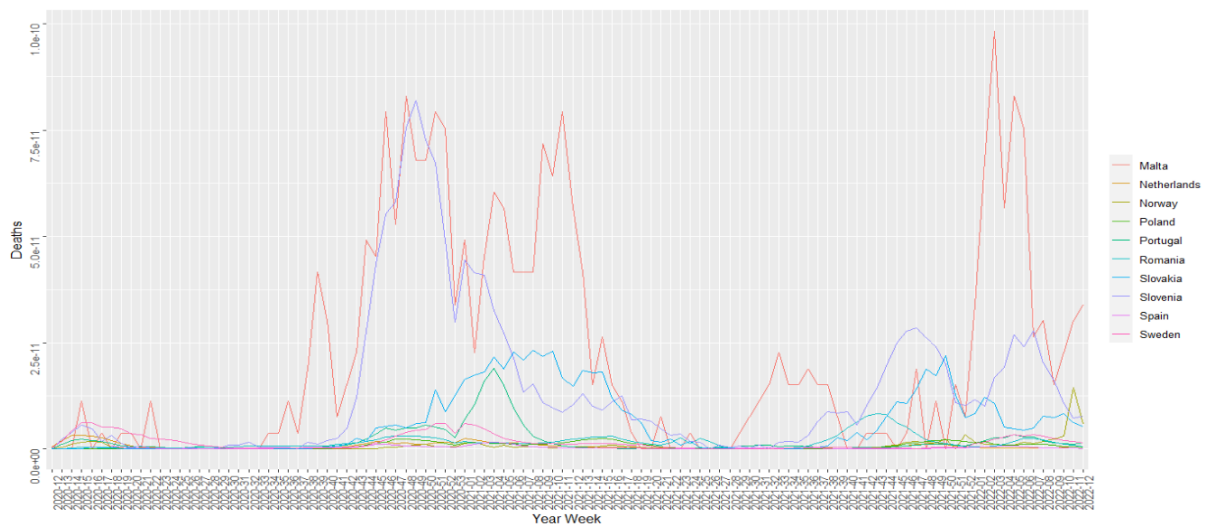
Gràfic 4.14. Representació gràfica de les morts en el grup 1



Gràfic 4.15. Representació gràfica de les morts en el grup 2



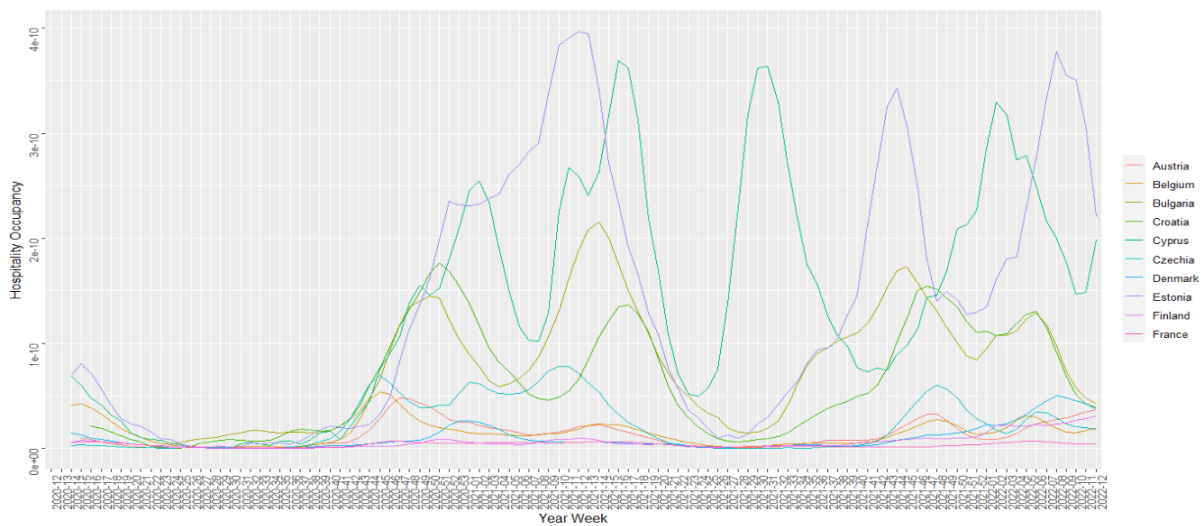
Gràfic 4.16. Representació gràfica de les morts en el grup 3



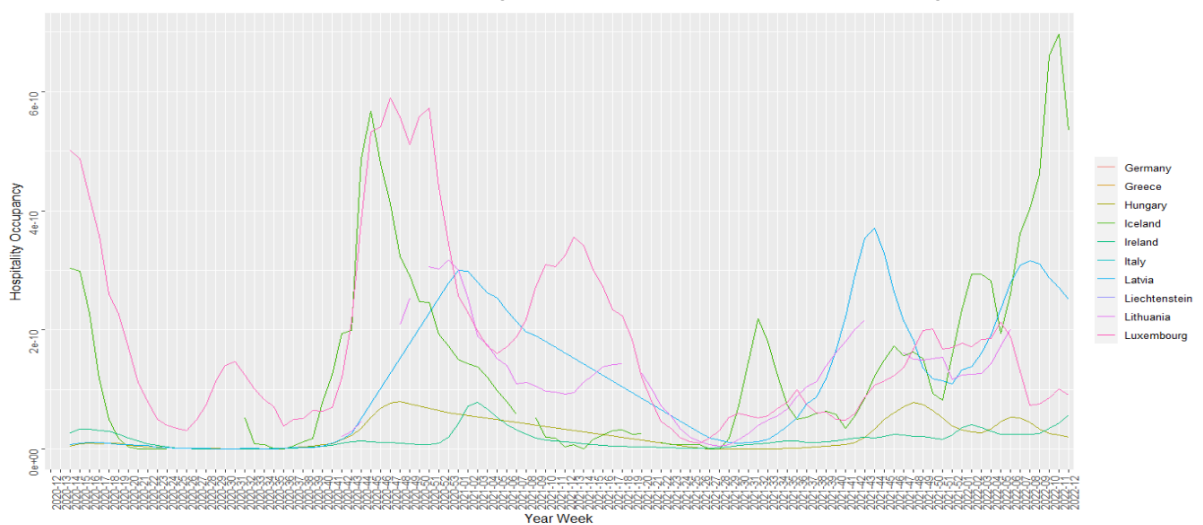
Les morts és una de les variables més observades i analitzades al llarg d'aquesta pandèmia i és que, aquest virus malauradament, n'ha comportat moltes. En aquestes representacions gràfiques es veu més varietat en les dades que del que hem vist fins ara. S'observen més diferències, i es podrien destacar els països que han tingut més morts per *Coronavirus*, com ara Croàcia, Xipre, Estònia, Bulgària, Liechtenstein, Malta i Eslovènia. Recordar que aquestes dades són totes proporcionals a la variable població, per tant, encara que en alguns països els nombres de morts poden ser superiors que a alguns dels països darrerament esmentats, aquests, han sigut més afectats proporcionalment a la població.

Hospitality occupancy

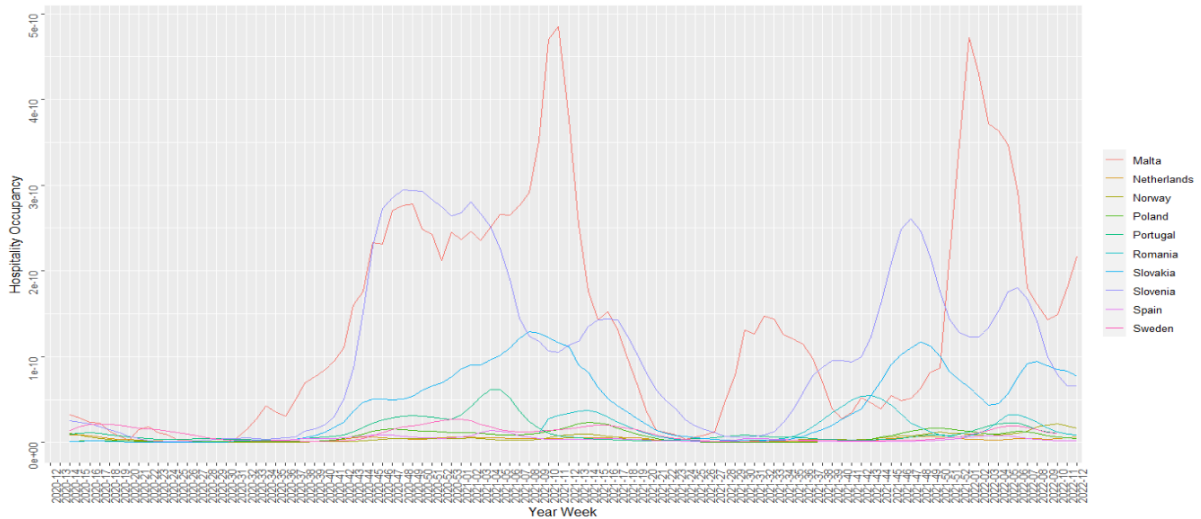
Gràfic 4.17. Representació gràfica de l'ocupació en els hospitals del grup 1



Gràfic 4.18. Representació gràfica de l'ocupació en els hospitals del grup 2



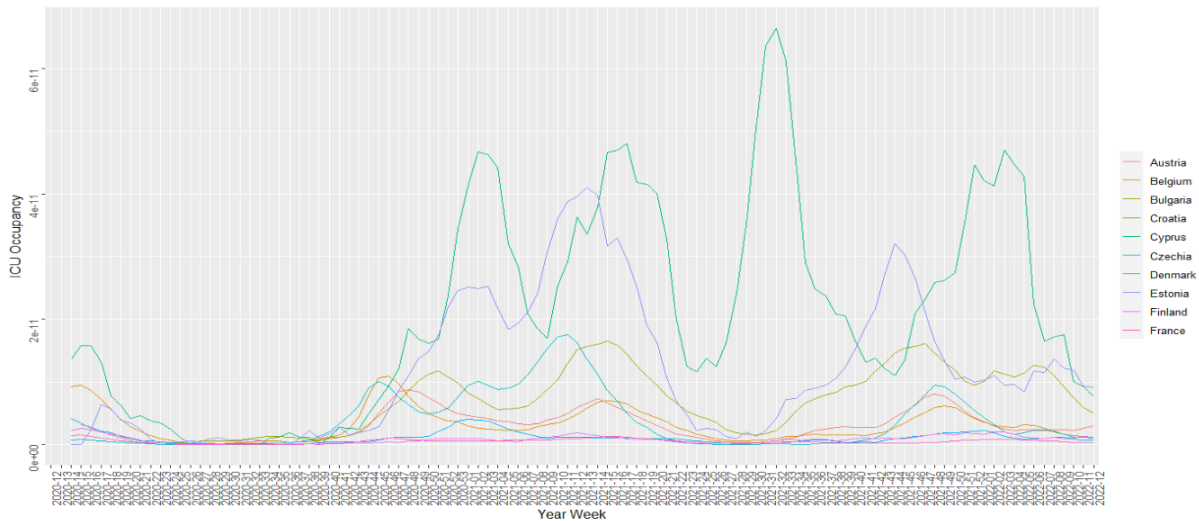
Gràfic 4.19. Representació gràfica de l'ocupació en els hospitals del grup 3



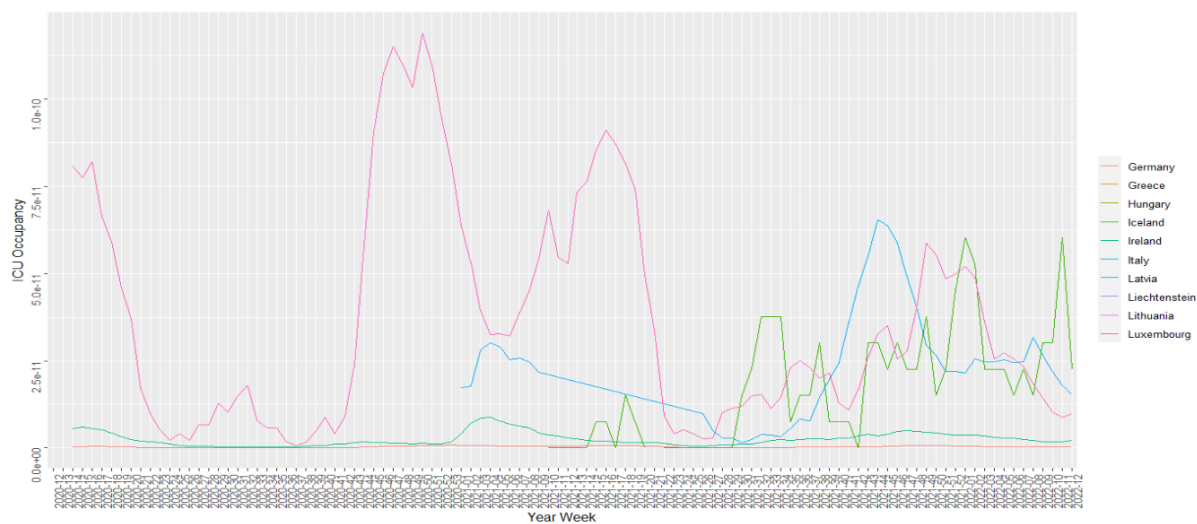
Respecte a l'ocupació dels hospitals, aquesta variable també presenta més varietat en el temps, com es pot observar en les tres representacions gràfiques. Els països amb major ocupació als hospitals setmanalment han estat Estònia, Xipre, Croàcia, Bulgària, Lituània, Hongria, Malta i Eslovènia.

ICU occupancy

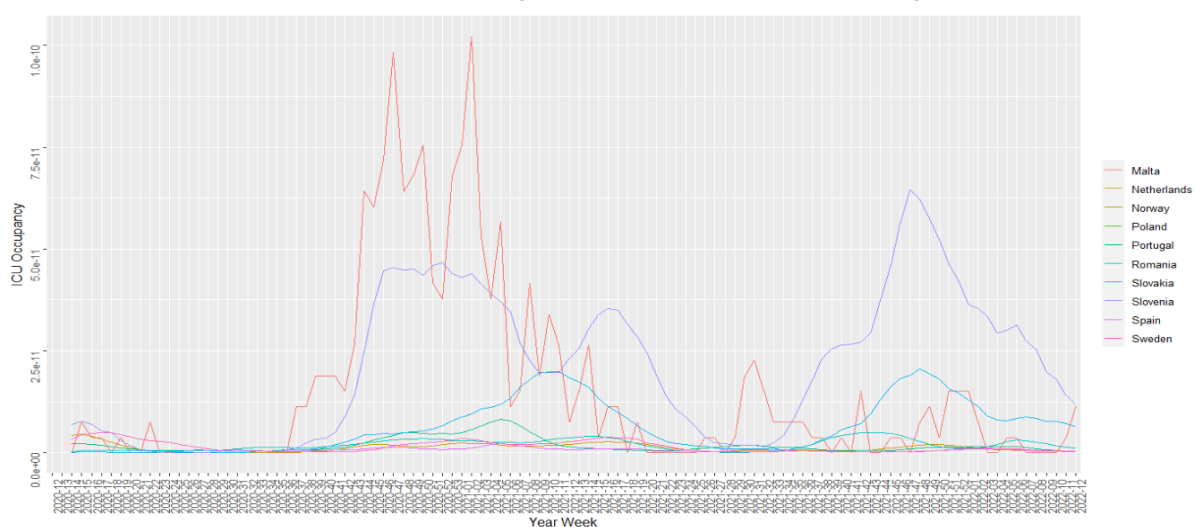
Gràfic 4.20. Representació gràfica de l'ocupació en les UCIs del grup 1



Gràfic 4.21. Representació gràfica de l'ocupació en les UCIs del grup 2



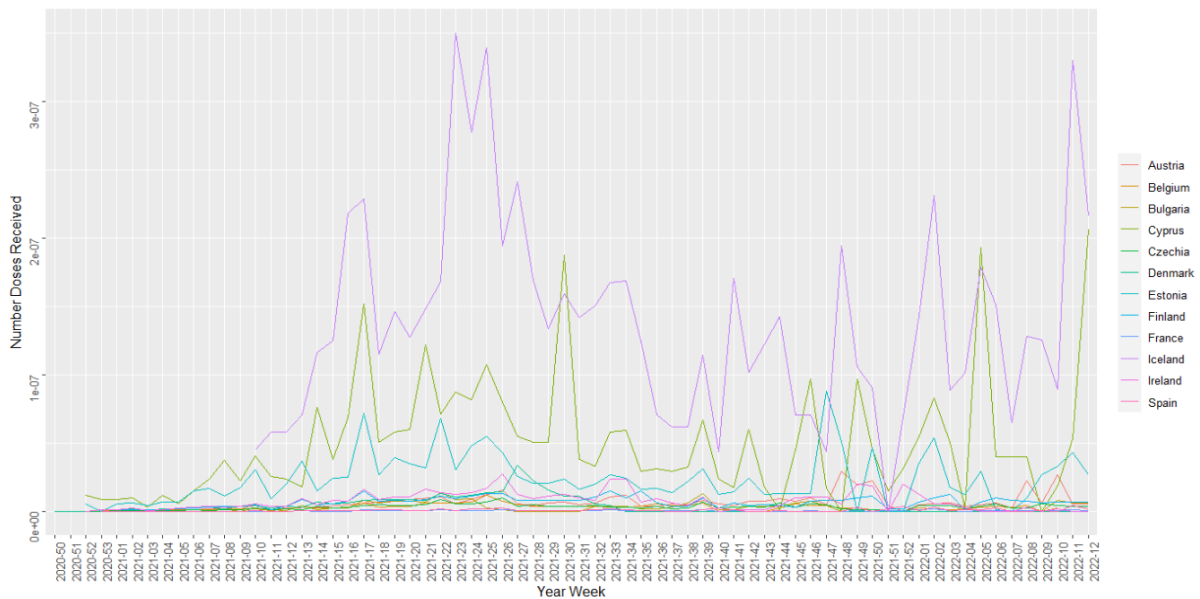
Gràfic 4.22. Representació gràfica de l'ocupació en les UCIs del grup 3



Relacionat amb l'anterior variable, els països com Xipre, Estònia, Lituània, Malta i Eslovènia, han tingut una taxa d'hospitalització a les unitats de cures intensives (UCI) major.

Number Doses Received

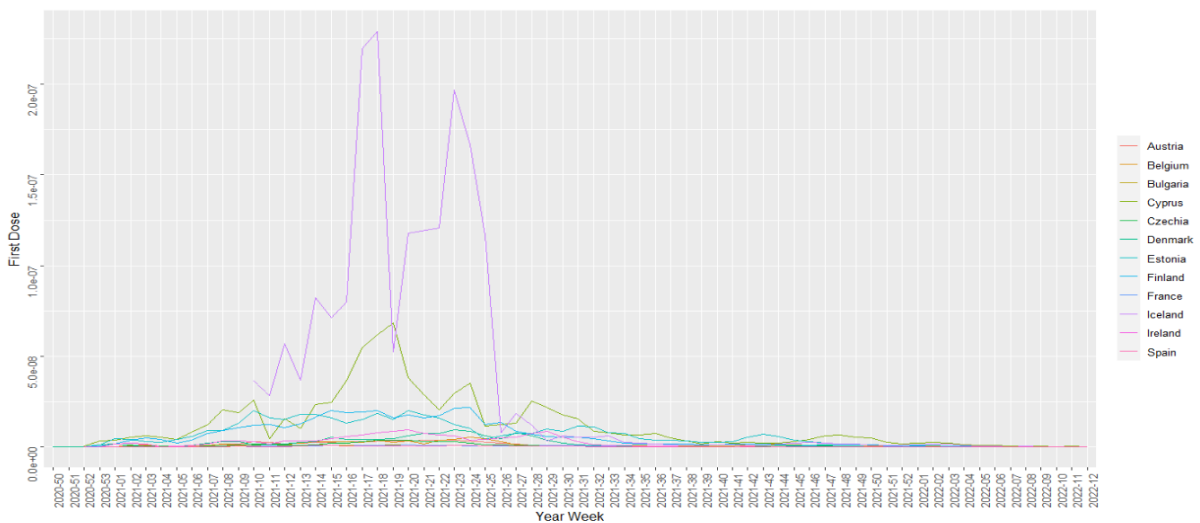
Gràfic 4.23. Representació gràfica del nombre de dosis rebudes per país



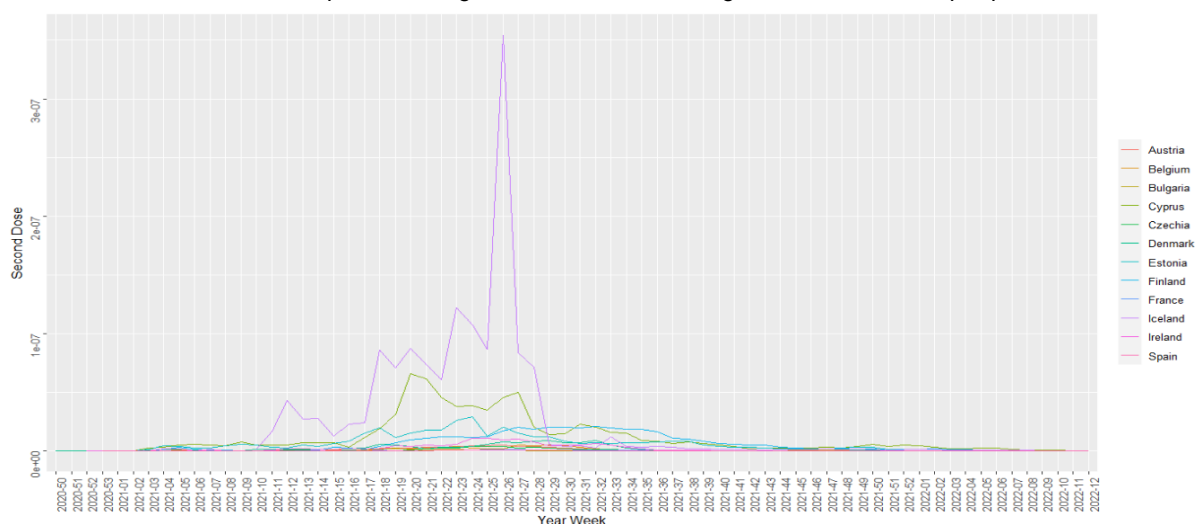
Aquesta variable, *Number Doses Recieved*, representa el nombre de dosis de vacuna distribuïdes pels fabricants al país durant la setmana d'informe. D'aquesta variable, falta la informació de la majoria dels països i a la representació gràfica es pot veure que Islàndia destaca amb majoria, seguit de Xipre i Estònia.

First Dose, Second Dose i Third Dose

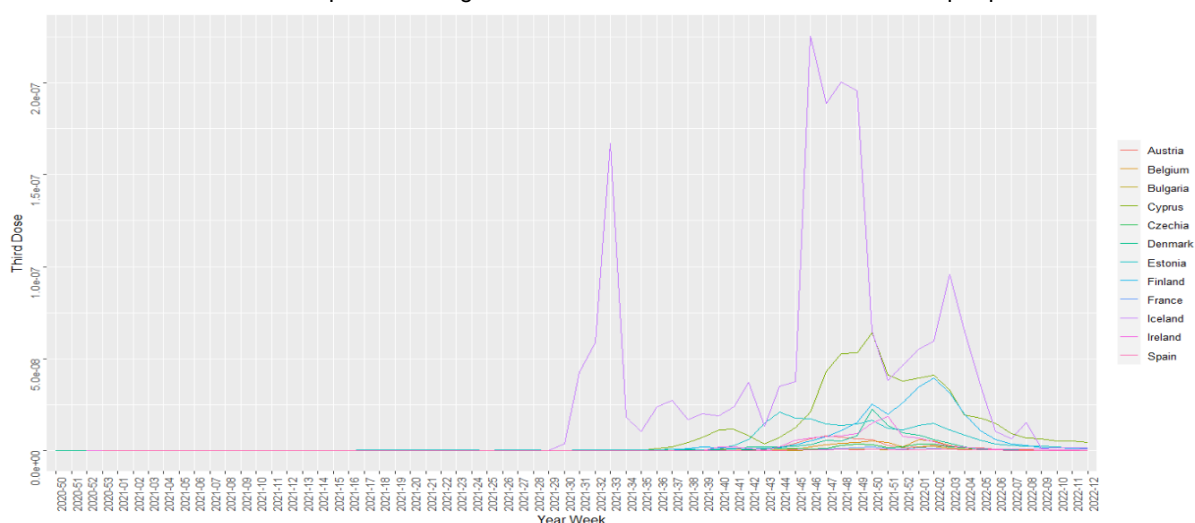
Gràfic 4.24. Representació gràfica del nombre de primeres dosis rebudes per país



Gràfic 4.25. Representació gràfica del nombre de segones dosis rebudes per país



Gràfic 4.26. Representació gràfica del nombre de terceres dosis rebudes per país



Respecte a la primera, segona i tercera dosi de vacunes, representades respectivament, només disposem d'informació dels països: Àustria, Bèlgica, Bulgària, la República Txeca, Xipre, Dinamarca, Estònia, Finlàndia, França, Islàndia, Irlanda i Espanya. Aquests gràfics comencen a la setmana 50 del 2020, ja que abans no hi havia vacunes. En els tres gràfics destaca Islàndia com al país que més vacunes ha rebut, en proporció a la població, seguit de Xipre.

Com s'ha comentat a l'inici de l'anàlisi, les dades estan estandarditzades per la població, és a dir, pel nombre d'habitants de cada país. És aquest el motiu pel qual es veu com països més petits tendeixen a tenir valors més alts en el gràfic. Conseqüentment, això indica que aquests països, a escala poblacional, han estat els més afectats, ja per raons socials o econòmiques o per capacitat de recursos.

Tractament de les dades mancants

Com s'ha comentat anteriorment, la base de dades es tracta d'una base de dades no balancejada, ja que tenim dades mancants i per conseqüència, no tenim les mateixes variables de tots els països. Això es veu reflectit sobretot a les variables de les vacunes, ja que les dades mancants abunden més.

Existeixen multitud de procediments per aplicar quan tenim valors *missings*. En el cas d'aquestes dades, ens plantegem dues aproximacions possibles:

- **Eliminar** les mostres o variables de les dades mancants
- **Imputar** aquests valors, és a dir, substituir-los per estimacions

Respecte a l'eliminació de les mostres, aquest procediment consisteix en eliminar les mostres no completades, és a dir, descartar totes aquelles que tenen valors mancants. Això ens pot perjudicar ja que al haver moltes dades mancants, sobretot les vacunes, es podria reduir considerablement el nombre d'observacions.

D'altra banda, la imputació de valors és una tècnica d'estimació del valor faltant. Hi ha diverses maneres de substituir aquests valors, ja sigui per la mitjana, la moda, els K-veïns més pròxims, etc.. S'ha de tenir en compte que es tracta d'una estimació i per tant, no es pot abusar ja que es podria estar tractant amb dades errònies.

Per tant, com en el nostre cas les dades faltants, és a dir, els *missings*, es troben bastant concentrats en les variables que fan referència a les vacunes, optarem per utilitzar el mètode de l'eliminació d'aquests en els models. En conclusió, en tots els models, representacions gràfiques o anàlisis que estiguem utilitzant informació referent a les vacunes, d'aquestes, s'eliminaran totes les files on no hi hagi informació. Es continuarà tenint una base de dades no balancejada, però els models emprats respecten aquest tipus de base de dades.

5. Modelització

S'inicia el capítol de la modelització recordant l'objectiu principal d'aquest anàlisi de dades, el qual és capturar l'heterogeneïtat no observable entre països. Aquesta heterogeneïtat no es pot detectar amb estudis de tall transversal ni tampoc amb sèries temporals.

Utilitzant aquesta tècnica amb les dades de panell, permet realitzar un anàlisi més dinàmic a l'incorporar la dimensió temporal en les dades, pel que enriqueix l'estudi, particularment en períodes de grans canvis (Wagle. M, 2022).

5.1. OLS i Pooled OLS

El primer model que s'aplica a l'*R* és el model **OLS** o el model de mínim quadrats ordinaris amb el qual es fa servir el paquet *lm*. Per aquest, en primer lloc s'intenta explicar les morts, és a dir, es té *deaths* com a variable dependent, i donat que el model, ignora l'estructura del panell, s'utilitzen les dades de les morts originals, és a dir, la base de dades *covid*.

Taula 5.1. Resultats del model OLS de l'*R*

OLS				
	deaths			
Predictors	Estimates	std. Error	CI	p
(Intercept)	50.30480	7.89577	34.81980 – 65.78981	<0.001
tests done	-0.00001	0.00001	-0.00002 – 0.00001	0.508
testing rate	0.00023	0.00065	-0.00104 – 0.00150	0.720
positivity rate	3.66684	0.47421	2.73683 – 4.59684	<0.001
new cases	-0.00050	0.00005	-0.00059 – -0.00040	<0.001
daily hosp occupancy	0.07763	0.00304	0.07167 – 0.08360	<0.001
daily ICU occupancy	0.30162	0.01924	0.26389 – 0.33935	<0.001
popData2020	-0.00001	0.00000	-0.00001 – -0.00001	<0.001
Observations	1964			
R ² / R ² adjusted	0.856 / 0.856			

El model explica la variable *deaths* així:

$$\begin{aligned}
 \text{deaths} = & 50.3 - 0.00001 * \text{tests done} + 0.00023 * \text{testing rate} + 0.366 * \text{positivity rate} \\
 & - 0.0005 * \text{new cases} + 0.07763 * \text{daily hosp occupancy} + 0.301 * \text{daily ICU occupancy} \\
 & - 0.00001 * \text{popData2020}
 \end{aligned}$$

El model mostra que, a part dels *tests_done* i el *testing_rate*, totes les altres variables són significatives i que el mateix model és significatiu, ja que els *p-valors* són més petits que 0.05. Com s'ha comentat anteriorment, aquest model ignora l'estructura de les dades, per tant, es considera que el terme independent és el mateix per tots els períodes de temps i per tots els països.

Com s'ha observat, en el model *OLS* s'ha afegit la variable *popData2020*, és a dir, la població per país del 2020 com a variable explicativa, i com s'ha mencionat anteriorment, es fa una segona base de dades amb totes les variables proporcionals a la població, anomenada *covid_pop*.

Per tant, es realitza aquest mateix model però amb la base de dades *covid_pop* i aquest cop, sense afegir la població com a variable explicativa. En aquest model, totes les variables, tant la dependent com les explicatives, són proporcionals a la població del país.

Taula 5.2. Model OLS population

OLS Population				
	deaths			
<i>Predictors</i>	<i>Estimates</i>	<i>std. Error</i>	<i>CI</i>	<i>p</i>
(Intercept)	-0.00000	0.00000	-0.00000 – -0.00000	0.042
tests done	-0.00002	0.00000	-0.00003 – -0.00001	<0.001
testing rate	-0.00005	0.00003	-0.00011 – 0.00001	0.121
positivity rate	0.04251	0.01729	0.00861 – 0.07641	0.014
new cases	-0.00013	0.00005	-0.00022 – -0.00004	0.005
daily hosp occupancy	0.12081	0.00317	0.11459 – 0.12702	<0.001
daily ICU occupancy	0.18122	0.01807	0.14577 – 0.21667	<0.001
Observations	1964			
R ² / R ² adjusted	0.799 / 0.798			

El model **OLS Population** queda explicat així:

$$deaths = -0.00002 * tests\ done - 0.00005 * testing\ rate + 0.00425 * positivity\ rate - 0.00013 * new\ cases + 0.1208 * daily\ hosp\ occupancy + 0.1812 * daily\ ICU\ occupancy$$

Com es pot observar, hi ha canvis respecte a l'anterior, ja que en aquest, tot és significatiu excepte les variables *testing_rate*, la qual coincideix amb el primer model i *positivity_rate*, la qual a l'anterior model si que era significatiu. De totes maneres, es veu en l'*R* que el *p-valor* del model és menor que 0.05 i per tant, aquest segon model també és significatiu.

A priori sembla estrany que les dades referents als pacients positius i als tests realitzats no siguin significatives, però el paquet d'*R* utilitzat per aquests dos models és el *lm*, el qual no respecta les dades no balancejades. És per això, que en la resta de models que es mostrarà a continuació, es farà servir el paquet *plm*, el qual sí que respecta les dades no balancejades (Panel data econometrics in R, s.f.).

Un cop fet aquests dos models, es repeteixen utilitzant el paquet *plm*, el qual s'utilitza especialment per l'anàlisi de dades de panell. En aquest paquet, s'indica que es vol utilitzar el model *pooling* i també s'indexen les variables *country* i *year_week*.

El model *pooling* és el que s'utilitza per fer regressions lineals amb *panel data*, i les variables indexores són els països i les setmanes de l'any, és a dir, aquelles variables que creuem en el *panel data*.

Per tant, en primer lloc es repeteix el model on s'afegeix la població com a variable explicativa però utilitzant el paquet *plm*.

Taula 5.3. Model Pooled OLS

Pooled OLS				
	Dependent variable			
<i>Predictors</i>	<i>Estimates</i>	<i>std. Error</i>	<i>CI</i>	<i>p</i>
(Intercept)	50.30480	7.89577	34.81980 – 65.78981	<0.001
tests_done	-0.00001	0.00001	-0.00002 – 0.00001	0.508
testing_rate	0.00023	0.00065	-0.00104 – 0.00150	0.720
positivity_rate	3.66684	0.47421	2.73683 – 4.59684	<0.001
new_cases	-0.00050	0.00005	-0.00059 – -0.00040	<0.001
daily_hosp_occupancy	0.07763	0.00304	0.07167 – 0.08360	<0.001
daily_ICU_occupancy	0.30162	0.01924	0.26389 – 0.33935	<0.001
popData2020	-0.00001	0.00000	-0.00001 – -0.00001	<0.001
Observations	1964			
R ² / R ² adjusted	0.856 / 0.856			

Com es pot observar, el model **Pooled OLS** coincideix amb el primer realitzat, ja que totes les variables són significatives amb excepció de *tests_done* i *testing_rate*. Es veu a la sortida del codi d'*R* que el *p-valor* del model és menor que 0.05, per tant, indica que les variables independents poden explicar la dependent, en aquest cas, els morts. Destacar que, malgrat haver fet ús del paquet *plm*, el qual respecta les dades no balancejades, els coeficients són molt semblants al model *OLS*.

Es pot observar també el *R-squared* o el coeficient de determinació igual a 0.85611, cosa que ens indica que el model és correcte, ja que quan més s'acosti a 1, més explicat està el model per aquestes variables independents.

Per les variables significatives, és a dir, les que presenten un *p*-valor menor que 0.05, s'observa que *daily_hosp_occupancy* i *daily_ICU_occupancy* són positives, pel qual ens indica que com major és aquesta ocupació als hospitals i a les UCIs, el nombre de morts també és major. Això pot ser donat als recursos mèdics, ja que quanta més gent hi ha hospitalitzada, els recursos són més limitats i les saturacions als centres sanitaris augmenten. A més a més, la variable amb major coeficient positiu és *positivity_rate*, pel que quan major sigui aquesta taxa de persones positives en el virus, majors són les morts.

I a continuació, es fa exactament el mateix amb el segon model, es calcula amb el paquet *plm* i s'observa si els resultats coincideixen o no.

Taula 5.4. Model Pooled OLS population

Pooled OLS Population				
	Dependent variable			
<i>Predictors</i>	<i>Estimates</i>	<i>std. Error</i>	<i>CI</i>	<i>p</i>
(Intercept)	0.00000	0.00000	0.00000 – 0.00000	<0.001
tests_done	0.00004	0.00001	0.00003 – 0.00006	<0.001
testing_rate	-0.00008	0.00006	-0.00020 – 0.00004	0.198
new_cases	0.00046	0.00005	0.00037 – 0.00055	<0.001
Observations	1964			
R ² / R ² adjusted	0.149 / 0.148			

En aquest cas es veu una clara diferència i és que el model, anomenat **Pooled OLS population**, està explicat així:

$$deaths = 5.01e^{-12} + 0.00004 * tests\ done - 0.00008 * testing\ rate + 0.0004 * new\ cases$$

S'observa que, la variable *testing_rate* no és significativa, com tampoc ho era en l'anterior model. Malgrat que es veu que el *p*-valor és menor que 0.05, és a dir, significatiu, s'observa que el coeficient de determinació és 0.1493, és a dir, molt menor que en l'anterior model que s'acaba de comentar. Aquest coeficient de determinació indica que amb aquestes variables el model només està explicat un 14,93%, el qual, comparat amb el 85,6% del model anterior és bastant inferior. Per tant, no seria un model vàlid per explicar les morts per Coronavirus.

En aquest darrer model, es veu que hi ha variables explicatives que no surten a l'executar-lo. Les variables excloses en el model, és a dir, aquelles variables que s'afegeixen al model però que a l'hora de veure els resultats no apareixen, són aquelles variables predictores que no s'han inclòs al model final. Això no vol dir que no siguin importants, i certament que no formen part del sistema que impulsa el comportament de la variable dependent, simplement vol dir que l'algoritme treballa i al final, aquestes variables no s'inclouen en el model de regressió. En altres paraules, no aporten informació significativa al model.

La regressió gradual (*stepwise regression*) és el mètode d'ajustament dels models de regressió en què l'elecció de les variables predictives es realitza de manera automàtica. És a dir, en cada pas es considera una variable per la suma o la resta del conjunt de variables explicatives en funció del criteri que haguem especificat amb anterioritat. Aquesta informació és aplicable per tots els models d'aquest treball.

D'aquesta primera secció de modelització, s'observa que el model més ben explicat és el **Pooled OLS**, on la seva equació queda explicada de la següent manera,

$$\begin{aligned} \text{deaths} = & 50.3 - 0.00001 * \text{tests done} - 0.00023 * \text{testing rate} + 3.666 * \text{positivity rate} \\ & - 0.0005 * \text{new cases} + 0.077 * \text{daily hosp occupancy} + 0.301 * \text{daily ICU occupancy} \\ & - 0.00001 * \text{popData2020} \end{aligned}$$

Aquest és el primer model, el qual consisteix en fer una regressió lineal simple per tot el període estimat amb el mètode de mínims quadrats ordinaris. Es fa la prova de la hipòtesi, on la hipòtesi nul·la és que les morts durant la pandèmia s'expliquen igualment bé amb un model que incorpora la mateixa sensibilitat per tots els països amb els canvis de les variables significatives. I a l'executar aquest model, el *p-valor* resultant ha estat menor que 0.05, per tant, rebutgem aquesta hipòtesi nul·la i podem concloure que existeix heterogeneïtat en el comportament d'aquesta pandèmia.

S'observa pels coeficients de les variables, que els ratis de positivitat i les hospitalitzacions, tant a hospitals com a centres de cures intensives (UCI) són les que més estan afectant a les morts, ja que són les variables amb coeficient positiu més elevat junt amb la taxa de població positiva en *Coronavirus*.

5.2. (One-Way) Fixed Effect Model

Explicar les dades amb el model d'efectes fixes considera que existeix un terme constant diferent per cada país, i suposa que els efectes individuals són independents entre si.

Amb aquest model es considera que les variables explicatives afecten per igual a les unitats de tall transversal i que aquestes es diferencien per característiques pròpies de cada una d'elles.

Per aquest primer model d'efectes fixes, el *One Way Fixed Effects Model*, també ens centrarem en explicar les morts, és a dir, la variable *deaths* com a variable dependent, i *tests_done*, *testing_rate*, *positivity_rate*, *new_cases*, *daily_hosp_occupancy*, *daily_ICU_occupancy* i *popData2020* com a variables independents.

Es torna a utilitzar el paquet *plm* de l'*R* però aquest cop amb indicant que l'efecte individual i el model *within*. Aquest model l'hem anomenat **Country Fixed Effects**, amb el qual observarem si la variable *country* té efectes fixos o no en les dades.

La sortida del model és la següent:

Taula 5.5. Model Country Fixed Effects

Country Fixed Effects				
<i>Predictors</i>	Dependent variable			
	<i>Estimates</i>	<i>std. Error</i>	<i>CI</i>	<i>p</i>
tests_done	-0.00004	0.00001	-0.00006 – -0.00002	< 0.001
testing_rate	0.00191	0.00078	0.00037 – 0.00345	0.015
positivity_rate	2.82906	0.47149	1.90438 – 3.75374	< 0.001
new_cases	-0.00044	0.00005	-0.00054 – -0.00034	< 0.001
daily_hosp_occupancy	0.08812	0.00337	0.08151 – 0.09473	< 0.001
daily_ICU_occupancy	0.26501	0.02003	0.22573 – 0.30429	< 0.001
Observations	1964			
R ² / R ² adjusted	0.790 / 0.787			

El model mostra que totes les variables dependents són significatives, cosa que fins ara, en els anteriors models no s'havia vist. També es veu que el model omet la variable que fa referència a la població, per tant, el model no troba prou significança en aquesta variable. S'observa clarament que el model està explicat en un 79%, ja que el *R-squared* és 0.79, per tant, es pot assumir que està bastant ben explicat. Però, abans de treure més conclusions, cal assegurar-se de que els països causen efecte en el nostre model. Es fa utilitzant la funció *plmtest()*.

Al realitzar aquest test al *R*, el *p-valor* suggereix la presència d'efectes del país, és a dir, de la variable *country*, el qual en el test és menor que 0.05, per tant, significatiu.

A continuació, es calculen els estimadors per país aplicant la funció *fixef* de *R*, ja que en aquest model d'efectes fixos, cada país té el seu propi estimador en l'explicació de les dades (Fixed effects, s.f.). La següent taula mostra el canvi de *deaths* al llarg del temps, per país, cada cop que les variables independents augmenten una unitat.

Taula 5.6. Taula amb dels països de la UE i el seu coeficient corresponent

Àustria	77.9540	Bèlgica	-4.1132	Bulgària	-82.8371
Xipre	-48.6869	R. Txeca	32.1125	Dinamarca	25.0273
Estònia	-35.2720	Finlàndia	-5.6620	França	-713.2979
Islàndia	-35.6290	Irlanda	-8.6436	Letònia	-32.4902
Luxemburgo	-29.3912	Malta	-12.5552	P.Baixos	19.3128
Portugal	45.6371	Romania	-61.1617	Eslovàquia	-8.7905
Eslovènia	-25.7625	Espanya	-394.6649	Suècia	22.4201

S'observa per tant, que els països amb més afectació de cara a les morts han estat aquells amb un major estimador, és a dir, Àustria, Portugal i la República Txeca. Recordar que en el model s'ha afegit la variable referent a la població del país, doncs, aquest model està explicat en referència a la població del 2020 d'aquests.

És sorprenent veure com països que semblaven ser els més afectats per la pandèmia com Espanya i França, tenen coeficients negatius. Això pot ser donat a que en relació a la població, aquests dos països són més grans i per tant, els recursos d'aquests també ho són.

Malgrat això, és molt probable que aquests estimadors canviïn a l'introduir les vacunes per a frenar la *Covid-19* en els models, ja que d'aquesta manera es veurà si aquestes han estat d'ajuda o no.

A més dels efectes fixos per país, una altra sèrie de factors que poden estar afectant la variable dependent *deaths* són els efectes fixos en el temps, els quals no són específics d'un país en concret, és a dir, no són específics d'un país per individual.

Per fer el model fixant el terme del temps, és a dir, *year_week*, el qual representa les setmanes de l'any, es pot utilitzar el mateix paquet *plm*, però en aquest cas es fixarà el terme del temps, és a dir, *effect = time*. Aquest nou model s'anomena **Time Fixed Effects**.

Taula 5.7. Model Time Fixed Effects

Time Fixed Effects					
<i>Predictors</i>	Dependent variable				
	<i>Estimates</i>	<i>std. Error</i>	<i>CI</i>	<i>p</i>	
tests_done	-0.00001	0.00001	-0.00003 – 0.00001	0.172	
testing_rate	0.00182	0.00064	0.00055 – 0.00308	0.005	
positivity_rate	4.35698	0.60401	3.17238 – 5.54159	<0.001	
new_cases	-0.00042	0.00005	-0.00052 – -0.00032	<0.001	
daily_hosp_occupancy	0.07487	0.00297	0.06904 – 0.08070	<0.001	
daily_ICU_occupancy	0.30582	0.01936	0.26785 – 0.34379	<0.001	
popData2020	-0.00001	0.00000	-0.00001 – -0.00001	<0.001	
Observations	1964				
R ² / R ² adjusted	0.862 / 0.854				

En primer lloc, s'observa clarament en la sortida que els tests fets, és a dir, *tests_done* és l'única variable independent no significativa en el model d'efectes fixes del temps. Malgrat això, es pot veure que el coeficient de determinació és major, ja que en aquest cas, el model està explicat un 86,19%. Es realitza també, el *plmtest* en aquest nou model en l'*R*, i el *p-valor* obtingut, el qual és menor que 0.05, indica que es pot rebutjar la hipòtesi nul·la i acceptar que sí hi ha efectes fixos en el temps presents en el nostre model.

En aquest model, d'efectes fixos en el temps, s'aplica la funció *fixef* de *R* de nou, i s'observa que cada setmana té un coeficient diferent, com s'ha vist anteriorment amb els països. D'aquests coeficients, els quals es podrà trobar la sortida del codi a l'annex, s'observa que durant les primeres setmanes aquests coeficients eren majors i positius, en canvi, a les darreres, aquests coeficients van disminuint i fins i tot són negatius. Per tant, es veu una clara evolució en l'afectació de les variables a les morts des del principi d'aquest període de dos anys fins al final.

En aquests dos models d'efectes fixos s'està assumint que cada variable explicativa té un sol coeficient per tots els països i un estimador diferent per cada un. Els tests realitzats ens permeten concloure que, encara agregant diferents estimadors, s'explica millor el comportament de la variable dependent quan s'estan estimant diferents coeficients per les variables explicatives.

En resum, els models i els tests realitzats, permeten rebutjar la hipòtesi de que les variables explicatives tenen un efecte igual sobre els morts, per tots els països. A més a més, cada país té un coeficient diferent, com hem observat en la Taula 5.6.

Ara, per a controlar d'aquests dos efectes, país i temps, es necessita estimar un model *Two-Ways*. Aquest, és el primer model que té en compte la dependència espaciotemporal en els models.

5.3. *Two-Way Individual Effects Models*

Com s'ha explicat anteriorment al marc teòric, aquest model bidireccional es pot representar de la següent manera:

$$y_{it} = \alpha + \beta_{xit} + \mu_i + \lambda_t + \epsilon_{it}$$

En aquest model, μ_i captura els efectes específics de l'individu no observables i en canvi, λ_t captura els efectes específics del temps no observables. Per una base de dades com la de la *Covid-19*, en la qual les dades varien en espai temporal, és un model molt interessant. Veiem també que μ_i no varia en el temps i en canvi λ_t no varia entre els individus, en el nostre cas, els països de la Unió Europea.

En aquest cas, es continua utilitzant el paquet *plm*, indicant l'efecte '*twoways*'. L'anomenem ***Two-Way Fixed Effects***. Es pot observar la sortida del model a continuació:

Taula 5.8. Model Two-Way Fixed Effects

<i>Predictors</i>	Dependent variable			
	<i>Estimates</i>	<i>std. Error</i>	<i>CI</i>	<i>p</i>
tests_done	-0.00005	0.00001	-0.00007 – -0.00003	<0.001
testing_rate	0.00473	0.00078	0.00319 – 0.00627	<0.001
positivity_rate	3.37344	0.59681	2.20294 – 4.54394	<0.001
new_cases	-0.00034	0.00005	-0.00044 – -0.00024	<0.001
daily_hosp_occupancy	0.08182	0.00331	0.07534 – 0.08830	<0.001
daily_ICU_occupancy	0.28720	0.01995	0.24807 – 0.32633	<0.001
Observations	1964			
R ² / R ² adjusted	0.783 / 0.767			

Aquest model a priori sembla que estigui molt ben explicat. En primer lloc, totes les variables independents són significatives, ja que tots els *p*-valors són menors que 0.05. també, es veu que els efectes d'aquestes són una mica més grans, i és que al tractar-se de valors molt petits com per exemple, el efecte de la ocupació diària dels hospitals (*daily_hosp_occupancy*) amb un 0.08182, en l'anterior model, on es fixa només el temps era de 0.074869, per tant, la diferència és mínima, però al tractar-se de dades massives, aquests valors són tots molt propers a 0, i per tant, els canvis també són mínims.

De nou, es torna a veure que les variables amb coeficient positiu són *positivity_rate*, *daily_hosp_occupancy* i *daily_ICU_occupancy*, variables les quals s'han vist amb coeficients positius en altres models vistos fins ara. També, respecte el coeficient de determinació, aquest és de 0.783, per tant, la variable *deaths* està ben explicada pel model.

A continuació, mostrem els resultats dels 3 models *Country Fixed Effects*, *Time Fixed Effects* i *Two-Way Fixed Effects*.

Taula 5.9. Resultats d'R dels tres models

<i>Predictors</i>	Dependent variable				Dependent variable				Dependent variable			
	<i>Estimates</i>	<i>std. Error</i>	<i>CI</i>	<i>p</i>	<i>Estimates</i>	<i>std. Error</i>	<i>CI</i>	<i>p</i>	<i>Estimates</i>	<i>std. Error</i>	<i>CI</i>	<i>p</i>
tests_done	-0.00004	0.00001	-0.00006 – -0.00002	<0.001	-0.00001	0.00001	-0.00003 – -0.00001	0.172	-0.00005	0.00001	-0.00007 – -0.00003	<0.001
testing_rate	0.00191	0.00078	0.00037 – 0.00345	0.015	0.00182	0.00064	0.00055 – 0.00308	0.005	0.00473	0.00078	0.00319 – 0.00627	<0.001
positivity_rate	2.82906	0.47149	1.90438 – 3.75374	<0.001	4.35698	0.60401	3.17238 – 5.54159	<0.001	3.37344	0.59681	2.20294 – 4.54394	<0.001
new_cases	-0.00044	0.00005	-0.00054 – -0.00034	<0.001	-0.00042	0.00005	-0.00052 – -0.00032	<0.001	-0.00034	0.00005	-0.00044 – -0.00024	<0.001
daily_hosp_occupancy	0.08812	0.00337	0.08151 – 0.09473	<0.001	0.07487	0.00297	0.06904 – 0.08070	<0.001	0.08182	0.00331	0.07534 – 0.08830	<0.001
daily_ICU_occupancy	0.26501	0.02003	0.22573 – 0.30429	<0.001	0.30582	0.01936	0.26785 – 0.34379	<0.001	0.28720	0.01995	0.24807 – 0.32633	<0.001
popData2020					-0.00001	0.00000	-0.00001 – -0.00001	<0.001				
Observations	1964				1964				1964			
R ² / R ² adjusted	0.790 / 0.787				0.862 / 0.854				0.783 / 0.767			

En primer lloc, s'observa que les variables amb més efecte explicatiu en el model són *positivity_rate*, *daily_hosp_occupancy* i *daily_ICU_occupancy*, ja que les 3 són les que presenten un efecte més gran i positiu en el model. Per tant, quan major sigui la taxa de positiu, i major siguin el nombre d'hospitalitzacions i d'ingressats en unitats de cures intensives, major serà també el nombre de morts.

Per últim, s'observen els coeficients de determinació dels 3 models d'efectes fixes. Es veu que el que presenta un *R-squared* major és el *Time Fixed Effects*, i per tant, el que millor està explicat, amb un 86%. També, aquest és l'únic model que no elimina la variable *popData2020*, la qual fa referència a la població de cada país a l'any 2020. Malgrat això, ambdós models restants també tenen un coeficient de determinació molt bo, amb un 79% i 78% per *Country Fixed Effects* i *Two-Way Fixed Effects*, respectivament.

Correlació en la sèrie

Per les dades de sèries temporals, s'ha d'abordar el potencial de correlació en sèrie en el terme d'error. Per tant, es provarà la correlació en sèrie amb la prova de *Breusch-Godfrey* (Tok.Wiki, 2020).

En estadística, el test *Breusch-Godfrey* s'utilitza com a mitjà per validar alguns supòsits aplicats a models de regressió de sèrie de dades. Més en concret, per detectar la presència de dependència en la sèrie que no ha estat considerat dins del model proposat.

El test de correlació serial de *Breusch-Godfrey LM* és un test d'autocorrelació en els errors i els residus estadístics en un model de regressió. S'utilitzen els errors generats al model de regressió i un test d'hipòtesis derivat d'aquest. Les hipòtesis nul·les, en aquest cas, és que no hi hagi correlació serial.

El procediment d'aquest test comença amb la consideració d'una regressió lineal de qualsevol forma, per exemple:

$$Y_t = \beta_1 + \beta_2 X_{t,1} + \beta_3 X_{t,2} + u_t$$

on els residus poden seguir un esquema autoregressiu $AR(p)$,

$$u_t = p_1 u_{t-1} + p_2 u_{t-2} + \dots + p_p u_{t-p} + \varepsilon_t .$$

El model de regressió sempre s'ajusta primer per Mínim Quadrat Ordinari per obtenir els residus mostrals u_t .

Per tant, *Breusch* i *Godfrey* diuen que, si la següent regressió auxiliar es ajustada:

$$\hat{u}_t = \alpha_0 + \alpha_1 X_{t,1} + \alpha_2 X_{t,2} + p_1 \hat{u}_{t-1} + p_2 \hat{u}_{t-2} + \dots + p_p \hat{u}_{t-p} + \varepsilon_t$$

i si el R^2 es calculat per aquest model, llavors la següent distribució pot ser utilitzada per la distribució d'aquest test estadístic:

$$nR^2 \sim X^2_p$$

Quan la hipòtesi nul·la $H_0 = \{p_i = 0 \text{ per tot } i\}$, llavors es compleix, és a dir, no existeix correlació serial de qualsevol ordre sobre p . En aquest càlcul, $n = T - p$, on T és el nombre d'observacions de la serie original.

Per tant, per fer aquest càlcul en l' R , s'utilitza la funció `pbgttest()` a l' R , on el p -valor és menor que 0.05, per tant, es rebutja la hipòtesi nul·la i es confirma la presència de correlació serial en el nostre terme d'error.

Per a corregir aquesta correlació serial, es pot utilitzar la funció `coefstest()` i es farà servir la funció `vcovHC()` de tal manera que s'obtindrà una matriu de covariàncies consistents amb l'heteroscedasticitat, ja que interessa corregir aquesta autocorrelació. Per a fer-ho, s'especificarà que el mètode empleat sigui *arellano*, el qual corregeix tant l'heteroscedasticitat com l'autocorrelació. La sortida del codi és la següent:

Taula 5.10. Taula amb els resultats d' R del `coefstest`

	Twoway Fixed Effects	Twoway Fixed Effects (HAC)
tests_done	-0.00 *** (0.00)	-0.00 * (0.00)
testing_rate	0.00 *** (0.00)	0.00 * (0.00)
positivity_rate	3.37 *** (0.60)	3.37 *** (0.84)
new_cases	-0.00 *** (0.00)	-0.00 *** (0.00)
daily_hosp_occupancy	0.08 *** (0.00)	0.08 *** (0.01)
daily_ICU_occupancy	0.29 *** (0.02)	0.29 *** (0.07)
R ²	0.78	
Adj. R ²	0.77	
Num. obs.	1964	

*** p < 0.001; ** p < 0.01; * p < 0.05

Es pot veure que, amb els errors estàndards d'heteroscedasticitat i d'autocorrelació (HAC), totes les variables continuen essent significatives en el model.

Dependència transversal

Abans d'acabar amb l'apartat dels models d'efectes fixos, cal tenir en compte que els models també poden tenir dependència transversal. Abans d'introduir aquest concepte, es descriuen les dades transversals com un tipus de dades recopilades mitjançant l'observació de molts subjectes, com en el nostre cas, molts països, al mateix temps o sense tenir en compte les diferències en el temps. L'anàlisi d'aquest tipus de dades normalment solen consistir en comparar les diferències entre els subjectes.

S'entén per dependència transversal que les dades de sèries en el temps per diferents unitats de secció transversal estan correlacionades, ja sigui com a resultat de factors no observats o d'efectes espacials o indirectes.

Per tant, com en aquest estudi s'analitzen les dades sobre una pandèmia mundial, és possible que es trobi aquesta dependència transversal, ja que un sol canvi podria afectar a tots els països.

Es pot verificar la dependència de la secció transversal utilitzant la prova de dependència de la secció transversal de *Pesaran (2004)*. Aquesta, és una prova aplicable als models de panell, que es basa en el promig dels coeficients de correlació de les parelles dels residus OLS de les regressions individuals en el panell (Colonescu, C., 2016).

$$y_{it} = \alpha_i + x'_{it}\beta_i + \varepsilon_{it}$$

$$\widehat{\rho}_{ij} = \frac{\sum_{t=1}^T e_{it} e_{jt}}{\sqrt{\sum_{t=1}^T e_{it}^2} \sqrt{\sum_{t=1}^T e_{jt}^2}} \text{ on } e_{it} = y_{it} - \hat{\alpha}_i - x'_{it}\beta_i$$

$$CD = \sqrt{\frac{2T}{n(n-1)}} \left(\sum_{i=1}^{n-1} \sum_{j=i+1}^n \widehat{\rho}_{ij} \right)$$

Sota $H_0: Cov(\varepsilon_{it}, \varepsilon_{ij}) = 0 \forall t, i \neq j$,

$$\sqrt{T} \widehat{\rho}_{ij} \Rightarrow N(0,1) \quad CD \Rightarrow N(0,1)$$

Per a la realització d'aquest test de Pesaran al R , s'utilitza la funció '*pcdtest()*'.

Com s'ha vist amb les proves realitzades fins ara, en aquest cas la hipòtesi nul·la és que no hi ha dependència transversal. Malgrat això, el *p-valor* del test de *Pesaran* és menor que 0.05, per tant, es rebutja la hipòtesi nul·la i s'accepta que existeix una dependència transversal i que per tant, s'ha de corregir. Per a corregir aquesta dependència transversal, hi ha diferents mètodes, però a l'estar tractant amb grans volums de dades, s'utilitza el mètode *Driscoll & Kraay (1998)*.

Driscoll i Kraay, proposen un estimador no paramètric de la matriu de covariància i robust a diverses formes de dependència transversal i temporal. Aquest, calcula els errors estàndards com l'arrel de la diagonal de la matriu de covariància asimptòtica

$$V(\hat{\beta}) = (X'X)^{-1} \hat{S}_T (X'X)^{-1}$$

on \hat{S}_T és $\hat{S}_T = \hat{\Omega}_0 + \sum_{j=1}^{m(T)} \omega(j, m) (\hat{\Omega}_j + \hat{\Omega}'_j)$, (*Newey West, 1987*)

$m(T)$ és fins a on els errors poden estar correlacionats

$$\omega(j, m) = 1 - \frac{j}{m(T) + 1}, \hat{\Omega}_j = \sum_{t=j+1}^T h_t(\hat{\beta}) h_{t-1}(\hat{\beta})'$$

i $h_t(\hat{\beta}) = \sum_{i=1}^{N(T)} x_{it} \varepsilon_{it}$ captura la correlació mitjana del panell t .

S'utilitza a l'R, la funció 'vcovSCC',
 Per últim, s'adjunten tots els resultats en una taula.

Taula 5.11. Taula amb els resultats d'R del resum dels models

	Country Effects	Time Effects	Two-Way Fixed Effects	SCC
tests_done	-0.00 *** (0.00)	-0.00 (0.00)	-0.00 *** (0.00)	-0.00 * (0.00)
testing_rate	0.00 * (0.00)	0.00 ** (0.00)	0.00 *** (0.00)	0.00 * (0.00)
positivity_rate	2.83 *** (0.47)	4.36 *** (0.60)	3.37 *** (0.60)	3.37 *** (0.72)
new_cases	-0.00 *** (0.00)	-0.00 *** (0.00)	-0.00 *** (0.00)	-0.00 *** (0.00)
daily_hosp_occupancy	0.09 *** (0.00)	0.07 *** (0.00)	0.08 *** (0.00)	0.08 *** (0.01)
daily_ICU_occupancy	0.27 *** (0.02)	0.31 *** (0.02)	0.29 *** (0.02)	0.29 *** (0.07)
popData2020		-0.00 *** (0.00)		
RA2	0.79	0.86	0.78	
Adj. RA2	0.79	0.85	0.77	
Num. obs.	1964	1964	1964	

*** p < 0.001; ** p < 0.01; * p < 0.05

5.4. (One-Way) Random Effects Model

Un cop finalitzat l'estudi dels models amb efectes fixos, queda analitzar els models amb efectes aleatoris. El *Random Effects Model* es distingeix per la seva estructura especial del terme d'error, ja que els errors tenen mitjana de 0, variància igual a $\sigma^2_u + \sigma^2_e$, sense correlació entre els individus i amb la covariància en el temps igual a σ^2_u .

Una altra característica important d'aquest model és que la correlació temporal dels error no disminueix en el temps.

$$\rho = \text{corr}(v_{it}, v_{is}) = \frac{\sigma^2_u}{\sigma^2_u + \sigma^2_e}$$

En primer lloc es realitza una prova, la qual equival a provar la hipòtesi de que no hi ha diferències entre els individus, el que implica que la variable aleatòria específica de l'individu té variància zero, és a dir,

$$H_0: \sigma^2_u = 0$$

$$H_1: \sigma^2_u \neq 0$$

Per a realitzar aquesta prova en R, s'utilitza la funció *plmtest* i el model analitzat serà el **OLS**.

Imatge 5.1. Sortida codi R del plmtest

Random effects test			alternative
statistic	p.value	method	
30.33896		0 Lagrange Multiplier Test - (Honda) for unbalanced panels	significant effects

S'observa que, com s'ha obtingut un *p*-valor inferior a 0.05, es rebutja la hipòtesi nul·la de variància zero en els errors específics de l'individu, és a dir, l'heterogeneïtat entre els individus pot ser significativa.

Per al modelatge d'aquest model d'efectes aleatoris a l'*R*, cal especificar en aquest cas que l'efecte és '*random*'. Es veu la sortida a continuació,

Taula 5.12. Model Random Effects

Random Effects Model					
<i>Predictors</i>	Dependent variable				
	<i>Estimates</i>	<i>std. Error</i>	<i>CI</i>		<i>p</i>
(Intercept)	-49.35106	15.09171	-78.94857 – -19.75355		0.001
tests_done	-0.00006	0.00001	-0.00009 – -0.00004		<0.001
testing_rate	0.00365	0.00075	0.00217 – 0.00513		<0.001
positivity_rate	3.18483	0.48052	2.24245 – 4.12721		<0.001
new_cases	-0.00038	0.00005	-0.00048 – -0.00028		<0.001
daily_hosp_occupancy	0.08119	0.00334	0.07465 – 0.08774		<0.001
daily_ICU_occupancy	0.26831	0.02030	0.22849 – 0.30812		<0.001
Observations	1964				
R ² / R ² adjusted	0.790 / 0.790				

Aquest model inclou la possibilitat de variacions entre països i també assumeix que aquesta variació és de la naturalesa aleatòria o que no està correlacionada amb les variables de l'estudi. A primera vista, es veu que totes les variables independents són significatives, ja que el seu *p*-valor és menor que 0.05 i veiem que el model està explicat amb un 79%.

La raó de l'ús del model d'efectes aleatoris és que, com s'ha indicat anteriorment, a diferència del model d'efectes fixos, s'assumeix que la variació entre les entitats és aleatòria i no està correlacionada amb les variables predictores o independents incloses en el model.

S'ha observat que hi ha diferències entre països a l'hora d'explicar les morts, per tant, es té raons per creure que les diferències entre les entitats, és a dir, els països, tenen alguna influència a la vostra variable dependent. És per això, que també es calcula el model d'efectes aleatoris. Els efectes aleatoris suposen que el terme d'error de l'entitat no està correlacionat amb els predictors que permeten que les variables invariants en el temps tinguin un paper explicatiu, com podem veure en la taula, observem que la variable *popData2020*, queda omesa del model.

5.5. Selecció dels models

Un cop vistos tots els models, en aquesta secció se'n farà una selecció, d'aquells que siguin millors que els altres, per tal manera de veure el model que millor explica la variable dependent *deaths*.

En primer lloc, es fa una ràpida comparació del model *Pooled OLS* amb els models vistos en efectes fixos, els quals són *Country Effects Model*, *Time Effects Model* i *Two-Way Effect Model*.

Per a fer aquesta comprovació, s'utilitza la funció de *l'R*, anomenada *pFtest*. En aquest test, amb el mètode *plm*, l'argument d'aquesta funció són dos objectes *plm*, és a dir, els dos models, el primer model és *within*, per tant, d'efectes fixos, i el segon és *pooling*, és a dir, OLS. Els efectes provats són individuals, temporals i bidireccionals, depenent dels efectes introduïts en el model d'efectes fixos (RDocumentation, s.f.). S'utilitza per a comprovar quin tipus de model és millor per a les dades de les quals es disposa.

En aquesta funció, les hipòtesis són les següents:

$$H_0 = \text{el model Pooled OLS és coherent}$$
$$H_1 = \text{el model d'efectes fixos és coherent}$$

Entenent com a coherent, que aquell és millor que l'altre. Per tant, es procedeix a realitzar les tres comparacions.

- *Pooled OLS vs Country Effects Model* on el *p-valor* d'aquesta prova és menor que 2.2e-16.
- *Pooled OLS vs Time Effects Model* on el *p-valor* d'aquesta prova és menor que 2.2e-16.
- *Pooled OLS vs Two-Way Effects Model* on el *p-valor* d'aquesta prova és menor que 2.2e-16.

Per tant, com es pot observar, en els tres casos, es veu que el *p-valor* és menor que 0.05, indicant així que els models d'efectes fixos són una millor opció. Aquests tests no indiquen que el model *Pooled OLS* no sigui vàlid, però al comparar-los, els models d'efectes fixos expliquen millor la variable *deaths*.

Malgrat això, encara cal comparar els models d'efectes fixos amb els d'efectes aleatoris, i per fer aquesta comparació, s'utilitzarà el *Test de Hausman*.

Test de Hausman

Com s'ha comentat, un cop en aquest punt, s'ha de decidir quin model explica millor les dades: els models d'efectes fixos o el model d'efectes aleatoris. Tots els models vists en els efectes fixos i aleatori estan ben explicats, ja que presenten un *R-squared* bastant elevat, cosa que indica que el model està majoritàriament explicat per les variables independents.

Tanmateix, la prova per comprovar-ho és el *Test de Hausman*. En aquest test s'assumeix que el model d'efectes fixos és consistent pels paràmetres veraders i que el model d'efectes aleatoris és una especificació eficient dels efectes individuals sota el supòsit que són aleatoris i segueixen una distribució normal.

Se suposa que en el model d'efectes fixos es calcula sempre l'estimador constant, mentre que en el model d'efectes aleatoris, es calcula l'estimador que és constant i eficient sota H_0 . Per a realitzar el test en *l'R*, s'utilitza la funció *phptest*, també del paquet *plm*. Les hipòtesis sobre les quals treballem en aquest test són les següents:

$H_0 =$ el model d'efectes aleatoris és coherent

$H_1 =$ el model d'efectes fixos és coherent

Entenent com a coherent, que aquell és millor que l'altre i a la inversa. Per tant, es procedeix de nou amb els tests. A continuació els resultats dels *p-valors*.

- *Random Effects Model vs Country Effects Model* on el *p-valor* d'aquesta prova és igual a 5.348e-06.
- *Random Effects Model vs Time Effects Model* on el *p-valor* d'aquesta prova és menor que 2.2e-16.
- *Random Effects Model vs Two-Way Effects Model* on el *p-valor* d'aquesta prova és menor que 2.2e-16.

Per tant, sota l'especificació actual, la hipòtesi inicial de que els efectes individuals són moderats adequadament amb un model d'efectes aleatoris és clarament rebutjada, ja que el *p-valor* en els tres casos és menor que 0.05. En conclusió, es trien els models d'efectes fixos per explicar les dades de la *Covid-19*.

D'aquests darrers models, se'n farà una nova secció afegint les vacunes, de les quals no hi ha dades de tots els països de la Unió Europea, per això no se n'havia fet ús fins al moment.

5.6. Modelització tenint en compte les vacunes contra la Covid-19

Les vacunes contra la Covid-19 són una eina poderosa per garantir la salut pública i controlar la pandèmia. Resultats d'estudis observacionals han demostrat que les vacunes autoritzades a la Unió Europea i a la EEA (*European Economic Area*) protegeixen actualment en gran mesura de la Covid-19 greu, les hospitalitzacions i la mort (Covid-19 vaccination, s.f.).

La primera vacuna contra la Covid-19 es va posar fora de l'entorn d'assaig clínic el 8 de desembre del 2020. Un any més tard, fins al 8 de desembre del 2021, es calcula que el 55,9% de la població va rebre almenys una dosi d'una vacuna, sense diferenciar per tipus, contra la Covid-19, el 45,5% s'estima haver rebut la segona dosi i un 4,3% la dosi de reforç (Staff, A., 2021).

Malgrat la increïble velocitat amb la qual es van administrar aquestes dosis arreu del món, amb el fi comú de parar els peus a aquesta pandèmia afectada mundialment, les morts per *coronavirus* seguien informant-se. Comprendre l'impacte de la vacunació per frenar la pandèmia i com aquestes, han influït en els morts o en les hospitalitzacions és clau per entendre com han funcionat al llarg del temps.

A continuació, una taula actualitzada amb el percentatge de població vacunada, a finals de juliol del 2022, per país de la Unió Europea.

Taula 5.13. Percentatge de població vacunada per país a finals de Juliol del 2022

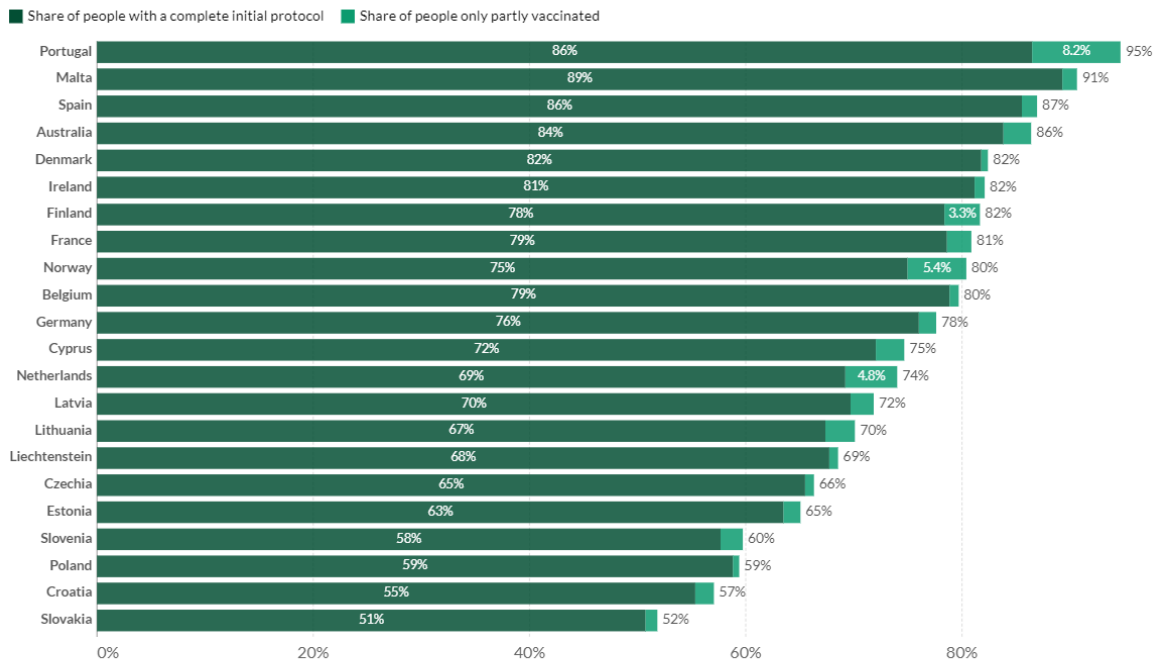
Àustria	76%	Bèlgica	80%	Bulgària	16%	Croàcia	57%	Xipre	75%
R.Txeca	66%	Dinamarca	82%	Estònia	65%	Finlàndia	82%	França	81%
Alemanya	78%	Grècia	76%	Hongria	66%	Islàndia	84%	Irlanda	82%
Itàlia	86%	Letònia	72%	Liechtenstein	69%	Lituània	70%	Luxemburg	75%
Malta	91%	P. Baixos	74%	Noruega	80%	Polònia	59%	Portugal	95%
Romania	28%	Eslovàquia	72%	Eslovènia	60%	Espanya	87%	Suècia	75%

També, respecte al conjunt de la Unió Europea un 76% de la població total està vacunada d'una dosi o més. Es veu a continuació, una representació gràfica d'alguns països de la Unió Europea on es pot observar amb un verd més fosc el percentatge de població vacunada completament, és a dir, primera, segona i dosi de reforç, i amb un verd més clar, es veu el percentatge de població parcialment vacunada (Ritchie, H. 2020).

Gràfic 5.1. Representació gràfica del percentatge de població vacunada per país

Share of people vaccinated against COVID-19, Jul 31, 2022

Our World in Data



Source: Official data collated by Our World in Data
 Note: Alternative definitions of a full vaccination, e.g. having been infected with SARS-CoV-2 and having 1 dose of a 2-dose protocol, are ignored to maximize comparability between countries. CC BY

Dec 27, 2020 — Aug 10, 2022

A continuació, es realitzen els models d'efectes fixos per país i per temps, com ja s'ha fet anteriorment en aquest treball, però, com s'ha esmentat, afegint les vacunes, en concret, les tres dosis subministrades acumulades, és a dir, s'utilitzaran les variables *FirstDoseAcc*, *SecondDoseAcc* i *ThirdDoseAcc*.

Durant aquest apartat en la modelització, no es treballa amb tots els països de la Unió Europea com s'ha estat fent actualment, ja que respecte les dades de vacunació no hi ha informació de tots els països. Per tant, s'ha de tenir en compte que per aquest anàlisi, estem treballant amb menys observacions.

Durant aquest capítol de modelització, els països dels quals disposem dades sobre les vacunes són Àustria, Bèlgica, Bulgària, la República Txeca, Xipre, Dinamarca, Estònia, Finlàndia, França, Islàndia, Irlanda i Espanya.

Country Effects Model Vaccine

Per aquest primer model, anomenat **Country Effects Model Vaccine**, ens centrem de nou en les morts, és a dir, la variable *deaths* és la variable dependent.

Taula 5.14. Country Effects Model Vaccine

<i>Predictors</i>	Dependent variable			
	<i>Estimates</i>	<i>std. Error</i>	<i>CI</i>	<i>p</i>
tests_done	0.00008	0.00001	0.00006 – 0.00011	< 0.001
testing_rate	-0.00522	0.00088	-0.00694 – -0.00350	< 0.001
positivity_rate	0.36427	0.46078	-0.54030 – 1.26885	0.429
new_cases	-0.00027	0.00005	-0.00036 – -0.00018	< 0.001
daily_hosp_occupancy	0.13335	0.00376	0.12597 – 0.14074	< 0.001
daily_ICU_occupancy	-0.14371	0.01977	-0.18252 – -0.10490	< 0.001
FirstDoseAcc	-0.00003	0.00000	-0.00003 – -0.00002	< 0.001
SecondDoseAcc	0.00003	0.00000	0.00003 – 0.00003	< 0.001
ThirdDoseAcc	-0.00002	0.00000	-0.00002 – -0.00001	< 0.001
Observations	765			
R ² / R ² adjusted	0.915 / 0.913			

En aquest primer model es veu que totes les variables són significatives, amb excepció de *Positivity Rate*. Malgrat això, és el model millor explicat vist fins ara, ja que el *R-squared* és 0.915, i també es pot observar en la sortida de l'*R* que el *p-valor* és menor que 0.05, per tant, el model és significatiu.

S'observa a més a més, que la primera i la tercera dosi tenen ambdues un coeficient negatiu, són coeficients baixos, però són negatius, el qual no s'esperaria, ja que aquestes vacunes, a priori, haurien d'ajudar a baixar el nombre de morts. També es veu que la segona dosi si que té un efecte negatiu, i és evident, que de les primeres dosis n'hi ha moltes més que de la segona i la tercera, pel que té un pes més important a l'hora de reduir el nombre de morts.

Abans de continuar, cal assegurar-se de nou de si hi ha algun efecte dels països en el nostre model, utilitzant la funció *plmtest*. A l'executar aquest test a l'*R*, el *p-valor* suggereix la presència d'efectes del país, pel qual es pot veure que s'ha obtingut un *p-valor* menor que 0.05, per tant significatiu.

Es veu a continuació, l'estimador individual per país calculat en aquest model, el qual a l'hora d'executar-lo, ha sortit significatiu, per tant, el model permet rebutjar la hipòtesi de que les variables explicatives tenen un efecte igual sobre les morts, per tots els països.

A continuació, es calculen els estimadors individuals per país. S'utilitza la funció *fixef* de l'*R*.

Taula 5.15. Coeficient del efecte per país al model Country Effects Model Vaccine

Àustria	5.1035	Bèlgica	77.0567	Bulgària	54.3073
Xipre	344.2897	R. Txeca	273.6893	Dinamarca	949.0614
Estònia	947.6173	Finlàndia	1499.3583	França	901.4073
Islàndia	2624.8494	Irlanda	2579.6997	Espanya	900.0918

Per acabar, aquesta darrera taula indica els estimadors individuals per país. En el model vist anteriorment *Country Fixed Effect*, on no s'havien incorporat les vacunes, els països amb major estimador van ser Àustria, Portugal i la República Txeca. En aquest cas, es veu que els estimadors més grans són Islàndia i Irlanda, i també en aquest cas, tots són positius.

Com major sigui aquest estimador, major tendirà a ser la variable dependent *deaths*, per tant, es pot observar que els països més afectats són Islàndia, Irlanda i Finlàndia. Aquests països no destaquen per ser abundants en població, per tant, poden estar més afectats de cara a tenir menys població vacunada o recursos sanitaris més limitats. Es veu un exemple contrari com és Àustria, amb l'estimador més petit, cosa que indica que les morts seran menors cada cop que les variables explicatives augmentin una unitat.

Com s'ha fet també anteriorment, al veure que sí que hi ha efectes fixos en els països, també es calcula el model d'efectes fixos en el temps, amb les vacunes incloses.

Time Effects Model Vaccine

S'utilitza el mateix paquet *plm* per a realitzar el model fixant l'efecte del temps, anomenat ***Time Effects Model Vaccine***, afegint també les tres dosis de les vacunes.

Taula 5.16. Time Effects Model Vaccine

<i>Predictors</i>	Dependent variable			
	<i>Estimates</i>	<i>std. Error</i>	<i>CI</i>	<i>p</i>
tests_done	0.00003	0.00001	0.00002 – 0.00005	< 0.001
testing_rate	-0.00181	0.00059	-0.00296 – -0.00066	0.002
positivity_rate	0.82573	0.81046	-0.76554 – 2.41700	0.309
new_cases	-0.00047	0.00004	-0.00055 – -0.00038	< 0.001
daily_hosp_occupancy	0.08849	0.00332	0.08198 – 0.09500	< 0.001
daily_ICU_occupancy	0.08542	0.02108	0.04404 – 0.12681	< 0.001
popData2020	-0.00001	0.00000	-0.00001 – -0.00000	< 0.001
FirstDoseAcc	-0.00002	0.00000	-0.00002 – -0.00002	< 0.001
SecondDoseAcc	0.00002	0.00000	0.00002 – 0.00003	< 0.001
ThirdDoseAcc	-0.00000	0.00000	-0.00000 – 0.00000	0.203
Observations	765			
R ² / R ² adjusted	0.923 / 0.914			

En aquest cas, es torna a veure que la variable *Positivity Rate* no és significativa, mentre que en el cas de les vacunes, la variable *Third Dose* no és significativa, cosa que en l'anterior model, sí que ho era. S'observa clarament que el model està millor explicat, ja que el coeficient de determinació és major, essent aquest un 0.923, i també és significatiu, ja que el *p-valor* és inferior a 0.05.

S'observa de nou, que la segona dosi continua tenint un coeficient positiu, però aquest cop una mica menor. En canvi, la primera i la tercera dosi segueixen tenint un coeficient negatiu. A l'estar tractant el model amb efectes fixos en el temps, és possible que la variable respecte a la tercera dosi no sigui significativa perquè aquesta comença a ser subministrada més tard que les altres dues. No obstant, ajuda a disminuir les morts al llarg del temps.

De nou, es realitza el *plmtest* en aquest nou model i el *p-valor* és menor que 0.05, per tant, es pot rebutjar la hipòtesi nul·la i acceptar que hi ha efectes fixos en el temps presents en el nostre model. A més a més, es calculen els estimadors en el temps per setmana, on es podrà veure la sortida del codi de l'*R* a l'annex. Allà, s'observa que, a diferència del model sense vacunes, els estimadors són tots positius.

Per tant, es pot confirmar de nou la presència d'efectes tant de país com de temps, per tant, s'estima un nou model *Two-Ways*.

Two-Way Effects Model Vaccine

En aquest darrer model, el qual s'anomena ***Two-Way Effects Model Vaccine***, s'indica en el mateix paquet l'efecte *twoways*.

Taula 5.17. Two-Way Effects Model Vaccine

Two-Ways Effects Model Vaccine					
<i>Predictors</i>	Dependent variable				
	<i>Estimates</i>	<i>std. Error</i>	<i>CI</i>	<i>p</i>	
tests_done	0.00007	0.00001	0.00004 – 0.00009	<0.001	
testing_rate	-0.00377	0.00091	-0.00556 – -0.00199	<0.001	
positivity_rate	-0.61342	0.66340	-1.91598 – 0.68914	0.355	
new_cases	-0.00023	0.00005	-0.00033 – -0.00013	<0.001	
daily_hosp_occupancy	0.12989	0.00380	0.12243 – 0.13735	<0.001	
daily_ICU_occupancy	-0.11993	0.02034	-0.15987 – -0.07998	<0.001	
FirstDoseAcc	-0.00002	0.00000	-0.00003 – -0.00002	<0.001	
SecondDoseAcc	0.00003	0.00000	0.00002 – 0.00003	<0.001	
ThirdDoseAcc	-0.00002	0.00000	-0.00002 – -0.00001	<0.001	
Observations	765				
R ² / R ² adjusted	0.907 / 0.895				

S'observa en la sortida del codi que, en aquest cas, la variable no significativa és *Positivity Rate*, és a dir, la mateixa variable que no eren significatives en el model *Country Effects Model Vaccine* i en el *Time Effects Model Vaccine*. No obstant això, té un coeficient de determinació de 0.907, bastant semblant als darrers models esmentats, i també és significatiu.

Abans de continuar, es mostren els resultats dels 3 models en una taula conjunta.

Taula 5.18. Taula amb els resultats d'R dels tres models

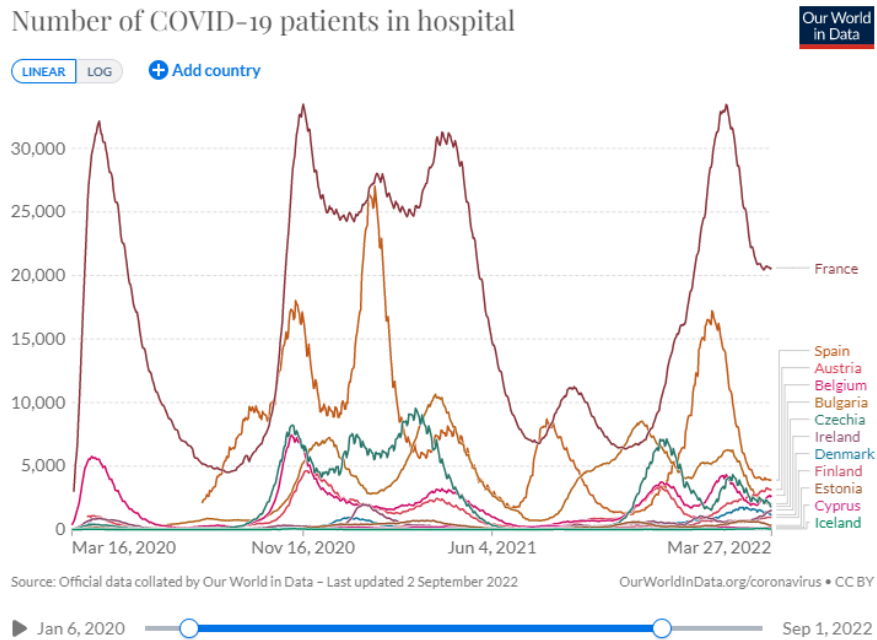
Predictors	Country Fixed Effects Vaccine				Time Fixed Effects Vaccine				Two-Ways Fixed Effects Vaccine			
	Estimates	std. Error	CI	p	Estimates	std. Error	CI	p	Estimates	std. Error	CI	p
tests_done	0.00008	0.00001	0.00006 – 0.00011	<0.001	0.00003	0.00001	0.00002 – 0.00005	<0.001	0.00007	0.00001	0.00004 – 0.00009	<0.001
testing_rate	-0.00522	0.00088	-0.00694 – -0.00350	<0.001	-0.00181	0.00059	-0.00296 – -0.00066	0.002	-0.00377	0.00091	-0.00556 – -0.00199	<0.001
positivity_rate	0.36427	0.46078	-0.54030 – 1.26885	0.429	0.82573	0.81046	-0.76554 – 2.41700	0.309	-0.61342	0.66340	-1.91598 – 0.68914	0.355
new_cases	-0.00027	0.00005	-0.00036 – -0.00018	<0.001	-0.00047	0.00004	-0.00055 – -0.00038	<0.001	-0.00023	0.00005	-0.00033 – -0.00013	<0.001
daily_hosp_occupancy	0.13335	0.00376	0.12597 – 0.14074	<0.001	0.08849	0.00332	0.08198 – 0.09500	<0.001	0.12989	0.00380	0.12243 – 0.13735	<0.001
daily_ICU_occupancy	-0.14371	0.01977	-0.18252 – -0.10490	<0.001	0.08542	0.02108	0.04404 – 0.12681	<0.001	-0.11993	0.02034	-0.15987 – -0.07998	<0.001
FirstDoseAcc	-0.00003	0.00000	-0.00003 – -0.00002	<0.001	-0.00002	0.00000	-0.00002 – -0.00002	<0.001	-0.00002	0.00000	-0.00003 – -0.00002	<0.001
SecondDoseAcc	0.00003	0.00000	0.00003 – 0.00003	<0.001	0.00002	0.00000	0.00002 – 0.00003	<0.001	0.00003	0.00000	0.00002 – 0.00003	<0.001
ThirdDoseAcc	-0.00002	0.00000	-0.00002 – -0.00001	<0.001	-0.00000	0.00000	-0.00000 – 0.00000	0.203	-0.00002	0.00000	-0.00002 – -0.00001	<0.001
popData2020					-0.00001	0.00000	-0.00001 – -0.00000	<0.001				
Observations	765				765				765			
R ² / R ² adjusted	0.915 / 0.913				0.923 / 0.914				0.907 / 0.895			

A simple vista, es veu que els tres models s'expliquen millor que els anteriors, on els quals no s'havien inclòs les vacunes. També es veu que els coeficients són més petits, i s'observa que el coeficient amb més pes és *Daily Hospitality Occupancy*, el qual ja ho era en els anteriors models acompanyat de les variables *Positivity Rate* i *Daily ICU Occupancy*.

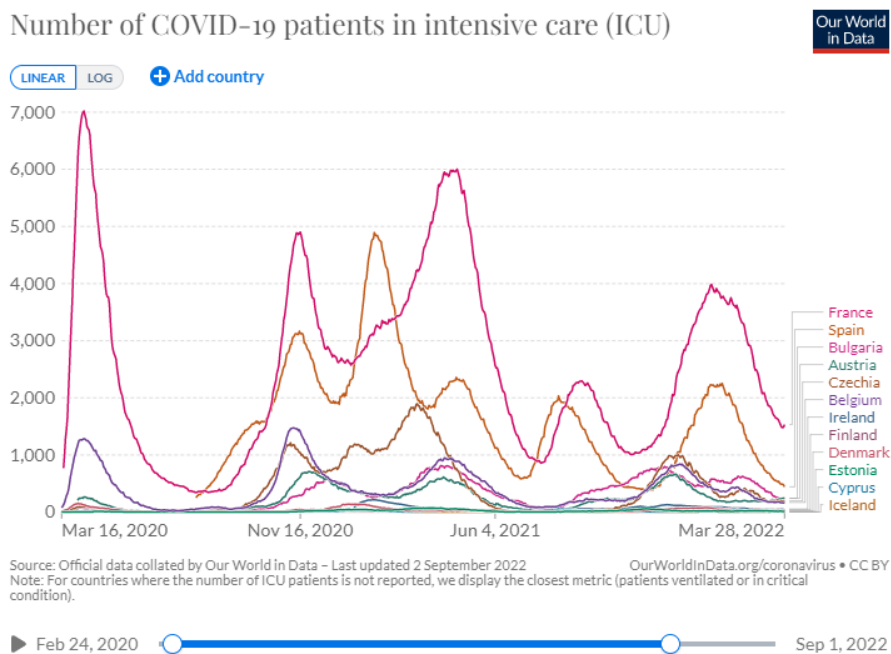
Per últim, s'observa que el model millor explicat és el *Time Effects Model Vaccine*, amb un 92,3%. En aquest, es pot observar que els coeficients més elevats són en les variables *Daily Hospitality Occupancy* i en *Daily ICU Occupancy*, és a dir, com major siguin les hospitalitzacions i els ingressats en unitats de cures intensives (UCI), major serà la taxa de morts. Aquest, és el model que millor està explicat i és on es veuen els coeficients més coherents per les variables, ja que la primera i la segona dosi tenen coeficient negatiu, pel que ens indica que les morts disminueixen i les hospitalitzacions, tant en hospitals com a les UCIs tenen coeficient positiu, pel que a mesura que aquestes augmenten, el nombre de morts també. Per últim, també s'observa que en aquest model, la variable referent a la població al 2020 és significativa, ja que en els altres dos models executats, aquesta variable quedava eliminada al realitzar el model, és a dir, el model no la trobava significativa.

Per a fer una representació de la realitat actual, a mitjans d'any del 2022, a continuació es veu una representació gràfica, dels ingressats en hospitals i en unitats de cures intensives (UCI), dels països de la Unió Europea que s'han estat tractant en aquests darrers models. Aquestes dades són agafades a partir del 27 de desembre del 2020, de la web oficial *Our World in Data* (Ritchie H, 2020).

Gràfic 5.2. Representació gràfica del nombre d'hospitalitzacions per Covid-19 (Ritchie H, 2020)



Gràfic 5.3. Representació gràfica del nombre d'hospitalitzacions a les UCI per Covid-19 (Ritchie H, 2020)



Detallar que, com en aquest treball s'està treballant amb el període de setmanes 12-2020 i 20-2022, informar que la setmana del 12 del 2020 comença el dia 16 de març i que la setmana 20 del 2022 acaba el 27 de març. Per tant, en la representació gràfica anterior, s'ha respectat aquest període.

6. Conclusions

Recordar que l'objectiu principal d'aquest treball de fi de grau és explicar el comportament de les dades de la *Covid-19*, específicament, dels morts deguts a aquesta pandèmia. En concret, l'objectiu és veure si les morts per aquest virus s'han comportat de manera diferent entre països i en el temps. Aquestes conclusions són realitzades sobre les dades d'un període total de 2 anys, d'entre les setmanes 12 del 2020 i la setmana 12 del 2022. Gràcies a les dades de panell i els models estadístics aplicats, s'han pogut extreure algunes conclusions que permeten explicar el comportament dels morts per aquest virus.

Amb aquesta petita introducció i els resultats dels models explicats anteriorment, es pot concloure que, a trets generals, s'ha obtingut molt bona explicació de tots els models, i gairebé totes les variables analitzades han estat significatives.

Malgrat la bona explicació en general, sí que s'ha vist una millora d'aquests models a l'hora d'afegir les vacunes, i és que, cal destacar que en aquesta modelització, s'ha tractat amb menys dades, ja que no disposàvem d'informació de tots els països de la Unió Europea. Per tant, aquest podria ser un efecte del per què l'explicació pot haver estat millor. No obstant, al realitzar el model amb efectes fixos en el temps, aquest ha estat explicat amb un 92,3%, el qual és molt bon resultat per un model amb tantes dades.

També, cal afegir que les variables amb més pes en l'explicació d'aquestes morts han estat les hospitalitzacions i els ingressos en unitats de cures intensives (UCI), ja que s'ha vist que aquests coeficients eren els més alts en els models. I destacar també, que al tractar-se de dades massives, és a dir al tractar-se de molts països i moltes setmanes, aquests coeficients han estat molt baixos i amb molt poca variació entre ells.

Respecte als models i a l'heterogeneïtat que es buscaven entre els països, s'ha pogut observar que en tots, aquesta era significativa. Això sí, també s'ha observat que en els models referents als efectes fixos els models estaven millor explicats. No obstant, tant els models d'efectes fixos per país, temps i d'ambdós efectes, s'ha comprovat que els països s'han comportat de manera diferent en el temps, ja que cap d'aquests donava significatiu en homogeneïtat. A més a més, s'ha comprovat que pels estimadors individuals de cada un, els països que han estat més afectats per les morts han estat Irlanda, Islàndia i Finlàndia.

No obstant això, sorprèn que països com Espanya i França no s'han vist tant afectats, ja que malgrat hagin tingut un volum major de morts, casos i hospitalitzacions, en proporció a la població del país, aquestes han estat menors. Justament, al tractar-se de països amb un nombre d'habitants superior, també disposa de més recursos mèdics. Per tant, malgrat que s'hagin vist uns nombres molt alts en el nombre de morts i de casos, aquests, en comparació amb països de menor població, no s'han vist tan afectats.

En conclusió, respecte als models d'efectes fixos en el temps i tenint en compte ambdós models, un explicat sense vacunes i l'altre explicat amb les vacunes, els models més ben explicats al llarg de tot l'estudi han estat aquests, i és que, les dades de la *Covid-19*, han estat molt variants en el temps. Això pot estar explicat per les diferents onades de casos que hem estat vivint aquests darrers dos anys, vistes gràficament en el primer anàlisi de les variables. També, s'ha vist certa millora a l'hora d'afegir les vacunes en aquest model, per tant, la conclusió és positiva, ja que aquestes han ajudat a fer caure el nombre de morts afectats pel *Coronavirus*.

Per últim, cal comentar que en aquest treball no ens s'ha volgut fer una predicció de les dades, sinó en entendre el comportament. Malgrat això, seria una bona opció fer-ne una segona part, de tal manera que es pugui ampliar aquesta informació i veure una comparació amb les dades reals. D'aquesta manera, es provaria l'eficàcia de la predicció dels models i podria ser útil de cara a entendre les onades que hem estat patint durant la pandèmia.

A més a més, en aquest treball s'ha treballat amb dades no balancejades, una altra possible millor d'aquest seria aplicar mètodes per balancejar les dades o trobar unes dades balancejades. Això permetria veure si un estudi amb les dades balancejades obtindria una millora en els models i en els resultats ja que d'aquesta manera potser els resultats dels models serien més significatius o les dades estarien millor explicades.

A nivell personal, ha estat tot un repte utilitzar dades de panell, ja que era un món completament desconegut per a mi, i a la vegada, molt interessant. Gràcies a una profunda recerca, he pogut descobrir un altre nou grup en el món de les dades i que sé segur, que serà molt útil en l'àmbit professional en un futur. El món de les dades no para de créixer i és important saber tractar-les per treure les millors conclusions i entendre el que està passant, sobretot en aspectes tan importants com una pandèmia que ha afectat globalment.

7. Bibliografía

Colonescu, C. (2016, 1 septiembre). Principles of Econometrics with R. Principles of Econometrics with R. <https://bookdown.org/ccolonescu/RPoE4/panel-data-models.html#mix-eqn-eqgrandefftest15>

COVID-19. (s. f.). European Centre for Disease Prevention and Control. <https://www.ecdc.europa.eu/en/covid-19>

COVID-19 vaccination. (s. f.). European Centre for Disease Prevention and Control. <https://www.ecdc.europa.eu/en/covid-19/prevention-and-control/vaccines>

Download COVID-19 data sets. (s. f.). European Centre for Disease Prevention and Control. <https://www.ecdc.europa.eu/en/covid-19/data>

E. (2021, 3 marzo). *Introduction to the Fundamentals of Panel Data*. Aptech. <https://www.aptech.com/blog/introduction-to-the-fundamentals-of-panel-data/#modeling-panel-data>

Econometrics Academy - Panel Data Models. (s. f.). Sites. <https://sites.google.com/site/econometricsacademy/econometrics-models/panel-data-models>

Fixed effects. (s. f.). Recuperado 5 de septiembre de 2022, de <https://ds4ps.org/PROG-EVAL-III/FixedEffects.html>

Gómez, R. S. (s. f.). Entorno de programación RStudio: regresión con datos de panel. Entorno de programación RStudio. http://www.ugr.es/%7Eromansg/material/WebEco/04-Eco2/Ordenador/R/03_PanelData.html

Introducción al análisis econométrico con datos de panel. (s. f.). CORE Reader. <https://core.ac.uk/reader/322620370>

How ECDC collects and processes COVID-19 data. (s. f.). European Centre for Disease Prevention and Control. <https://www.ecdc.europa.eu/en/covid-19/data-collection>

Modelos econométricos. Datos de panel | | UPV. (2015, 27 julio). YouTube. https://www.youtube.com/watch?v=kjwXXfDeLmA&ab_channel=UniversitatPolit%C3%A8cnicaUPV

Pacheco, J. (2021, 6 octubre). Dependencia temporal: Definición, Qué es y Ejemplos | 2022. Economía360. <https://www.economia360.org/dependencia-temporal/>

Panel data econometrics in R: (s. f.). Panel data econometrics in R: https://cran.r-project.org/web/packages/plm/vignettes/A_plmPackage.html

pFtest function - RDocumentation. (s. f.). RDocumentation. <https://www.rdocumentation.org/packages/plm/versions/1.6-5/topics/pFtest>

Research Gate. (s. f.). Cross Sectional Dependence in Panel Data Analysis. https://www.researchgate.net/publication/41494980_Cross-Sectional_Dependence_in_Panel_Data_Analysis

Ritchie, H. (2020, 5 marzo). Coronavirus (COVID-19) Hospitalizations. Our World in Data. <https://ourworldindata.org/covid-hospitalizations>

Ritchie, H. (2020, 5 marzo). Coronavirus (COVID-19) Vaccinations. Our World in Data. <https://ourworldindata.org/covid-vaccinations>

RPubs - R Tutorial: Panel Data Analysis 1. (2021, 15 enero). Panel Data Analysis 1. https://rpubs.com/phle/r_tutorial_panel_data_analysis

Santibáñez, J. (s. f.). Verificación del supuesto de homocedasticidad. Cursos Ciencias. http://sigma.iimas.unam.mx/jsantibanez/Cursos/Ciencias/2018_1/08_homocedasticidad.html

Science Direct. (s. f.). Econometrics and Finance. https://www.researchgate.net/publication/41494980_Cross-Sectional_Dependence_in_Panel_Data_Analysis

Siebert, M. (2019, 17 diciembre). Heterogeneity: what is it and why does it matter? Students 4 Best Evidence. <https://s4be.cochrane.org/blog/2018/11/29/what-is-heterogeneity/>

Staff, A. (2021, 3 agosto). A Timeline of COVID-19 Vaccine Developments in 2021. AJMC. <https://www.ajmc.com/view/a-timeline-of-covid-19-vaccine-developments-in-2021>

The Fixed Effects Regression Model For Panel Data Sets. (2022, 26 marzo). Time Series Analysis, Regression and Forecasting. <https://timeseriesreasoning.com/contents/the-fixed-effects-regression-model-for-panel-data-sets/>

tok.wiki. (2020). *Prueba de Breusch-Godfrey FondoyProcedimiento*. Prueba de Breusch-Godfrey FondoyProcedimiento. https://hmong.es/wiki/Breusch%E2%80%93Godfrey_test

Wagle, M. (2022, 6 enero). *How to deal with Panel Data. Practical Application using R*. Medium. <https://medium.com/@manilwagle/how-to-deal-with-panel-data-practical-application-using-r-18ef95ae99c6>

Zaghdoudi, T. (2019, 1 mayo). Dynamic Panel Data Models. Dynamic Panel Data Models.
<https://rdr.io/cran/dynpanel/man/dynpanel-package.html#heading-1>

Bases de dades

Vaccine:

https://opendata.ecdc.europa.eu/covid19/vaccine_tracker/csv/data.csv

Weekcase:

<https://opendata.ecdc.europa.eu/covid19/subnationalcaseweekly/csv>

Hospital:

<https://opendata.ecdc.europa.eu/covid19/hospitalicuadmissionrates/csv/data.csv>

Test:

<https://opendata.ecdc.europa.eu/covid19/testing/csv>

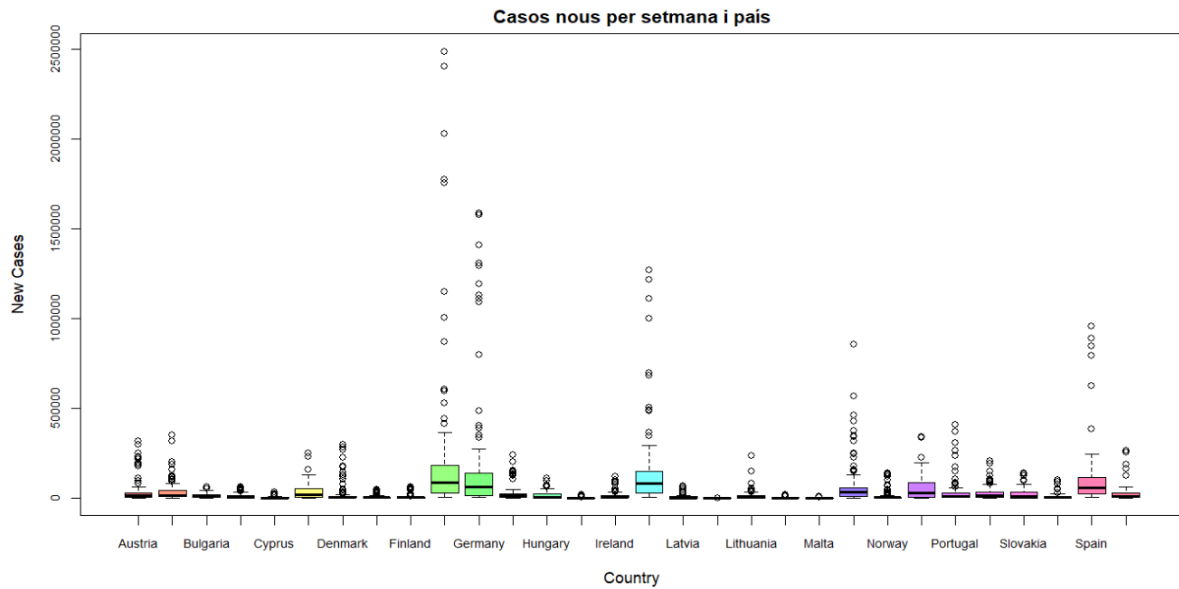
Deaths:

https://opendata.ecdc.europa.eu/covid19/nationalcasedeath_eueea_daily_ei/csv

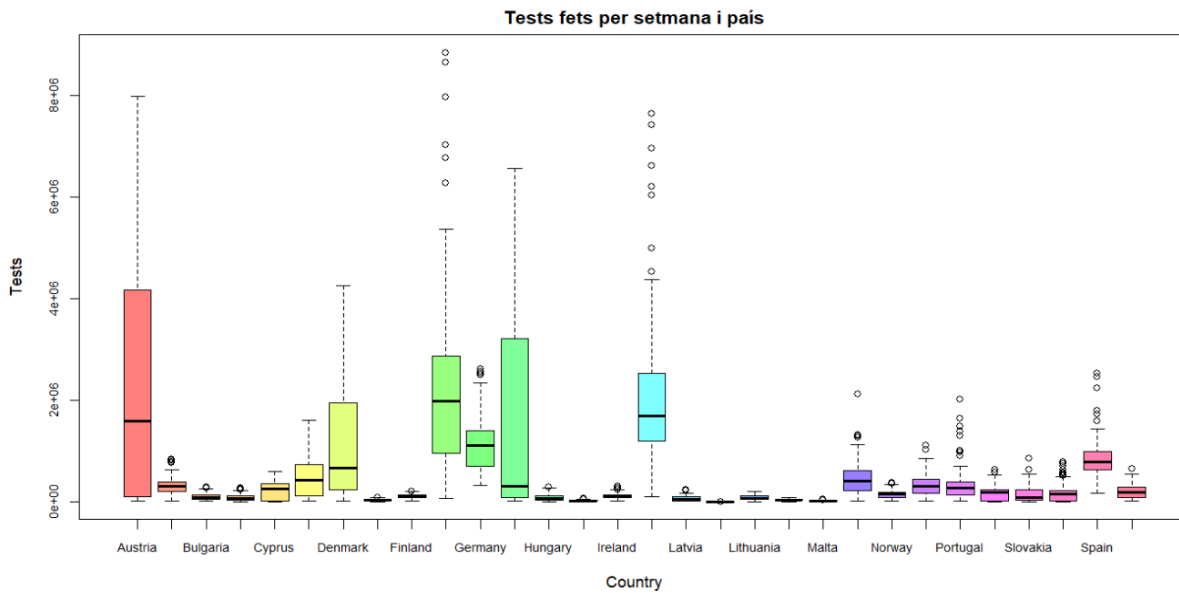
8. Annex

8.1. Representacions gràfiques

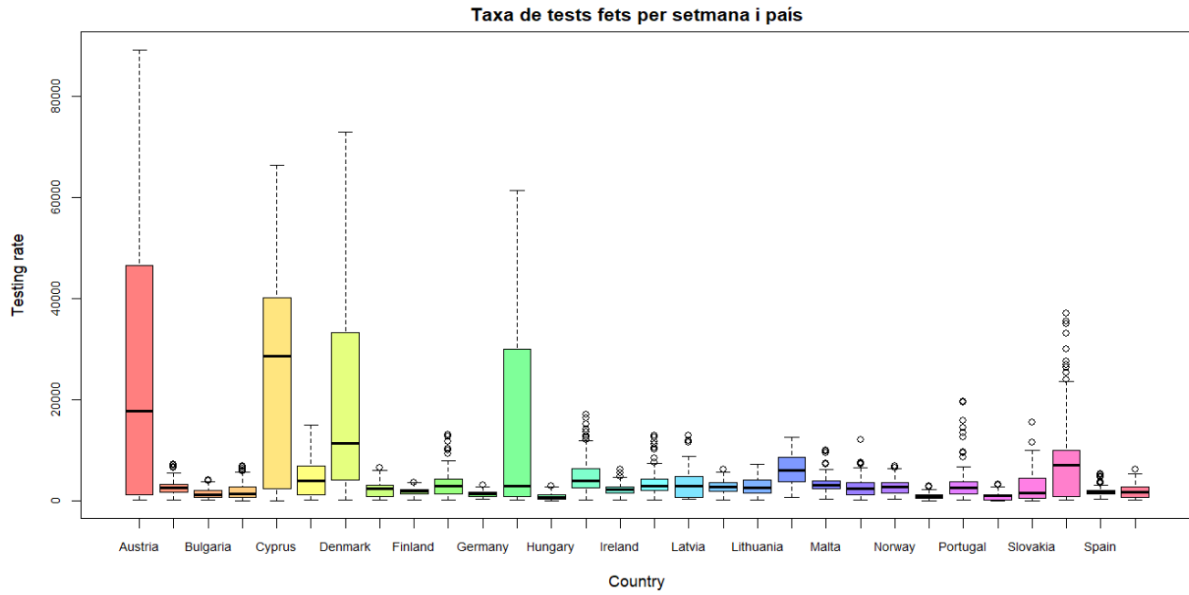
Gràfic 8.1.1 Representació gràfica del nombre de nous casos



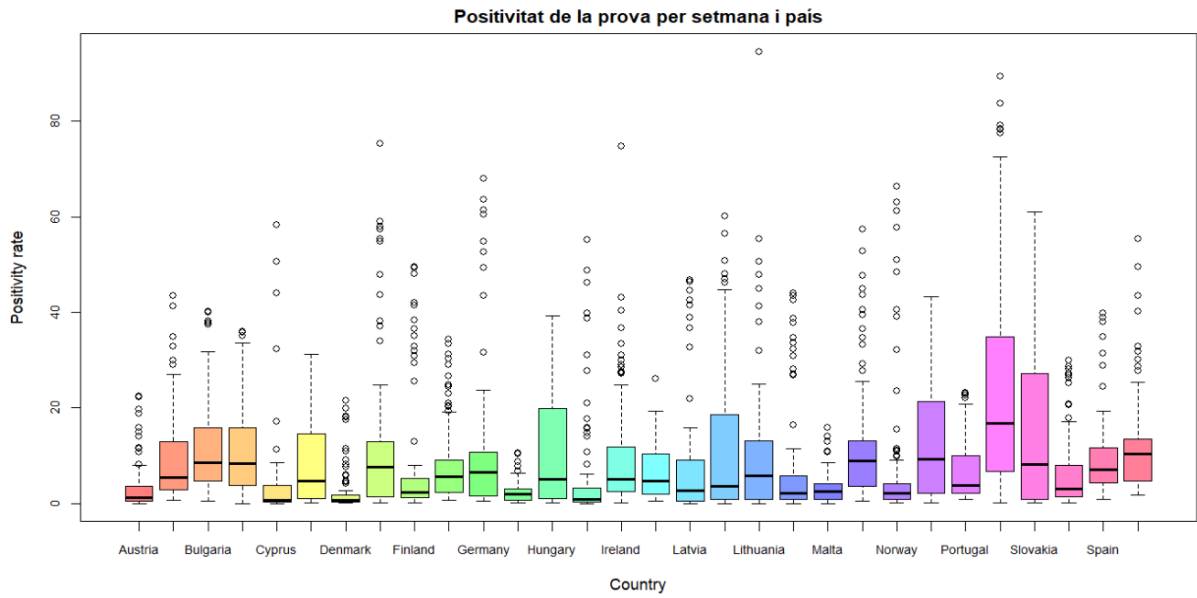
Gràfic 8.1.2 Representació gràfica del nombre de test fets



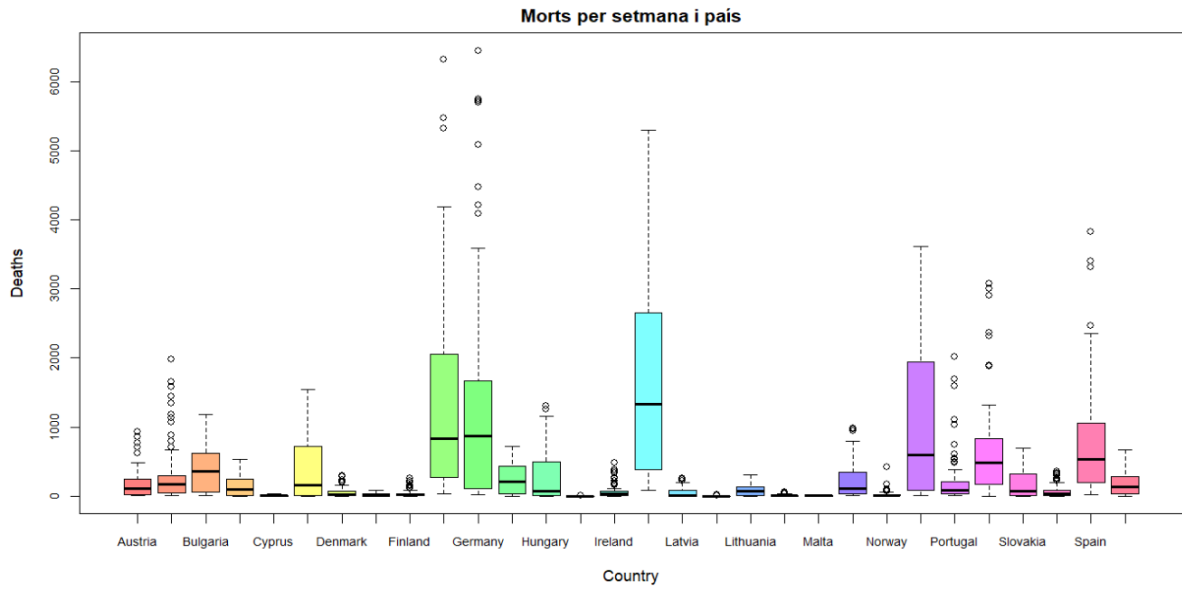
Gràfic 8.1.3 Representació gràfica de la taxa de tests realitzats



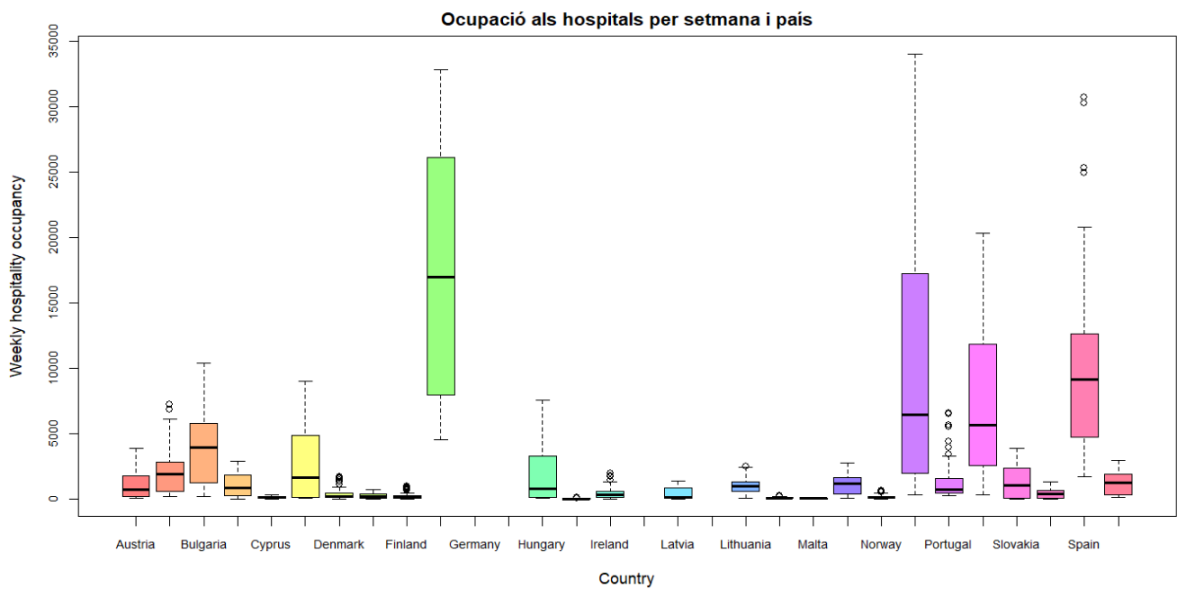
Gràfic 8.1.4 Representació gràfica del nombre de la taxa de població positiva



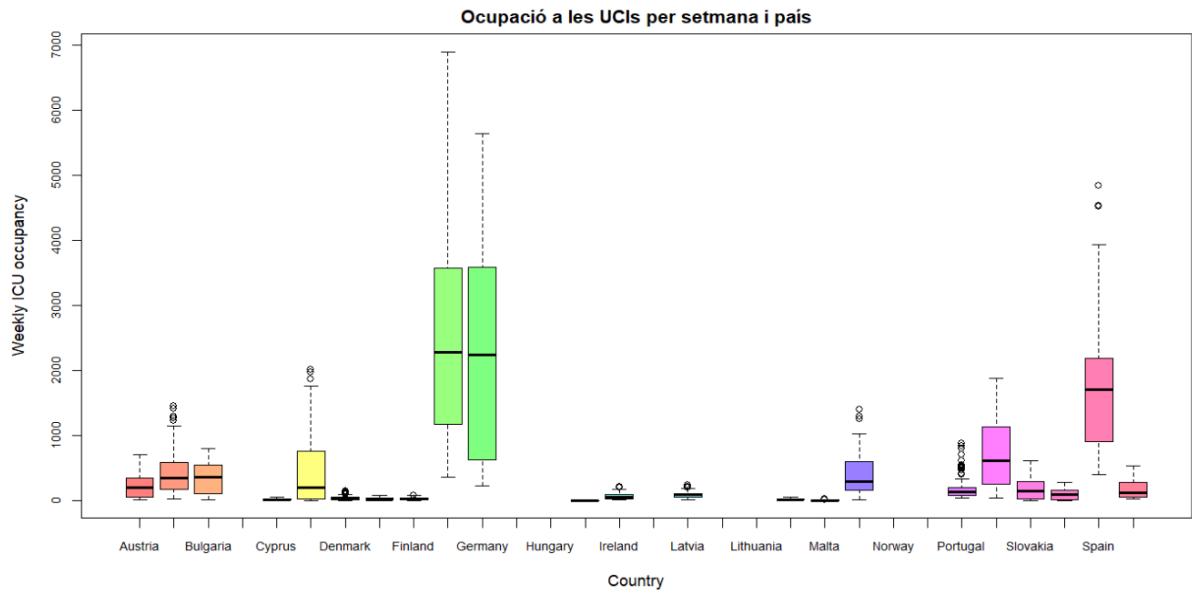
Gràfic 8.1.5 Representació gràfica del nombre de morts



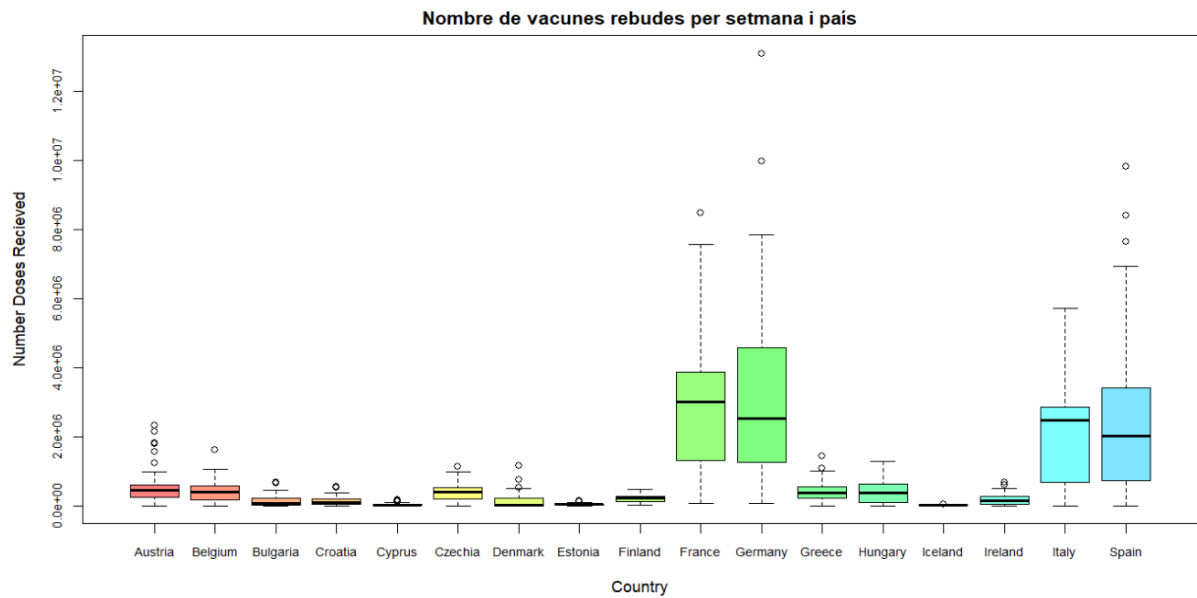
Gràfic 8.1.6. Representació gràfica de l'hospitalització



Gràfic 8.1.7. Representació gràfica de l'hospitalització a les UCIs

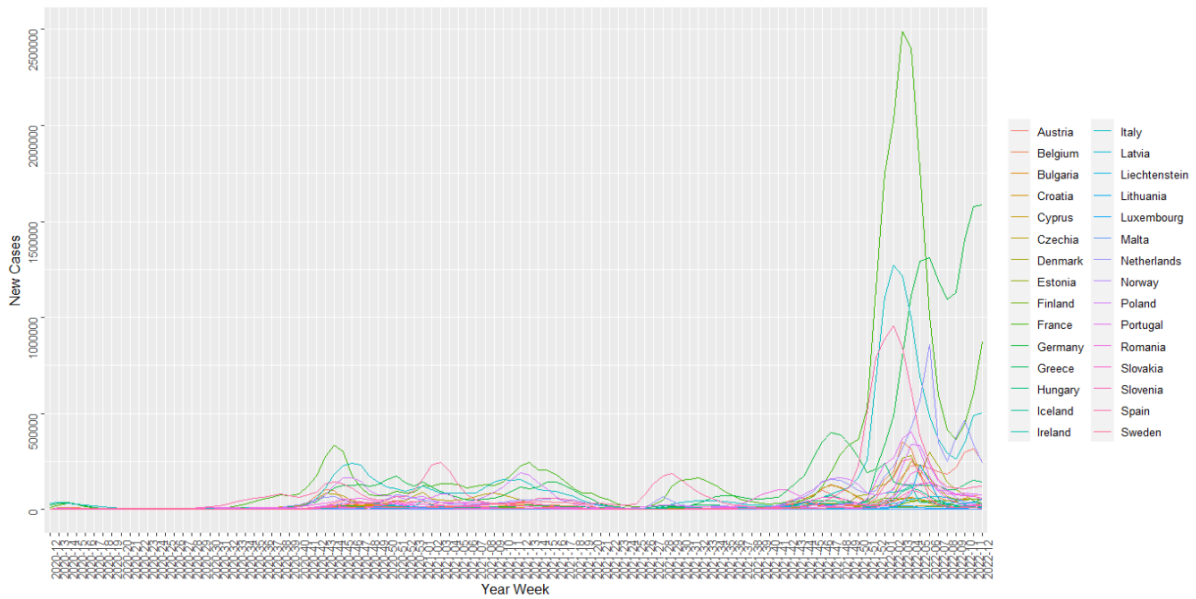


Gràfic 8.1.8. Representació gràfica del nombre vacunes rebudes per subministradors

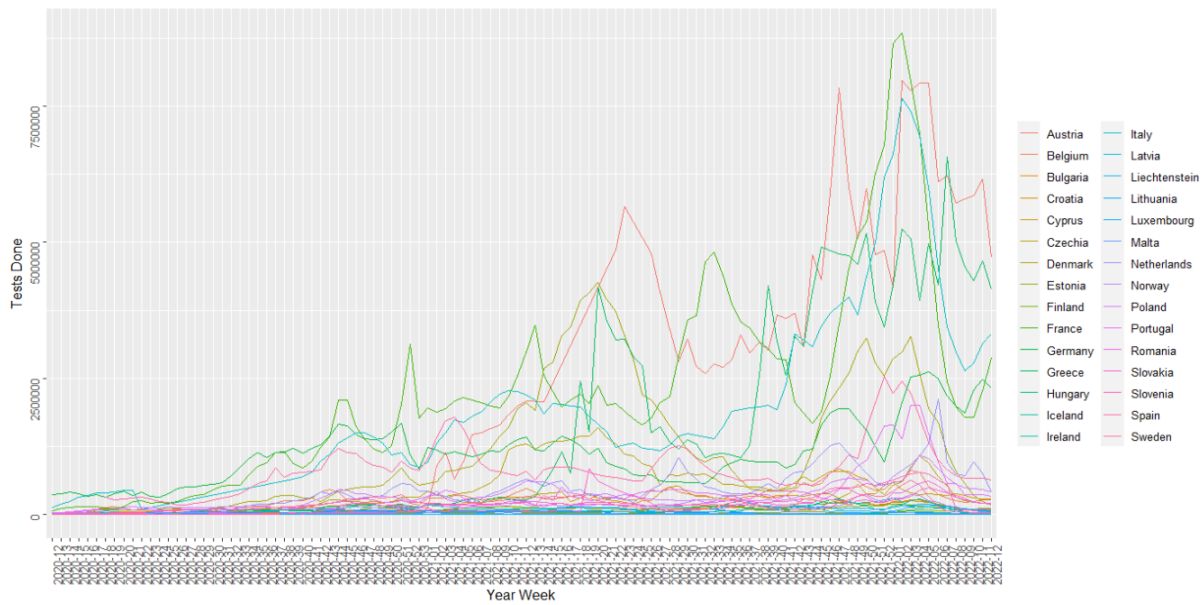


8.2. Representacions gràfiques en el temps de les variables per setmana i país

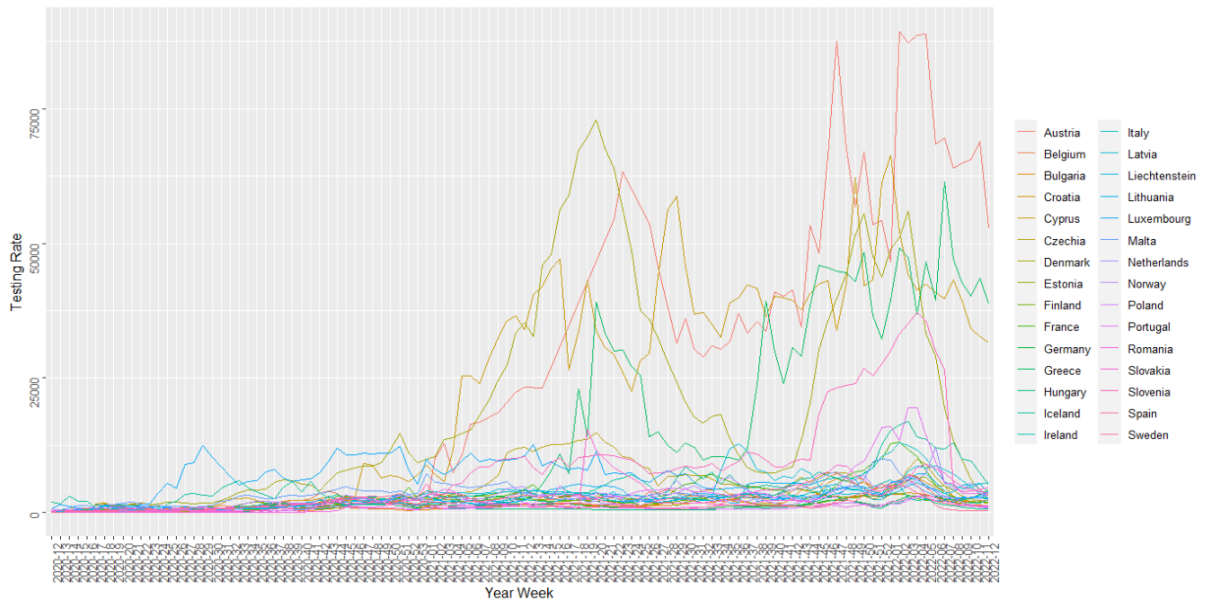
Gràfic 8.2.1. Representació gràfica en el temps del nombre de nous casos



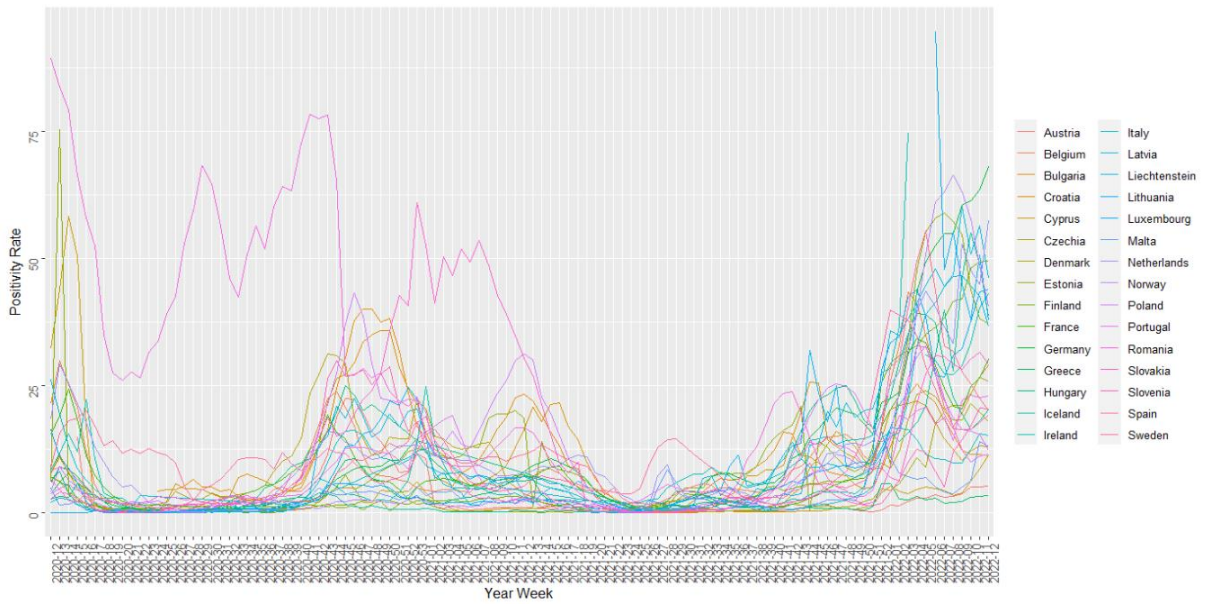
Gràfic 8.2.2. Representació gràfica en el temps del nombre de tests



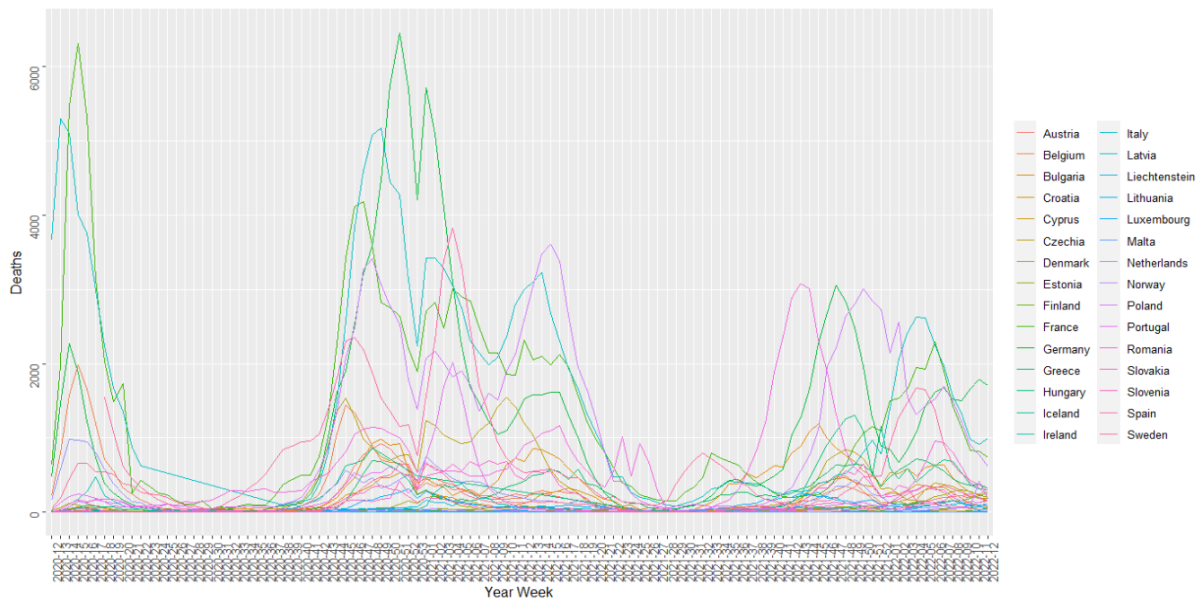
Gràfic 8.2.3. Representació gràfica en el temps de la taxa de tests



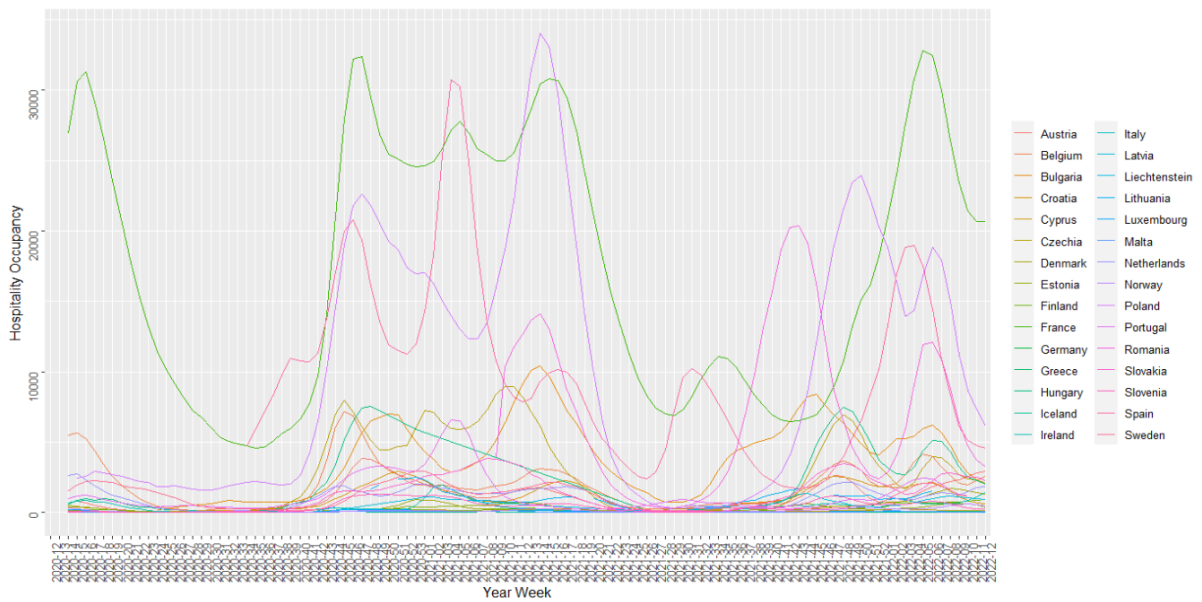
Gràfic 8.2.4. Representació gràfica en el temps de la taxa de positivitat



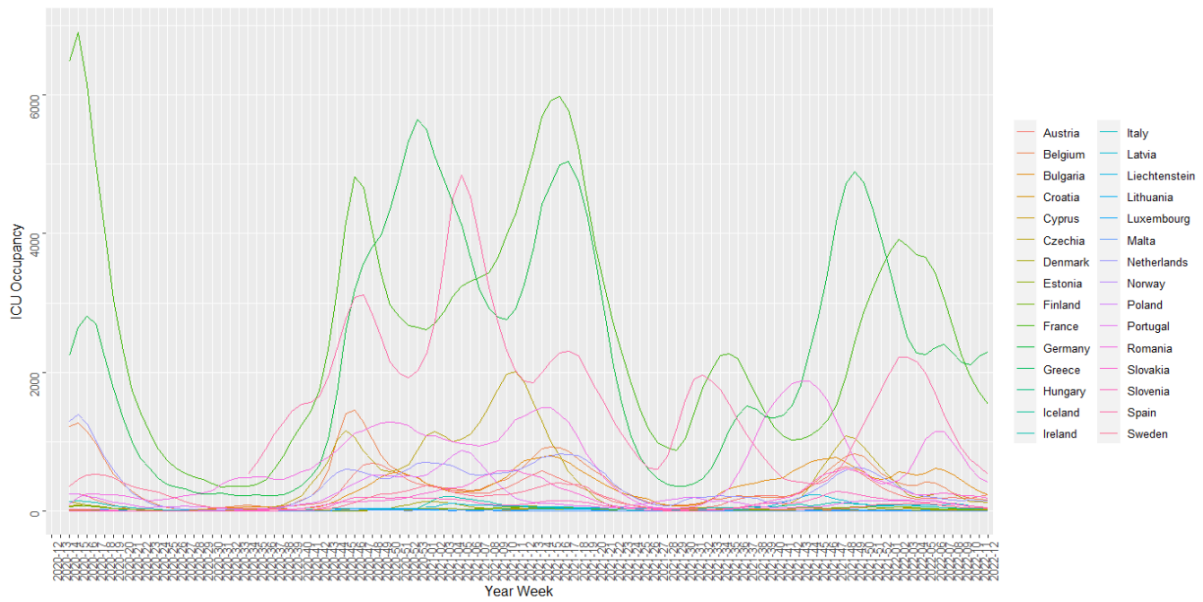
Gràfic 8.2.5. Representació gràfica en el temps del nombre de morts



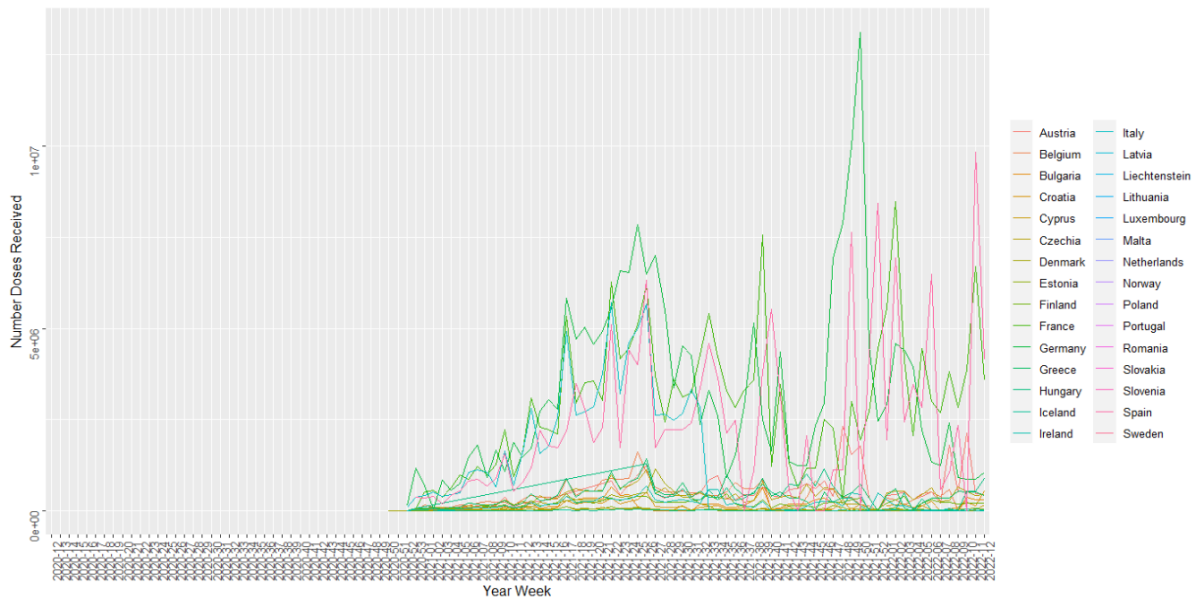
Gràfic 8.2.6. Representació gràfica en el temps del nombre d'hospitalitzacions



Gràfic 8.2.7. Representació gràfica en el temps del nombre d'hospitalitzacions a les UCIs



Gràfic 8.2.8. Representació gràfica en el temps del nombre de vacunes rebudes per subministradors



Imatge 8.2.1. Sortida codi *R* de la funció *fixef* per setmana del model *Time Effects Model*

2020-14	2020-15	2020-16	2020-17	2020-18	2020-19	2020-20	2020-21
230.1731	291.0789	244.0358	131.1425	53.6567	38.2216	62.4505	-7.6741
2020-22	2020-23	2020-24	2020-25	2020-26	2020-27	2020-28	2020-29
18.6827	24.2116	31.6973	37.5499	38.2797	40.7653	44.8948	48.1697
2020-30	2020-31	2020-32	2020-33	2020-34	2020-35	2020-36	2020-37
46.4446	50.2532	49.9593	52.9294	62.4822	59.7695	55.9562	57.1981
2020-38	2020-39	2020-40	2020-41	2020-42	2020-43	2020-44	2020-45
55.4227	46.4618	45.8398	37.3967	31.8613	54.5084	94.7641	160.7537
2020-46	2020-47	2020-48	2020-49	2020-50	2020-51	2020-52	2020-53
160.7895	158.9977	171.5507	149.3288	153.4293	168.6449	117.2939	-32.5395
2021-01	2021-02	2021-03	2021-04	2021-05	2021-06	2021-07	2021-08
162.2102	199.7184	213.8952	212.1905	138.2522	88.2782	32.3153	13.8184
2021-09	2021-10	2021-11	2021-12	2021-13	2021-14	2021-15	2021-16
1.8645	-32.7557	-64.8599	-69.1674	-89.8286	-113.0138	-126.9479	-120.6429
2021-17	2021-18	2021-19	2021-20	2021-21	2021-22	2021-23	2021-24
-125.6976	-122.2143	-106.8379	-91.2915	-67.5189	-16.6056	20.2599	23.0499
2021-25	2021-26	2021-27	2021-28	2021-29	2021-30	2021-31	2021-32
57.3082	55.5204	41.7612	25.9550	23.8135	17.0351	12.3350	11.6479
2021-33	2021-34	2021-35	2021-36	2021-37	2021-38	2021-39	2021-40
16.6393	8.6404	6.5614	6.1012	13.5343	11.7181	22.4291	42.1308
2021-41	2021-42	2021-43	2021-44	2021-45	2021-46	2021-47	2021-48
54.4346	73.9907	85.6562	94.9213	80.1055	64.0459	55.6143	38.8056
2021-49	2021-50	2021-51	2021-52	2022-01	2022-02	2022-03	2022-04
29.2697	24.8522	-2.1580	-56.3450	-14.1539	-19.4159	-22.6875	-27.5821
2022-05	2022-06	2022-07	2022-08	2022-09	2022-10	2022-11	2022-12
-47.5888	-26.0928	-46.2607	-55.0527	-50.7319	-54.4893	-77.3201	-74.1190

Imatge 8.2.1. Sortida codi *R* de la funció *fixef* per setmana del model *Time Effects Model Vaccine*

2020-50	2020-51	2020-52	2020-53	2021-01	2021-02	2021-03	2021-04	2021-05	2021-06
76.1929	113.7962	115.1601	26.4872	254.0403	302.4086	304.1035	303.5176	243.0038	196.5462
2021-07	2021-08	2021-09	2021-10	2021-11	2021-12	2021-13	2021-14	2021-15	2021-16
142.3851	131.5261	125.2996	111.5117	86.1677	87.2237	55.5895	33.6863	9.0549	17.4843
2021-17	2021-18	2021-19	2021-20	2021-21	2021-22	2021-23	2021-24	2021-25	2021-26
8.4041	5.9743	9.7013	26.6615	45.0713	72.5795	74.3248	104.3337	102.9196	98.9768
2021-27	2021-28	2021-29	2021-30	2021-31	2021-32	2021-33	2021-34	2021-35	2021-36
95.6062	86.9293	80.2363	75.0820	72.8719	76.2495	84.8683	68.0909	67.1087	64.6662
2021-37	2021-38	2021-39	2021-40	2021-41	2021-42	2021-43	2021-44	2021-45	2021-46
72.8832	64.5754	70.8292	74.7892	75.5926	90.7752	96.6434	111.0153	130.4216	124.4542
2021-47	2021-48	2021-49	2021-50	2021-51	2021-52	2022-01	2022-02	2022-03	2022-04
116.8876	116.6616	108.5281	91.9025	82.7677	49.3419	127.2380	112.7518	123.3435	128.6529
2022-05	2022-06	2022-07	2022-08	2022-09	2022-10	2022-11	2022-12		
92.9757	109.3721	67.1491	36.6570	44.0575	37.3468	30.4269	35.1864		

8.3. Codi

Dades

```
## Script amb les BBDD amb dades setmanals que utilitzarem

- Vaccine (sumatori de vacunes)
- Weekcase --> sumar-los per setmanes
- Hospital
- Test
- deaths

**Base de dades VACINNE**

```{r}
vacinne <-
read.csv("https://opendata.ecdc.europa.eu/covid19/vaccine_tracker/cs
v/data.csv", na.strings = "", fileEncoding = "UTF-8-BOM")
summary(vacinne)

vacinne$YearWeekISO <- gsub("W", "", vacinne$YearWeekISO,
ignore.case = TRUE)

vacinne <- subset.data.frame(vacinne, vacinne$TargetGroup == 'ALL',
drop = FALSE)

library(sqldf)

vacinne <- sqldf(" SELECT YearWeekISO, ReportingCountry,
SUM(NumberDosesReceived) as NumberDosesReceived, SUM(FirstDose) as
FirstDose, SUM(SUM(FirstDose)) OVER (ORDER BY ReportingCountry,
YearWeekISO) AS FirstDoseAcc,
SUM(SecondDose) as SecondDose, SUM(SUM(SecondDose)) OVER (ORDER BY
ReportingCountry, YearWeekISO) AS SecondDoseAcc,
SUM(DoseAdditional1) as ThirdDose, SUM(SUM(DoseAdditional1)) OVER
(ORDER BY ReportingCountry, YearWeekISO) AS ThirdDoseAcc, Population
FROM vacinne
GROUP BY YearWeekISO, ReportingCountry, Population")

vacinne

library(sqldf)
vacinne_prova <- sqldf("SELECT YearWeekISO, ReportingCountry,
FirstDose,
SUM(SUM(FirstDose)) OVER (ORDER BY ReportingCountry,
YearWeekISO) AS FirstDoseAcc, SecondDose,
```



```

 SUM(SUM(SecondDose)) OVER (ORDER BY ReportingCountry,
YearWeekISO) AS SecondDoseAcc, ThirdDose,
 SUM(SUM(ThirdDose)) OVER (ORDER BY ReportingCountry,
YearWeekISO) AS ThirdDoseAcc
 FROM vacinne
 GROUP BY YearWeekISO, ReportingCountry")

```

```

vacinne_prova
```

```

```

**Base de dades WEEKCASE**

```

```

```{r}

```

```

weekcase<-

```

```

read.csv("https://opendata.ecdc.europa.eu/covid19/subnationalcaseweekly/csv", stringsAsFactors=T, na.strings = "", fileEncoding = "UTF-8-BOM")

```

```

weekcase$country <- gsub("Austria", "AT", weekcase$country, ignore.case = TRUE)
weekcase$country <- gsub("Belgium", "BE", weekcase$country, ignore.case = TRUE)
weekcase$country <- gsub("Bulgaria", "BG", weekcase$country, ignore.case = TRUE)
weekcase$country <- gsub("Cyprus", "CY", weekcase$country, ignore.case = TRUE)
weekcase$country <- gsub("Czechia", "CZ", weekcase$country, ignore.case = TRUE)
weekcase$country <- gsub("Germany", "DE", weekcase$country, ignore.case = TRUE)
weekcase$country <- gsub("Denmark", "DK", weekcase$country, ignore.case = TRUE)
weekcase$country <- gsub("Estonia", "EE", weekcase$country, ignore.case = TRUE)
weekcase$country <- gsub("Greece", "EL", weekcase$country, ignore.case = TRUE)
weekcase$country <- gsub("Spain", "ES", weekcase$country, ignore.case = TRUE)
weekcase$country <- gsub("Finland", "FI", weekcase$country, ignore.case = TRUE)
weekcase$country <- gsub("France", "FR", weekcase$country, ignore.case = TRUE)
weekcase$country <- gsub("Croatia", "HR", weekcase$country, ignore.case = TRUE)
weekcase$country <- gsub("Hungary", "HU", weekcase$country, ignore.case = TRUE)
weekcase$country <- gsub("Ireland", "IE", weekcase$country, ignore.case = TRUE)

```

```

weekcase$country <- gsub("Iceland", "IS", weekcase$country,
ignore.case = TRUE)
weekcase$country <- gsub("Italy", "IT", weekcase$country,
ignore.case = TRUE)
weekcase$country <- gsub("Liechtenstein", "LI", weekcase$country,
ignore.case = TRUE)
weekcase$country <- gsub("Lithuania", "LT", weekcase$country,
ignore.case = TRUE)
weekcase$country <- gsub("Luxembourg", "LU", weekcase$country,
ignore.case = TRUE)
weekcase$country <- gsub("Latvia", "LV", weekcase$country,
ignore.case = TRUE)
weekcase$country <- gsub("Malta", "MT", weekcase$country,
ignore.case = TRUE)
weekcase$country <- gsub("Netherlands", "NL", weekcase$country,
ignore.case = TRUE)
weekcase$country <- gsub("Norway", "NO", weekcase$country,
ignore.case = TRUE)
weekcase$country <- gsub("Poland", "PL", weekcase$country,
ignore.case = TRUE)
weekcase$country <- gsub("Portugal", "PT", weekcase$country,
ignore.case = TRUE)
weekcase$country <- gsub("Romania", "RO", weekcase$country,
ignore.case = TRUE)
weekcase$country <- gsub("Sweden", "SE", weekcase$country,
ignore.case = TRUE)
weekcase$country <- gsub("Slovenia", "SI", weekcase$country,
ignore.case = TRUE)
weekcase$country <- gsub("Slovakia", "SK", weekcase$country,
ignore.case = TRUE)

weekcase <- sqldf("SELECT year_week, country, SUM(weekly_count) as
weekly_count, SUM(population) as population
FROM weekcase
GROUP BY year_week, country")

weekcase
```



```

Base de dades HOSPITAL

```{r}
hospital<-
read.csv("https://opendata.ecdc.europa.eu/covid19/hospitalicuadmissi
onrates/csv/data.csv", stringsAsFactors=T, na.strings = "",
fileEncoding = "UTF-8-BOM")

summary(hospital)

```


```

```

hospital$year_week <- gsub("W", "", hospital$year_week, ignore.case
= TRUE)

hospital$country <- gsub("Austria", "AT", hospital$country,
ignore.case = TRUE)
hospital$country <- gsub("Belgium", "BE", hospital$country,
ignore.case = TRUE)
hospital$country <- gsub("Bulgaria", "BG", hospital$country,
ignore.case = TRUE)
hospital$country <- gsub("Cyprus", "CY", hospital$country,
ignore.case = TRUE)
hospital$country <- gsub("Czechia", "CZ", hospital$country,
ignore.case = TRUE)
hospital$country <- gsub("Germany", "DE", hospital$country,
ignore.case = TRUE)
hospital$country <- gsub("Denmark", "DK", hospital$country,
ignore.case = TRUE)
hospital$country <- gsub("Estonia", "EE", hospital$country,
ignore.case = TRUE)
hospital$country <- gsub("Greece", "EL", hospital$country,
ignore.case = TRUE)
hospital$country <- gsub("Spain", "ES", hospital$country,
ignore.case = TRUE)
hospital$country <- gsub("Finland", "FI", hospital$country,
ignore.case = TRUE)
hospital$country <- gsub("France", "FR", hospital$country,
ignore.case = TRUE)
hospital$country <- gsub("Croatia", "HR", hospital$country,
ignore.case = TRUE)
hospital$country <- gsub("Hungary", "HU", hospital$country,
ignore.case = TRUE)
hospital$country <- gsub("Ireland", "IE", hospital$country,
ignore.case = TRUE)
hospital$country <- gsub("Iceland", "IS", hospital$country,
ignore.case = TRUE)
hospital$country <- gsub("Italy", "IT", hospital$country,
ignore.case = TRUE)
hospital$country <- gsub("Liechtenstein", "LI", hospital$country,
ignore.case = TRUE)
hospital$country <- gsub("Lithuania", "LT", hospital$country,
ignore.case = TRUE)
hospital$country <- gsub("Luxembourg", "LU", hospital$country,
ignore.case = TRUE)
hospital$country <- gsub("Latvia", "LV", hospital$country,
ignore.case = TRUE)
hospital$country <- gsub("Malta", "MT", hospital$country,
ignore.case = TRUE)

```

```

hospital$country <- gsub("Netherlands", "NL", hospital$country,
ignore.case = TRUE)
hospital$country <- gsub("Norway", "NO", hospital$country,
ignore.case = TRUE)
hospital$country <- gsub("Poland", "PL", hospital$country,
ignore.case = TRUE)
hospital$country <- gsub("Portugal", "PT", hospital$country,
ignore.case = TRUE)
hospital$country <- gsub("Romania", "RO", hospital$country,
ignore.case = TRUE)
hospital$country <- gsub("Sweden", "SE", hospital$country,
ignore.case = TRUE)
hospital$country <- gsub("Slovenia", "SI", hospital$country,
ignore.case = TRUE)
hospital$country <- gsub("Slovakia", "SK", hospital$country,
ignore.case = TRUE)

hospital$date2 <- as.Date(hospital$date)
hospital <- hospital[hospital$date2 >= "2020-04-01" & hospital$date2
<= "2022-03-27",]

hospital_1 <- subset.data.frame(hospital, hospital$indicator ==
"Daily hospital occupancy", drop = FALSE)
hospital_2 <- subset.data.frame(hospital, hospital$indicator ==
"Daily ICU occupancy", drop = FALSE)

library(sqldf)

hospital_1 <- sqldf(" SELECT country, year_week, AVG(value) as
daily_hosp_occupancy
FROM hospital_1
GROUP BY country, year_week")
hospital_1

hospital_2 <- sqldf(" SELECT country, year_week, AVG(value) as
daily_ICU_occupancy
FROM hospital_2
GROUP BY country, year_week")
hospital_2
```


**Base de dades TEST**



```

```{r}
test<-
read.csv("https://opendata.ecdc.europa.eu/covid19/testing/csv",
stringsAsFactors=T, na.strings = "", fileEncoding = "UTF-8-BOM")

```


```

```

test$year_week <- gsub("W", "", test$year_week, ignore.case = TRUE)

test <- subset.data.frame(test, test$level == 'national', drop =
FALSE)

test <- sqldf("SELECT country, country_code, year_week, new_cases,
tests_done, population, testing_rate, positivity_rate
FROM test")

test
```

```

****Base de dades DEATHS****

```

```{r}
deaths <-
read.csv("https://opendata.ecdc.europa.eu/covid19/nationalcasedeath_
eueea_daily_ei/csv", na.strings = "",
stringsAsFactors=T,fileEncoding = "UTF-8-BOM")

```

```

deaths$countriesAndTerritories <- gsub("Austria", "AT",
deaths$countriesAndTerritories, ignore.case = TRUE)
deaths$countriesAndTerritories <- gsub("Belgium", "BE",
deaths$countriesAndTerritories, ignore.case = TRUE)
deaths$countriesAndTerritories <- gsub("Bulgaria", "BG",
deaths$countriesAndTerritories, ignore.case = TRUE)
deaths$countriesAndTerritories <- gsub("Cyprus", "CY",
deaths$countriesAndTerritories, ignore.case = TRUE)
deaths$countriesAndTerritories <- gsub("Czechia", "CZ",
deaths$countriesAndTerritories, ignore.case = TRUE)
deaths$countriesAndTerritories <- gsub("Germany", "DE",
deaths$countriesAndTerritories, ignore.case = TRUE)
deaths$countriesAndTerritories <- gsub("Denmark", "DK",
deaths$countriesAndTerritories, ignore.case = TRUE)
deaths$countriesAndTerritories <- gsub("Estonia", "EE",
deaths$countriesAndTerritories, ignore.case = TRUE)
deaths$countriesAndTerritories <- gsub("Greece", "EL",
deaths$countriesAndTerritories, ignore.case = TRUE)
deaths$countriesAndTerritories <- gsub("Spain", "ES",
deaths$countriesAndTerritories, ignore.case = TRUE)
deaths$countriesAndTerritories <- gsub("Finland", "FI",
deaths$countriesAndTerritories, ignore.case = TRUE)
deaths$countriesAndTerritories <- gsub("France", "FR",
deaths$countriesAndTerritories, ignore.case = TRUE)
deaths$countriesAndTerritories <- gsub("Croatia", "HR",
deaths$countriesAndTerritories, ignore.case = TRUE)

```

```

deaths$countriesAndTerritories <- gsub("Hungary", "HU",
deaths$countriesAndTerritories, ignore.case = TRUE)
deaths$countriesAndTerritories <- gsub("Ireland", "IE",
deaths$countriesAndTerritories, ignore.case = TRUE)
deaths$countriesAndTerritories <- gsub("Iceland", "IS",
deaths$countriesAndTerritories, ignore.case = TRUE)
deaths$countriesAndTerritories <- gsub("Italy", "IT",
deaths$countriesAndTerritories, ignore.case = TRUE)
deaths$countriesAndTerritories <- gsub("Liechtenstein", "LI",
deaths$countriesAndTerritories, ignore.case = TRUE)
deaths$countriesAndTerritories <- gsub("Lithuania", "LT",
deaths$countriesAndTerritories, ignore.case = TRUE)
deaths$countriesAndTerritories <- gsub("Luxembourg", "LU",
deaths$countriesAndTerritories, ignore.case = TRUE)
deaths$countriesAndTerritories <- gsub("Latvia", "LV",
deaths$countriesAndTerritories, ignore.case = TRUE)
deaths$countriesAndTerritories <- gsub("Malta", "MT",
deaths$countriesAndTerritories, ignore.case = TRUE)
deaths$countriesAndTerritories <- gsub("Netherlands", "NL",
deaths$countriesAndTerritories, ignore.case = TRUE)
deaths$countriesAndTerritories <- gsub("Norway", "NO",
deaths$countriesAndTerritories, ignore.case = TRUE)
deaths$countriesAndTerritories <- gsub("Poland", "PL",
deaths$countriesAndTerritories, ignore.case = TRUE)
deaths$countriesAndTerritories <- gsub("Portugal", "PT",
deaths$countriesAndTerritories, ignore.case = TRUE)
deaths$countriesAndTerritories <- gsub("Romania", "RO",
deaths$countriesAndTerritories, ignore.case = TRUE)
deaths$countriesAndTerritories <- gsub("Sweden", "SE",
deaths$countriesAndTerritories, ignore.case = TRUE)
deaths$countriesAndTerritories <- gsub("Slovenia", "SI",
deaths$countriesAndTerritories, ignore.case = TRUE)
deaths$countriesAndTerritories <- gsub("Slovakia", "SK",
deaths$countriesAndTerritories, ignore.case = TRUE)

deaths$date2 <- as.Date(deaths$dateRep, format = "%d/%m/%Y")
deaths <- deaths[deaths$date2 >= "2020/03/16" & deaths$date2 <=
"2022/03/27",]

deaths$dateRep <- factor(c(deaths$dateRep))
deaths$dateRep <- as.Date(deaths$dateRep, format = "%d/%m/%Y")
deaths$dateRep <- strftime(c(deaths$dateRep), format = "%Y-%V")

deaths <- sqldf("SELECT dateRep, SUM(cases) as cases, SUM(deaths) as
deaths, countriesAndTerritories, popData2020
FROM deaths

```

```

GROUP BY dateRep, countriesAndTerritories,
popData2020")

deaths
` ``

Modificacions

Corespondències de country_code amb country

"AT" Austria
"BE" Belgium
"BG" Bulgaria
"CY" Cyprus
"CZ" Czechia
"DE" Germany
"DK" Denmark
"EE" Estonia
"EL" Greece
"ES" Spain
"FI" Finland
"FR" France
"HR" Croatia
"HU" Hungary
"IE" Ireland
"IS" Iceland
"IT" Italy
"LI" Liechtenstein
"LT" Lithuania
"LU" Luxembourg
"LV" Latvia
"MT" Malta
"NL" Netherlands
"NO" Norway
"PL" Poland
"PT" Portugal
"RO" Romania
"SE" Sweden
"SI" Slovenia
"SK" Slovakia

#subset amb els que coincideixen

weekdead2 <- subset.data.frame(weekdead2, weekdead2$country ==
c("AT", "BE", "BG", "CY", "CZ", "DE", "DK", "EE",
"EL", "ES", "FI", "FR", "HR", "HU", "IE", "IS",
"IT", "LI", "LT", "LU", "LV", "MT", "NL", "NO", "PL",
"PT", "RO", "SE", "SI", "SK"), drop = FALSE)

```

```

```{r}
# table(vacinne$YearWeekISO)
# table(weekcase$year_week)
# table(hospital$year_week)
# table(test$year_week)
# table(deaths$dateRep)
```

```

Tots els països coincideixen

```

```{r}
# table(vacinne$ReportingCountry)
# table(weekcase$country)
# table(hospital$country)
# table(test$country)
# table(deaths$countriesAndTerritories)
```

```

# Merge de les dades

Merge de totes les bdd anteriors

```

```{r}
library(sqldf)

dd <- sqldf("SELECT test.country_code, test.year_week, test.country,
new_cases, tests_done, test.population, testing_rate,
positivity_rate,
deaths.cases, deaths.deaths, popData2020,
hospital_1.daily_hosp_occupancy,
hospital_2.daily_ICU_occupancy,
NumberDosesReceived, FirstDose, FirstDoseAcc, SecondDose,
SecondDoseAcc, ThirdDose, ThirdDoseAcc, vacinne.Population
weekly_count, weekcase.population
FROM test
LEFT JOIN deaths ON
test.country_code = deaths.countriesAndTerritories
AND test.year_week = deaths.dateRep
LEFT JOIN hospital_1 ON
test.country_code = hospital_1.country
AND test.year_week = hospital_1.year_week
LEFT JOIN hospital_2 ON
test.country_code = hospital_2.country
AND test.year_week = hospital_2.year_week
LEFT JOIN vacinne ON
test.country_code = vacinne.ReportingCountry
AND test.year_week = vacinne.YearWeekISO")

```



```

LEFT JOIN weekcase ON
  test.country_code = weekcase.country
  AND test.year_week = weekcase.year_week")

dd

head(dd)
```

DADES
```{r}
dades <- dd
#install.packages("openxlsx")
library(openxlsx)
write.xlsx(dades, "dades.xlsx")

library(readxl)
dades <- read_excel("dades.xlsx")
View(dades)

write.csv(dades, "dades_def_2.csv")

table(dades_def_2$country)
```

```

## Anàlisi descriptiu

```
ANÀLISI DESCRIPTIU

```{r}
d <- dades_def_2
summary(covid)

table(covid$year_week)
table(covid$country)
```

Anàlisi de les variables en funció de la població

```{r}
covid_pop <- covid
covid_pop <- covid_pop[, c(-7, -10, -22, -23)]
covid_pop$new_cases <- covid_pop$new_cases/covid_pop$popData2020
covid_pop$tests_done <- covid_pop$tests_done/covid_pop$popData2020
covid_pop$testing_rate <-
covid_pop$testing_rate/covid_pop$popData2020
covid_pop$positivity_rate <-
covid_pop$positivity_rate/covid_pop$popData2020
covid_pop$deaths <- covid_pop$deaths/covid_pop$popData2020
covid_pop$daily_hosp_occupancy <-
covid_pop$daily_hosp_occupancy/covid_pop$popData2020
covid_pop$daily_ICU_occupancy <-
covid_pop$daily_ICU_occupancy/covid_pop$popData2020
covid_pop$NumberDosesReceived <-
covid_pop$NumberDosesReceived/covid_pop$popData2020
covid_pop$FirstDose <- covid_pop$FirstDose/covid_pop$popData2020
covid_pop$FirstDoseAcc <-
covid_pop$FirstDoseAcc/covid_pop$popData2020
covid_pop$SecondDose <- covid_pop$SecondDose/covid_pop$popData2020
covid_pop$SecondDoseAcc <-
covid_pop$SecondDoseAcc/covid_pop$popData2020
covid_pop$ThirdDose <- covid_pop$ThirdDose/covid_pop$popData2020
head(covid_pop)
table(covid_pop$country)

#Primer grup
covid_pop_1 <- covid_pop[(covid_pop$country == "Austria") |
(covid_pop$country == "Belgium") | (covid_pop$country == "Bulgaria")
| (covid_pop$country == "Croatia") | (covid_pop$country == "Cyprus")
| (covid_pop$country == "Czechia") | (covid_pop$country ==
"Denmark") | (covid_pop$country == "Estonia") | (covid_pop$country
== "Finland") | (covid_pop$country == "France") , ]
```

```

#Segon grup
covid_pop_2 <- covid_pop[(covid_pop$country == "Germany") |
(covid_pop$country == "Greece") | (covid_pop$country == "Hungary") |
(covid_pop$country == "Iceland") | (covid_pop$country == "Ireland")
| (covid_pop$country == "Italy") | (covid_pop$country == "Latvia") |
(covid_pop$country == "Liechtenstein") | (covid_pop$country ==
"Luxembourg") | (covid_pop$country == "Lithuania " ) , ]

#Tercer grup
covid_pop_3 <- covid_pop[(covid_pop$country == "Malta") |
(covid_pop$country == "Netherlands") | (covid_pop$country ==
"Norway") | (covid_pop$country == "Poland") | (covid_pop$country ==
"Portugal") | (covid_pop$country == "Romania") | (covid_pop$country
== "Slovakia") | (covid_pop$country == "Slovenia") |
(covid_pop$country == "Spain") | (covid_pop$country == "Sweden") , ]

library("plyr")
covid_pop_4 <- rbind.fill(covid_pop_2, covid_pop_litu)
table(covid_pop_4$country)

table(covid$country)
table(covid_pop$country)
```



```

*Gràfiques per covid_pop_EU*

```{r}
boxplot(covid_pop_1$new_cases ~ covid_pop_1$country, xlab =
'Country', ylab = 'New Cases', main = 'Casos nous per setmana i
país', cex.axis=0.8, col = rainbow(10, alpha=0.5))
boxplot(covid_pop$tests_done ~ covid_pop$country, xlab = 'Country',
ylab = 'Tests', main = 'Tests fets per setmana i país',
cex.axis=0.8, col = rainbow(30, alpha=0.5))
boxplot(covid_pop$cases ~ covid_pop$country, xlab = 'Country',
ylab = 'Cases', main = 'Casos nous per setmana i país',
cex.axis=0.8, col = rainbow(30, alpha=0.5))
boxplot(covid_pop$testing_rate ~ covid_pop$country, xlab =
'Country', ylab = 'Testing rate', main = 'Taxa de tests fets per
setmana i país', cex.axis=0.8, col = rainbow(30, alpha=0.5))
boxplot(covid_pop$positivity_rate ~ covid_pop$country, xlab =
'Country', ylab = 'Positivity rate', main = 'Positivitat de la prova
per setmana i país', cex.axis=0.8, col = rainbow(30, alpha=0.5))
boxplot(covid_pop$deaths ~ covid_pop$country, xlab = 'Country', ylab
= 'Deaths', main = 'Morts per setmana i país', cex.axis=0.8, col =
rainbow(30, alpha=0.5))

```


```

```

boxplot(covid_pop$daily_hosp_occupancy ~ covid_pop$country, xlab =
'Country', ylab = 'Weekly hospitality occupancy', main = 'Ocupació
als hospitals per setmana i país', cex.axis=0.8, col = rainbow(30,
alpha=0.5))
boxplot(covid_pop$daily_ICU_occupancy ~ covid_pop$country, xlab =
'Country', ylab = 'Weekly ICU occupancy', main = 'Ocupació a les
UCIs per setmana i país', cex.axis=0.8, col = rainbow(30,
alpha=0.5))
boxplot(d$NumberDosesReceived ~ d$country, xlab = 'Country', ylab =
'Number Doses Recieved', main = 'Nombre de vacunes rebudes per
setmana i país', cex.axis=0.8, col = rainbow(30, alpha=0.5))
` ``

```

Graficant amb ggplot2

New cases

```
````{r}
```

```
library(ggplot2)
```

```
ggplot(data = covid_pop_1, mapping = aes(x = country, y =
new_cases), cex.axis=0.5) + geom_boxplot(alpha=0.3, col = 'black',
fill='blue') + facet_grid(scales = 'free') + labs(x = 'Country', y
= 'New Cases', color = '') + coord_flip() # + theme(axis.text =
element_text(angle = 90))
```

```
ggplot(data = covid_pop_2, mapping = aes(x = country, y =
new_cases), cex.axis=0.5) + geom_boxplot(alpha=0.3, col = 'black',
fill='blue') + facet_grid(scales = 'free') + labs(x = 'Country', y
= 'New Cases', color = '') + coord_flip() # + theme(axis.text =
element_text(angle = 90))
```

```
ggplot(data = covid_pop_3, mapping = aes(x = country, y =
new_cases), cex.axis=0.5) + geom_boxplot(alpha=0.3, col = 'black',
fill='blue') + facet_grid(scales = 'free') + labs(x = 'Country', y
= 'New Cases', color = '') + coord_flip() # + theme(axis.text =
element_text(angle = 90))
` ``

```

tests done

```
````{r}
```

```
library(ggplot2)
```

```
ggplot(data = covid_pop_1, mapping = aes(x = country, y =
tests_done), cex.axis=0.5) + geom_boxplot(alpha=0.3, col = 'black',
fill='green') + facet_grid(scales = 'free') + labs(x = 'Country', y
= 'Tests Done', color = '') + coord_flip() # + theme(axis.text =
element_text(angle = 90))
```

```
ggplot(data = covid_pop_2, mapping = aes(x = country, y =
tests_done), cex.axis=0.5) + geom_boxplot(alpha=0.3, col = 'black',
fill='green') + facet_grid(scales = 'free') + labs(x = 'Country', y
= 'Tests Done', color = '') + coord_flip() # + theme(axis.text =
element_text(angle = 90))
```

```
ggplot(data = covid_pop_3, mapping = aes(x = country, y =
tests_done), cex.axis=0.5) + geom_boxplot(alpha=0.3, col = 'black',
fill='green') + facet_grid(scales = 'free') + labs(x = 'Country', y
= 'Tests Done', color = '') + coord_flip() # + theme(axis.text =
element_text(angle = 90))
```

```
```
```

positivity rate

```
```{r}
```

```
library(ggplot2)
```

```
ggplot(data = covid_pop_1, mapping = aes(x = country, y =
positivity_rate), cex.axis=0.5) + geom_boxplot(alpha=0.3, col =
'black', fill='red') + facet_grid(scales = 'free') + labs(x =
'Country', y = 'Positivity Rate', color = '') + coord_flip() # +
theme(axis.text = element_text(angle = 90))
```

```
ggplot(data = covid_pop_2, mapping = aes(x = country, y =
positivity_rate), cex.axis=0.5) + geom_boxplot(alpha=0.3, col =
'black', fill='red') + facet_grid(scales = 'free') + labs(x =
'Country', y = 'Positivity Rate', color = '') + coord_flip() # +
theme(axis.text = element_text(angle = 90))
```

```
ggplot(data = covid_pop_3, mapping = aes(x = country, y =
positivity_rate), cex.axis=0.5) + geom_boxplot(alpha=0.3, col =
'black', fill='red') + facet_grid(scales = 'free') + labs(x =
'Country', y = 'Positivity Rate', color = '') + coord_flip() # +
theme(axis.text = element_text(angle = 90))
```

```
```
```

deaths

```
```{r}
```

```
library(ggplot2)
```

```
ggplot(data = covid_pop_1, mapping = aes(x = country, y = deaths),
cex.axis=0.5) + geom_boxplot(alpha=0.3, col = 'black',
fill='orange') + facet_grid(scales = 'free') + labs(x = 'Country',
y = 'Deaths', color = '') + coord_flip() # + theme(axis.text =
element_text(angle = 90))
```

```
ggplot(data = covid_pop_2, mapping = aes(x = country, y = deaths),
cex.axis=0.5) + geom_boxplot(alpha=0.3, col = 'black',
fill='orange') + facet_grid(scales = 'free') + labs(x = 'Country',
y = 'Deaths', color = '') + coord_flip() # + theme(axis.text =
element_text(angle = 90))
```

```
ggplot(data = covid_pop_3, mapping = aes(x = country, y = deaths),
cex.axis=0.5) + geom_boxplot(alpha=0.3, col = 'black',
fill='orange') + facet_grid(scales = 'free') + labs(x = 'Country',
y = 'Deaths', color = '') + coord_flip() # + theme(axis.text =
element_text(angle = 90))
```

```
```
```

```
daily hosp occupancy
```

```
```{r}
```

```
library(ggplot2)
```

```
ggplot(data = covid_pop_1, mapping = aes(x = country, y =
daily_hosp_occupancy), cex.axis=0.5) + geom_boxplot(alpha=0.3, col =
'black', fill='violet') + facet_grid(scales = 'free') + labs(x =
'Country', y = 'Daily Hospitality Occupancy', color = '') +
coord_flip() # + theme(axis.text = element_text(angle = 90))
```

```
ggplot(data = covid_pop_2, mapping = aes(x = country, y =
daily_hosp_occupancy), cex.axis=0.5) + geom_boxplot(alpha=0.3, col =
'black', fill='violet') + facet_grid(scales = 'free') + labs(x =
'Country', y = 'Daily Hospitality Occupancy', color = '') +
coord_flip() # + theme(axis.text = element_text(angle = 90))
```

```
ggplot(data = covid_pop_3, mapping = aes(x = country, y =
daily_hosp_occupancy), cex.axis=0.5) + geom_boxplot(alpha=0.3, col =
'black', fill='violet') + facet_grid(scales = 'free') + labs(x =
'Country', y = 'Daily Hospitality Occupancy', color = '') +
coord_flip() # + theme(axis.text = element_text(angle = 90))
```

```
```
```

```
daily ICU occupancy
```

```
```{r}
```

```
library(ggplot2)
```

```
ggplot(data = covid_pop_1, mapping = aes(x = country, y =
daily_ICU_occupancy), cex.axis=0.5) + geom_boxplot(alpha=0.3, col =
'black', fill='lightblue') + facet_grid(scales = 'free') + labs(x =
'Country', y = 'Daily ICU Occupancy', color = '') + coord_flip() #
+ theme(axis.text = element_text(angle = 90))
```

```
ggplot(data = covid_pop_2, mapping = aes(x = country, y =
daily_ICU_occupancy), cex.axis=0.5) + geom_boxplot(alpha=0.3, col =
'black', fill='lightblue') + facet_grid(scales = 'free') + labs(x =
'Country', y = 'Daily ICU Occupancy', color = '') + coord_flip() #
+ theme(axis.text = element_text(angle = 90))
```

```
ggplot(data = covid_pop_3, mapping = aes(x = country, y =
daily_ICU_occupancy), cex.axis=0.5) + geom_boxplot(alpha=0.3, col =
'black', fill='lightblue') + facet_grid(scales = 'free') + labs(x =
'Country', y = 'Daily ICU Occupancy', color = '') + coord_flip() #
+ theme(axis.text = element_text(angle = 90))
```
```

\*Graficant al llarg de les setmanes\*

```
```{r}
library(ggplot2)
ggplot(covid_pop_1, aes(x = year_week, y = new_cases,
color = country, group = country,
cex.axis=0.5)) +
geom_line() +
labs(x = 'Year Week', y = 'New Cases', color = '', cex.axis=0.3) +
theme(axis.text = element_text(angle = 90))
```

```
ggplot(covid_pop_4, aes(x = year_week, y = new_cases,
color = country, group = country,
cex.axis=0.5)) +
geom_line() +
labs(x = 'Year Week', y = 'New Cases', color = '', cex.axis=0.3) +
theme(axis.text = element_text(angle = 90))
```

```
ggplot(covid_pop_3, aes(x = year_week, y = new_cases,
color = country, group = country,
cex.axis=0.5)) +
geom_line() +
labs(x = 'Year Week', y = 'New Cases', color = '', cex.axis=0.3) +
theme(axis.text = element_text(angle = 90))
```
```

```
```{r}
ggplot(covid_pop_1, aes(x = year_week, y = tests_done,
color = country, group = country)) +
geom_line() +
labs(x = 'Year Week', y = 'Tests Done', color = '') +
theme(axis.text = element_text(angle = 90))
```

```
ggplot(covid_pop_4, aes(x = year_week, y = tests_done,
```

```

        color = country, group = country)) +
  geom_line() +
  labs(x = 'Year Week', y = 'Tests Done', color = '') +
  theme(axis.text = element_text(angle = 90))

ggplot(covid_pop_3, aes(x = year_week, y = tests_done,
        color = country, group = country)) +
  geom_line() +
  labs(x = 'Year Week', y = 'Tests Done', color = '') +
  theme(axis.text = element_text(angle = 90))
```

```{r}
ggplot(covid_pop_1, aes(x = year_week, y = testing_rate,
        color = country, group = country)) +
  geom_line() +
  labs(x = 'Year Week', y = 'Testing Rate', color = '') +
  theme(axis.text = element_text(angle = 90))

ggplot(covid_pop_4, aes(x = year_week, y = testing_rate,
        color = country, group = country)) +
  geom_line() +
  labs(x = 'Year Week', y = 'Testing Rate', color = '') +
  theme(axis.text = element_text(angle = 90))

ggplot(covid_pop_3, aes(x = year_week, y = testing_rate,
        color = country, group = country)) +
  geom_line() +
  labs(x = 'Year Week', y = 'Testing Rate', color = '') +
  theme(axis.text = element_text(angle = 90))
```

```{r}
ggplot(covid_pop_1, aes(x = year_week, y = positivity_rate,
        color = country, group = country)) +
  geom_line() +
  labs(x = 'Year Week', y = 'Positivity Rate', color = '') +
  theme(axis.text = element_text(angle = 90))

ggplot(covid_pop_4, aes(x = year_week, y = positivity_rate,
        color = country, group = country)) +
  geom_line() +
  labs(x = 'Year Week', y = 'Positivity Rate', color = '') +
  theme(axis.text = element_text(angle = 90))

ggplot(covid_pop_3, aes(x = year_week, y = positivity_rate,
        color = country, group = country)) +

```



```

    geom_line() +
    labs(x = 'Year Week', y = 'Positivity Rate', color = '') +
theme(axis.text = element_text(angle = 90))
```

```{r}
ggplot(covid_pop_1, aes(x = year_week, y = deaths,
                        color = country, group = country)) +
  geom_line() +
  labs(x = 'Year Week', y = 'Deaths', color = '') + theme(axis.text
= element_text(angle = 90))

ggplot(covid_pop_4, aes(x = year_week, y = deaths,
                        color = country, group = country)) +
  geom_line() +
  labs(x = 'Year Week', y = 'Deaths', color = '') + theme(axis.text
= element_text(angle = 90))

ggplot(covid_pop_3, aes(x = year_week, y = deaths,
                        color = country, group = country)) +
  geom_line() +
  labs(x = 'Year Week', y = 'Deaths', color = '') + theme(axis.text
= element_text(angle = 90))
```

```{r, warning=FALSE}
ggplot(covid_pop_1, aes(x = year_week, y = daily_hosp_occupancy,
                        color = country, group = country)) +
  geom_line() +
  labs(x = 'Year Week', y = 'Hospitality Occupancy', color = '') +
theme(axis.text = element_text(angle = 90))

ggplot(covid_pop_4, aes(x = year_week, y = daily_hosp_occupancy,
                        color = country, group = country)) +
  geom_line() +
  labs(x = 'Year Week', y = 'Hospitality Occupancy', color = '') +
theme(axis.text = element_text(angle = 90))

ggplot(covid_pop_3, aes(x = year_week, y = daily_hosp_occupancy,
                        color = country, group = country)) +
  geom_line() +
  labs(x = 'Year Week', y = 'Hospitality Occupancy', color = '') +
theme(axis.text = element_text(angle = 90))
```

```{r, warning=FALSE}
ggplot(covid_pop_1, aes(x = year_week, y = daily_ICU_occupancy,
                        color = country, group = country)) +

```

```

    geom_line() +
    labs(x = 'Year Week', y = 'ICU Occupancy', color = '') +
    theme(axis.text = element_text(angle = 90))

ggplot(covid_pop_4, aes(x = year_week, y = daily_ICU_occupancy,
                        color = country, group = country)) +
    geom_line() +
    labs(x = 'Year Week', y = 'ICU Occupancy', color = '') +
    theme(axis.text = element_text(angle = 90))

ggplot(covid_pop_3, aes(x = year_week, y = daily_ICU_occupancy,
                        color = country, group = country)) +
    geom_line() +
    labs(x = 'Year Week', y = 'ICU Occupancy', color = '') +
    theme(axis.text = element_text(angle = 90))
```

```{r, warning=FALSE}
ggplot(na.omit(covid_pop), aes(x = year_week, y =
NumberDosesReceived,
                             color = country, group = country)) +
    geom_line() +
    labs(x = 'Year Week', y = 'Number Doses Received', color = '') +
    theme(axis.text = element_text(angle = 90))
```

Primera segona i tercera dosi
```{r, warning=FALSE}
ggplot(na.omit(covid_pop), aes(x = year_week, y = FirstDose,
                              color = country, group = country)) +
    geom_line() +
    labs(x = 'Year Week', y = 'First Dose', color = '') +
    theme(axis.text = element_text(angle = 90))

ggplot(na.omit(covid_pop), aes(x = year_week, y = SecondDose,
                              color = country, group = country)) +
    geom_line() +
    labs(x = 'Year Week', y = 'Second Dose', color = '') +
    theme(axis.text = element_text(angle = 90))

ggplot(na.omit(covid_pop), aes(x = year_week, y = ThirdDose,
                              color = country, group = country)) +
    geom_line() +
    labs(x = 'Year Week', y = 'Third Dose', color = '') +
    theme(axis.text = element_text(angle = 90))
```

```

Guardem bbdd

```
```{r}
write.csv(covid, "covid.csv")
write.csv(covid_pop, "covid_pop.csv")
write.csv(covid_pop_1, "covid_pop_1.csv")
write.csv(covid_pop_2, "covid_pop_2.csv")
write.csv(covid_pop_3, "covid_pop_3.csv")
write.csv(covid_pop_4, "covid_pop_4.csv")
write.csv(covid_pop_litu, "covid_pop_litu.csv")
```
```

```
```{r}
barplot(table(covid$year_week), main = 'Nº observacions per
setmanes', col = 'lightblue', ylim=c(0,30), xlab = 'Setmanes', ylab
= 'Observacions')
```
```

## Models

\*MODELITZACIÓ\*

Bases de dades

```
`` `{r}
setwd("D:/4t/TFG")
dd2 <- read.csv("covid_pop.csv")
```

```
library(knitr)
library(kableExtra)
library(sjPlot)
library(sjmisc)
library(sjlabelled)
library(stargazer)
`` `
```

```
`` `{r}
if(!require('plm')){
 install.packages("plm")
 library(somepackage)
}
```

```
if(!require('foreign')){
 install.packages("foreign")
}
library("foreign")
```

```
library(plm)
library(knitr)
library(broom)
library(tidyverse)
library(stargazer)
library(lmtest)
library(gplots)
library(wooldridge)
library(ggplot2)
library(dplyr)
library(plm)
library(lfe)
library(lmtest)
library(car)
library(geepack)
`` `
```

### \*Models\*

- Pooled Cross Sections
- Fixed Effects Model
- Random Effects Model
- Including the Time Dimension

### \*OLS / Regresión lineal múltiple\*

```
deaths
```{r}
ols =
lm(deaths~tests_done+testing_rate+positivity_rate+new_cases+daily_ho
sp_occupancy+daily_ICU_occupancy + popData2020, data = covid)
summary(ols)
tab_model(ols, show.se = TRUE, digits = 5, title = 'OLS')
stargazer(ols, title = 'OLS', align=TRUE,header = FALSE,
out="table1.txt")
```
```

```
deaths by covid_pop
```{r}
ols_pop =
lm(deaths~tests_done+testing_rate+positivity_rate+new_cases+daily_ho
sp_occupancy+daily_ICU_occupancy, data = covid_pop)
summary(ols_pop)
tab_model(ols_pop, show.se = TRUE, digits = 5, title = 'OLS
Population')
```
```

```
deaths pooled
```{r}
library(plm)
ols_2 =
plm(deaths~tests_done+testing_rate+positivity_rate+new_cases+daily_h
osp_occupancy+daily_ICU_occupancy + popData2020, data = covid, model
= 'pooling', index = c("country", "index"))
summary(ols_2)
tab_model(ols_2, show.se = TRUE, digits = 5, title = 'Pooled OLS')
```
```

```

deaths_pop pooled
```{r}
ols_pop_2 =
plm(deaths~tests_done+testing_rate+positivity_rate+new_cases+daily_hosp_occupancy+daily_ICU_occupancy, data = covid_pop, model = 'pooling', index = c("country", "year_week"))
summary(ols_pop_2)
tab_model(ols_pop_2, show.se = TRUE, digits = 5, title = 'Pooled OLS Population')
```

```

```

<!-- daily_hosp -->
<!-- ```{r} -->
<!-- ols_hosp =
lm(daily_hosp_occupancy~deaths+tests_done+testing_rate+positivity_rate+new_cases+daily_ICU_occupancy + popData2020, data = covid) -->
<!-- summary(ols_hosp) -->
<!-- ``` -->

```

\*Fixed effects ONE WAY\*

```

deaths effect individual
```{r}
library(plm)
country_effects <- plm(deaths~
tests_done+testing_rate+positivity_rate+new_cases+daily_hosp_occupancy+daily_ICU_occupancy + popData2020,
data = covid,
index = c("country", "year_week"),
model = "within",
effect = "individual")

```

```

summary(country_effects)
tab_model(country_effects, show.se = TRUE, digits = 5, title = 'Country Fixed Effects')
```

```

```

```{r}
plmtest(country_effects, effect="individual")
```

```

```

```{r}

```

```
fixef(country_effects)
```
```

El coeficiente de x1 indica cuánto cambia Y a lo largo del tiempo, en promedio por país, cuando X aumenta en una unidad.

```
deaths time effect
```{r}
time_effects <- plm(deaths~
tests_done+testing_rate+positivity_rate+new_cases+daily_hosp_occupan
cy+daily_ICU_occupancy + popData2020,
                    data = covid,
                    index = c("country", "year_week"),
                    model = "within",
                    effect = "time")
```

```
summary(time_effects)
tab_model(time_effects, show.se = TRUE, digits = 5, title = 'Time
Fixed Effects')
```
```

```
```{r}
plmtest(time_effects, effect="time")
```
```

```
```{r}
fixef(time_effects)
```
```

El coeficiente de x1 indica cuánto cambia Y a lo largo del tiempo, en promedio por país, cuando X aumenta en una unidad.

**\*Fixed effects TWO WAYS\***

```
deaths by two ways
```{r}
twoway_effects <- plm(deaths~
tests_done+testing_rate+positivity_rate+new_cases+daily_hosp_occupan
cy+daily_ICU_occupancy + popData2020,
                    data = covid,
                    index = c("country", "year_week"),
                    model = "within",
```

```

        effect = "twoways")

summary(twoway_effects)
tab_model(twoway_effects, show.se = TRUE, digits = 5, title = 'Two-
Way Fixed Effects')
```

```{r}
library(lmtest)
library(texreg)
screenreg(list(country_effects, time_effects, twoway_effects),
          custom.model.names = c("Country Fixed Effects",
                                "Time Fixed Effects",
                                "Two-Way Fixed Effects"))

tab_model(country_effects, time_effects, twoway_effects, show.se =
TRUE, digits = 5, title = 'Country Fixed Effects, Time Fixed Effects
and Two-Way Fixed Effects')
```

Correlació en la serie
```{r}
pbgttest(twoway_effects)
```

```{r}
twoway_effects_hac <- coeftest(twoway_effects,
                             vcov = vcovHC(twoway_effects,
                                             method = "arellano",
                                             type = "HC3"))

screenreg(list(twoway_effects, twoway_effects_hac),
          custom.model.names = c("Twoway Fixed Effects",
                                "Twoway Fixed Effects (HAC)"))

#tab_model(twoway_effects, twoway_effects_hac, show.se = TRUE,
title = 'Two-Way Fixed Effects and Two-Way Fixed Effects (HAC)')
```

Dependencia transversal
```{r}
pcdtest(twoway_effects)
```

```



```

Driscoll i Kraay (1998)
```{r}
twoway_effects_scc <- coeftest(twoway_effects,
                             vcov = vcovSCC(twoway_effects,
                                             type="HC3",
                                             cluster = "group"))

twoway_effects_scc

as.table(twoway_effects_scc)
```

```{r}
screenreg(list(country_effects,
              time_effects,
              twoway_effects,
              twoway_effects_scc),
          custom.model.names = c("Country Effects",
                                "Time Effects",
                                "Two-Way Fixed Effects",
                                "SCC"))
```

Regresión con efectos aleatorios / Random Effects Model

```{r}
library(plm)
library(knitr)
library(broom)
wageReTest <- plmtest(ols_2, effect="individual")
kable(tidy(wageReTest), caption=
      "Random effects test")
```

```{r}
library(Ecdat)
library(plm)
random <- plm(deaths ~
tests_done+testing_rate+positivity_rate+new_cases+daily_hosp_occupan
cy+daily_ICU_occupancy, index=c("country", "year_week"),
model="random", data=covid)

```

```
summary(random)
tab_model(random, show.se = TRUE, digits = 5, title = 'Random
Effects Model')
```

```

OLS pooled vs fixed

```
```{r}
pFtest(country_effects, ols_2)
pFtest(time_effects, ols_2)
pFtest(twoway_effects, ols_2)
```

```

Test de Hausman : fixed vs random

```
```{r}
phptest(country_effects, random)
phptest(time_effects, random)
phptest(twoway_effects, random)
```

```

**\*\*VACCINE ACC\*\***

Utilitzem Fixed Effects

deaths effect individual + vaccine acumulades

```
```{r}
library(plm)
country_effects_vaccine_acc <- plm(deaths~
tests_done+testing_rate+positivity_rate+new_cases+daily_hosp_occupan
cy+daily_ICU_occupancy + popData2020 + FirstDoseAcc + SecondDoseAcc
+ ThirdDoseAcc,
                                data = covid,
                                index = c("country", "year_week"),
                                model = "within",
                                effect = "individual")

```

```
summary(country_effects_vaccine_acc)
tab_model(country_effects_vaccine_acc, show.se = TRUE, digits = 5,
title = 'Country Effects Model Vaccine')
```

```

```
```{r}
plmtest(country_effects_vaccine_acc, effect="individual")

```

```
```
```

```
```{r}
fixef(country_effects_vaccine_acc)
```
```

El coeficiente de x1 indica cuánto cambia Y a lo largo del tiempo, en promedio por país, cuando X aumenta en una unidad.

```
deaths time effect
```{r}
time_effects_vaccine_acc <- plm(deaths~
tests_done+testing_rate+positivity_rate+new_cases+daily_hosp_occupan
cy+daily_ICU_occupancy + popData2020 + FirstDoseAcc + SecondDoseAcc
+ ThirdDoseAcc,
                                data = covid,
                                index = c("country", "year_week"),
                                model = "within",
                                effect = "time")
```

```
summary(time_effects_vaccine_acc)
tab_model(time_effects_vaccine_acc, show.se = TRUE, digits = 5,
title = 'Time Effects Model Vaccine')
```
```

```
```{r}
library(plm)
plmtest(time_effects_vaccine_acc, effect="time")
```
```

```
```{r}
fixef(time_effects_vaccine_acc)
```
```

El coeficiente de x1 indica cuánto cambia Y a lo largo del tiempo, en promedio por país, cuando X aumenta en una unidad.

\*Fixed effects TWO WAYS\*

```
deaths by two ways
```{r}
```

```

twoway_effects_vaccine_acc <- plm(deaths~
tests_done+testing_rate+positivity_rate+new_cases+daily_hosp_occupan
cy+daily_ICU_occupancy + popData2020 + FirstDoseAcc + SecondDoseAcc
+ ThirdDoseAcc,
                                data = covid,
                                index = c("country", "year_week"),
                                model = "within",
                                effect = "twoways")

summary(twoway_effects_vaccine_acc)
tab_model(twoway_effects_vaccine_acc, show.se = TRUE, digits = 5,
title = 'Two-Ways Effects Model Vaccine')
#stargazer(twoway_effects_vaccine_acc, title = 'Two-Ways Effects
Model Vaccine')
```

```{r}
library(lmtest)
library(texreg)
screenreg(list(country_effects_vaccine_acc,
time_effects_vaccine_acc, twoway_effects_vaccine_acc),
          custom.model.names = c("Country Fixed Effects Vaccine",
                                "Time Fixed Effects Vaccine",
                                "Two-Way Fixed Effects Vaccine"))

tab_model(country_effects_vaccine_acc, time_effects_vaccine_acc,
twoway_effects_vaccine_acc, show.se = TRUE, digits = 5, title =
'Country Fixed Effects Vaccine, Time Fixef Effects Vaccine and Two-
Ways Fixed Effects Vaccine')
```

```