

Grado en Estadística

Título: Modelos predictivos para la distribución del número de casos diarios de la COVID-19 para los diferentes distritos de Barcelona

Autor: Pablo Pemán Rodríguez

Director: Dr. David Moraña Soler

Departamento: Econometría, Estadística y Economía Aplicada

Convocatoria: 2n 2020-2021



UNIVERSITAT DE
BARCELONA



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Facultat de Matemàtiques i Estadística

Resumen

La COVID-19 ha afectado a toda la sociedad, es por ello por lo que hemos considerado necesario ajustar modelos que nos permitan explicar y predecir la serie temporal de los casos detectados. Al ser la COVID-19 una enfermedad infecciosa, donde los casos de un día vienen influenciados por los días anteriores, consideramos que los mejores modelos para explicar la serie serán los autorregresivos, en concreto, los AR y los INAR. Analizaremos la serie en función del sexo y el distrito. A nivel de resultados obtenemos diferentes mecanismos en ambos modelos en función del sexo, y también vemos como los barrios con mayor población o densidad poblacional son aquellos con más tendencia a detectar más casos. Comparando los modelos, se observa claramente que el modelo INAR es superior en todos los aspectos al modelo AR, teniendo en cuenta que nuestra serie no es estacionaria, y, por tanto, ninguno es perfecto.

Palabras clave: COVID-19, Modelización de enfermedades infecciosas, Análisis estocástico de series temporales, Series temporales de valores enteros

Clasificación AMS:

- 37M10 Time series analysis
- 62M10 Time series, auto-correlation, regression, etc.
- 62M20 Prediction
- 62P10 Applications to biology and medical sciences

Abstract

COVID-19 has affected the whole of society, which is why we have considered it necessary to adjust models that allow us to explain and predict the time series of the detected cases. As COVID-19 is an infectious disease, where the cases of one day are influenced by the previous days, we consider that the best models to explain the series will be the autoregressive ones, specifically, the AR and the INAR. We will analyze the series as a function of sex and district. At the level of results, we obtain different mechanisms in both models as a function of sex, and we also see how the districts with higher population or population density are those with a greater tendency to detect more cases. Comparing the models, it is clearly observed that the INAR model is superior in all aspects to the AR model, taking into account that our series is not stationary, and, therefore, neither is perfect.

Key words: COVID-19, Infectious disease modeling, Stochastic time series analysis, Integer valued time series.

Agradecimientos

Primero de todo, agradecer a mi tutor el Dr. David Moraña Soler por haber aceptado ser mi tutor en este trabajo, y por guiarme y ayudarme con todas las dudas que me encontré en la realización de este.

Doy las gracias a los profesores del grado de Estadística por hacerme crecer en mi la pasión por el análisis de datos.

Finalmente, doy las gracias a mi familia y amigos, por siempre estar a mi lado y apoyarme en los momentos difíciles, sin ellos no estaría donde me encuentro hoy en día.

ÍNDICE

I. INTRODUCCIÓN	1
I.1. COVID-19	1
I.2. Análisis de series temporales	3
I.3. Objetivos del trabajo	5
I.4. Hipótesis	6
II. METODOLOGÍA	7
II.1. Base de datos	7
II.2. Modelos AR	8
II.3. Modelos INAR	10
II.3. Indicadores para las predicciones	12
III. RESULTADOS	13
III.1. Descriptiva de los Datos	13
III.2. Identificación de los modelos	16
III.3. Modelos AR	19
<i>III.3.1 Estimación de los parámetros</i>	19
<i>III.3.2. Validación del modelo</i>	20
<i>III.3.3 Predicciones</i>	21
III.4. Modelos INAR	26
<i>III.4.1 Estimación de los parámetros</i>	26
<i>III.4.2 Predicciones</i>	28
III.5. Comparación de los modelos	33
IV. CONCLUSIONES	35
V. BIBLIOGRAFIA	37
VI. ANNEXOS	38

I. INTRODUCCIÓN

I.1. COVID-19

La COVID-19 es la enfermedad infecciosa causada por el coronavirus que se ha descubierto más recientemente (SARS-CoV2). Tanto este nuevo virus, como la enfermedad que provoca, eran desconocidos antes de que estallara el brote en Wuhan (China) en diciembre de 2019.

Su rápida propagación, su transmisión aérea y la necesidad del mínimo contacto para la infección, añadido a la posibilidad de no manifestarse, o retrasar su manifestación, provocó que en marzo de 2020 la COVID-19 se hubiera propagado por todo el mundo y la Organización Mundial de la Salud (OMS) lo clasificara como pandemia. [1]

La COVID-19 ha afectado a todo el mundo, pero sobre todo ha tenido mayor violencia en los países y zonas más desarrolladas. Wuhan es el centro político, cultural, social y económico de China. En Estados Unidos los estados que más sufrieron fueron los de Nueva York, Nueva Jersey, Michigan o California. En España, la Comunidad Autónoma de Madrid y Cataluña sufrieron los mayores ataques. Esto se explica a que todas estas zonas acumulan grandes concentraciones de personas, ya sean residentes o turistas. Todas estas aglomeraciones facilitan la propagación del virus, no solo a nivel regional, sino también mundial.

La COVID-19 no solo ha afectado a nivel sanitario, provocando colapsos de los hospitales y millones de muertes alrededor del mundo. También ha afectado a la sociedad y la economía a nivel mundial.

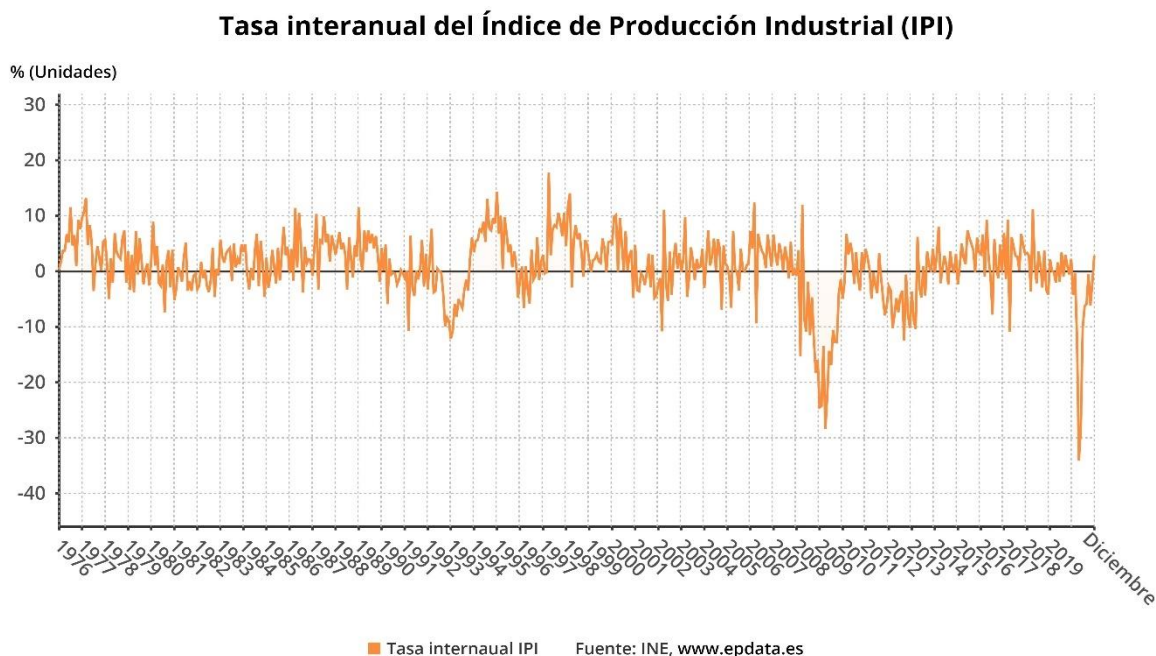


Ilustración 1. Tasa interanual del IPI en España

A nivel español, podemos comprobar como la producción industrial llegó a mínimos históricos, superando incluso los datos correspondientes a la crisis económica que sufrió el país en 2009 (*Ilustración 1*).

Variación anual de la compraventa de viviendas en España

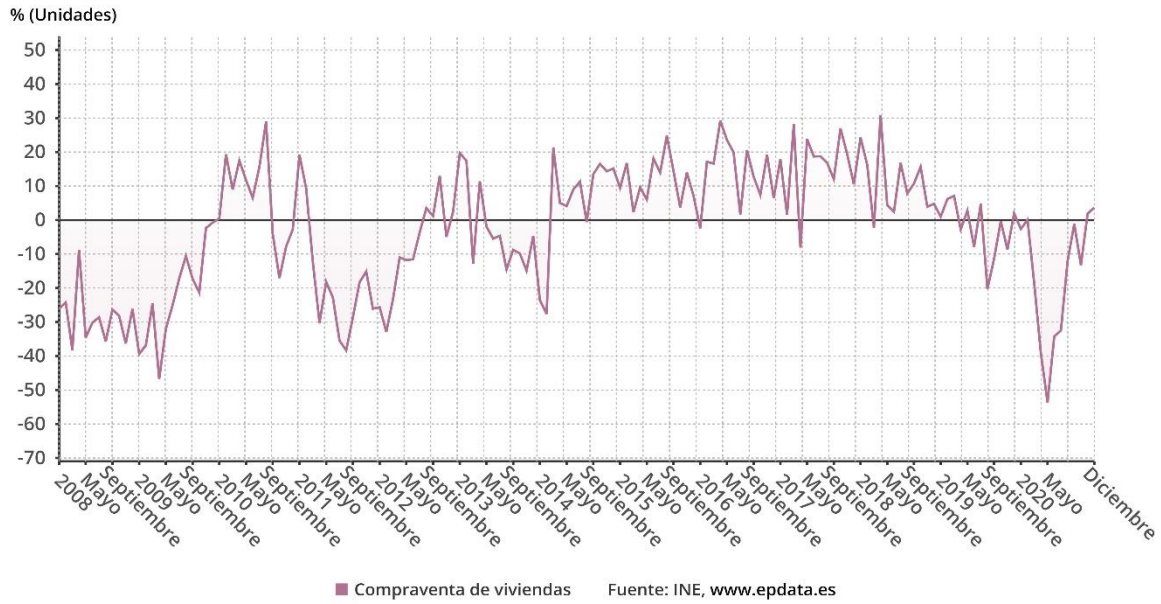


Ilustración 2. Variación anual de la compraventa de viviendas en España

También podemos comprobar como el número de compraventas de viviendas ha bajado muchísimo, superando también a los valores mínimos que hubo en España durante la crisis económica de 2009 y los posteriores años (*Ilustración 2*).

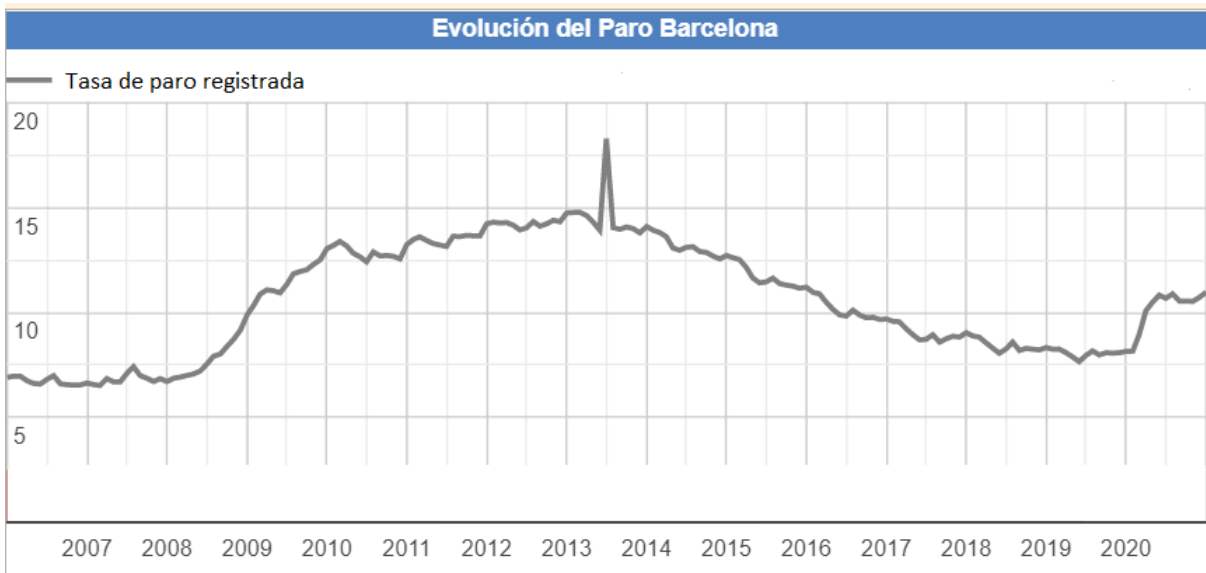


Ilustración 3. Evolución del paro en la ciudad de Barcelona. Fuente: <https://datosmacro.expansion.com/paro/espana/municipios/cataluna/barcelona/barcelona>

En el gráfico superior podemos ver la afectación social que ha tenido la COVID-19 en la ciudad de Barcelona. Observamos como el paro ha aumentado a partir del primer trimestre de 2020, inicio de la pandemia, y sigue subiendo. (*Ilustración 3*)

En conclusión, la pandemia provocada por la COVID-19 ha afectado negativamente al mundo actual tal y como lo conocemos, y aunque los gobiernos estén intentando paliar los efectos adversos, está claro que la COVID-19 cambiará para siempre la sociedad y la forma de vida actual. Es por ello por lo que creemos que conseguir estimar un modelo capaz de explicar y predecir el comportamiento de la propagación del virus es vital para poder llevar a cabo las medidas preventivas y estrategias necesarias para mitigar al mínimo los daños, tanto sociales, económicos o humanos, que pueda provocar.

I.2. Análisis de series temporales

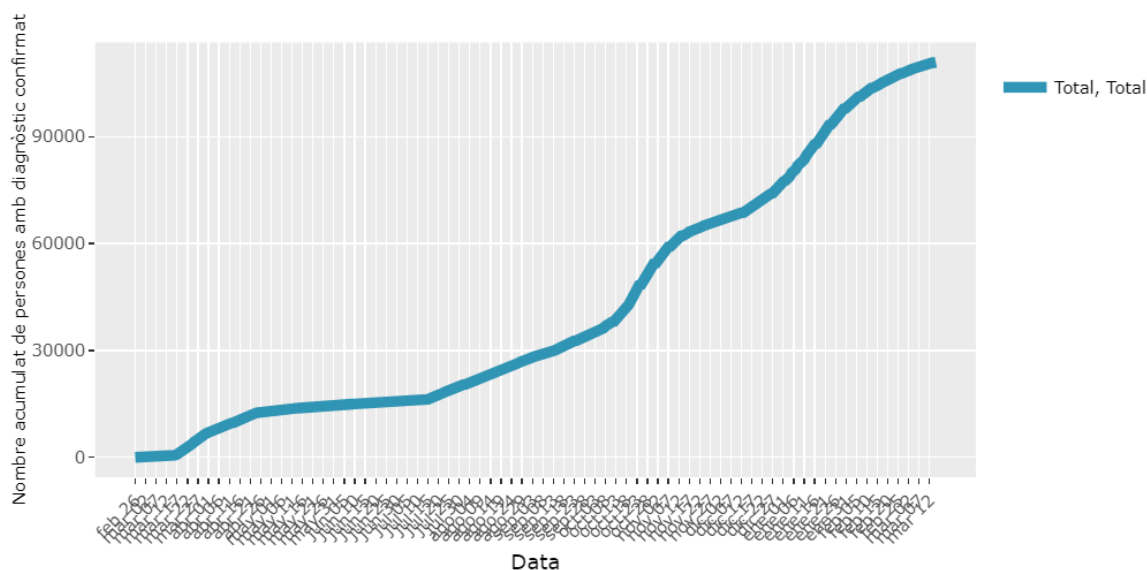
Las series temporales se definen como una sucesión de valores de una variable aleatoria ordenados cronológicamente en el tiempo. Su expresión es la siguiente:

$$x_1, x_2, x_3, \dots, x_t$$

donde t representa la posición temporal que ocupa el valor dentro de la serie.

Las series temporales pueden constar de varios componentes, los cuales son importantes para su posterior análisis. Dichos componentes son los siguientes:

- **Tendencia:** Componente de la serie temporal que representa la evolución a largo plazo de la serie, ya sea de forma ascendente o decreciente.
- **Componente cíclico:** Recoge los movimientos oscilatorios por encima y por debajo de la tendencia que se mantienen en periodos superiores al año.
- **Componente estacional:** Recoge los movimientos oscilatorios que se repiten en periodos menores a un año y siempre en el mismo momento temporal.
- **Componente irregular:** Recoge las variaciones esporádicas de la serie, que no están incluidas en ninguno de los componentes estacionales, y, por tanto, aquellas con un carácter residual.



Font: Registre COVID-19, Departament de Salut, Generalitat de Catalunya.

Ilustración 4. Número acumulado de casos de COVID-19 en la ciudad de Barcelona

El gráfico superior (*Ilustración 4*) hace referencia a la serie del número de casos de COVID-19 acumulados en la ciudad de Barcelona, y es un muy buen ejemplo para ver el componente de la tendencia en una serie temporal. Además, usaremos este gráfico para introducir un nuevo término, la estacionariedad.

Una serie es estacionaria cuando la media y la varianza son constantes a lo largo de toda la serie, y las autocovarianzas entre dos variables aleatorias solo dependen de la distancia temporal que las separa. En nuestro gráfico (*Ilustración 4*), podemos ver como claramente, al tener una tendencia, esto hace que la media no sea constante a lo largo de la serie. Un componente estacional fuerte podría provocar una serie heterocedástica, es decir, con una varianza no constante. Para paliar estos problemas, normalmente es suficiente con aplicar el logaritmo, o una diferencia estacional, respectivamente.

Tenemos dos formas básicas para el análisis de las series temporales dentro del análisis univariante: el análisis determinista y el análisis estocástico.

El análisis determinista se basa en gran parte en la descomposición de la serie temporal en todos los componentes explicados anteriormente. En función de la presencia o no de tendencia o estacionalidad, las series se pueden clasificar de la siguiente forma:

Tabla 1.1 Tipos de series en función de los componentes presentes

		Tendencia	
		Si	No
Estacionalidad	Si	Tipo IV	Tipo II
	No	Tipo III	Tipo I

La presencia o no de estos componentes puede comprobarse, además de por la observación de la propia serie, por diferentes estadísticos de contraste, en concreto el contraste de Daniel para la tendencia, y el contraste de Kruskal-Wallis para la estacionalidad.

Cada tipo de serie tiene diferentes métodos para predecir su comportamiento en función de los componentes entre los que se puede descomponer. Algunos de estos métodos son: método de las medias móviles, método de la tendencia lineal, método ingenuo estacional, método de descomposición El problema de este análisis es cuando la descomposición de la serie en alguno de estos componentes no describe con precisión la propia serie. En estas situaciones es cuando el análisis estocástico entra en juego.

El análisis estocástico no intenta descomponer la serie, sino que se centra en las autocorrelaciones de los valores. Como bien sabemos, el valor de una serie temporal puede venir influenciado por los valores anteriores de la serie. Por ejemplo, el precio de una acción de bolsa no solo viene influenciado por las acciones que hayan podido ocurrir ese día, sino también por los sucesos de días anteriores. Este tipo de modelos, donde un valor temporal concreto viene influenciado por sus valores pasados se conoce como modelo autorregresivo, AR. Hay otros tipos de modelos, en función de cómo sea la serie y sus autocorrelaciones. Si el valor no viene influenciado por los valores pasados se conoce como modelo de medias móviles MA, y si es mixto se conoce como modelo ARMA. Todos estos modelos comentados son modelos para series continuas, en cambio, para las series que siguen una distribución discreta tenemos los mismos modelos, pero adaptados para utilizar una distribución marginal discreta. Dichos modelos son: modelo autorregresivo para valores enteros no negativos, INAR, modelo de medias móviles para valores enteros no negativos, INMA....

La razón para usar estos modelos, los cuales son la contraparte discreta de los modelos continuos explicados anteriormente, es que dichos modelos utilizan la distribución normal como distribución marginal, y nuestra serie al ser un simple conteo, con valores pequeños, es muy difícil que sea explicada por una distribución marginal normal, y es por ello por lo que también utilizamos los modelos con distribución marginal discreta. La gran diferencia es que los modelos discretos introducen un operador de refinamiento, el cual no está presente en los modelos continuos, y que nos permite asegurar que las estimaciones del modelo serán valores enteros.

En concreto, nos centraremos en los modelos AR e INAR, cuya diferencia principal, como hemos comentado anteriormente, es el tipo de series para los que se utilizan. El modelo AR hace referencia a series continuas, mientras que el modelo INAR hace referencia a series discretas. Para una mayor explicación de ambos modelos véase el capítulo **II. METODOLOGÍA**.

I.3. Objetivos del trabajo

El análisis temporal de los casos de la COVID-19 en Barcelona, en función del distrito analizado y el sexo, se pretende llevar a cabo mediante la parametrización de modelos capaces de explicar la serie, ajustarse correctamente a la volatilidad de los posibles valores, y predecir con éxito su evolución. Es por ello por lo que los objetivos de este trabajo son:

analizar los datos correspondientes al número de contagios, para ver si existe diferencia entre los distintos distritos de Barcelona, y dentro de los distritos, diferencias entre los sexos de los individuos, ajustar modelos predictivos, en concreto modelos AR e INAR, y discutir sobre las diferencias expresadas anteriormente, decidir cuál de los dos modelos se ajusta mejor a la serie y es capaz de explicarla, mediante las predicciones y los índices de eficacias respectivos.

I.4. Hipótesis

Nuestra serie temporal es una serie discreta ya que hace referencia al contaje del número de casos de COVID-19 en los distintos distritos de Barcelona. A nivel de modelos creemos que el modelo autorregresivo para valores enteros no negativos, INAR, puede ser un buen candidato para explicar la serie, ya que como hemos comentado anteriormente el modelo INAR se usa para series temporales discretas. Aun así, cabe destacar que la serie no es estacionaria, y, por tanto, sabemos que ninguno de los dos modelos será perfecto. Con relación a la distribución que sigue tenemos varias opciones: binomial negativa, Poisson, hipergeométrica... Nosotros vamos a suponer que sigue una Poisson para facilitar los cálculos y estimaciones, ya que las otras distribuciones presentan muchas complejidades para parametrizar. Además, una Poisson tiene sentido, ya que como hemos dicho anteriormente, nuestra serie es un contaje de sucesos ocurridos en un momento del tiempo concreto. Además, debido al teorema del límite central y la ley de los grandes números, los cuales demuestran que, si tenemos una sucesión de variables aleatorias con una media conocida y una varianza no nula pero finita, dicha distribución se puede aproximar a una normal. Al aproximarse a una normal, la serie puede considerarse continua y por tanto podemos utilizar también el modelo AR simple. Estimaremos ambos modelos, AR e INAR, y los compararemos para ver cuál de ellos explica mejor la serie.

También consideramos que cada distrito es distinto. Dentro de Barcelona hay barrios con más o menos población, más renta por familia o menos, y es por ello por lo que creemos que dichas diferencias serán suficientes para observar diferencias de modelos dentro de los barrios de Barcelona. Es por ello por lo que estimaremos los dos tipos de modelo para cada distrito por separado. A nivel de diferenciación de sexo creemos que se pueden observar diferencias, no tanto a nivel epidemiológico, sino a nivel social, ya que se ha observado que los diferentes tipos de vidas sociales pueden afectar al número de contagios detectados[2]. Es por ello por lo que dentro de nuestro análisis desagregaremos los números de casos diarios detectados por sexo

II. METODOLOGÍA

En este apartado describiremos la base de datos que utilizaremos, y explicaremos en profundidad los modelos que usaremos y los métodos que utilizaremos para estimarlos. Todos los resultados de este trabajo se han obtenido mediante el programa R [3] (*Véase Anexo 9*), y algunos de los gráficos mostrados usando el paquete ggplot2 [4] .

II.1. Base de datos

La base de datos usada en este trabajo ha sido proporcionada por el Ayuntamiento de Barcelona, concretamente por la Agencia de Salud Pública de Barcelona (ASPB) y se puede encontrar en el siguiente enlace: <https://analisi.transparenciacatalunya.cat/Salut/Registre-de-casos-de-COVID-19-realitzats-a-Catalun/xuwf-dxjd/data>

Las variables que podemos encontrar son:

- **TipusCasData:** Fecha donde se registra un positivo
- **RegioSanitariaCodi:** Código localizador de la región sanitaria
- **RegioSanitariaDescripció:** Región sanitaria
- **SectorSanitariCodi:** Código localizador del sector sanitario dentro de la región
- **SectorSanitariaDescripcio:** Sector sanitario dentro de la región
- **ABSCodi:** Código localizador del área básica de salud (ABS) dentro del sector
- **ABSDescripcio:** Área básica de salud dentro del sector
- **SexeCodi:** 1: hombre. 0: mujer
- **SexeDescripció:** Género del individuo
- **TipusCasDescripció:** Test utilizado para la detección del positivo
- **NumCasos:** Número de casos que comparten todas las características anteriores

En dicha base de datos tenemos valores para toda Cataluña, pero como hemos dicho anteriormente, simplemente nos centraremos en los distritos de Barcelona ciudad, los cuales son: *Ciutat Vella, Eixample, Gràcia, Sàrria-Sant Gervasi, Horta-Guinardó, Sant Martí, Sant Andreu, Les Corts, Nou Barris i Sants-Montjuïc.* (*Véase Tabla Anexos.1*)

Aparte, únicamente tenemos interés en la fecha registrada, el número de positivos y el sexo. Por tanto, llevaremos a cabo una limpieza de la base de datos hasta obtener la tabla deseada.

En nuestra base de datos no solamente tenemos todos los datos para todas las regiones de Cataluña, si no que el número de casos por día puede estar repartido en varias filas, ya que como hemos comentado anteriormente, cada fila hace referencia al número de casos con características iguales. Es decir, el positivo de un hombre por PCR el 27/02/2021, ocupa una fila diferente al positivo de una mujer que ha dado positivo por una prueba de antígenos el 27/02/2021, y otra fila puede ser una mujer que ha dado positivo por PCR el 27/02/2021, perteneciendo los 3 a la misma región sanitaria. Como pueden ver la cantidad de datos es bastante extensa, y la limpieza a realizar, para conseguir la base de datos deseada, es amplia.

Para empezar la limpieza filtramos los datos en función del código de la región sanitaria. Como solo nos interesan los datos de Barcelona ciudad, seleccionamos aquellas filas que tienen como

código en la variable *RegioSanitariaCodi* el 7803. Una vez tenemos únicamente los datos de la ciudad de Barcelona, ahora nos toca agrupar el número de casos por distrito, día y sexo. Como hemos dicho anteriormente, la base de datos original diferencia entre las pruebas realizadas para detectar el positivo. Para nosotros esta información no es relevante, así que procedemos contabilizar los casos detectados cada día, en cada distrito en función del sexo. Hay que tener en cuenta que, si algún día no se detectó ningún caso, no encontraremos dicha entrada en la base de datos, por ello tendremos que crearla nosotros. Decidimos trabajar con única tabla con todos los distritos para optimizar el tiempo. Tendremos que aplicar los mismos procesos a todos los distritos, por ello consideramos oportuno trabajar con una única tabla de dimensiones superiores, a trabajar con 10 tablas distintas con dimensiones inferiores.

Después de llevar a cabo todos estos procesos tenemos una única tabla con entradas para todos los distritos, días, y diferenciando entre sexo donde podemos visualizar los casos detectados. A parte, también tenemos tablas individuales para cada distrito, y una tabla con todos los casos detectados diarios en cada distrito sin distinción de sexo.

II.2. Modelos AR

Como hemos ido comentando a lo largo del trabajo, las series temporales se pueden explicar mediante modelos, los cuales varían en función del tipo de distribución de nuestra variable de estudio. En este apartado observaremos uno de los modelos más comunes para las series continuas.

El modelo autorregresivo de orden p , $AR(p)$, es un proceso estocástico y lineal que se expresa de la siguiente manera:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \varepsilon_t \quad (1)$$

También lo podemos escribir con el operador de retardo en forma de polinomio, y quedaría tal que así:

$$\phi(L)x_t = \varepsilon_t$$

En esta fórmula, $\phi(L)$ hace referencia a un polinomio, el cual se podría expresar de la siguiente forma:

$$\phi(L)x_t = (1 - \phi)Lx_t$$

El operador ϕ es el operador de retardo, y su función es la de retrasar en el tiempo el valor de una variable.

$$\phi^k x_t = x_{t-k}$$

Podemos observar que el orden del operador es el número de posiciones que retrasamos la variable.

Los diversos parámetros de ϕ se pueden interpretar como las personas contagiadas el día $t-j$ y que siguen contagiadas el día t . En nuestro caso, como las series son simples contagios diarios, esta interpretación no nos sirve, y las ϕ son simples valores teóricos.

El componente ε_t se entiende como las innovaciones del nuevo día que no dependen de los valores anteriores de la serie, y son un componente aleatorio. Este parámetro está considerado como ruido blanco. El ruido blanco es un proceso estocástico que cumple las siguientes características y sigue una distribución normal:

- $E(\varepsilon_t) = 0 \quad \forall t \in Z$
- $var(\varepsilon_t) = \sigma_\varepsilon^2 \quad \forall t \in Z$
- $cov(\varepsilon_s, \varepsilon_t) = 0 \quad \forall s \neq t \in Z$

Debido a su definición, los procesos AR son siempre invertibles, ya que cumplen la condición siguiente para todos los valores posibles de los parámetros.

$$\text{Condición invertibilidad: } \varepsilon_t = \sum_{j=0}^{\infty} \phi_j x_{t-j}$$

En cambio, para comprobar la estacionariedad debemos observar que los parámetros cumplan las siguientes condiciones:

- **Condición necesaria pero no suficiente:** $\left| \sum_{j=1}^p \phi_j \right| < 1$
- **Condición suficiente pero no necesaria:** $\sum_{j=1}^p |\phi_j| < 1$

A partir de estas fórmulas, podemos representar todas las condiciones para los modelos AR. A continuación, veremos las condiciones para los modelos más comunes, AR (1) y AR (2).

- **AR (1):** $|\phi_1| < 1$
- **AR (2):** $\phi_2 - \phi_1 < 1, \phi_2 + \phi_1 < 1, |\phi_2| < 1$

A nivel de predicciones, a partir de la expresión (1) podemos deducir la forma que tomarán las futuras observaciones. La expresión sería la siguiente:

$$x_{t+k} = \phi_1 x_{t-1+k} + \phi_2 x_{t-2+k} + \dots + \phi_p x_{t-p+k} + \varepsilon_{t+k} \quad (2)$$

Esta fórmula, para un AR (1), se puede generalizar de la siguiente forma:

$$x_{t+k} = \phi^{k+1} * x_{t-1} + \sum_{n=0}^k \phi^{k-n} * \varepsilon_{t+n} \quad (3)$$

Al seguir el modelo una distribución normal, podemos suponer que el intervalo de confianza de una predicción se puede expresar de la siguiente manera:

$$P \left[\bar{x} - z_{\alpha/2} \frac{\delta}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \frac{\delta}{\sqrt{n}} \right] = 1 - \alpha$$

Donde \bar{x} hace referencia a la predicción, $z_{\alpha/2}$ al valor de una $N(0,1)$ con una probabilidad de $\frac{\alpha}{2}$ a su derecha, y $\frac{\delta}{\sqrt{n}}$ representa el error estándar.

De dicha fórmula podemos sacar nuestra región de predicción:

$$\left[\bar{x} - z_{\alpha/2} \frac{\delta}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\delta}{\sqrt{n}} \right]$$

II.3. Modelos INAR

En este apartado abordaremos uno de los modelos utilizados para las series temporales discretas. Dicho modelo se conoce como el proceso autorregresivo de valores enteros no negativos, INAR. El modelo $\text{INAR}(p)$ viene definido por la siguiente fórmula: [5]

$$X_t = \alpha_1 \circ X_{t-1} + \alpha_2 \circ X_{t-2} + \alpha_3 \circ X_{t-3} + \dots + \alpha_p \circ X_{t-p} + \varepsilon_t \quad (4)$$

Donde $\alpha_1, \dots, \alpha_p$ son parámetros fijos y ε_t se asume que sigue una distribución Poisson con una media fija correspondiente a λ [6]. Además, los parámetros X_{t-1} y ε_t se consideran independientes entre sí en cualquier momento de t . El símbolo \circ hace referencia al operador binomial de refinamiento, el cual se expresa de la siguiente manera:

$$\alpha_j \circ X_{t-j} = \sum_{i=1}^{X_{t-j}} Y_i \quad (5)$$

Donde Y_i son variables aleatorias independientes e idénticamente distribuidas (i.i.d) que siguen una Bernoulli con una probabilidad de éxito equivalente a α_j . Nuestra variable Y_i adopta valor 1, si un individuo i que da positivo en la semana $t-j$, sigue siendo positivo en la semana j , o 0 si no se cumple esta condición.

Los parámetros del modelo $\text{INAR}(p)$ puede tener varias interpretaciones. Una interpretación puede ser la siguiente: el parámetro α_j hace referencia a la proporción de individuos que dan positivo en la semana $t-j$ y siguen siéndolo en la semana j . Este no es nuestro caso ya que los casos diarios no son acumulados, sino que simplemente nuestra base de datos nos indica los casos que ha habido ese día. Para nosotros es un simple valor teórico. Por otro lado, el componente ε_j hace referencia a aquellos individuos que han dado positivo, por primera vez, en la semana j .

Suponiendo estacionariedad, la esperanza de un proceso $\text{INAR}(p)$ viene dada por la siguiente fórmula: [7]

$$E[X_t] = \frac{\mu_\varepsilon}{1 - \sum_{i=1}^p \alpha_i}$$

Los coeficientes de autocorrelación vienen dados de la siguiente forma: [8]

$$\rho_0 = 1,$$

$$\rho_1 = \alpha_1,$$

$$\rho_k = \alpha_1 \rho_{k-1} + \alpha_2 \rho_{k-2}, k \geq 2$$

A nivel de predicciones, nos centraremos en las predicciones a corto período.[6]

A partir de la fórmula (4) podemos descifrar la fórmula que seguirán las predicciones para los modelos INAR. La expresión es la siguiente:

$$X_{t+k} = \alpha_1 \circ X_{t-1+k} + \alpha_2 \circ X_{t-2+k} + \alpha_3 \circ X_{t-3+k} + \dots + \alpha_p \circ X_{t-p+k} + \varepsilon_t \quad (6)$$

En este mismo artículo [6] se explican las variaciones que se han de aplicar a la fórmula (6) para expresar la estacionalidad que pueda presentar la serie.

$$X_{t+1} = \alpha_1 \circ X_t + W_t(\lambda_{t+1}) \quad (7)$$

A partir de esta ecuación, podemos generalizar la forma que tendrán todas las predicciones para $k \geq 1$, las cuales se pueden obtener expresando X_{t+k} de la siguiente manera:

$$X_{t+k} = f_k \circ X_t + W(\lambda_k^*)$$

Donde $f_k = \alpha_1^k$ y $W(\lambda_k^*) = \lambda_{t+k}$. A partir de aquí podemos calcular las relaciones que siguen las demás predicciones, para $k \geq 1$:

- $f_{k+1} = \alpha_1 * f_k$
- $\lambda_{k+1}^* = \alpha_1 * \lambda_k^* + \lambda_{t+k+1}$

A partir de estas relaciones, podemos generalizar la función que siguen las predicciones de un INAR (1):

$$X_{t+k} = p_1^k \circ X_t + \sum_{n=1}^k p_1^{k-n} * \lambda_{t+n} \quad (8)$$

Reemplazando los parámetros de las fórmulas superiores por sus estimaciones, podemos aproximar las regiones de predicción con un tamaño $1 - \alpha$, encontrando los límites superiores e inferiores a partir de las siguientes fórmulas:

$$\sum_{j=0}^{l_1} P(X_{t+k} = j) \approx \alpha/2, \sum_{j=l_1}^{l_2} P(X_{t+k} = j) \approx 1 - \alpha$$

En estas ecuaciones vemos cómo vamos sumando las probabilidades de que las predicciones tomen el valor j , hasta que llegamos a la probabilidad $\alpha/2$. Dicha j , será nuestro límite inferior, a partir del cual calcularemos el límite superior de la misma manera. Sumaremos probabilidades hasta alcanzar la probabilidad $1 - \alpha$.

Con estas fórmulas, también podremos calcular la mediana de las predicciones. La única variación es la de sumar probabilidades hasta llegar a la probabilidad 0.50.

$$\sum_{j=0}^{l_1} P(X_{t+k} = j) \approx 0.50$$

Los apartados correspondientes a la estimación de los parámetros, el ajuste del modelo y las predicciones se podrán consultar a continuación en el capítulo III

II.3. Indicadores para las predicciones

La eficacia de las predicciones se puede analizar de varias formas. Aquí explicaremos los 3 indicadores que utilizaremos para determinar si las predicciones del modelo se ajustan correctamente a la serie o no. Dichos indicadores serán: los errores de predicción, la cobertura, y el rango de las regiones de predicción.

Los errores que estudiaremos son; el error cuadrático medio (EQM), el error absoluto medio (EAM), y el error porcentual absoluto (EPAM).

- **EQM** = $\frac{1}{k} \sum_{j=1}^k (\widehat{X}_T(j) - X_{t+j})^2$
- **EAM** = $\frac{1}{k} \sum_{j=1}^k |\widehat{X}_T(j) - X_{t+j}|$ (9)
- **EPAM** = $\frac{1}{k} \sum_{j=1}^k \frac{|\widehat{X}_T(j) - X_{t+j}|}{X_{t+j}}$

Donde $\widehat{X}_T(j)$ hace referencia a la predicción $t+j$, y k hace referencia al número de predicciones calculadas.

En relación con la cobertura, simplemente hemos contado las observaciones entre enero y marzo de 2021 que se encuentran dentro de las regiones de predicción calculadas para dicho tiempo muestral.

Para calcular el rango de las regiones de predicción, hemos calculado el rango de cada predicción, restando el límite inferior al límite superior, y una vez hemos calculado el rango para cada predicción hemos calculado la media de todas ellas. El rango lo podemos utilizar para analizar la precisión de los modelos a la hora de predecir.

III. RESULTADOS

III.1. Descriptiva de los Datos

Nuestra base de datos está dividida por los diversos distritos de Barcelona, y en este apartado procederemos a comentar y analizar cada distrito por separado.

Primero de todo observaremos de manera gráfica cómo se comportan las series.

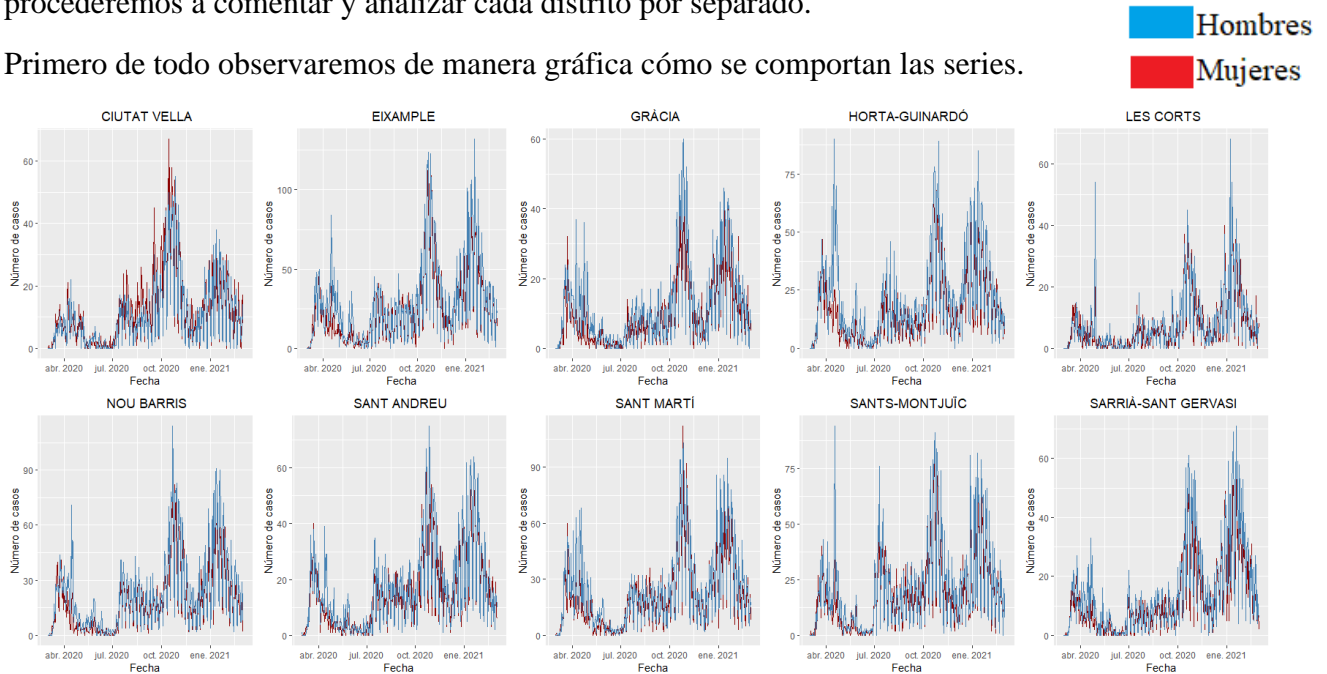


Ilustración 5. Representación gráfica del número de casos detectados en los distritos de Barcelona

En todos los distritos podemos ver como el virus se ha comportado de manera similar, con una ola inicial alrededor de marzo-abril de 2020, inicio de la pandemia. Posteriormente, a finales de julio de 2020 podemos ver como se inicia una segunda ola, que durará hasta inicios de diciembre, y alcanza su pico, y el pico máximo de la serie, en octubre-noviembre. Finalmente, entre mediados y finales de diciembre empieza una tercera ola, muy similar a la segunda e incluso sobrepasándola en algunos distritos, Les Corts o Sarrià-Sant Gervasi, la cual perdura hasta marzo de 2021, fecha final de nuestro estudio. *(Para ver las representaciones individuales de cada distrito ir a los Anexos)*

A nivel de sexos vemos que la COVID-19 se comporta más o menos igual en hombres como en mujeres, con alguna diferencia puntual, por ejemplo, el pico de casos en hombres de Les Corts alrededor de mayo 2020. Esto tiene sentido ya que no se han detectado evidencias para afirmar que la COVID-19 se comporta distinto en función del sexo. Las posibles diferencias que se observen se pueden deber más aun tema social o psicológico, que hace que un sexo respete mejor las medidas que otro, o tenga más propensión a infectarse. Pero aquí no podemos afirmar ni siquiera que estas diferencias sean significativas. Más adelante cuando estimemos los parámetros del modelo podremos ahondar en esta hipótesis.

Considero importante explicar, que, aunque la primera ola sea la menor en cuanto a número de casos detectados, fue la más mortal. Este bajo número de casos detectados se puede explicar fácilmente. Al inicio de la pandemia no había métodos necesarios para dicha detección, y, además, los hospitales estuvieron colapsados debido al enorme número de gente infectada, lo

que provocó que no se pudiera atender a todo el mundo, en especial, aquellos que no mostraban síntomas claros del virus. Es por ello por lo que el número de casos es inferior a las demás olas, las cuales sucedieron cuando había métodos eficaces de detección, y dichos métodos estaban disponibles para todos aquellos sospechosos de estar infectados. *(Para ver una proyección de la mortalidad ir Ilustración Anexos.1)*

Ahora procederemos a analizar la serie desde un punto de vista estadístico. Como hemos ido comentado anteriormente, para realizar un buen análisis de la serie y poder estimar los modelos que la explican, necesitamos que nuestras series sean estacionarias, para poder analizar de una forma correcta los coeficientes de autocorrelación. La estacionariedad viene dada por una varianza y media constantes.

Por desgracia, en nuestros gráficos podemos ver como los casos van aumentando a medida que se detectan más, una pequeña tendencia, y, por tanto, podemos afirmar con claridad que nuestra serie no es estacionaria con relación a la media.

A nivel de varianza no encontramos muchos problemas. El mayor problema de la varianza es la estacionalidad ya que produciría valores anómalos, lejanos a la media, sin que haya aumentado o disminuido el número de casos. Un ejemplo claro sería una serie temporal con el número de casos por la gripe. Durante todo el año el número de casos sería similar, pero al llegar el invierno aumentará de forma drástica. En nuestro caso vemos como el número de casos se mantiene similar a sus anteriores.

Una vez hemos visto la representación gráfica de las series, y hemos determinado que no existe estacionariedad, vamos a analizar más en profundidad cada distrito. Para ello calcularemos el mínimo y máximo de contagios en un día, la media de contagios por día y su desviación estándar, el número de casos totales en el distrito desde el inicio hasta el final del estudio (1/03/2020 – 3/03/2021), y finalmente calcularemos la incidencia de casos por 10.000 habitantes, y relacionaremos todos estos valores con la población total del distrito, su densidad y la renta.

Tabla 3.1 Estadística descriptiva de los casos detectados por distrito

	Media	Desviación estándar	Máximo Casos Diarios	Número de Casos Totales	Casos por 10.000 habitantes	Población	Densidad Poblacional	Renta familiar en miles de €
CIUTAT VELLA	11,03	11,20	67	8.116	749,19	108.331	23.908	84.3
EIXAMPLE	26,20	24,38	132	19.282	712,32	270.694	35.556	122.4
GRÀCIA	10,49	10,73	60	7.722	624,50	123.651	29.130	105.3
HORTA-GUINARDÓ	18,33	16,70	90	13.493	771,92	174.799	14.155	78.0
LES CORTS	7,54	9,23	68	5.551	671,75	82.635	13.697	137.3
NOU BARRIS	20,76	19,61	114	15.280	878,10	174.012	20.820	55.0
SANT ANDREU	15,73	14,61	75	11.574	761,57	151.976	22.459	74.6
SANT MARTÍ	22,34	19,94	112	16.443	680,33	241.691	22.129	88.1
SANTS-MONTJUÏC	19,88	18,62	94	14.629	779,86	187.584	8.024	84.6
SARRIÀ-SANT GERVASI	13,35	13,82	71	9.828	650,18	151.157	7.235	182.8

En términos generales, podemos observar como el distrito que ha sufrido un mayor impacto por la COVID-19 ha sido el distrito de l'Eixample, superior en todas las categorías a los demás distritos. Hay que tener en cuenta que también es el distrito con mayor población y densidad, y, por tanto, tiene sentido que sea el distrito con mayor afectación. En términos generales

podemos ver como el número de casos totales se encuentra en torno a los 8.000-14.000 casos. Dentro de este intervalo encontramos la mayoría de los distritos como: Ciutat Vella, Horta-Guinardó, Sant Andreu, Sants-Montjuïc y Sarrià- Sant Gervasi. Sorprendentemente, Les Corts es el distrito, no solo con menor población, sino también con una menor afectación, ya que el total de casos detectados es de 5.551, bastante más bajo que en el siguiente en la lista, Gràcia con 7.772. Por encima de la media no solo encontramos l'Eixample, si no también encontramos Nou Barris y Sant Martí, con 15.280 y 16.443, respectivamente. *(Para ver tablas específicas en función del sexo mirar tablas Anexos 3 y 4).*

A nivel socioeconómico, nuestra hipótesis es que aquellos distritos con una renta familiar superior tendrán menos casos. No solo porque en teoría la población ha de ser menor, ya que no todo el mundo se puede permitir las viviendas en dichos distritos, si no también, porque a medida que aumenta la renta las viviendas tienden a ser más espaciales, y las familias no se encuentran con la necesidad de compartir residencia con otras burbujas familiares, o incluso alquilar habitaciones. Si observamos los distritos con mayor renta, estos son: Sarrià-Sant Gervasi, Les Corts y L'Eixample. Tanto Les Corts como Sarrià son los distritos con mayores rentas, y los distritos con menos casos, pero también son los distritos con menor población y densidad. Si miramos L'Eixample, aun siendo el tercer distrito con mayor renta, es el distrito con mayor número de casos, población y densidad poblacional. Estas desavenencias en el distrito de l'Eixample se pueden explicar gracias a los barrios que forman l'Eixample. Los barrios de l'Esquerra de l'Eixample, y, sobre todo, el barrio de La Dreta de l'Eixample, donde encontramos Paseo de Gràcia, son barrios de alto nivel adquisitivo, que hacen que la renta del distrito total aumente, aunque en los demás distritos la renta no sea tan alta. Es por ello por lo que, aunque tengamos una renta alta, el número de casos es más elevado de lo que se podía suponer. Aún no podemos afirmar que la renta realmente afecte al número de contagios, porque los distritos que hemos comprobado distaban bastante en cuanto a niveles de población y densidad poblacional. Si miramos otros distritos con características similares, tanto en renta como en población y densidad poblacional, encontramos los distritos de Sant Andreu, Sant Martí y Ciutat Vella. En todos observamos una renta próxima a 80.000€, y una densidad poblacional de 22.000 personas por km². Observamos como Sant Andreu es el distrito con menos casos totales, y menor renta, pero es el distrito con mayor incidencia de casos por 10.000 habitantes. Esto nos reafirma en la creencia de que, en distritos con menor renta, en las viviendas tienden a vivir más personas, por tanto, parece que se detectan muchos casos en los núcleos familiares y eso hace que la incidencia sea alta, pero después el virus no se contagia hacia otras residencias, reduciendo el número de casos totales, como puede suceder en los otros distritos, donde la incidencia es menor, pero el número de casos totales es mayor.

En conclusión, parece ser que la renta, más que afectar al número de casos totales, afecta a la incidencia de casos totales por 10.000 habitantes, ya que las bajas rentas promueven que se creen núcleos residenciales más acinados. Mientras que la población total y, en concreto, su densidad, afectan al número de casos totales detectados, ya que, a más población o densidad poblacional, existe más probabilidad de contagio.

Probablemente el distrito más curioso sea el de Sants-Montjuïc, ya es el segundo distrito con más población, pero vemos que es el segundo distrito con una menor densidad. Esto se puede explicar en gran parte al complejo de Montjuïc, el cual no es habitable. Por tanto, podemos intuir que esta densidad es un poco irreal, ya que, aunque el distrito sea muy grande, no todo es zona habitable, y resulta en una densidad poblacional baja. Esto lo que provoca, son núcleos de

viviendas concentrados, y provoca que los casos en Sants-Montjuïc no acaben de seguir la lógica establecida por la densidad de población y la población total. Tiene un número elevado de casos por la baja densidad de población, pero si miramos la población total, se observan menos casos que en otros distritos con menos residentes.

Otra curiosidad que no se muestra en la tabla, es que en todos los distritos el mínimo de casos detectados en un día fue 0. Esta información no es muy relevante ya que la detección no significa que no hubiera gente nueva contagiada ese día.

En definitiva, podemos observar en la tabla como existen diferencias a nivel de contagio en los diferentes distritos, y entre sexos, lo que defiende nuestras hipótesis iniciales de que realmente existen diferencias entre géneros y tipos de distrito. En general, los distritos con mayor población y densidad son los distritos con más casos, mientras que, a nivel de sexo, parece que son las mujeres aquellas con más propensión al contagio (*Véanse Tablas Anexos 3 y 4*).

III.2. Identificación de los modelos

Para poder determinar el tipo de modelo y su orden, se procederá a analizar los coeficientes de autocorrelación (ACF) y los coeficientes de autocorrelación parciales (PACF) de cada distrito. Estos gráficos nos indican el grado de correlación que tiene las observaciones de la serie. Para ver si existe correlación tenemos que observar cuantos coeficientes son superiores a 0, dichos coeficientes nos dirán el grado de correlación entre las observaciones. En nuestro caso, serán el número de días anteriores que afectan al número de casos del día analizado.

Debido a que la COVID-19 es una enfermedad contagiosa, donde los contagios de un día dependen de los casos detectados en los días anteriores, creemos correcto suponer que los modelos autorregresivos son los indicados para su análisis. Estos modelos se caracterizan por tener un ACF donde el número de coeficientes estadísticamente diferentes de 0, decrecen de forma exponencial hacia 0, y unos PACF con p coeficientes superiores a 0, los cuales nos indican el orden del modelo.

Cosa muy importante para destacar que sabemos, y era esperado, es que nuestra serie no es estacionaria, y, por tanto, esperamos que los resultados que obtengamos no serán perfectos al 100%.

- ACF

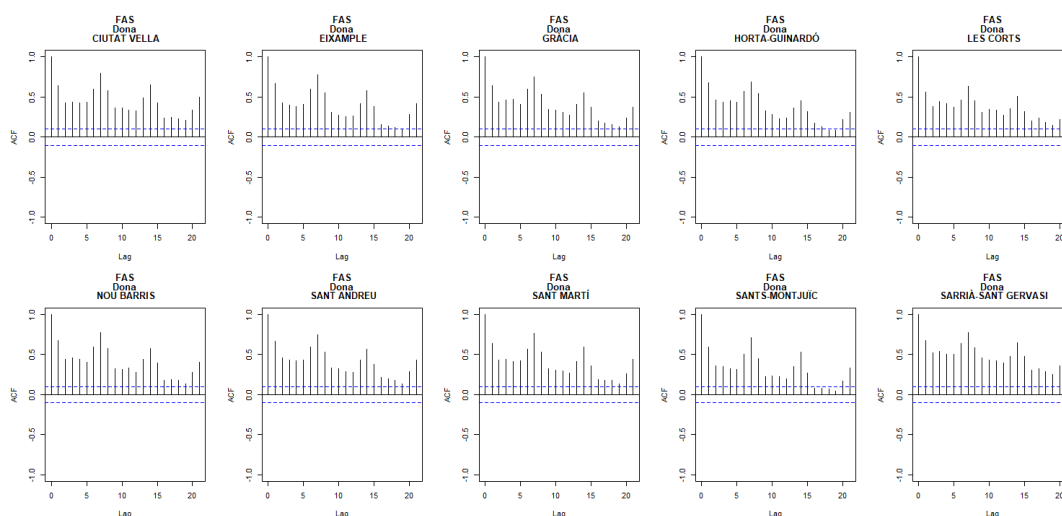


Ilustración 6. FAS para las mujeres en los diferentes distritos de Barcelona

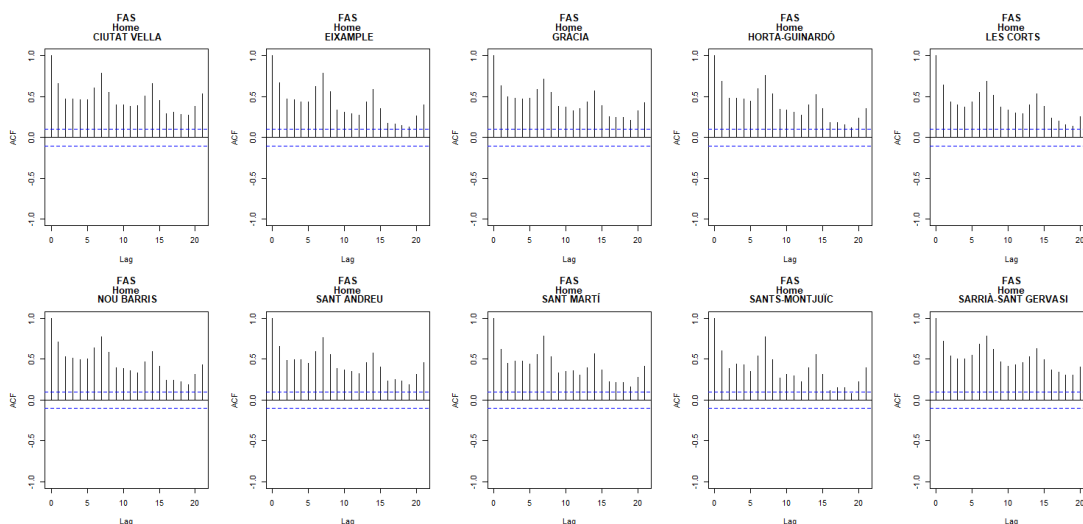


Ilustración 7. FAS de los hombres en los distintos distritos de Barcelona

Si observamos las ilustraciones 6 y 7, podemos ver claramente como los coeficientes de autocorrelación en ningún distrito tienden a 0 a lo largo del tiempo, pero vemos como descienden exponencialmente con el tiempo.

También es importante destacar el comportamiento de dichos coeficientes y la forma que dibujan. El primer coeficiente es siempre 1, pero luego observamos que entre los coeficientes 2 y 8, desciende y asciende formando una especie de valle, un comportamiento simétrico. Esta simetría nos indica que puede existir cierta estacionalidad de orden 7, es decir, semanal. Dicha estacionalidad se podrá confirmar cuando analicemos los PACF. El hecho de que los coeficientes no hayan alcanzado el valor 0, se explica por esta posible estacionalidad. Si realmente existe, al haber solo 21 coeficientes solo hemos dado 3 “vueltas”, y, por eso es posible que aún no se puedan considerar 0. Si calculamos más coeficientes seguramente viéramos a los coeficientes alcanzar la región significativa.

- PACF

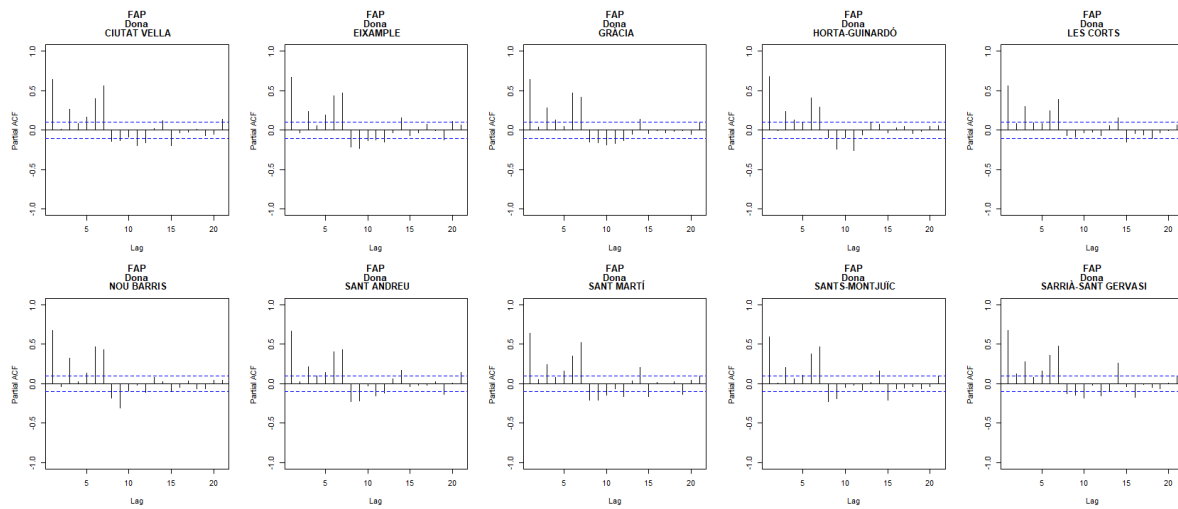


Ilustración 8. PACF para las mujeres distintos distritos de Barcelona

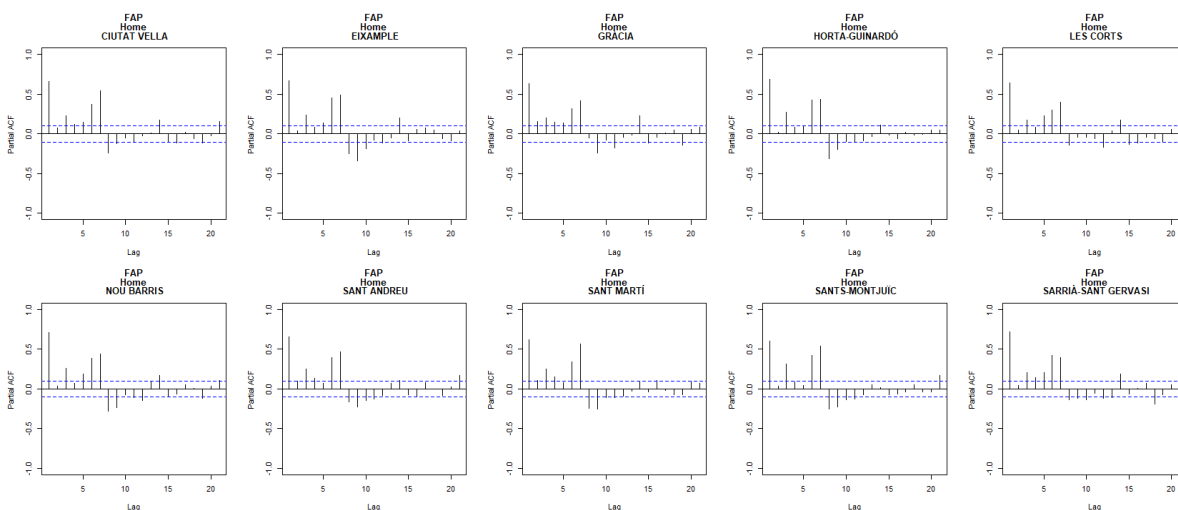


Ilustración 9. PACF para los hombres en los distintos distritos de Barcelona

Al contrario que en el gráfico de la ACF, podemos ver claramente, como en ambos sexos, el número de coeficientes no descienden exponencialmente a 0, si no que únicamente observamos como el primer coeficiente es estadísticamente diferente de 0, y luego los demás coeficientes son estadísticamente equivalentes a 0, exceptuando algún coeficiente específico. Esto nos reafirma definitivamente que nuestros modelos son del tipo AR. (Véase tabla Anexos.2)

En la mayoría de los distritos el número de coeficientes superiores a 0, antes de equivaler a 0, son 1. Se pueden apreciar ciertos coeficientes que superan el 0 en momentos puntuales, lo cual nos podría indicar cierta estacionalidad. Es decir, que cada x días hubiera más probabilidad de contagio, por ejemplo, los fines de semana es probable que la gente socialice más, y, por tanto, hay más posibilidades de que los lunes el número de contagios aumente. Esto podría provocar una pequeña modificación en los modelos, ya que el día de la semana afectaría al número de innovaciones que se suceden.

Observamos que dichos coeficientes se repiten en los coeficientes 1, 7 y 14, lo que nos indica una estacionalidad de orden 7, tal y como sospechamos anteriormente. Los demás coeficientes que son estadísticamente diferentes de 0, y no ocupan una posición múltiple de 7, se explican mediante el componente aleatorio de la serie, que influye en que algunos coeficientes salgan significativos, pero no tiene mayor importancia.

En conclusión, gracias al análisis de los ACF y PACF hemos podido confirmar que realmente los modelos a parametrizar son del tipo autorregresivo, y, además, estos modelos serán de orden 1 con una estacionalidad de orden 7.

III.3. Modelos AR

Durante todo este apartado, al igual que en los anteriores, hay que recordar que trabajamos con el conteo de casos de la COVID-19 en Barcelona, con los datos desagregados por sexo y distrito.

III.3.1 Estimación de los parámetros

El modelo autorregresivo sigue la *fórmula (1)* explicada anteriormente.

Como podemos observar, al haber determinado que el orden de nuestros modelos es 1, únicamente tendremos que estimar el parámetro ϕ_1 . Para ello usaremos el método de la máxima verosimilitud. Una vez aplicada el código de R obtenemos los parámetros siguientes:

Tabla 3.2 Parámetros modelo AR

	Mujeres				Hombres			
	ϕ_1	Desviación estándar	Intercept	Desviación estándar	ϕ_1	Desviación estándar	Intercept	Desviación estándar
CIUTAT VELLA	0,6448	0,0397	10,3741	0,1114	0,6573	0,0391	11,5436	0,1212
EIXAMPLE	0,6660	0,0388	28,3597	0,1108	0,6710	0,0385	23,6629	0,0968
GRÀCIA	0,6417	0,0399	11,6624	0,1065	0,6342	0,0402	9,1789	0,0822
HORTA-GUINARDÓ	0,6828	0,0380	20,6341	0,1255	0,6849	0,0379	15,6833	0,0985
LES CORTS	0,5661	0,0428	7,9256	0,1155	0,6412	0,0398	7,0891	0,1168
NOU BARRIS	0,6754	0,0383	22,7039	0,1445	0,7160	0,0363	18,3635	0,1272
SANT ANDREU	0,6634	0,0389	16,7437	0,1182	0,6564	0,0392	14,4677	0,1007
SANT MARTÍ	0,6370	0,0401	24,1985	0,0973	0,6249	0,0406	20,1791	0,0812
SANTS-MONTJUÏC	0,5919	0,0420	21,3601	0,1101	0,6015	0,0416	18,1391	0,093
SARRIÀ-SANT GERVASI	0,6717	0,0385	14,6853	0,1163	0,7214	0,036	11,7370	0,1046

No se observan grandes diferencias en ϕ_1 a nivel de sexos, parece que los hombres obtienen valores ligeramente superiores. La mayor separación se encuentra en el distrito de Nou Barris, donde los casos del día anterior afectan mucho menos a las mujeres que a los hombres.

El parámetro intercept hace referencia a las innovaciones, es decir, los casos nuevos detectados diariamente. Vemos que, en todos los distritos, excepto Ciutat Vella, las mujeres tienden a tener más casos diarios que los hombres. Esto se podría entender como que las mujeres se contagian en mayor abundancia que los hombres. Esta observación se complementa con lo observado en la descriptiva desagregada por sexos (*tablas Anexo.3 y Anexos.4*), donde las mujeres eran el sexo con más casos detectados, y refuerzan nuestra hipótesis de que existen diferencias entre sexos a la hora de casos detectados.

A nivel de significación de los parámetros, hemos calculado su error estándar y hemos comprobado que ϕ_1 es significativamente diferente de 0.

III.3.2. Validación del modelo

Una vez tenemos un posible modelo, hemos de validarlo, comprobar que realmente cumple las condiciones necesarias para poder afirmar que este modelo explica la serie, y, por tanto, poder aplicarlo para intentar predecir los valores futuros de la serie. Hay que destacar que nuestra serie no es estacionaria, y, por tanto, puede fallar alguna validación.

La validación consta de diversos pasos:

1. Normalidad de los residuos
2. Autocorrelación residuos

III.3.2.1. Normalidad de los residuos

Para asegurar la normalidad de los residuos llevaremos a cabo la prueba de Shapiro-Wilk.

Tabla 3.3. Valores P-valor prueba de Shapiro_Wilk

	P-Valor	
	Hombres	Mujeres
CIUTAT VELLA	1,45E-15	1,40E-15
EIXAMPLE	9,91E-17	1,27E-16
GRÀCIA	2,99E-16	3,88E-18
HORTA-GUINARDÓ	1,46E-14	1,83E-12
LES CORTS	1,75E-20	1,56E-23
NOU BARRIS	7,52E-17	4,88E-16
SANT ANDREU	8,31E-16	9,35E-15
SANT MARTÍ	5,79E-19	7,08E-18
SANTS-MONTJUÏC	8,17E-18	2,76E-16
SARRIÀ-SANT GERVASI	1,39E-16	3,35E-16

Observamos como todos los p-valores de las pruebas son inferiores a 0.05, y, por tanto, tenemos evidencias suficientes para rechazar la hipótesis nula de normalidad, y afirmar que la distribución de la serie no es una normal. Este resultado tiene todo el sentido del mundo ya que nuestra serie es un conteo de valores bajos, y, por tanto, es muy difícil que su

distribución marginal siga una normal. Además, sabemos que la serie no es estacionaria y por ello el modelo sabemos que no será perfecto.

III.3.2.2. Independencia de los residuos

Tenemos que asegurar que los residuos del modelo se comportan como ruido blanco. Para ello llevaremos a cabo la prueba de Box-Pierce.

Tabla 3.4. P-valor prueba de Box-Pierce

	P-Valor	
	Hombres	Mujeres
CIUTAT VELLA	0,3226	0,8570
EIXAMPLE	0,6071	0,6860
GRÀCIA	0,0502	0,6101
HORTA-GUINARDÓ	0,7243	0,9480
LES CORTS	0,5741	0,3496
NOU BARRIS	0,6489	0,6655
SANT ANDREU	0,1975	0,6889
SANT MARTÍ	0,2015	0,5350
SANTS-MONTJUÏC	0,6671	0,9109
SARRIÀ-SANT GERVASI	0,5314	0,0969

Hemos obtenido todos los p-valores superiores a 0.05, por tanto, con un nivel de significación del 95%, no tenemos suficientes evidencias para rechazar la hipótesis nula de independencia, y, entonces podemos afirmar que los residuos son independientes entre sí.

En conclusión, hemos validado correctamente el modelo, a excepción de la normalidad de los residuos, y, a continuación, podemos llevar a cabo las predicciones.

III.3.3 Predicciones

Para las predicciones, no solo comprobaremos la capacidad predictiva del modelo para valores conocidos (Predicción ex - post) sino que también comprobaremos la eficacia del modelo para valores futuros desconocidos (Predicción ex - ante). Las predicciones seguirán la forma expresada en la fórmula (3).

III.3.3.1 Predicción Ex - Post

Para la predicción Ex - Post, cogeremos la serie desde el 1/03/2020 hasta el 31/12/2020, y calcularemos las predicciones diarias para los meses de enero y febrero de 2021. Los parámetros de la *Tabla 3.2.* hacen referencia a los modelos con toda la serie como período

muestral. Al haber modificado la muestra estimaremos nuevos modelos. (Ver parámetros en *Tabla Anexos.5*)

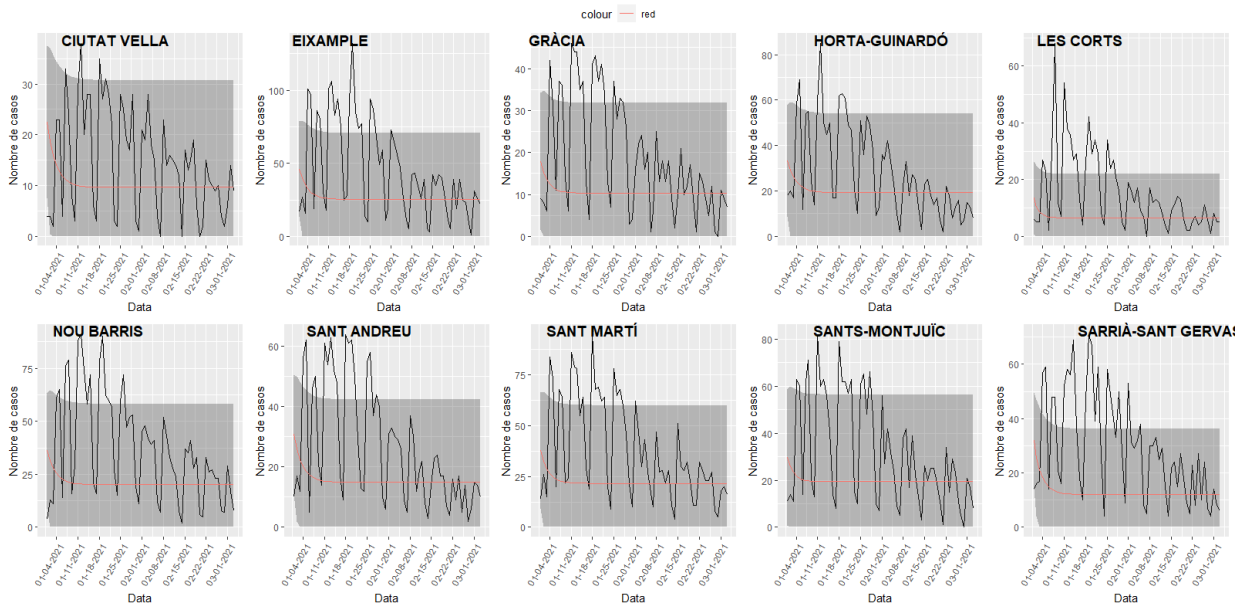


Ilustración 10. Predicciones Ex-Post del modelo AR para las mujeres

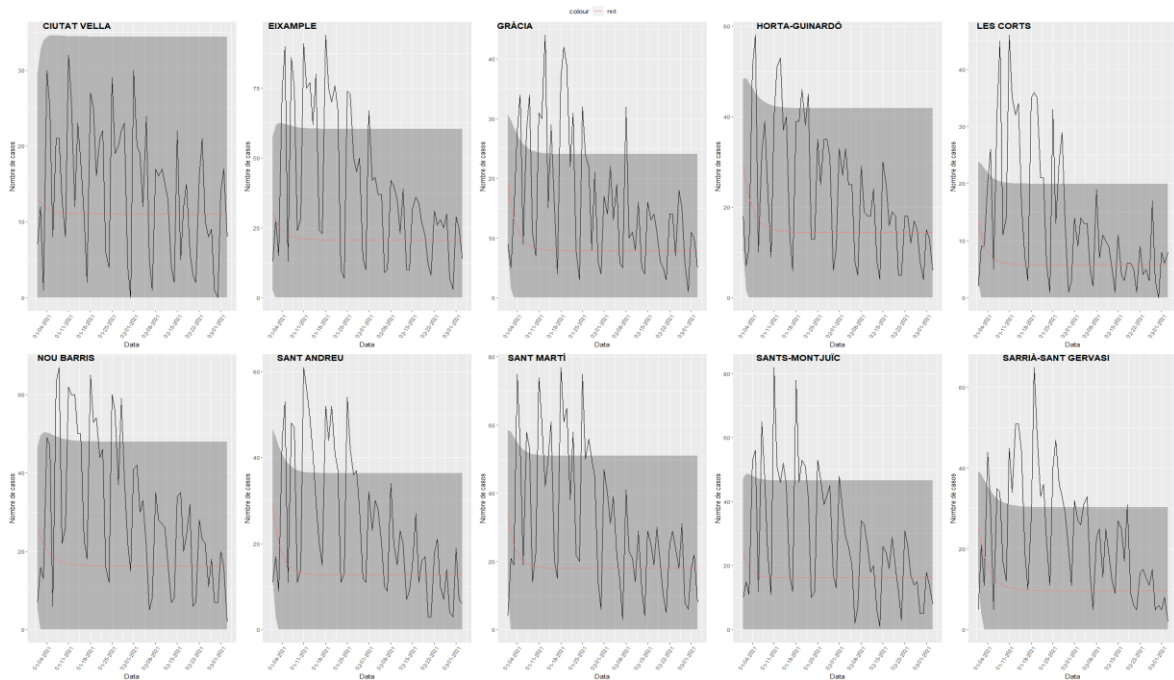


Ilustración 11. Predicciones Ex-Post del modelo AR para los hombres

En ambos sexos vemos como las predicciones acaban convergiendo a la media de casos diarios en cada distrito (Véanse *Ilustraciones 10* y *11*). Esta convergencia, viene explicada simplemente por la expresión del modelo AR, y, nos indica que nuestro modelo AR no es capaz de captar en su totalidad la volatilidad de nuestros datos.

A nivel de región de predicción, observamos como en la mayoría de los distritos, tanto en hombres como en mujeres, las regiones de predicción abarcan la mitad o la totalidad de las observaciones en nuestro espacio temporal. Algo a destacar es que vemos como a medida que disminuyen el número de casos, aumenta la cobertura del modelo. Esto nos puede estar indicando que el modelo puede ser bastante bueno para los valores futuros que desconocemos.

Para asegurarnos la capacidad predictiva calcularemos los diferentes índices de eficacia explicados en **II.3 Indicadores para las predicciones**.

Al haber predicho valores conocidos, podemos calcular el error de predicción y determinar de una forma numérica la eficacia de nuestro modelo (*Véanse fórmulas 9*). Como nuestras predicciones convergen en la media de la incidencia, dichos errores nos mostrarán la habilidad del modelo para aproximarse a la media.

Los errores obtenidos son los siguientes:

Tabla 3.5. Errores de predicción para el modelo AR

	EQM		EAM		EPAM	
	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres
CIUTAT VELLA	85,4073	137,2581	7,8756	9,7796	124,73%	118,58%
EIXAMPLE	1001,9673	1406,8686	23,7728	27,9355	71,18%	123,43%
GRÀCIA	184,9680	254,8740	9,8840	11,9327	67,97%	106,61%
HORTA-GUINARDÓ	268,2529	466,8363	12,9124	16,3456	67,88%	91,83%
LES CORTS	201,7701	263,1441	9,5900	10,8218	92,08%	82,97%
NOU BARRIS	526,1812	891,9871	17,8455	23,3002	75,62%	94,58%
SANT ANDREU	377,4403	500,4232	14,4231	16,6157	67,37%	81,98%
SANT MARTÍ	603,1306	848,2204	18,2062	21,3312	77,50%	63,17%
SANTS-MONTJUÏC	506,1836	674,0423	16,7662	19,9486	97,70%	103,99%
SARRIÀ-SANT GERVAZI	376,5833	632,4495	14,8299	19,3945	67,26%	67,23%

Realmente, nos centraremos el EPAM, el cual es superior a 60% en todos los distritos, lo cual nos indica una capacidad predictiva muy baja. (*Véase tabla Anexos.6*).

A continuación, para aumentar el análisis de la capacidad predictiva de los modelos calcularemos la cobertura de las predicciones, porcentaje de observaciones dentro de las regiones de predicción, y el rango de dichas regiones de predicción.

Los resultados obtenidos para la cobertura son:

Tabla 3.6. Porcentaje de Cobertura para las predicciones del modelo AR

	Cobertura	
	Hombres	Mujeres
CIUTAT VELLA	100,00%	91,94%
EIXAMPLE	72,58%	72,58%
GRÀCIA	77,42%	74,19%
HORTA-GUINARDÓ	90,32%	88,71%
LES CORTS	74,19%	72,58%
NOU BARRIS	79,03%	77,42%
SANT ANDREU	72,58%	72,58%
SANT MARTÍ	77,42%	72,58%

SANTS-MONTJUÏC	79,03%	75,81%
SARRIÀ-SANT GERVASI	66,13%	66,13%

Observamos una cobertura muy buena, la mayoría de los modelos han sido capaces de ajustarse a la volatilidad de los datos, e incluir más del 65% de las observaciones reales en sus predicciones. Esto coincide con lo observado en los gráficos anteriores (*Ilustraciones 10 y 11*), donde observábamos como en todos los distritos las regiones de predicción eran capaces de incluir más de la mitad de los valores de las observaciones.

Los resultados obtenidos con relación al rango de las regiones de predicción son:

Tabla 3.7. Rangos de las regiones de predicción modelos AR

	Rangos de predicción	
	Hombres	Mujeres
CIUTAT VELLA	34,33	31,11
EIXAMPLE	60,56	71,05
GRÀCIA	24,37	31,96
HORTA-GUINARDÓ	42,23	54,26
LES CORTS	20,20	22,13
NOU BARRIS	48,14	58,57
SANT ANDREU	36,78	42,75
SANT MARTÍ	51,33	60,42
SANTS-MONTJUÏC	46,74	56,63
SARRIÀ-SANT GERVASI	30,79	36,74

Observamos todo tipo de rangos. Hay distritos como Sarrià-Sant Gervasi, Gràcia, Les Corts o Ciutat Vella donde los rangos no superan los 35 casos, lo que, añadido a sus grandes porcentajes de cobertura, nos están indicando que sus modelos son capaces de explicar las series con una capacidad y precisión aceptables. Mientras que otros distritos como Sant Martí, Nou Barris o Eixample tenemos rangos superiores a 50 casos, lo que nos indica que sus buenos porcentajes de cobertura son un poco “falsos” ya que tienen una región de predicción muy grande, lo que facilita que las observaciones entren dentro del modelo.

III.3.3.2. Predicción Ex - Ante

Para la predicción Ex - Ante cogemos la totalidad de nuestra serie, desde el 1 de marzo de 2020 hasta el 3 de marzo de 2021, e intentamos predecir las observaciones que se obtendrían en los próximos 20 días.

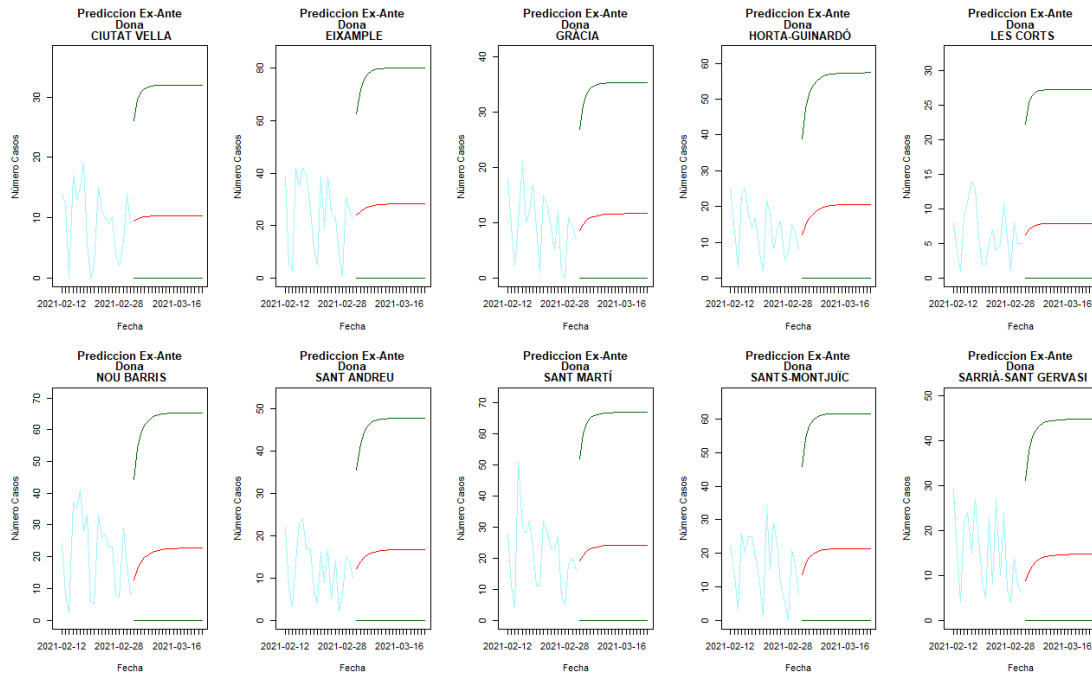


Ilustración 12. Predicción Ex-Ante del modelo AR para mujeres

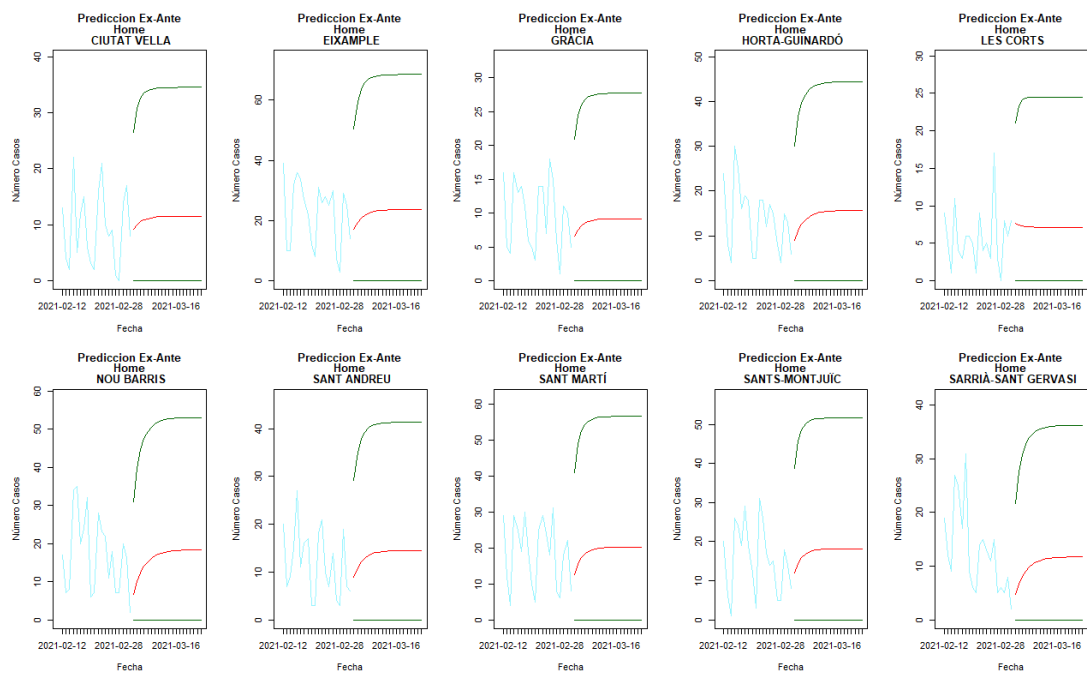


Ilustración 13. Predicción Ex-Ante del modelo AR para los hombres

Volvemos a observar la convergencia de las predicciones hacia la media, nada extraño ni importante a comentar. Una cosa para destacar es que vemos como el límite superior es bastante más alto que las últimas observaciones, por ello podemos creer que las regiones de predicción serían capaces de incluir las observaciones futuras. (Véanse Ilustraciones 12 y 13)

En conclusión, parece ser que nuestro modelo tiene una baja capacidad predictiva, tanto, para las predicciones Ex - Post como para las predicciones Ex - Ante, pero aun así presenta

coberturas “muy buenas”, las cuales vienen influenciadas por las grandes regiones de predicción que presentan los modelos.

III.4. Modelos INAR

Durante todo este apartado, al igual que en los anteriores, hay que recordar que trabajamos con el conteo de casos de la COVID-19 en Barcelona, con los datos desagregados por sexo y distrito.

III.4.1 Estimación de los parámetros

Los modelos INAR se explican mediante la fórmula (2). Como hemos comentado anteriormente (II.2. Identificación del modelo), hemos detectado que nuestros modelos son de orden 1 y presentan una estacionalidad de orden 7 en la serie, es por ello que la fórmula inicial sufre una pequeña modificación y un modelo INAR (1) con estacionalidad queda de la siguiente manera: [9]

$$X_n = \alpha_1 \circ X_1 + \varepsilon_t (\lambda_t)$$

Donde el parámetro α_1 se mantiene igual, pero ahora las innovaciones ε_t se asume que siguen varias distribuciones Poisson con diferentes medias, de forma que $\lambda_t = \lambda_{t+\eta}$, donde η es igual al orden de la estacionalidad, en nuestro caso 7. Esto significa que tendremos que estimar 8 parámetros; $\alpha_1, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6, \lambda_7$

Para estimar los parámetros del modelo INAR explicados en el párrafo anterior usaremos el mismo método que con los modelos AR, el método de la máxima verosimilitud. Para ello usaremos la fórmula de la máxima verosimilitud expresada a continuación: [10]

$$P(X_t | X_{t-1}, \dots, X_{t-p}) = \sum_{i_1=0}^{\min(X_{t-1}, X_t)} \binom{X_{t-1}}{i_1} \alpha_1^{i_1} (1 - \alpha_1)^{X_{t-1}-i_1} \sum_{i_2=0}^{\min(X_{t-2}, X_{t-1}-i_1)} \binom{X_{t-2}}{i_2} \alpha_2^{i_2} (1 - \alpha_2)^{X_{t-2}-i_2} \dots \sum_{i_p=0}^{\min(X_{t-p}, X_t - (i_1 + i_2 + \dots + i_p))} \binom{X_{t-p}}{i_p} \alpha_p^{i_p} (1 - \alpha_p)^{X_{t-p}-i_p} \frac{e^{-\lambda} \lambda^{X_t - (i_1 + i_2 + \dots + i_p)}}{[X_t - (i_1 + i_2 + \dots + i_p)]!} \quad (10)$$

En nuestro caso, al tener un INAR (1) y una estacionalidad de orden 7, solamente tendríamos el primer sumatorio, y luego la λ iría variando en función del día de la semana. Es decir, en total tendríamos 7 fórmulas, una por cada día de la semana.

Hay que comentar que hemos decidido no aplicar la fórmula de la máxima verosimilitud sobre el número de casos, si no aplicarla sobre la incidencia de casos cada 10.000 habitantes, y de esta manera evitar la influencia que puede tener la población de cada distrito. El hecho de trabajar con la incidencia provoca, que en la función de máxima verosimilitud tenemos que multiplicar las λ por $\frac{\text{población distrito}}{10.000}$. [10] (Véase fórmula 10)

Las estimaciones de los parámetros del modelo descrito para cada distrito pueden verse en las tablas 8 y 9:

Tabla 3.8. Parámetros Modelo INAR mujeres

	α_1	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7
CIUTAT VELLA	0,45902	1,20772	0,59634	0,55200	0,58917	0,56827	0,04712	0,08811
EIXAMPLE	0,46705	1,24089	0,68980	0,62186	0,64822	0,51765	0,10915	0,10845
GRÀCIA	0,45129	1,09104	0,62350	0,60147	0,62213	0,55003	0,06812	0,09154
HORTA-GUINARDÓ	0,47732	1,17340	0,78293	0,64675	0,80589	0,62010	0,16061	0,17271
LES CORTS	0,42415	1,28668	0,48313	0,71950	0,58550	0,61007	0,04917	0,15087
NOU BARRIS	0,48779	1,44354	0,85524	0,64995	0,90831	0,63667	0,06105	0,16647
SANT ANDREU	0,47567	1,24632	0,67959	0,61958	0,67038	0,63854	0,09754	0,12139
SANT MARTÍ	0,49095	1,20100	0,51942	0,61265	0,55062	0,55581	0,05000	0,10224
SANTS-MONTJUÏC	0,41491	1,43953	0,68734	0,78363	0,78061	0,77795	0,06411	0,15943
SARRIÀ-SANT GERVASI	0,45494	0,56476	0,67711	0,04198	0,90762	0,65070	0,30783	0,60088

Tabla 3.9. Parámetros Modelo INAR para hombres

	α_1	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7
CIUTAT VELLA	0,4726	1,2902	0,5866	0,6461	0,5563	0,6441	0,0892	0,1502
EIXAMPLE	0,4721	1,0080	0,5547	0,5349	0,4603	0,5262	0,0449	0,1265
GRÀCIA	0,4661	0,8639	0,4771	0,4414	0,3879	0,4435	0,0752	0,1042
HORTA-GUINARDÓ	0,5132	0,9762	0,5436	0,4416	0,4778	0,5149	0,0321	0,1002
LES CORTS	0,4545	1,1582	0,4361	0,6062	0,4558	0,4590	0,0733	0,1038
NOU BARRIS	0,5518	1,1931	0,5046	0,4917	0,5341	0,4605	0,0388	0,1281
SANT ANDREU	0,4947	1,0926	0,5232	0,5036	0,4778	0,5706	0,0731	0,1545
SANT MARTÍ	0,4938	1,0628	0,4281	0,4293	0,4498	0,4245	0,0715	0,1132
SANTS-MONTJUÏC	0,4765	1,2527	0,5442	0,4660	0,6092	0,5553	0,0405	0,0977
SARRIÀ-SANT GERVASI	0,5000	0,4339	0,4233	0,0695	0,7293	0,4159	0,2369	0,4509

Observando las tablas superiores, podemos ver como los parámetros α son un poco más elevados para los hombres que para las mujeres en todos los distritos. Una interpretación podría ser que los hombres arrastran más casos una vez contagiados, pero no es nuestro caso ya que nuestra serie solo tiene en cuenta los nuevos casos diarios detectados, y no los acumulados. Simplemente podemos afirmar que parece que existe diferentes comportamientos en función del sexo.

A nivel de las innovaciones, tenemos que fijarnos en las diferentes λ . Hay que recordar que, debido a la estacionalidad, tenemos una λ para cada día. ($\lambda_1 = \text{Lunes}$, $\lambda_2 = \text{Martes}$, $\lambda_3 = \text{Miércoles}$, $\lambda_4 = \text{Jueves}$, $\lambda_5 = \text{Viernes}$, $\lambda_6 = \text{Sábado}$, $\lambda_7 = \text{Domingo}$)

Podemos ver como a medida que pasa la semana las innovaciones van descendiendo. A nivel social estos resultados tienen mucho sentido. Durante el fin de semana es cuando hay más actividad social y más probabilidad de propagar el virus, es por ello por lo que los lunes el número de casos suele ser más alto, y a medida que la actividad social desciende durante la semana, debido a la actividad laboral, el número de casos detectados también desciende. Así hasta llegar al fin de semana e iniciar el ciclo de nuevo. Otra cosa para comentar es el problema que tienen la notificación. No todos los métodos de detección son instantáneos,

algunos tardan un par horas o días, y eso hace que un caso salga detectado, por ejemplo, un jueves cuando en realidad el contagio se produjo el martes.

A nivel de sexos, observamos como las mujeres presentan innovaciones mucho más abundantes que los hombres. En la mayoría de los días y distritos, se detectan más nuevos casos diarios de mujeres que de hombres. Esto coincide con lo observado en la descriptiva de la serie (*Véanse tablas Anexos.3 y Anexo.4*), donde las mujeres eran el sexo con más casos detectados.

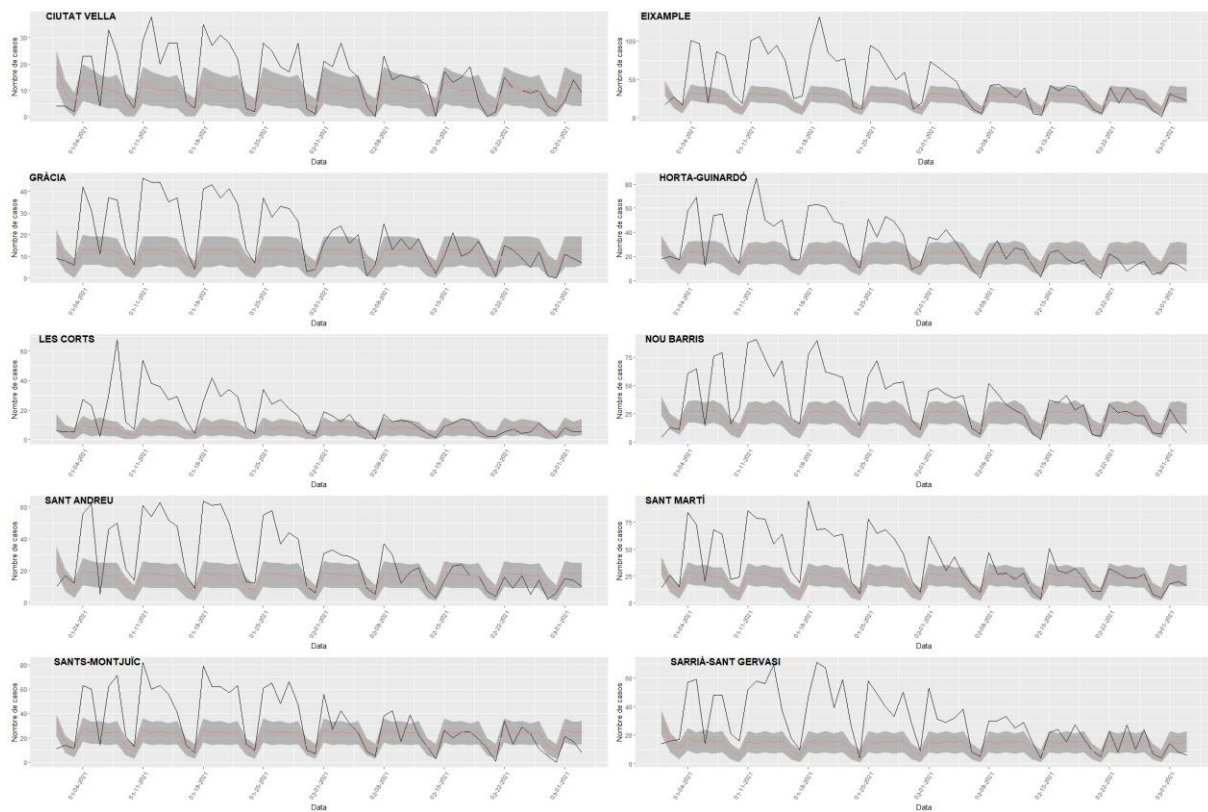
En conclusión, dichos resultados nos refuerzan la hipótesis que existen diferencias a nivel de contagio entre los sexos.

III.4.2 Predicciones

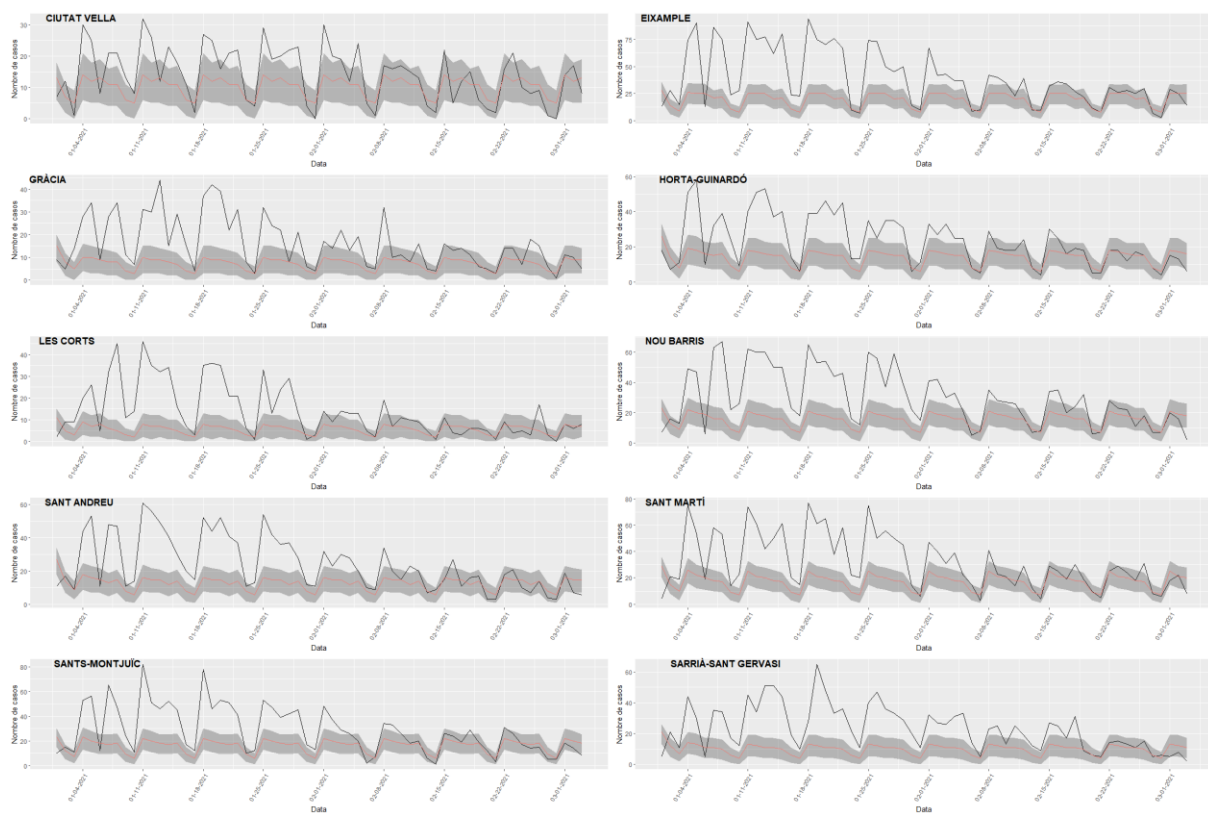
Para las predicciones del modelo INAR seguiremos la misma estructura que en los modelos AR. En este caso las predicciones estarán basadas en la fórmula (7).

III.4.2.1. Predicciones Ex - Post

Para las predicciones Ex - Post usaremos las observaciones desde el 1 de marzo de 2020 hasta el 31 de diciembre de 2020, y predeciremos los valores para los meses de enero y febrero de 2021, y los compararemos con los valores reales observados. Como el caso de los modelos AR, volveremos a parametrizar nuevos modelos, ya que la muestra ha cambiado. (*Ver Tabla Anexos.7 y Anexos.8. Parámetros modelo INAR predicciones Ex-Post.*)



Il·lustració 14. Prediccions Ex-Post del modelo INAR para las mujeres



Il·lustració 15. Predicciones Ex -Post del modelo INAR para los hombres

Observamos como el modelo es capaz de adaptarse perfectamente a las tendencias que sigue la serie. Una cosa muy destacable es que podemos observar como al principio de la serie las predicciones son inferiores, pero a medida que pasa el tiempo y los casos detectados descienden las predicciones se ajustan a la perfección. Este evento se podría explicar debido a que el inicio de las predicciones es enero y es donde nos encontramos con la tercera ola. El aumento exponencial durante enero hace que el modelo no sea capaz de explicar a la perfección este período muestral, ya que las observaciones de diciembre no indicaban que se pudiera dar estos valores extremos. Por otro lado, a medida que el número de casos disminuye y toman un comportamiento más lógico, el modelo es capaz de explicar perfectamente la serie y predecirla.

Al igual que en los modelos AR, calcularemos los errores de predicción, la cobertura, y, el rango de las regiones de predicción.

Los errores de predicción obtenidos son: (*Véanse fórmulas 9*)

Tabla 3.10. Errores de predicción para el modelo INAR

	EQM		EAM		EPAM	
	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres
CIUTAT VELLA	61,10	110,19	6,33	7,85	66,69%	56,07%
EIXAMPLE	899,92	1181,92	21,31	23,92	46,32%	64,54%
GRÀCIA	174,66	209,56	9,27	10,28	48,04%	64,90%
HORTA-GUINARDÓ	228,16	372,94	10,87	14,06	41,14%	63,32%
LES CORTS	194,51	237,61	9,33	9,97	65,08%	62,54%
NOU BARRIS	484,15	680,92	16,66	18,95	62,76%	63,30%
SANT ANDREU	340,27	426,98	13,47	14,69	53,88%	59,18%
SANT MARTÍ	528,63	705,32	16,66	18,74	54,33%	42,39%
SANTS-MONTJUÏC	414,76	496,48	14,56	15,89	58,12%	61,18%
SARRIÀ-SANT GERVASI	353,95	544,18	14,69	17,44	66,22%	52,75%

Observamos que los EPAM son superiores al 10%, por tanto, nos indican bajas capacidades predictivas de los modelos INAR estimados. Algo destacable es que los EPAM son inferiores para los hombres que para las mujeres. Esto quiere decir que dentro de los modelos INAR, aquellos para los hombres son los más capaces para explicar la serie. (*Véase Tabla Anexos.6*)

A continuación, para aumentar el análisis de la capacidad predictiva de los modelos calcularemos la cobertura de las predicciones, porcentaje de observaciones dentro de las regiones de predicción, y el rango de dichas regiones de predicción.

Los resultados obtenidos para la cobertura son:

Tabla 3.11. Porcentaje de Cobertura para las predicciones del modelo INAR

	Cobertura	
	Hombres	Mujeres
CIUTAT VELLA	61,29%	56,45%
EIXAMPLE	40,32%	43,55%
GRÀCIA	51,61%	56,45%
HORTA-GUINARDÓ	51,61%	43,55%
LES CORTS	51,61%	54,84%
NOU BARRIS	32,26%	41,94%
SANT ANDREU	46,77%	45,16%
SANT MARTÍ	43,55%	48,39%
SANTS-MONTJUÏC	38,71%	48,39%
SARRIÀ-SANT GERVASI	27,42%	32,26%

Observamos coberturas aceptables, superiores al 50%, en los distritos de Ciutat Vella, Gràcia, Horta-Guinardó y Les Corts. Pero el problema lo tenemos en los otros distritos, con porcentajes inferiores al 50% alcanzando cotas mínimas del 27% en el distrito de Sarrià-Sant Gervàsi. Dichas coberturas no son especialmente altas, y esto concuerda con lo obtenido en los EPAM, donde se observaba una baja capacidad predictiva. Además, observando las *Ilustraciones 14 y 15*, sobre todo al principio de la serie, se ve con claridad como las regiones de predicción no incluyen las observaciones, y esto provoca bajas coberturas.

A continuación, para poder explicar un poco mejor estas coberturas calcularemos el rango de las regiones de predicción. Los resultados obtenidos son:

Tabla 3.12. Rangos de las regiones de predicción modelos INAR

	Rangos de predicción	
	Hombres	Mujeres
CIUTAT VELLA	12,26	11,50
EIXAMPLE	16,34	18,55
GRÀCIA	10,08	12,26
HORTA-GUINARDÓ	14,11	16,69
LES CORTS	8,77	9,66
NOU BARRIS	14,44	17,56
SANT ANDREU	13,29	14,63
SANT MARTÍ	15,32	17,56
SANTS-MONTJUÏC	14,71	17,24
SARRIÀ-SANT GERVASI	12,19	13,47

Observamos rangos muy pequeños. El rango más amplio lo encontramos en 18.55 casos para las mujeres en l'Eixample. Estos rangos nos están indicando que nuestros modelos INAR intentan ser muy precisos en sus predicciones, aunque parece que si observamos la cobertura dichos modelos no son capaces de adaptarse bien a los valores de la serie.

III.4.2.2. Predicciones Ex - Ante

Para la predicción Ex - Ante cogemos la totalidad de nuestra serie, desde el 1 de marzo de 2020 hasta el 3 de marzo de 2021, e intentamos predecir las observaciones que se obtendrían en los próximos 20 días.

Hay que tener en cuenta que tenemos estacionalidad de orden 7, y, por tanto, las lambdas se comportaran siguiendo la relación: $\lambda_n = \lambda_{n+7}$

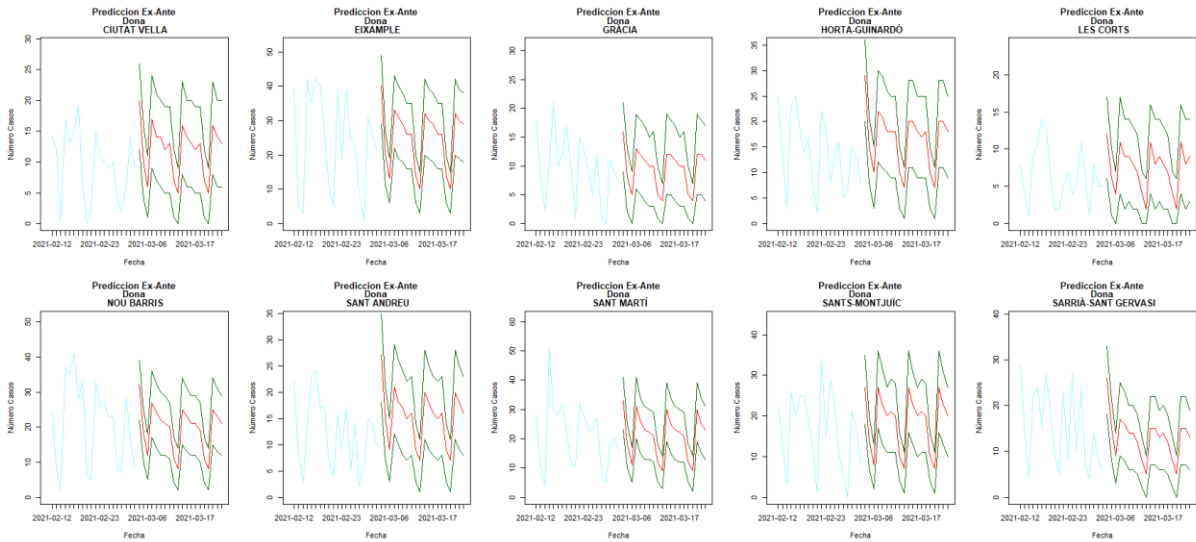


Ilustración 16. Predicciones Ante del modelo INAR para las mujeres

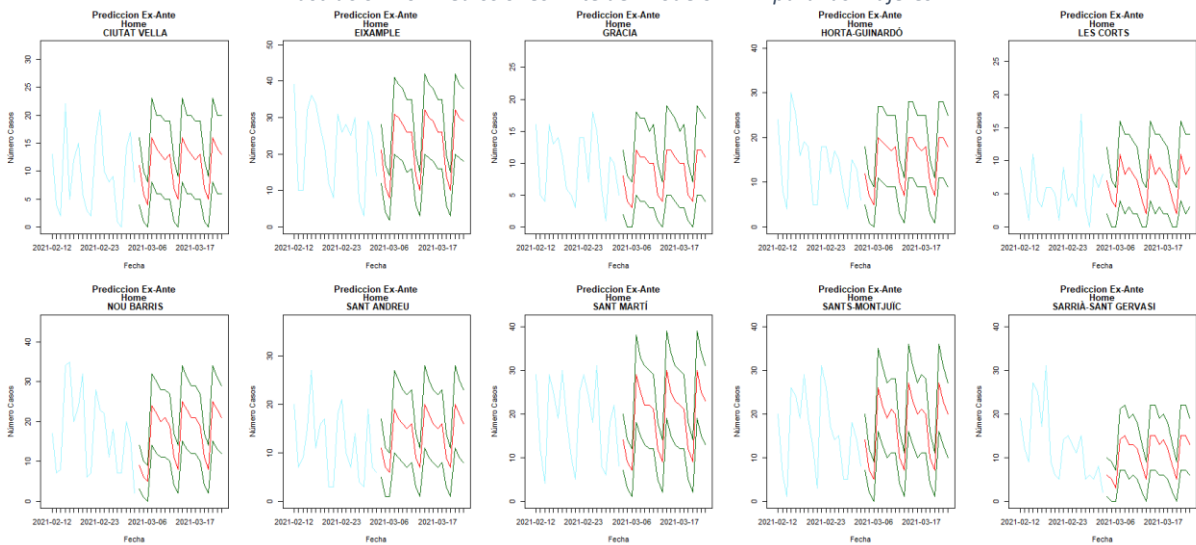


Ilustración 17. Predicciones Ante del modelo INAR para los hombres

Se puede observar cómo claramente los modelos han sido capaces de adaptarse a la tendencia de las últimas observaciones y ofrecer unas predicciones plausibles, que nos pueden permitir llevar estrategias eficientes. También se puede destacar como las mujeres obtienen predicciones

superiores a los hombres, lo cual tiene sentido ya que como vimos (*Véanse tablas Anexos 3 y 4*), las mujeres son aquellas con más casos detectados.

En conclusión, parece ser que nuestro modelo tiene una baja capacidad predictiva, tanto, para las predicciones Ex - Post como para las predicciones Ex - Ante, y presenta unas coberturas bajas en algunos distritos, lo cual se debe a que los modelos INAR son muy precisos, con regiones de predicción bastante estrechas, y la tercera ola con valores extremos que encontramos en enero de 2021.

III.5. Comparación de los modelos

Una vez hemos estimado los dos tipos de modelos: AR e INAR, consideramos oportuno comparar ambos modelos, tanto a nivel de parámetros (*Véanse tablas 3.2, 3.8 y 3.9*), como a nivel de predicciones (*Véanse tablas 3.5, 3.6, 3.7, 3.10, 3.11, 3.12*) para discernir cuál de los dos explica mejor la serie temporal de número de casos diarios en los diferentes distritos de Barcelona.

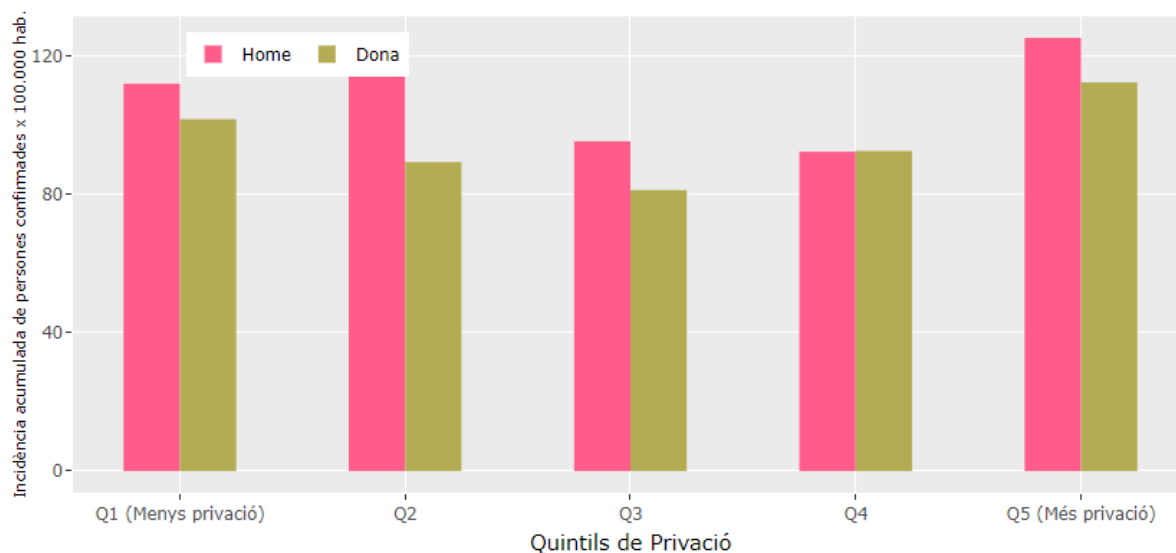
A nivel de parámetros vamos a dividirlos en dos categorías: resultados en función del sexo y en función del distrito.

- **Sexo:**

A nivel del parámetro α observamos mayores valores para los hombres, lo que nos empieza a indicar la existencia de diferentes mecanismos para los sexos. Luego a nivel de innovaciones, casos que no depende de los días anteriores, es en las mujeres en las que se detectan más casos, es decir que las mujeres tienen más propensión a contagiarse. Estos resultados concuerdan con los obtenidos en la descriptiva realizada donde observábamos que más de la mitad de los casos detectados eran de mujeres. (*Véanse tablas Anexos.3 y Anexos.4*)

- **Distritos:**

En cuanto a la diferenciación de los parámetros en los diferentes distritos, en ambos modelos obtenemos que el número de casos, tanto los dependientes de días anteriores como las innovaciones aumentan cuanto mayor es la densidad poblacional del distrito. También hemos observado como la renta afecta a los casos a nivel de que, en distritos con renta baja, las agrupaciones familiares suelen ser más extensas que en los distritos de renta más alta, y por ello, cuando un miembro familiar se contagia, rápidamente se contagian más personas que en otros distritos donde las agrupaciones pueden ser inferiores. Renta y densidad poblacional están bastante relacionadas. Distritos con renta alta tienen menor densidad poblacional, ya que poca gente puede permitirse el coste de vida, y, por ende, en los distritos de renta más baja se producen mayores agrupaciones de viviendas, lo que aumenta la densidad poblacional, y facilita la propagación del virus.



Il·lustració 18. Número de casos de la COVID-19 en relació a los quintiles privació socioeconómica[11]

Con la imagen superior (*Ilustración 18*) no observamos una tendencia ascendente, desde el primer quintil hasta el último, pero si podemos ver como en el último quintil, aquel con mayor privación socioeconómica, y, por tanto, menor renta, alcanzamos el máximo de casos. En conclusión, observamos como la renta influye en el número de casos detectados, a menor renta, mayor número de casos de la COVID-19.

A nivel de predicciones, obtenemos resultados malos en ambos modelos. En relación con los errores de predicción, en ambos modelos obtenemos EPAM elevados, inferiores en los modelos INAR, indicando bajas capacidades predictivas. En adición, los modelos AR presentan porcentajes de cobertura muy altos, superiores al 60%, a diferencia de los modelos INAR, donde solo 4 distritos son capaces de superar el 50% y ni siquiera llegan al 60%, pero hay que comentar que dichos porcentajes de los modelos AR son un poco “irreales”, ya que los rangos de las regiones de predicción de los modelos AR son mucho más grandes que los modelos INAR. En los modelos AR el rango varía desde 20 casos diarios en un distrito hasta más de 50 casos en varios distritos, mientras que en los modelos INAR el máximo rango que obtenemos es de 19 casos en un solo distrito, mientras que los demás distritos se mueven en valores en torno a los 10-15 casos de rango.

En conclusión, hemos podido ver como ambos modelos se comportan de la misma manera a la hora de estimar de los parámetros, diferenciando correctamente entre sexos y distritos, en la parte regresiva los hombres obtienen valores superiores, mientras que en la parte de las innovaciones es al revés, mientras que en los distritos se observa claramente como los barrios con mayor densidad poblacional obtienen parámetros con valores más altos. A nivel de predicciones obtenemos que ambos modelos tienen una capacidad predictiva muy baja, siendo los modelos INAR los que obtienen un mejor EPAM y más precisos, mientras que los AR son los menos eficaces y más imprecisos.

IV. CONCLUSIONES

La COVID-19 ha tenido un impacto en todos los niveles de la sociedad, (*Véanse Ilustraciones 1, 2, 3, 4, y Anexos.1*) es por ello por lo que hemos considerado necesario parametrizar modelos que puedan ajustarse a la serie y así disponer de instrumentos estadísticos que nos permitan predecir el comportamiento del virus para llevar a cabo estrategias eficientes, y mitigar su impacto.

Nuestra hipótesis inicial, es que al ser la COVID-19 una enfermedad infecciosa, donde los casos diarios dependen de los casos detectados los días anteriores, los modelos más adecuados para parametrizar serían los modelos autorregresivos. Para confirmar esta hipótesis hemos observado los coeficientes de autocorrelación (*Véanse Ilustraciones 6, 7, 8, y 9*). Al analizar los coeficientes también hemos descubierto la existencia de una estacionalidad de orden 7, es decir, semanal. Aunque no fuera una hipótesis inicial, el hecho de que el contagio de la COVID-19 se comporte diferente en función del día de la semana tiene todo el sentido del mundo. Los fines de semana, al haber más vida social, la probabilidad de contagio aumenta con respecto a otro día de la semana. En conclusión, utilizaremos los modelos tipo autorregresivos, más concreto el modelo AR, para distribuciones continuas, y el modelo INAR, para distribuciones discretas. Dichos modelos serán de orden 1, y presentarán una estacionalidad de 7 días.

Una vez tenemos el tipo de modelos a estimar, hemos de comparar los resultados. Tal y como comentamos al inicio del trabajo, nuestra hipótesis es que tanto el sexo, como el distrito a analizar influyen en el número de casos. En relación con el sexo, dicha hipótesis se defiende al ver como las mujeres son aquellas con más de la mitad de los casos (*Véanse Tablas Anexos.3 y 4*), y también observando las diferencias en los parámetros del modelo, tanto a nivel regresivo, donde los hombres presentan valores superiores, como a nivel de las innovaciones, donde vemos como las mujeres llevan la delantera (*Véanse Tablas 3.2, 3.8, 3.9, Anexos.5, 7, y 8*). A nivel de distritos también se observa como la densidad poblacional del distrito es un elemento clave para analizar los contagios. Aquellos distritos con más densidad poblacional son los distritos con mayores casos detectados y parámetros del modelo. También hemos concluido que la renta influye de una forma indirecta, distritos con alta renta son poco poblados, y, por tanto, se dificulta la propagación del virus. (*Véase Tabla 3.1*)

Los modelos AR obtienen resultados optimistas, pero un poco irreales. Obtienen unos EPAM alrededor del 70-100%, lo que nos indica una bajísima capacidad predictiva (*Véase Tabla 3.5*), pero, por otro lado, observando la representación gráfica se observa como los valores predichos son incluidos en las regiones de predicción (*Véanse Ilustraciones 10 y 11*). Esto supone una cobertura mínima del 65% en todos los distritos (*Véanse tabla 3.6*), la cual está influenciada por las enormes regiones de predicción del modelo AR, donde el rango varía desde los 20 casos diarios hasta los 50. (*Véase tabla 3.7*)

En cambio, los modelos INAR presentan mejores EPAM (*Véase Tabla 3.10*), y por las representaciones gráficas se observa como también se adaptan mucho mejor al comportamiento de la serie (*Véanse Ilustraciones 14 y 15*), mientras que los modelos AR acaban convergiendo a la media, el modelo INAR es capaz de dibujar las tendencias que sigue la serie. A nivel de cobertura es inferior para los modelos INAR (*Véase Tabla 3.11*), pero también lo es la región de predicción (*Véase Tabla 3.12*), por lo que el modelo INAR tiende a ser más preciso que los modelos AR.

Tal y como se ha comentado a lo largo del trabajo, esperábamos que la serie no fuera estacionaria y así ha resultado, por tanto, destacamos que ninguno de los dos modelos ha cumplido todas las validaciones y, por ende, ninguna estimación será perfecta. Aun así, recomendamos enormemente el modelo INAR sobre el AR, no solo por los mejores resultados, y el mejor comportamiento observado, sino porque al ser nuestra serie un conteo de valores bajos, un modelo que siga una distribución discreta (INAR), tiene mejor expectativa para explicar la serie que un modelo con distribución marginal continua (AR).

V. BIBLIOGRAFIA

- [1] “Coronavirus: La OMS declara el brote de Covid-19 pandemia.” <https://www.redaccionmedica.com/secciones/sanidad-hoy/coronavirus-pandemia-brote-de-covid-19-nivel-mundial-segun-oms-1895> (accessed Mar. 06, 2021).
- [2] M. Woodward, S. A. E. Peters, and K. Harris, “Social deprivation as a risk factor for COVID-19 mortality among women and men in the UK Biobank: nature of risk and context suggests that social interventions are essential to mitigate the effects of future pandemics,” *J. Epidemiol. Community Health*, p. jech-2020-215810, Apr. 2021, doi: 10.1136/jech-2020-215810.
- [3] R Core Team, “R: A Language and Environment for Statistical Computing,” 2019, [Online]. Available: <https://www.r-project.org/>.
- [4] H. Wickham., “ggplot2: Elegant Graphics for Data Analysis.,” *Springer-Verlag New York*, 2016.
- [5] D. Jin-Guan and L. Yuan, “THE INTEGER-VALUED AUTOREGRESSIVE (INAR(p)) MODEL,” *J. Time Ser. Anal.*, vol. 12, no. 2, pp. 129–142, 1991, doi: 10.1111/j.1467-9892.1991.tb00073.x.
- [6] D. Moriña, P. Puig, J. Ríos, A. Vilella, and A. Trilla, “A statistical model for hospital admissions caused by seasonal diseases,” *Stat. Med.*, vol. 30, no. 26, pp. 3125–3136, Nov. 2011, doi: 10.1002/sim.4336.
- [7] F. De, C. Matemáticas, and J. A. T. Cazorla, *UNIVERSIDAD COMPLUTENSE DE MADRID MODELO “INARMA” PARA SERIES TEMPORALES DE VALORES ENTEROS: ANÁLISIS, PROPIEDADES ASINTOMÁTICAS Y ESTIMACIÓN. MEMORIA PARA OPTAR AL GRADO DE DOCTOR PRESENTADA POR José Nerys Funes Torres Bajo la dirección del doctor.* 2010.
- [8] C. Felipe González López, A. : María, and E. Correal Núñez, “Modelos de Series de Tiempo para Datos Enteros.”
- [9] D. Moriña, P. Puig, J. Ríos, A. Vilella, and A. Trilla, “A statistical model for hospital admissions caused by seasonal diseases,” *Stat. Med.*, vol. 30, no. 26, pp. 3125–3136, Nov. 2011, doi: 10.1002/sim.4336.
- [10] R. Bu and B. McCabe, “Model selection, estimation and forecasting in INAR(p) models: A likelihood-based Markov Chain approach,” *Int. J. Forecast.*, vol. 24, no. 1, pp. 151–162, Jan. 2008, doi: 10.1016/j.ijforecast.2007.11.002.
- [11] “#COVID19aldiaBCN.” https://aspb.shinyapps.io/COVID19_BCN/#Nivell_socioeconòmic_de_l'àrea_de_residència (accessed Jun. 17, 2021).

VI. ANNEXOS

Tabla Anexos.1 Códigos Identificativos distritos de Barcelona

Códigos Identificativos para los Sector Sanitarios dentro de Barcelona				
7846: Ciutat Vella	7847: Sant Martí	7848: Les Corts	7849: Sants-Montjuic	7850: Sarrià-Sant Gervasi
7851: Eixample	7852: Gracia	7853: Horta-Guinardó	7854: Nou Barris	7855: Sant Andreu

Tabla Anexos.2 Relación gráficos ACF/PACF vs Tipo de Modelo

	ACF	PACF
AR	∞	P
MA	Q	∞
ARMA	∞	∞

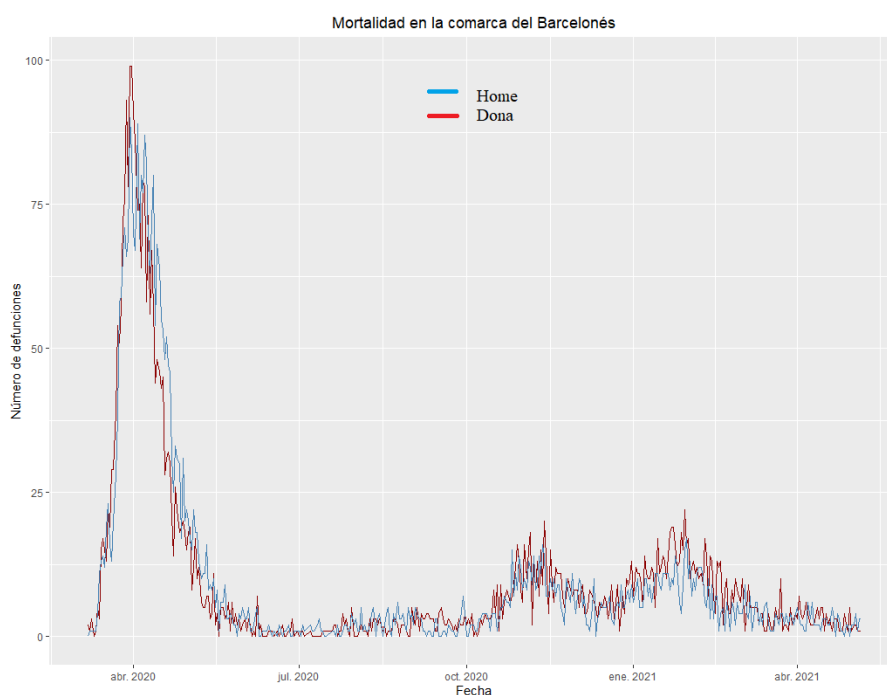


Ilustración Anexos.1 Mortalidad por la COVID-19 en la comarca del Barcelonés

Tabla Anexos.3 Estadística descriptiva para los hombres

	Media	Desviación estándar	Máximo Casos Diarios	Número de Casos Totales	Población	Casos por 10.000 habitantes	Densidad Poblacional
CIUTAT VELLA	11,62	11,53	67	4.277	108.331	394,81	23.908
EIXAMPLE	23,85	22,47	124	8.776	270.694	324,20	35.556
GRÀCIA	9,24	9,29	44	3.401	123.651	275,05	29.130
HORTA-GUINARDÓ	15,83	14,38	68	5.827	174.799	333,35	14.155
LES CORTS	7,12	8,72	46	2.620	82.635	317,06	13.697
NOU BARRIS	18,61	17,41	82	6.847	174.012	393,48	20.820
SANT ANDREU	14,59	13,46	61	5.368	151.976	353,21	22.459
SANT MARTÍ	20,33	18,20	112	7.480	241.691	309,49	22.129
SANTS-MONTJUÏC	18,26	16,79	82	6.718	187.584	358,13	8.024
SARRIÀ-SANT GERVAZI	11,89	12,31	65	4.376	151.157	289,50	7.235

Tabla Anexos.4 Estadística descriptiva para las mujeres

	Media	Desviación estándar	Máximo Casos Diarios	Número de Casos Totales	Población	Casos por 10.000 habitantes	Densidad Poblacional
CIUTAT VELLA	10,43	10,84	55	3.839	108.331	354,38	23.908
EIXAMPLE	28,55	25,96	132	10.506	270.694	388,11	35.556
GRÀCIA	11,74	11,88	60	4.321	123.651	349,45	29.130
HORTA-GUINARDÓ	20,83	18,41	90	7.666	174.799	438,56	14.155
LES CORTS	7,96	9,69	68	2.931	82.635	354,69	13.697
NOU BARRIS	22,92	21,38	114	8.433	174.012	484,62	20.820
SANT ANDREU	16,86	15,62	75	6.206	151.976	408,35	22.459
SANT MARTÍ	24,36	21,37	103	8.963	241.691	370,85	22.129
SANTS-MONTJUÏC	21,5	20,17	94	7.911	187.584	421,73	8.024
SARRIÀ-SANT GERVAZI	14,82	15,06	71	5.452	151.157	360,68	7.235

Tabla 5. Parámetros modelo AR para predicciones Ex-Post

	Mujeres				Hombres			
	ϕ_1	Desviación estándar	Intercept	Desviación estándar	ϕ_1	Desviación estándar	Intercept	Desviación estándar
CIUTAT VELLA	0,7026	0,0408	9,6387	1,4601	0,7109	0,0400	11,0621	1,6446
EIXAMPLE	0,7002	0,0408	25,2605	3,1217	0,7202	0,0395	20,6894	2,8502
GRÀCIA	0,6505	0,0434	10,3157	1,3516	0,6945	0,0414	7,9169	1,1010
HORTA-GUINARDÓ	0,7088	0,0403	19,2440	2,4319	0,7207	0,0397	14,4178	1,9685
LES CORTS	0,5582	0,0476	6,5050	0,8424	0,6689	0,0426	5,8254	0,9209
NOU BARRIS	0,7023	0,0407	20,0643	2,6376	0,7511	0,0376	16,2721	2,4289
SANT ANDREU	0,6889	0,0416	14,9675	1,8478	0,6767	0,0424	12,7583	1,5580
SANT MARTÍ	0,6670	0,0426	21,6457	2,4901	0,6643	0,0428	18,0540	2,1228
SANTS-MONTJUÏC	0,6217	0,0447	19,3837	2,2239	0,6367	0,0440	16,2164	1,8700
SARRIÀ-SANT GERVAZI	0,6832	0,0423	12,0440	1,6140	0,7358	0,0390	9,7164	1,5253

Tabla Anexos.6. Capacidad Predictiva EPAM

	Capacidad Predictiva
$EPAM \leq 1\%$	Muy buena capacidad predictiva
$1\% < EPAM \leq 3\%$	Buena capacidad predictiva
$3\% < EPAM \leq 5\%$	Mala capacidad predictiva
$5\% < EPAM$	Muy baja capacidad predictiva

Tabla Anexos.7. Parámetros modelo INAR predicciones Ex-Post Mujeres

	α_1	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7
CIUTAT VELLA	0,45902	1,20772	0,59634	0,55200	0,58917	0,56827	0,04712	0,08811
EIXAMPLE	0,46705	1,24089	0,68980	0,62186	0,64822	0,51765	0,10915	0,10845
GRÀCIA	0,45129	1,09104	0,62350	0,60147	0,62213	0,55003	0,06812	0,09154
HORTA-GUINARDÓ	0,47732	1,17340	0,78293	0,64675	0,80589	0,62010	0,16061	0,17271
LES CORTS	0,42415	1,28668	0,48313	0,71950	0,58550	0,61007	0,04917	0,15087
NOU BARRIS	0,48779	1,44354	0,85524	0,64995	0,90831	0,63667	0,06105	0,16647
SANT ANDREU	0,47567	1,24632	0,67959	0,61958	0,67038	0,63854	0,09754	0,12139
SANT MARTÍ	0,49095	1,20100	0,51942	0,61265	0,55062	0,55581	0,05000	0,10224
SANTS-MONTJUÏC	0,41491	1,43953	0,68734	0,78363	0,78061	0,77795	0,06411	0,15943
SARRIÀ-SANT GERVAZI	0,45494	0,56476	0,67711	0,04198	0,90762	0,65070	0,30783	0,60088

Tabla Anexos.8. Parámetros modelo INAR predicciones Ex-Post Hombres

	α_1	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7
CIUTAT VELLA	0,45902	1,20772	0,59634	0,55200	0,58917	0,56827	0,04712	0,08811
EIXAMPLE	0,46705	1,24089	0,68980	0,62186	0,64822	0,51765	0,10915	0,10845
GRÀCIA	0,45129	1,09104	0,62350	0,60147	0,62213	0,55003	0,06812	0,09154
HORTA-GUINARDÓ	0,47732	1,17340	0,78293	0,64675	0,80589	0,62010	0,16061	0,17271
LES CORTS	0,42415	1,28668	0,48313	0,71950	0,58550	0,61007	0,04917	0,15087
NOU BARRIS	0,48779	1,44354	0,85524	0,64995	0,90831	0,63667	0,06105	0,16647
SANT ANDREU	0,47567	1,24632	0,67959	0,61958	0,67038	0,63854	0,09754	0,12139
SANT MARTÍ	0,49095	1,20100	0,51942	0,61265	0,55062	0,55581	0,05000	0,10224
SANTS-MONTJUÏC	0,41491	1,43953	0,68734	0,78363	0,78061	0,77795	0,06411	0,15943
SARRIÀ-SANT GERVAZI	0,45494	0,56476	0,67711	0,04198	0,90762	0,65070	0,30783	0,60088

Anexo 9. Código R

```

library("tidyverse")
library("ggplot2")
library('tseries')
library('forecast')
library("stats")
library("lattice")
library("fBasics")
library("ggpubr")
library("lemon")
library("normtest")
library("nortest")
library("dplyr")

```

```

dades      <-      read.csv("C:/Users/pablo/OneDrive/Escritorio/TFG/Registre_de_casos_de_COVID-
19_realitzats_a_Catalunya._Segregaci__per_sexe_i__rea_b_sica_de_salut__ABS_.csv",header=T,
sep=",")

dadesBCN <- dades[dades$RegioSanitariaCodi==7803, ]

dadesBCN <- dadesBCN[dadesBCN$SectorSanitariCodi!=7844 &

dadesBCN$SectorSanitariCodi!=7867, ]

dadesBCN$TipusCasData <- as.Date(dadesBCN$TipusCasData, format="%d/%m/%Y")
dadesBCN <- dadesBCN[, c(1,5,8,10,11)]

dadesBCN2 <- dadesBCN %>% group_by(TipusCasData, SectorSanitariDescripcio, SexeCodi) %>%
summarise(Casos=sum(NumCasos))
dadesBCN2$SexeCodi <- factor(dadesBCN2$SexeCodi, levels=c(0, 1), labels=c("Home", "Dona"))
dadesBCN2$SectorSanitariDescripcio <- factor(dadesBCN2$SectorSanitariDescripcio)
dadesBCN2 <- dadesBCN2[order(dadesBCN2$TipusCasData, dadesBCN2$SectorSanitariDescripcio,
dadesBCN2$SexeCodi), ]
dadesBCN3 <- dadesBCN2 %>% ungroup() %>% complete(TipusCasData,
nesting(SectorSanitariDescripcio, SexeCodi))

dadesBCN3$Casos[is.na(dadesBCN3$Casos)] <- 0

dadesBCN4 <- dadesBCN3 %>% group_by(TipusCasData, SectorSanitariDescripcio, SexeCodi) %>%
summarise(TCasos=sum(Casos))

dadesBCN4$Poblacio <- NA
dadesBCN4$Poblacio[dadesBCN4$SectorSanitariDescripcio=="BARCELONA CIUTAT VELLA"] <- 108331
dadesBCN4$Poblacio[dadesBCN4$SectorSanitariDescripcio=="BARCELONA EIXAMPLE"] <- 270694
dadesBCN4$Poblacio[dadesBCN4$SectorSanitariDescripcio=="BARCELONA GRÀCIA"] <- 123651
dadesBCN4$Poblacio[dadesBCN4$SectorSanitariDescripcio=="BARCELONA HORTA-GUINARDÓ"] <- 174799
dadesBCN4$Poblacio[dadesBCN4$SectorSanitariDescripcio=="BARCELONA LES CORTS"] <- 82635
dadesBCN4$Poblacio[dadesBCN4$SectorSanitariDescripcio=="BARCELONA NOU BARRIS"] <- 174012
dadesBCN4$Poblacio[dadesBCN4$SectorSanitariDescripcio=="BARCELONA SANT ANDREU"] <- 151976
dadesBCN4$Poblacio[dadesBCN4$SectorSanitariDescripcio=="BARCELONA SANT MARTÍ"] <- 241691
dadesBCN4$Poblacio[dadesBCN4$SectorSanitariDescripcio=="BARCELONA SANTS-MONTJUÏC"] <- 187584
dadesBCN4$Poblacio[dadesBCN4$SectorSanitariDescripcio=="BARCELONA SARRIÀ-SANT GERVASI"] <-
151157

rm(dades, dadesBCN, dadesBCN2)

l <- split(dadesBCN3, dadesBCN3$SectorSanitariDescripcio)

```

```
l2 <- split(dadesBCN4, dadesBCN4$SectorSanitariDescripcio)
```

#Representación gráfica número de casos por barrio

```
gcv <- ggplot(CVH, aes(x=TipusCasData)) +
geom_line(aes(y=CVH$Casos), color="darkred")+geom_line(aes(y=CVD$Casos), color="steelblue") +
ggtitle(substring(unique(dadesBCN3$SectorSanitariDescripcio)[1], first=10))+labs(x = "Fecha", y =
"Número de casos") +theme(plot.title = element_text(hjust = 0.5), legend.position = "none")

geix <- ggplot(EIXH, aes(x=TipusCasData,)) +
geom_line(aes(y=EIXH$Casos), color="darkred")+geom_line(aes(y=EIXD$Casos), color="steelblue") +
ggtitle(substring(as.character(unique(dadesBCN3$SectorSanitariDescripcio)[2]), first=10))+labs(
x = "Fecha", y = "Número de casos")+theme(plot.title = element_text(hjust = 0.5), legend.position
= "none")

ggracia <- ggplot(GRACIAH, aes(x=TipusCasData)) +
geom_line(aes(y=GRACIAH$Casos), color="darkred")+geom_line(aes(y=GRACIAD$Casos), color="steelblu
e") +
ggtitle(substring(as.character(unique(dadesBCN3$SectorSanitariDescripcio)[3]), first=10))+labs(
x = "Fecha", y = "Número de casos")+theme(plot.title = element_text(hjust = 0.5), legend.position
= "none")

ghg <- ggplot(HGH, aes(x=TipusCasData)) +
geom_line(aes(y=HGH$Casos), color="darkred")+geom_line(aes(y=HGD$Casos), color="steelblue") +
ggtitle(substring(as.character(unique(dadesBCN3$SectorSanitariDescripcio)[4]), first=10))+labs(
x = "Fecha", y = "Número de casos")+theme(plot.title = element_text(hjust = 0.5), legend.position
= "none")

glc <- ggplot(LCH, aes(x=TipusCasData)) +
geom_line(aes(y=LCH$Casos), color="darkred")+geom_line(aes(y=LCD$Casos), color="steelblue") +
ggtitle(substring(as.character(unique(dadesBCN3$SectorSanitariDescripcio)[5]), first=10))+labs(
x = "Fecha", y = "Número de casos")+theme(plot.title = element_text(hjust = 0.5), legend.position
= "none")

gnb <- ggplot(NBH, aes(x=TipusCasData)) +
geom_line(aes(y=NBH$Casos), color="darkred")+geom_line(aes(y=NBD$Casos), color="steelblue") +
ggtitle(substring(as.character(unique(dadesBCN3$SectorSanitariDescripcio)[6]), first=10))+labs(
x = "Fecha", y = "Número de casos")+theme(plot.title = element_text(hjust = 0.5), legend.position
= "none")

gsa <- ggplot(SAH, aes(x=TipusCasData)) +
geom_line(aes(y=SAH$Casos), color="darkred")+geom_line(aes(y=SAD$Casos), color="steelblue") +
ggtitle(substring(as.character(unique(dadesBCN3$SectorSanitariDescripcio)[7]), first=10))+labs(
x = "Fecha", y = "Número de casos")+theme(plot.title = element_text(hjust = 0.5), legend.position
= "none")

gsmrt <- ggplot(SMRTH, aes(x=TipusCasData)) +
geom_line(aes(y=SMRTH$Casos), color="darkred")+geom_line(aes(y=SMRTD$Casos), color="steelblue")
+
ggtitle(substring(as.character(unique(dadesBCN3$SectorSanitariDescripcio)[8]), first=10))+labs(
x = "Fecha", y = "Número de casos")+theme(plot.title = element_text(hjust = 0.5), legend.position
= "none")

gsm <- ggplot(SMH, aes(x=TipusCasData)) +
geom_line(aes(y=SMH$Casos), color="darkred")+geom_line(aes(y=SMD$Casos), color="steelblue") +
ggtitle(substring(as.character(unique(dadesBCN3$SectorSanitariDescripcio)[9]), first=10))+labs(
x = "Fecha", y = "Número de casos")+theme(plot.title = element_text(hjust = 0.5), legend.position
= "none")

gssg <- ggplot(SSGH, aes(x=TipusCasData)) +
geom_line(aes(y=SSGH$Casos), color="darkred")+geom_line(aes(y=SSGD$Casos), color="steelblue") +
ggtitle(substring(as.character(unique(dadesBCN3$SectorSanitariDescripcio)[10]), first=10))+labs(
x = "Fecha", y = "Número de casos")+theme(plot.title = element_text(hjust =
0.5), legend.position = "none")

ggarrange(gcv, geix, ggracia, ghg, glc, gnb, gsa, gsmrt, gsm, gssg, nrow=2, ncol=5, common.legend =
TRUE, legend="bottom")
```

#Estacionariedad de las series

```
pv<-c()

for (s in levels(dadesBCN3$SexeCodi)){

df <- dadesBCN3[dadesBCN3$SexeCodi==s,]

for(i in 1:10){

dd <- df$Casos
dft <- adf.test(dd, alternative = "stationary")
pv <- c(pv,round(dft$p.value,30))

}

}

dftH <- data.frame("P-valor"=pv[1:10],row.names=unique(dadesBCN3$SectorSanitariDescripicio))

dftD <- data.frame("P-valor"=pv[11:20],row.names=unique(dadesBCN3$SectorSanitariDescripicio))

write.csv(dftH,"C:/Users/pablo/OneDrive/Escritorio/TFG/EstacionariedadHombres.csv", row.names
= TRUE)

write.csv(dftD,"C:/Users/pablo/OneDrive/Escritorio/TFG/EstacionariedadMujeres.csv", row.names
= TRUE)

#Descriptiva de los barrios

poblacio <- c(108331,270694,123651,174799,82635,174012,151976,241691,187584,151157)
dens <- c(23908,35556,29130,14155,13697,20820,22459,22129,8024,7235)
renta <- c(84.3,122.4,105.3,78.0,137.3,55.0,74.6,88.1,84.6,182.8)
ed <- c()

for (i in 1:10){

df
dadesBCN3[dadesBCN3$SectorSanitariDescripicio==unique(dadesBCN3$SectorSanitariDescripicio)[i],] <-

for (s in levels(df$SexeCodi)){
```

```

casos <- df[df$SexeCodi==s,]
casos <- casos$Casos
ed<-c(ed,summary(casos),sum(casos),sd(casos),(sum(casos)/poblacio[i])*10000)
}

}

edT <- c()
for (i in 1:10){

df <- dadesBCN3[dadesBCN3$SectorSanitariDescripcio==unique(dadesBCN3$SectorSanitariDescripcio)[i],]

casos <- df$Casos
edT<-c(edT,summary(casos),sum(casos),sd(casos),(sum(casos)/poblacio[i])*10000)

}

ed<-matrix(ed,nrow =10,byrow=TRUE)

edH <- data.frame("Mean"=ed[,4],"Desviación estándar"=ed[,8],"Max"=ed[,6],"Num
Casos"=ed[,7],"Població"=poblacio,"Casos por 10.000 habitantes"=ed[,9],"Densidad poblacional"=
dens,row.names = unique(dadesBCN3$SectorSanitariDescripcio))

edD <- data.frame("Mean"=ed[,13],"Desviación estándar"=ed[,17],"Max"=ed[,15],"Num
Casos"=ed[,16],"Població"=poblacio,"Casos por 10.000 habitantes"=ed[,18],"Densidad
poblacional"=dens ,row.names = unique(dadesBCN3$SectorSanitariDescripcio))

edT <- matrix(edT,nrow=10,byrow=TRUE)

edT <- data.frame("Media"=edT[,4],"Desviación estándar"=edT[,8],"Máximo"=edT[,6],"Número de
Casos Totales"=edT[,7],"Población"=poblacio,"Casos por 10.000 habitantes"=edT[,9],"Densidad
poblacional"=dens,"Renta familiar en miles de ???"= renta,row.names =
unique(dadesBCN3$SectorSanitariDescripcio))

write.csv(edT,"C:/Users/pablo/OneDrive/Escritorio/TFG/DescriptivaTotal.csv", row.names = TRUE)

write.csv(edH,"C:/Users/pablo/OneDrive/Escritorio/TFG/DescriptivaHombres.csv", row.names =
TRUE)

```



```
write.csv(edD,"C:/Users/pablo/OneDrive/Escritorio/TFG/DescriptivaMujeres.csv", row.names =
TRUE)
```

#FAS de cada barrio

```
for (s in levels(dadesBCN3$SexeCodi)){

df <- dadesBCN3[dadesBCN3$SexeCodi==s,]
par(mfrow=c(2,5))

for(i in 1:10){

dd <- df[df$SectorSanitariDescripcio==unique(df$SectorSanitariDescripcio)[i],]
dd <- ts(dd$Casos)

acf(dd,ylim=c(-1,1),lag=21,
main=c('FAS',s,substring(unique(dadesBCN3$SectorSanitariDescripcio)[i],first=10)))
}

}
```

#FAP de cada barrio

```
for (s in levels(dadesBCN3$SexeCodi)){

df <- dadesBCN3[dadesBCN3$SexeCodi==s,]
par(mfrow=c(2,5))

for(i in 1:10){

dd <- df[df$SectorSanitariDescripcio==unique(df$SectorSanitariDescripcio)[i],]
dd <- ts(dd$Casos)

pacf(dd,ylim=c(-1,1),lag=21,
main=c('FAP',s,substring(unique(dadesBCN3$SectorSanitariDescripcio)[i],first=10)))
}

}
```

```
#Modelo AR(1)
```

```
#Estimación
```

```
arc<- c()
```

```
ars <- c()
```

```
arn <- c()
```

```
ari <-c()
```

```
arcv <- c()
```

```
par(mfrow=c(2,5))
```

```
for (s in levels(dadesBCN3$SexeCodi)){
```

```
df <- dadesBCN3[dadesBCN3$SexeCodi==s,]
```

```
for (i in 1:10){
```

```
casos <- df[df$SectorSanitariDescripcio==unique(df$SectorSanitariDescripcio)[i],]
```

```
model<-arima(casos$Casos,c(1,0,0),seasonal=list(order=c(0,0,0),period=7))
```

```
arc <- c(arc,model$coef)
```

```
ars <- c(ars,2*pnorm(c(abs(model$coef)/sqrt(diag(model$var.coef))), mean=0, sd=1, lower.tail=FALSE))
```

```
arn<-c(arn,shapiro.test(model$residuals)$p.value)
```

```
ari<- c(ari,Box.test(model$residuals)$p.value)
```

```
arcv <- c(arcv,sqrt(diag(model$var.coef)))
```

```
#Predicciones Ex -Ante
```

```
serief<-predict(model,n.ahead=20)
```

```
inf = serief$pred - 2*serief$se
```

```
inf<- ifelse(inf<0,0,inf)
```

```
sup = serief$pred + 2*serief$se
```

```

x<-casos$TipusCasData[349:368]
y<-seq.Date(as.Date("2021-03-04"),as.Date("2021-03-23"),1)

plot(casos$Casos[349:368],col="cadetblue1",          main=c('Prediccion Ex-
Ante',s,substring(unique(dadesBCN3$SectorSanitariDescripccio)[i],first=10)),type="lines",ylab="
Número Casos",xlab="Fecha",xaxt="n",xlim=c(0,40),ylim=c(0,max(sup)+5))

lines(seq(21,40,1),serief$pred, col="red" )
lines(seq(21,40,1),inf, col="darkgreen")
lines(seq(21,40,1),sup, col="darkgreen")
axis(1,at=seq(1,40,1),labels=c(x,y))

}

}

```

#Parámetros significativos y residuos independientes, no normalidad, no hay estacionariedad

```

arc <- matrix(arc,ncol=2,byrow = TRUE)

arcH <- data.frame("Alpha"=arc[1:10,1],"Intercept" = arc[1:10,2],row.names =
unique(dadesBCN3$SectorSanitariDescripccio))

arcD <- data.frame("Alpha"=arc[11:20,1],"Intercept" = arc[11:20,2],row.names =
unique(dadesBCN3$SectorSanitariDescripccio))

arcv <- matrix(arcv,ncol=2,byrow = TRUE)

arcvH <- data.frame("Alpha"=arcv[1:10,1],"Intercept" = arcv[1:10,2],row.names =
unique(dadesBCN3$SectorSanitariDescripccio))

arcvD <- data.frame("Alpha"=arcv[11:20,1],"Intercept" = arcv[11:20,2],row.names =
unique(dadesBCN3$SectorSanitariDescripccio))

ars <- matrix(ars,ncol=2,byrow = TRUE)

arsH <- data.frame("Alpha"=ars[1:10,1],"Intercept" = ars[1:10,2],row.names =
unique(dadesBCN3$SectorSanitariDescripccio))

arsD <- data.frame("Alpha"=ars[11:20,1],"Intercept" = ars[11:20,2],row.names =
unique(dadesBCN3$SectorSanitariDescripccio))

```

```

arn <- matrix(arn,ncol=1,byrow = TRUE)
arnH <- data.frame("Alpha"=arn[1:10,],row.names = unique(dadesBCN3$SectorSanitariDescripicio))
arnD <- data.frame("Alpha"=arn[11:20,],row.names = unique(dadesBCN3$SectorSanitariDescripicio))

ari <- matrix(ari,ncol=1,byrow = TRUE)
ariH <- data.frame("Alpha"=ari[1:10,],row.names = unique(dadesBCN3$SectorSanitariDescripicio))
ariD <- data.frame("Alpha"=ari[11:20,],row.names = unique(dadesBCN3$SectorSanitariDescripicio))

write.csv(arcH,"C:/Users/pablo/OneDrive/Escritorio/TFG/ParametrosARHombres.csv", row.names =
TRUE)
write.csv(arcD,"C:/Users/pablo/OneDrive/Escritorio/TFG/ParametrosARMujeres.csv", row.names =
TRUE)

write.csv(arcvH,"C:/Users/pablo/OneDrive/Escritorio/TFG/ParametrosDEARHombres.csv", row.names
= TRUE)
write.csv(arcvD,"C:/Users/pablo/OneDrive/Escritorio/TFG/ParametrosDEARMujeres.csv", row.names
= TRUE)

write.csv(arsH,"C:/Users/pablo/OneDrive/Escritorio/TFG/SignificacionARHombres.csv", row.names
= TRUE)
write.csv(arsD,"C:/Users/pablo/OneDrive/Escritorio/TFG/SignificacionARMujeres.csv", row.names
= TRUE)

write.csv(arnH,"C:/Users/pablo/OneDrive/Escritorio/TFG/normalidadARHombres.csv", row.names =
TRUE)
write.csv(arnD,"C:/Users/pablo/OneDrive/Escritorio/TFG/normalidadARMujeres.csv", row.names =
TRUE)

write.csv(ariH,"C:/Users/pablo/OneDrive/Escritorio/TFG/independenciaARHombres.csv", row.names
= TRUE)
write.csv(ariD,"C:/Users/pablo/OneDrive/Escritorio/TFG/independenciaARMujeres.csv", row.names
= TRUE)

#Predicciones Ex-Post

cobertura <- function(li,ls,obs){

kk<-c()

```

```

k<-0

for(i in 1:length(ls)){

  if(li[i] <= obs[i] && obs[i] <= ls[i]){

    k<-k+1

  }

}

kk <- c(kk,k)
return(kk)
}

arc<- c()
ars <- c()
arn <- c()
ari <-c()
arcv <- c()
eqmAR<- c()
eamAR <- c()
epamAR <- c()
cobsAR<-c()
riAR<-c()
par(mfrow=c(2,5))

for (s in levels(dadesBCN3$SexeCodi)){

df <- dadesBCN3[dadesBCN3$SexeCodi==s,]

for (i in 1:10){

casos <- df[df$SectorSanitariDescripcio==unique(df$SectorSanitariDescripcio)[i],]

serie<-casos$Casos[1:306]
model2<-arima(serie,c(1,0,0),seasonal=list(order=c(0,0,0),period=7))
serie2f<-predict(model2,n.ahead=62,interval="confidence")

```

```

inf = serie2f$pred - 1.96*serie2f$se
inf <- ifelse(inf<0,0,inf)
sup = serie2f$pred + 1.96*serie2f$se

arc <- c(arc,model2$coef)
ars <- c(ars,2*pnorm(c(abs(model2$coef)/sqrt(diag(model2$var.coef))), mean=0, sd=1,
lower.tail=FALSE))
arn<-c(arn,shapiro.test(model2$residuals)$p.value)
ari<- c(ari,Box.test(model2$residuals)$p.value)
arcv <- c(arcv,sqrt(diag(model2$var.coef)))

preds <- data.frame(TCasos=serie2f$pred, LL=inf,
UL=sup, TipusCasData=unique(dadesBCN4$TipusCasData[dadesBCN4$TipusCasData>"2020-12-31"]))

d<-
dadesBCN4[dadesBCN4$SectorSanitariDescripcio==names(table(dadesBCN4$SectorSanitariDescripcio))
[i] & dadesBCN4$SexeCodi==s &
dadesBCN4$TipusCasData>"2020-12-31", ]

assign(paste0("p",s, i), ggplot(d, aes(x=TipusCasData, y=TCasos)) +
geom_line()+scale_x_date(date_breaks = "1 week", date_labels = "%m-%d-%Y")+theme(axis.text.x=element_text(angle=60, hjust=1), legend.position = "none")+
geom_ribbon(data=preds,aes(ymin=LL,ymax=UL),alpha=0.3)+geom_line(data=preds,aes(x=TipusCasData
, y=TCasos, color = "yellow"))+
xlab("Data")+ylab("Nombre de casos"))

#Errores predicció

op<-data.frame("Obs"=casos[307:368,]$Casos,"Pred"=serie2f$pred)
op <- op[op$Obs!=0,]
errors<- op$Obs-op$Pred
eqmAR<-c(eqmAR,sum(errors*errors)/length(op$Obs))
eamAR<-c(eamAR,sum(abs(errors))/length(op$Obs))
epamAR<-c(epamAR,sum(abs(errors)/abs(op$Obs))/length(op$Obs))

#Cobertura
cobsAR<-c(cobsAR,cobertura(inf,sup,casos$Casos[307:368]))

#Rangs Predicció

```

```

riAR<-c(riAR,mean(sup-inf))
}

}

ggarrange(pHome1,pHome2,pHome3,pHome4,pHome5,pHome6,pHome7,pHome8,pHome9,pHome10,labels=substr
ing(unique(dadesBCN3$SectorSanitariDescripcio),first=10),ncol=5,nrow=2,common.legend = TRUE)

ggarrange(pDona1,pDona2,pDona3,pDona4,pDona5,pDona6,pDona7,pDona8,pDona9,pDona10,labels=substr
ing(unique(dadesBCN3$SectorSanitariDescripcio),first=10),ncol=5,nrow=2,common.legend = TRUE)

#Parametres

arc <- matrix(arc,ncol=2,byrow = TRUE)

arcH <- data.frame("Alpha"=arc[1:10,1],"Intercept" = arc[1:10,2],row.names =
unique(dadesBCN3$SectorSanitariDescripcio))

arcD <- data.frame("Alpha"=arc[11:20,1],"Intercept" = arc[11:20,2],row.names =
unique(dadesBCN3$SectorSanitariDescripcio))

arcv <- matrix(arcv,ncol=2,byrow = TRUE)

arcvH <- data.frame("Alpha"=arcv[1:10,1],"Intercept" = arcv[1:10,2],row.names =
unique(dadesBCN3$SectorSanitariDescripcio))

arcvD <- data.frame("Alpha"=arcv[11:20,1],"Intercept" = arcv[11:20,2],row.names =
unique(dadesBCN3$SectorSanitariDescripcio))

ars <- matrix(ars,ncol=2,byrow = TRUE)

arsH <- data.frame("Alpha"=ars[1:10,1],"Intercept" = ars[1:10,2],row.names =
unique(dadesBCN3$SectorSanitariDescripcio))

arsD <- data.frame("Alpha"=ars[11:20,1],"Intercept" = ars[11:20,2],row.names =
unique(dadesBCN3$SectorSanitariDescripcio))

arn <- matrix(arn,ncol=1,byrow = TRUE)

arnH <- data.frame("Alpha"=arn[1:10,],row.names = unique(dadesBCN3$SectorSanitariDescripcio))
arnD <- data.frame("Alpha"=arn[11:20,],row.names = unique(dadesBCN3$SectorSanitariDescripcio))

ari <- matrix(ari,ncol=1,byrow = TRUE)

ariH <- data.frame("Alpha"=ari[1:10,],row.names = unique(dadesBCN3$SectorSanitariDescripcio))
ariD <- data.frame("Alpha"=ari[11:20,],row.names = unique(dadesBCN3$SectorSanitariDescripcio))

```

```

write.csv(arcH,"C:/Users/pablo/OneDrive/Escritorio/TFG/ParametrosARHombres.csv", row.names =
TRUE)

write.csv(arcD,"C:/Users/pablo/OneDrive/Escritorio/TFG/ParametrosARMujeres.csv", row.names =
TRUE)

write.csv(arcvH,"C:/Users/pablo/OneDrive/Escritorio/TFG/ParametrosDEARHombres.csv", row.names
= TRUE)

write.csv(arcvD,"C:/Users/pablo/OneDrive/Escritorio/TFG/ParametrosDEARMujeres.csv", row.names
= TRUE)

write.csv(arsH,"C:/Users/pablo/OneDrive/Escritorio/TFG/SignificacionARHombres.csv", row.names
= TRUE)

write.csv(arsD,"C:/Users/pablo/OneDrive/Escritorio/TFG/SignificacionARMujeres.csv", row.names
= TRUE)

write.csv(arnH,"C:/Users/pablo/OneDrive/Escritorio/TFG/normalidadARHombres.csv", row.names =
TRUE)

write.csv(arnD,"C:/Users/pablo/OneDrive/Escritorio/TFG/normalidadARMujeres.csv", row.names =
TRUE)

write.csv(ariH,"C:/Users/pablo/OneDrive/Escritorio/TFG/independenciaARHombres.csv", row.names
= TRUE)

write.csv(ariD,"C:/Users/pablo/OneDrive/Escritorio/TFG/independenciaARMujeres.csv", row.names
= TRUE)

```

#Errors Predicción

```

eqmAR <- matrix(eqmAR,ncol=2)

eqmAR <- data.frame("EQM Hombres"=eqmAR[1:10,1],"EQM Mujeres"=eqmAR[1:10,2],row.names =
unique(dadesBCN3$SectorSanitariDescripcio))

write.csv(eqmAR,"C:/Users/pablo/OneDrive/Escritorio/TFG/EQMAR.csv", row.names = TRUE)

```

```

eamAR <- matrix(eamAR,ncol=2)

eamAR <- data.frame("EAM Hombres"=eamAR[1:10,1],"EAM Mujeres"=eamAR[1:10,2],row.names =
unique(dadesBCN3$SectorSanitariDescripcio))

write.csv(eamAR,"C:/Users/pablo/OneDrive/Escritorio/TFG/EAMAR.csv", row.names = TRUE)

```

```

epamAR <- matrix(epamAR,ncol=2)

epamAR <- data.frame("EPAM Hombres"=epamAR[1:10,1],"EPAM Mujeres"=epamAR[1:10,2],row.names =
unique(dadesBCN3$SectorSanitariDescripcio))

```



```
write.csv(epamAR,"C:/Users/pablo/OneDrive/Escritorio/TFG/EPAMARs.csv", row.names = TRUE)
```

#Cobertura

```
cobsAR<- matrix(cobsAR,ncol=2)
```

```
cobsAR <- data.frame("Cobertura Hombres"=cobsAR[1:10,1],"Cobertura  
Mujeres"=cobsAR[1:10,2],row.names = unique(dadesBCN3$SectorSanitariDescripccio))
```

```
write.csv(cobsAR,"C:/Users/pablo/OneDrive/Escritorio/TFG/CoberturaARs.csv", row.names = TRUE)
```

#Rang Interval Predicció

```
riAR<- matrix(riAR,ncol=2)
```

```
riAR <- data.frame("Rang Hombres"=riAR[1:10,1],"Rang Mujeres"=riAR[1:10,2],row.names =  
unique(dadesBCN3$SectorSanitariDescripccio))
```

```
write.csv(riAR,"C:/Users/pablo/OneDrive/Escritorio/TFG/RangARs.csv", row.names = TRUE)
```

#Modelo INAR(1)

#Función Verosimilitud y estimación

```
inar <- function(param){
```

```
lv <- 0
```

```
for (t in 2:length(x)){
```

```
if(weekdays(d[t])=="lunes"){
```

```
xt <- min(x[t-1],x[t])
```

```
bin <- 0
```

```
for (i in 0:xt ){
```

```
bin <- bin + (choose(x[t-1],i) * param[1]^i * (1- param[1])^(x[t-1] - i )) * ( exp(-  
param[2]*pob/10000) * (param[2]*pob/10000)^(x[t]- i ) ) / factorial(x[t]- i )
```

```
}
```

```
lv <- lv + log(bin)
```

```
}
```

```
if(weekdays(d[t])=="martes"){
```

```
xt <- min(x[t-1],x[t])
```

```
bin <- 0
```

```
for (i in 0:xt ){
```

```
bin <- bin + (choose(x[t-1],i) * param[1]^i * (1- param[1])^(x[t-1] - i )) * ( exp(-  
param[3]*pob/10000) * (param[3]*pob/10000)^(x[t]- i ) ) / factorial(x[t]- i )
```

```
}
```

```

lv <- lv + log(bin)
}

if(weekdays(d[t])=="miércoles"){
xt <- min(x[t-1],x[t])
bin <- 0
for (i in 0:xt ){
bin <- bin + (choose(x[t-1],i) * param[1]^i * (1- param[1])^(x[t-1] - i )) * ( exp(-
param[4]*pob/10000) * (param[4]*pob/10000)^(x[t]- i ) ) / factorial(x[t]- i )
}
lv <- lv + log(bin)
}

if(weekdays(d[t])=="jueves"){
xt <- min(x[t-1],x[t])
bin <- 0
for (i in 0:xt ){
bin <- bin + (choose(x[t-1],i) * param[1]^i * (1- param[1])^(x[t-1] - i )) * ( exp(-
param[5]*pob/10000) * (param[5]*pob/10000)^(x[t]- i ) ) / factorial(x[t]- i )
}
lv <- lv + log(bin)
}

if(weekdays(d[t])=="viernes"){
xt <- min(x[t-1],x[t])
bin <- 0
for (i in 0:xt ){
bin <- bin + (choose(x[t-1],i) * param[1]^i * (1- param[1])^(x[t-1] - i )) * ( exp(-
param[6]*pob/10000) * (param[6]*pob/10000)^(x[t]- i ) ) / factorial(x[t]- i )
}
lv <- lv + log(bin)
}

if(weekdays(d[t])=="sábado"){
xt <- min(x[t-1],x[t])
bin <- 0
for (i in 0:xt ){
bin <- bin + (choose(x[t-1],i) * param[1]^i * (1- param[1])^(x[t-1] - i )) * ( exp(-
param[7]*pob/10000) * (param[7]*pob/10000)^(x[t]- i ) ) / factorial(x[t]- i )
}
lv <- lv + log(bin)
}

```

```

}

if(weekdays(d[t])=="domingo"){
xt <- min(x[t-1],x[t])

bin <- 0

for (i in 0:xt ){

bin <- bin + (choose(x[t-1],i) * param[1]^i * (1- param[1])^(x[t-1] - i )) * ( exp(-
param[8]*pob/10000) * (param[8]*pob/10000)^(x[t]- i ) ) / factorial(x[t]- i )

}

lv <- lv + log(bin)

}

}

return(lv*(-1))

}

```

#Dataframes únicos para cada barrio y su verosimilitud y errores

```
poblacio <- c(108331,270694,123651,174799,82635,174012,151976,241691,187584,151157)
```

```
CV <- dadesBCN3 [dadesBCN3$SectorSanitariDescripcio == "BARCELONA CIUTAT VELLA",]
```

```
x <- ts(CV[CV$SexeCodi=="Home",4])
```

```
d <- unique(CV$TipusCasData)
```

```
pob<-poblacio[1]
```

```
lmvCVH <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
```

```
deCVH<- sqrt(diag(solve(lmvCVH$hessian)))
```

```
x <- ts(CV[CV$SexeCodi=="Dona",4])
```

```
lmvCVD <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
```

```
deCVD<- sqrt(diag(solve(lmvCVD$hessian)))
```

```
EIX <- dadesBCN3 [dadesBCN3$SectorSanitariDescripcio == "BARCELONA EIXAMPLE",]
```

```
x <- ts(EIX[EIX$SexeCodi=="Home",4])
```

```
d <- unique(EIX$TipusCasData)
```

```
pob <- poblacio[2]
```

```
lmvEIXH <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
```

```
deEIXH<- sqrt(diag(inv(lmvEIXH$hessian)))
```

```

x <- ts(EIX[EIX$SexeCodi=="Dona",4])
lmvEIXD <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deEIXD<- sqrt(diag(inv(lmvEIXD$hessian)))

GRACIA <- dadesBCN3 [dadesBCN3$SectorSanitariDescripcio == "BARCELONA GRÀCIA",]
x <- ts(GRACIA[GRACIA$SexeCodi=="Home",4])
d <- unique(GRACIA$TipusCasData)
pob<-poblacio[3]
lmvGRACIAH <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deGRACIAH <- sqrt(diag(inv(lmvGRACIAH$hessian)))

x <- ts(GRACIA[GRACIA$SexeCodi=="Dona",4])
lmvGRACIAD <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deGRACIAD <- sqrt(diag(inv(lmvGRACIAD$hessian)))

HG <- dadesBCN3 [dadesBCN3$SectorSanitariDescripcio == "BARCELONA HORTA-GUINARDÓ",]
x <- ts(HG[HG$SexeCodi=="Home",4])
d <- unique(HG$TipusCasData)
pob<-poblacio[4]
lmvHGH <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deHGH <- sqrt(diag(inv(lmvHGH$hessian)))

x <- ts(HG[HG$SexeCodi=="Dona",4])
lmvHGD <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deHGD <- sqrt(diag(inv(lmvHGD$hessian)))

LESC <- dadesBCN3 [dadesBCN3$SectorSanitariDescripcio == "BARCELONA LES CORTS",]
x <- ts(LESC[LESC$SexeCodi=="Home",4])
d <- unique(LESC$TipusCasData)
pob<-poblacio[5]
lmvLESCH <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deLESCH <- sqrt(diag(inv(lmvLESCH$hessian)))

x <- ts(LESC[LESC$SexeCodi=="Dona",4])
lmvLESCD <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)

```

```

deLESCD <- sqrt(diag(inv(lmvLESCD$hessian)))

NB <- dadesBCN3 [dadesBCN3$SectorSanitariDescripcio == "BARCELONA NOU BARRIS",]
x <- ts(NB[NB$SexeCodi=="Home",4])
d <- unique(NB$TipusCasData)
pob<-poblacio[6]
lmvNBH <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deNBH <- sqrt(diag(inv(lmvNBH$hessian)))

x <- ts(NB[NB$SexeCodi=="Dona",4])
lmvNBD <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deNBD <- sqrt(diag(inv(lmvNBD$hessian)))

SA <- dadesBCN3 [dadesBCN3$SectorSanitariDescripcio == "BARCELONA SANT ANDREU",]
x <- ts(SA[SA$SexeCodi=="Home",4])
d <- unique(SA$TipusCasData)
pob<-poblacio[7]
lmvSAH <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deSAH <- sqrt(diag(inv(lmvSAH$hessian)))

x <- ts(SA[SA$SexeCodi=="Dona",4])
lmvSAD <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deSAD <- sqrt(diag(inv(lmvSAD$hessian)))

SMRT <- dadesBCN3 [dadesBCN3$SectorSanitariDescripcio == "BARCELONA SANT MARTÍ",]
x <- ts(SMRT[SMRT$SexeCodi=="Home",4])
d <- unique(SMRT$TipusCasData)
pob<-poblacio[8]
lmvSMRTH <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deSMRTH <- sqrt(diag(inv(lmvSMRTH$hessian)))

x <- ts(SMRT[SMRT$SexeCodi=="Dona",4])
lmvSMRTD <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deSMRTD <- sqrt(diag(inv(lmvSMRTD$hessian)))

SM <- dadesBCN3 [dadesBCN3$SectorSanitariDescripcio == "BARCELONA SANTIS-MONTJUÏC",]
x <- ts(SM[SM$SexeCodi=="Home",4])
d <- unique(SM$TipusCasData)

```

```

pob<-poblacio[9]

lmvSMH <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deSMH <-sqrt(diag(solve(lmvSMH$hessian)))

x <- ts(SM[SM$SexeCodi=="Dona",4])
lmvSMD <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deSMD <-sqrt(diag(solve(lmvSMD$hessian)))

SSG <- dadesBCN3 [dadesBCN3$SectorSanitariDescripcio == "BARCELONA SARRIÀ-SANT GERVASI",]
x <- ts(SSG[SSG$SexeCodi=="Home",4])
d <- unique(SSG$TipusCasData)
pob<-poblacio[10]
lmvSSGH <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deSSGH <- sqrt(diag(inv(lmvSSGH$hessian)))

x <- ts(SSG[SSG$SexeCodi=="Dona",4])
lmvSSGD <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deSSGD <- sqrt(diag(inv(lmvSSGD$hessian)))

peHante <- c(lmvCVH$estimate,
lmvEIXH$estimate,lmvGRACIAH$estimate,lmvHGH$estimate,lmvLESCH$estimate,lmvNBH$estimate,lmvSAH$
estimate,lmvSMRTH$estimate,lmvSMH$estimate,lmvSSGH$estimate)

peHante <- matrix(peHante,ncol=8,byrow = TRUE)

peHante <- data.frame(Alpha=peHante[,1],Lambda1=peHante[,2],Lambda2=peHante[,3],Lambda3=peHante[,4],Lambd
a4=peHante[,5],Lambda5=peHante[,6],Lambda6=peHante[,7],Lambda7=peHante[,8],row.names=
unique(dadesBCN3$SectorSanitariDescripcio))

deTHante <- c(deCVH,deEIXH,deGRACIAH,deHGH,deLESCH,deNBH,deSAH,deSMRTH,deSMH,deSSGH)
deTHante <- matrix(deTHante,byrow=TRUE,ncol=8)

deTHante <- data.frame("Alpha error"=deTHante[,1],"Lambda1 error"= deTHante[,2],"Lambda2
error"=deTHante[,3],"Lambda3 error"=deTHante[,4],"Lambda4 error"=deTHante[,5],"Lambda5
error"=deTHante[,6],"Lambda6 error"=deTHante[,7],"Lambda7 error"=deTHante[,8],row.names =
unique(dadesBCN3$SectorSanitariDescripcio))

peDante <- c(lmvCVD$estimate,
lmvEIXD$estimate,lmvGRACIAD$estimate,lmvHGD$estimate,lmvLESCD$estimate,lmvNBD$estimate,lmvSAD$
estimate,lmvSMRTD$estimate,lmvSMD$estimate,lmvSSGD$estimate)

peDante <- matrix(peDante,ncol=8,byrow=TRUE)

peDante <- data.frame(Alpha=peDante[,1],Lambda1=peDante[,2],Lambda2=peDante[,3],Lambda3=peDante[,4],Lambd
a4=peDante[,5],Lambda5=peDante[,6],Lambda6=peDante[,7],Lambda7=peDante[,8],row.names=
unique(dadesBCN3$SectorSanitariDescripcio))

deTDante <- c(deCVD,deEIXD,deGRACIAD,deHGD,deLESCD,deNBD,deSAD,deSMRTD,deSMD,deSSGD)

```

```

deTDante <- matrix(deTDante,byrow=TRUE,ncol=8)

deTDante <- data.frame("Alpha error"=deTDante[,1],"Lambda1 error"= deTDante[,2],"Lambda2
error"=deTDante[,3],"Lambda3 error"=deTDante[,4],"Lambda4 error"=deTDante[,5],"Lambda5
error"=deTDante[,6],"Lambda6 error"=deTDante[,7],"Lambda7 error"=deTDante[,8],row.names =
unique(dadesBCN3$SectorSanitariDescripcio))

write.csv(peHante,"C:/Users/pablo/OneDrive/Escritorio/TFG/ParametrosInarH.csv", row.names =
TRUE)

write.csv(peDante,"C:/Users/pablo/OneDrive/Escritorio/TFG/ParametrosInarD.csv", row.names =
TRUE)

write.csv(deTHante,"C:/Users/pablo/OneDrive/Escritorio/TFG/ErroresInarH.csv", row.names = TRUE)

write.csv(deTDante,"C:/Users/pablo/OneDrive/Escritorio/TFG/ErroresInard.csv", row.names = TRUE)

#Verosimilitud cada barrio

lmvp2 <- c()
for (s in levels(CV$SexeCodi) ){

df <- dadesBCN3[dadesBCN3$SexeCodi==s,]
for (i in 1:10){

df <- df[df$SectorSanitariDescripcio==df$SectorSanitariDescripcio[i],]

x <- ts(df$Casos)
d <- df$TipusCasData
pob <- poblacio[i]
lmv2 <- nlm(inar, p=c(0.7,1,2,3,4,5,6,7),hessian=TRUE)
lmvp2 <- c(lmvp,lmv2$estimate)

}

}

z1<-matrix(lmvp2,byrow=TRUE,nrow=20)

```

```

z1<- cbind(z2,deT)

z1 <- data.frame("Alpha1"=z2[,1],"Alpha2"= z2[,2],"Lambda"=z2[,3], "SD Alpha1"=z2[,4],"SD
Alpha2"=z2[,5],"SD Lambda"=z2[,6],row.names = unique(dadesBCN32$SectorSanitariDescripccio))

write.csv(z1,"C:/Users/pablo/OneDrive/Escritorio/TFG/EstimacionInar.csv", row.names = TRUE)

```

#Predicciones

```

probfunc <- function(xn, prob, p1, lamb, pob){
xt <- min(xn, prob)

bin <- 0

for (i in 0:xt){
bin <- bin + (choose(xn,i)*p1^i*(1-p1)^(xn-i))*(exp(-lamb*pob/10000)*(lamb*pob/10000)^(prob-
i))/factorial(prob-i)
}

return(bin)
}

```

```

pred <- function(xn, p1, lambda, pob,num){
mediana <- vector()
li <- vector()
ls <- vector()

for (npred in 1:num){
sumprob <- 0
k <- 0
while (sumprob < 0.50){
probk <- probfunc(xn, k, p1[npred], lambda[npred], pob)
sumprob <- probk + sumprob
mediana[npred] <- k
k <- k + 1
}

sumprob <- 0
k <- 0
while (sumprob < 0.95){
probk <- probfunc(xn, k, p1[npred], lambda[npred], pob)
sumprob <- probk + sumprob
if (sumprob < 0.025){
li[npred] <- k

```



```

}
k <- k + 1
}
k <- 0
sumprob <- 0
while (sumprob < 0.95){
probk <- probfunc(xn, k, pl[npred], lambda[npred], pob)
sumprob <- probk + sumprob
ls[npred] <- k
k <- k + 1
}
}
li[is.na(li)] <- 0
resultat=list(LowerLim=li, Median=mediana, UpperLim=ls)
return(resultat)
}

```

#Ex-Ante

```

homes <- vector()
for (i in 1:length(table(dadesBCN4$SectorSanitariDescripcio))){
nou_valor <- dadesBCN4$TCasos[dadesBCN4$SexeCodi=="Home" &
dadesBCN4$SectorSanitariDescripcio==names(table(dadesBCN4$SectorSanitariDescripcio))[i] &
dadesBCN4$TipusCasData=="2021-03-03"]
homes <- c(homes, nou_valor)
}

predH_ante <- list()
for (j in 1:length(table(dadesBCN4$SectorSanitariDescripcio))){
poblacio
unique(dadesBCN4$Poblacio[dadesBCN4$SectorSanitariDescripcio==names(table(dadesBCN4$SectorSani
tariDescripcio))[j]])
lambda <- vector()

for (i in 1:20){
if (i %in% seq(1, 20, 7)) lambda[i] <- pars[[j]][1, 5]
if (i %in% seq(2, 20, 7)) lambda[i] <- pars[[j]][1, 6]
if (i %in% seq(3, 20, 7)) lambda[i] <- pars[[j]][1, 7]
}
}

```

```

if (i %in% seq(4, 20, 7)) lambda[i] <- pars[[j]][1, 8]
if (i %in% seq(5, 20, 7)) lambda[i] <- pars[[j]][1, 2]
if (i %in% seq(6, 20, 7)) lambda[i] <- pars[[j]][1, 3]
if (i %in% seq(7, 20, 7)) lambda[i] <- pars[[j]][1, 4]
}

f <- vector()
g <- vector()
l <- vector()
f[1] <- pars[[j]][1, 1]
l[1] <- lambda[1]

for (i in 2:20){
f[i] <- f[1]*f[i-1]
l[i] <- f[1]*l[i-1] + lambda[i]
}

predH_ante[[j]] <- pred(homes[j], f, l, poblacio,20)
}

#Dones

dones <- vector()
for (i in 1:length(table(dadesBCN4$SectorSanitariDescripcio))){
nou_valor <- dadesBCN4$TCasos[dadesBCN4$SexeCodi=="Dona" &
dadesBCN4$SectorSanitariDescripcio==names(table(dadesBCN4$SectorSanitariDescripcio))[i] &
dadesBCN4$TipusCasData=="2021-03-03"]
dones <- c(dones, nou_valor)
}

predD_ante <- list()
for (j in 1:length(table(dadesBCN4$SectorSanitariDescripcio))){
poblacio <- unique(dadesBCN4$Poblacio[dadesBCN4$SectorSanitariDescripcio==names(table(dadesBCN4$SectorSanitariDescripcio))[j]])
lambda <- vector()
s
for (i in 1:20){
if (i %in% seq(1, 20, 7)) lambda[i] <- pars[[j]][1, 5]
if (i %in% seq(2, 20, 7)) lambda[i] <- pars[[j]][1, 6]
}
}

```

```

if (i %in% seq(3, 20, 7)) lambda[i] <- pars[[j]][1, 7]
if (i %in% seq(4, 20, 7)) lambda[i] <- pars[[j]][1, 8]
if (i %in% seq(5, 20, 7)) lambda[i] <- pars[[j]][1, 2]
if (i %in% seq(6, 20, 7)) lambda[i] <- pars[[j]][1, 3]
if (i %in% seq(7, 20, 7)) lambda[i] <- pars[[j]][1, 4]
}

f <- vector()
g <- vector()
l <- vector()
f[1] <- pars[[j]][1, 1]
l[1] <- lambda[1]

for (i in 2:20){
f[i] <- f[1]*f[i-1]
l[i] <- f[1]*l[i-1] + lambda[i]
}
predD_ante[[j]] <- pred(dones[j], f, l, poblacio,20)
}

par(mfrow=c(2,5))

for (s in levels(dadesBCN4$SexeCodi)){

df2<-dadesBCN3[dadesBCN3$SexeCodi==s,]

for(i in 1:10){

casos <- df2[df2$SectorSanitariDescripcio==unique(df2$SectorSanitariDescripcio)[i],]
x<-casos$TipusCasData[349:368]
y<-seq.Date(as.Date("2021-03-04"),as.Date("2021-03-23"),1)

if (s=="Home"){

```

```

plot(casos$Casos[349:368], col="cadetblue1", main=c('Prediccion Ex-
Ante', s, substring(unique(dadesBCN3$SectorSanitariDescripcio)[i], first=10)), type="lines", ylab="
Número Casos", xlab="Fecha", xaxt="n", xlim=c(0, 40), ylim=c(0, max(casos$Casos[349:368]+10)))

lines(seq(21, 40, 1), predH_ante[[i]]$Median, col="red" )
lines(seq(21, 40, 1), predH_ante[[i]]$LowerLim, col="darkgreen")
lines(seq(21, 40, 1), predH_ante[[i]]$UpperLim, col="darkgreen")
axis(1, at=seq(1, 40, 1), labels=c(x, y))

}else{

plot(casos$Casos[349:368], col="cadetblue1", main=c('Prediccion Ex-
Ante', s, substring(unique(dadesBCN3$SectorSanitariDescripcio)[i], first=10)), type="lines", ylab="
Número Casos", xlab="Fecha", xaxt="n", xlim=c(0, 40), ylim=c(0, max(casos$Casos[349:368]+10)))

lines(seq(21, 40, 1), predD_ante[[i]]$Median, col="red" )
lines(seq(21, 40, 1), predD_ante[[i]]$LowerLim, col="darkgreen")
lines(seq(21, 40, 1), predD_ante[[i]]$UpperLim, col="darkgreen")
axis(1, at=seq(1, 40, 1), labels=c(x, y))

}}
}

```

#Ex-Post

```

CV <- dadesBCN3 [dadesBCN3$SectorSanitariDescripcio == "BARCELONA CIUTAT VELLA",]
x <- ts(CV[CV$SexeCodi=="Home",]$Casos[1:306])
d <- unique(CV$TipusCasData)[1:306]
pob<-poblacio[1]
lmvCVH <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1), hessian=TRUE)
deCVH<- sqrt(diag(solve(lmvCVH$hessian)))

x <- ts(CV[CV$SexeCodi=="Dona",]$Casos[1:306])
lmvCVD <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1), hessian=TRUE)
deCVD<- sqrt(diag(solve(lmvCVD$hessian)))

EIX <- dadesBCN3 [dadesBCN3$SectorSanitariDescripcio == "BARCELONA EIXAMPLE",]
x <- ts(EIX[EIX$SexeCodi=="Home",]$Casos[1:306])
d <- unique(EIX$TipusCasData)[1:306]
pob <- poblacio[2]
lmvEIXH <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1), hessian=TRUE)
deEIXH<- sqrt(diag(inv(lmvEIXH$hessian)))

```

```

x <- ts(EIX[EIX$SexeCodi=="Dona",]$Casos[1:306])
lmvEIXD <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deEIXD<- sqrt(diag(inv(lmvEIXD$hessian)))

GRACIA <- dadesBCN3 [dadesBCN3$SectorSanitariDescripcio == "BARCELONA GRÀCIA",]
x <- ts(GRACIA[GRACIA$SexeCodi=="Home",]$Casos[1:306])
d <- unique(GRACIA$TipusCasData)[1:306]
pob<-poblacio[3]
lmvGRACIAH <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deGRACIAH <- sqrt(diag(inv(lmvGRACIAH$hessian)))

x <- ts(GRACIA[GRACIA$SexeCodi=="Dona",]$Casos[1:306])
lmvGRACIAD <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deGRACIAD <- sqrt(diag(inv(lmvGRACIAD$hessian)))

HG <- dadesBCN3 [dadesBCN3$SectorSanitariDescripcio == "BARCELONA HORTA-GUINARDÓ",]
x <- ts(HG[HG$SexeCodi=="Home",]$Casos[1:306])
d <- unique(HG$TipusCasData)[1:306]
pob<-poblacio[4]
lmvHGH <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deHGH <- sqrt(diag(inv(lmvHGH$hessian)))

x <- ts(HG[HG$SexeCodi=="Dona",]$Casos[1:306])
lmvHGD <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deHGD <- sqrt(diag(inv(lmvHGD$hessian)))

LESC <- dadesBCN3 [dadesBCN3$SectorSanitariDescripcio == "BARCELONA LES CORTS",]
x <- ts(LESC[LESC$SexeCodi=="Home",]$Casos[1:306])
d <- unique(LESC$TipusCasData)[1:306]
pob<-poblacio[5]
lmvLESCH <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deLESCH <- sqrt(diag(inv(lmvLESCH$hessian)))

x <- ts(LESC[LESC$SexeCodi=="Dona",]$Casos[1:306])
lmvLESCD <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deLESCD <- sqrt(diag(inv(lmvLESCD$hessian)))

```

```

NB <- dadesBCN3 [dadesBCN3$SectorSanitariDescripcio == "BARCELONA NOU BARRIS",]
x <- ts(NB[NB$SexeCodi=="Home",]$Casos[1:306])
d <- unique(NB$TipusCasData) [1:306]
pob<-poblacio[6]
lmvNBH <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deNBH <- sqrt(diag(inv(lmvNBH$hessian)))

x <- ts(NB[NB$SexeCodi=="Dona",]$Casos[1:306])
lmvNBD <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deNBD <- sqrt(diag(inv(lmvNBD$hessian)))

SA <- dadesBCN3 [dadesBCN3$SectorSanitariDescripcio == "BARCELONA SANT ANDREU",]
x <- ts(SA[SA$SexeCodi=="Home",]$Casos[1:306])
d <- unique(SA$TipusCasData) [1:306]
pob<-poblacio[7]
lmvSAH <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deSAH <- sqrt(diag(inv(lmvSAH$hessian)))

x <- ts(SA[SA$SexeCodi=="Dona",]$Casos[1:306])
lmvSAD <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deSAD <- sqrt(diag(inv(lmvSAD$hessian)))

SMRT <- dadesBCN3 [dadesBCN3$SectorSanitariDescripcio == "BARCELONA SANT MARTÍ",]
x <- ts(SMRT[SMRT$SexeCodi=="Home",]$Casos[1:306])
d <- unique(SMRT$TipusCasData) [1:306]
pob<-poblacio[8]
lmvSMRTH <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deSMRTH <- sqrt(diag(inv(lmvSMRTH$hessian)))

x <- ts(SMRT[SMRT$SexeCodi=="Dona",]$Casos[1:306])
lmvSMRTD <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deSMRTD <- sqrt(diag(inv(lmvSMRTD$hessian)))

SM <- dadesBCN3 [dadesBCN3$SectorSanitariDescripcio == "BARCELONA SANTSMONTJUÏC",]

```

```

x <- ts(SM[SM$SexeCodi=="Home",]$Casos[1:306])
d <- unique(SM$TipusCasData)[1:306]
pob<-poblacio[9]
lmvSMH <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deSMH <-sqrt(diag(solve(lmvSMH$hessian)))

x <- ts(SM[SM$SexeCodi=="Dona",]$Casos[1:306])
lmvSMD <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deSMD <-sqrt(diag(solve(lmvSMD$hessian)))

SSG <- dadesBCN3 [dadesBCN3$SectorSanitariDescripcio == "BARCELONA SARRIÀ-SANT GERVAZI",]
x <- ts(SSG[SSG$SexeCodi=="Home",]$Casos[1:306])
d <- unique(SSG$TipusCasData)[1:306]
pob<-poblacio[10]
lmvSSGH <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deSSGH <- sqrt(diag(inv(lmvSSGH$hessian)))

x <- ts(SSG[SSG$SexeCodi=="Dona",]$Casos[1:306])
lmvSSGD <- nlm(inar, p=c(0.5,1,1,1,1,1,1,1),hessian=TRUE)
deSSGD <- sqrt(diag(inv(lmvSSGD$hessian)))

peHpost <- c(lmvCVH$estimate,
lmvEIXH$estimate,lmvGRACIAH$estimate,lmvHGH$estimate,lmvLESCH$estimate,lmvNBH$estimate,lmvSAH$
estimate,lmvSMRTH$estimate,lmvSMH$estimate,lmvSSGH$estimate)
peHpost <- matrix(peHpost,ncol=8,byrow = TRUE)

peHpost <-
data.frame(Alpha=peHpost[,1],Lambda1=peHpost[,2],Lambda2=peHpost[,3],Lambda3=peHpost[,4],Lambd
a4=peHpost[,5],Lambda5=peHpost[,6],Lambda6=peHpost[,7],Lambda7=peHpost[,8],row.names=
unique(dadesBCN3$SectorSanitariDescripcio))

write.csv(peHpost,"C:/Users/pablo/OneDrive/Escritorio/TFG/ParametresHPostINAR.csv", row.names
= TRUE)

deTHpost <- c(deCVH,deEIXH,deGRACIAH,deHGH,deLESCH,deNBH,deSAH,deSMRTH,deSMH,deSSGH)
deTHpost <- matrix(deTHpost,byrow=TRUE,ncol=8)

deTHpost <- data.frame("Alpha error"=deTHpost[,1],"Lambda1 error"= deTHpost[,2],"Lambda2
error"=deTHpost[,3],"Lambda3 error"=deTHpost[,4],"Lambda4 error"=deTH[,5],"Lambda5
error"=deTH[,6],"Lambda6 error"=deTH[,7],"Lambda7 error"=deTH[,8],row.names
= unique(dadesBCN3$SectorSanitariDescripcio))

write.csv(deTHpost,"C:/Users/pablo/OneDrive/Escritorio/TFG/DesviacioHPostINAR.csv", row.names
= TRUE)

```

```

peDpost <- c(lmvCVD$estimate,
lmvEIXD$estimate, lmvGRACIAD$estimate, lmvHGD$estimate, lmvLESCD$estimate, lmvNBD$estimate, lmvSAD$
estimate, lmvSMRTD$estimate, lmvSMD$estimate, lmvSSGD$estimate)

peDpost <- matrix(peDpost, ncol=8, byrow=TRUE)

peDpost <- data.frame(Alpha=peDpost[,1], Lambda1=peDpost[,2], Lambda2=peDpost[,3], Lambda3=peDpost[,4], Lambd
a4=peDpost[,5], Lambda5=peDpost[,6], Lambda6=peDpost[,7], Lambda7=peDpost[,8], row.names=
unique(dadesBCN3$SectorSanitariDescripcio))

write.csv(peDpost, "C:/Users/pablo/OneDrive/Escritorio/TFG/ParametresDPostINAR.csv", row.names
= TRUE)

deTDpost <- c(deCVD, deEIXD, deGRACIAD, deHGD, deLESCD, deNBD, deSAD, deSMRTD, deSMD, deSSGD)

deTDpost <- matrix(deTDpost, byrow=TRUE, ncol=8)

deTDpost <- data.frame("Alpha error"=deTDpost[,1], "Lambda1 error"= deTDpost[,2], "Lambda2
error"=deTDpost[,3], "Lambda3 error"=deTDpost[,4], "Lambda4 error"=deTDpost[,5], "Lambda5
error"=deTDpost[,6], "Lambda6 error"=deTDpost[,7], "Lambda7 error"=deTDpost[,8], row.names =
unique(dadesBCN3$SectorSanitariDescripcio))

write.csv(deTDpost, "C:/Users/pablo/OneDrive/Escritorio/TFG/desviacioDPostINAR.csv", row.names
= TRUE)

# Homes

homes <- vector()

for (i in 1:length(table(dadesBCN4$SectorSanitariDescripcio))){
nou_valor <- dadesBCN4$TCasos[dadesBCN4$SexeCodi=="Home" &
dadesBCN4$SectorSanitariDescripcio==names(table(dadesBCN4$SectorSanitariDescripcio))[i] &
dadesBCN4$TipusCasData=="2020-12-31"]
homes <- c(homes, nou_valor)
}

predH_post <- list()

for (j in 1:length(table(dadesBCN4$SectorSanitariDescripcio))){

poblacio <- unique(dadesBCN4$Poblacio[dadesBCN4$SectorSanitariDescripcio==names(table(dadesBCN4$SectorSani
tariDescripcio))[j]])

lambda <- vector()

for (i in 1:62){
if (i %in% seq(1, 62, 7)) lambda[i] <- peHpost[j,6]
if (i %in% seq(2, 62, 7)) lambda[i] <- peHpost[j,7]
if (i %in% seq(3, 62, 7)) lambda[i] <- peHpost[j,8]
if (i %in% seq(4, 62, 7)) lambda[i] <- peHpost[j,2]
if (i %in% seq(5, 62, 7)) lambda[i] <- peHpost[j,3]

```



```

if (i %in% seq(6, 62, 7)) lambda[i] <- peHpost[j,4]
if (i %in% seq(7, 62, 7)) lambda[i] <- peHpost[j,5]
}

```

```

f <- vector()
g <- vector()
l <- vector()
f[1] <- pars[[j]][1, 1]
l[1] <- lambda[1]

```

```

for (i in 2:62){
f[i] <- f[1]*f[i-1]
l[i] <- f[1]*l[i-1] + lambda[i]
}
predH_post[[j]] <- pred(homes[j], f, l, poblacio,62)
}

```

#Dones

```

dones <- vector()
for (i in 1:length(table(dadesBCN4$SectorSanitariDescripcio))){
nou_valor <- dadesBCN4$TCasos[dadesBCN4$SexeCodi=="Dona" &
dadesBCN4$SectorSanitariDescripcio==names(table(dadesBCN4$SectorSanitariDescripcio))[i] &
dadesBCN4$TipusCasData=="2020-12-31"]
dones <- c(dones, nou_valor)
}

```

```

predD_post <- list() ### Cada element de la llista Ã©s un vector amb les prediccions per cada barri

```

```

for (j in 1:length(table(dadesBCN4$SectorSanitariDescripcio))
{
poblacio <- unique(dadesBCN4$Poblacio[dadesBCN4$SectorSanitariDescripcio==names(table(dadesBCN4$SectorSanitariDescripcio))[j]])
lambda <- vector()
# Les prediccions comencen en divendres
for (i in 1:62)
{
if (i %in% seq(1, 62, 7)) lambda[i] <- peDpost[j,6]
if (i %in% seq(2, 62, 7)) lambda[i] <- peDpost[j,7]

```

```

if (i %in% seq(3, 62, 7)) lambda[i] <- peDpost[j,8]
if (i %in% seq(4, 62, 7)) lambda[i] <- peDpost[j,2]
if (i %in% seq(5, 62, 7)) lambda[i] <- peDpost[j,3]
if (i %in% seq(6, 62, 7)) lambda[i] <- peDpost[j,4]
if (i %in% seq(7, 62, 7)) lambda[i] <- peDpost[j,5]
}

f <- vector()
g <- vector()
l <- vector()
f[1] <- pars[[j]][1, 1]
l[1] <- lambda[1]

for (i in 2:62)
{
f[i] <- f[1]*f[i-1]
l[i] <- f[1]*l[i-1] + lambda[i]
}
predD_post[[j]] <- pred(dones[j], f, l, poblacio,62)
}

#Gràfics

par(mfrow=c(2,5))

for (s in levels(dadesBCN4$SexeCodi)){

for(i in 1:10){

if(s=="Home"){

preds <- data.frame(TCasos=predH_post[[i]]$Median, LL=predH_post[[i]]$LowerLim,
UL=predH_post[[i]]$UpperLim,
TipusCasData=unique(dadesBCN4$TipusCasData[dadesBCN4$TipusCasData>"2020-12-31"]))

d<-
dadesBCN4[dadesBCN4$SectorSanitariDescripcio==names(table(dadesBCN4$SectorSanitariDescripcio))
[i] & dadesBCN4$SexeCodi==s &
dadesBCN4$TipusCasData>"2020-12-31", ]

```

```

assign(paste0("p",s, i), ggplot(d, aes(x=TipusCasData, y=TCasos)) +
geom_line()+scale_x_date(date_breaks = "1 week", date_labels = "%m-%d-%Y")+theme(axis.text.x=element_text(angle=60, hjust=1), legend.position = "none")+
geom_ribbon(data=preds,aes(ymin=LL,ymax=UL),alpha=0.3)+geom_line(data=preds,aes(x=TipusCasData, y=TCasos, color = "red"))+
xlab("Data")+ylab("Nombre de casos"))
}else{

preds <- data.frame(TCasos=predD_post[[i]]$Median, LL=predD_post[[i]]$LowerLim,
UL=predD_post[[i]]$UpperLim,
TipusCasData=unique(dadesBCN4$TipusCasData[dadesBCN4$TipusCasData>"2020-12-31"]))

d<-
dadesBCN4[dadesBCN4$SectorSanitariDescripcio==names(table(dadesBCN4$SectorSanitariDescripcio))
[i] & dadesBCN4$SexeCodi==s &
dadesBCN4$TipusCasData>"2020-12-31", ]

assign(paste0("p",s, i), ggplot(d, aes(x=TipusCasData, y=TCasos)) +
geom_line()+scale_x_date(date_breaks = "1 week", date_labels = "%m-%d-%Y")+theme(axis.text.x=element_text(angle=60, hjust=1), legend.position = "none")+
geom_ribbon(data=preds,aes(ymin=LL,ymax=UL),alpha=0.3)+geom_line(data=preds,aes(x=TipusCasData, y=TCasos, color = "red"))+
xlab("Data")+ylab("Nombre de casos"))

}

}

}

ggarrange(pHome1, pHome2, pHome3, pHome4, pHome5, pHome6, pHome7, pHome8, pHome9, pHome10,
labels = substring(as.character(unique(dadesBCN3$SectorSanitariDescripcio)),first=10),
ncol = 2, nrow = 5)

ggarrange(pDona1, pDona2, pDona3, pDona4, pDona5, pDona6, pDona7, pDona8, pDona9, pDona10,
labels=substring(as.character(unique(dadesBCN3$SectorSanitariDescripcio)),first=10),
ncol = 2, nrow = 5)

```

#Errors

```
eamINAR<-c()
eqmINAR<-c()
epamINAR<-c()

for (s in levels(dadesBCN3$SexeCodi)){

df<- dadesBCN3[dadesBCN3$SexeCodi==s,]

for (i in 1:10){

casos<- df[df$SectorSanitariDescripcio==unique(df$SectorSanitariDescripcio)[i],]

if(s=="Home"){

op<-data.frame("Obs"=casos[307:368,]$Casos,"Pred"=predH_post[[i]]$Median)
op <- op[op$Obs!=0,]
errors<- op$Obs-op$Pred
eqmINAR<-c(eqmINAR,sum(errors*errors)/length(op$Obs))
eamINAR<-c(eamINAR,sum(abs(errors))/length(op$Obs))
epamINAR<-c(epamINAR,sum(abs(errors)/abs(op$Obs))/length(op$Obs))

}else{

op<-data.frame("Obs"=casos[307:368,]$Casos,"Pred"=predD_post[[i]]$Median)
op <- op[op$Obs!=0,]
errors<- op$Obs-op$Pred
eqmINAR<-c(eqmINAR,sum(errors*errors)/length(op$Obs))
eamINAR<-c(eamINAR,sum(abs(errors))/length(op$Obs))
epamINAR<-c(epamINAR,sum(abs(errors)/abs(op$Obs))/length(op$Obs))

}

}

}

}
```

#Cobertura

```
cobsINAR<-c()
for (s in levels(dadesBCN3$SexeCodi)){

df <- dadesBCN3[dadesBCN3$SexeCodi==s,]

for (i in 1:10){

casos <- df[df$SectorSanitariDescripcio==unique(df$SectorSanitariDescripcio)[i],]

if(s=="Home"){

cobsINAR <-
c(cobsINAR, cobertura(predH_post[[i]]$LowerLim, predH_post[[i]]$UpperLim, casos$Casos[307:368]))

}else{

cobsINAR <-
c(cobsINAR, cobertura(predD_post[[i]]$LowerLim, predD_post[[i]]$UpperLim, casos$Casos[307:368]))

}

}

}
```

#Rang Interval Predicció

```
riINAR<-c()
for (s in levels(dadesBCN3$SexeCodi)){

df <- dadesBCN3[dadesBCN3$SexeCodi==s,]

for (i in 1:10){

casos <- df[df$SectorSanitariDescripcio==unique(df$SectorSanitariDescripcio)[i],]

if(s=="Home"){
```

```

riINAR <- c(riINAR,mean(predH_post[[i]]$UpperLim-predH_post[[i]]$LowerLim))

}else{

riINAR <- c(riINAR,mean(predD_post[[i]]$UpperLim-predD_post[[i]]$LowerLim))

}

}

}

}

eqmINAR <- matrix(eqmINAR,ncol=2)

eqmINAR <- data.frame("EQM Hombres"=eqmINAR[1:10,1],"EQM Mujeres"=eqmINAR[1:10,2],row.names =
unique(dadesBCN3$SectorSanitariDescripcio))

write.csv(eqmINAR,"C:/Users/pablo/OneDrive/Escritorio/TFG/EQMINAR.csv", row.names = TRUE)

eamINAR <- matrix(eamINAR,ncol=2)

eamINAR <- data.frame("EAM Hombres"=eamINAR[1:10,1],"EAM Mujeres"=eamINAR[1:10,2],row.names =
unique(dadesBCN3$SectorSanitariDescripcio))

write.csv(eamINAR,"C:/Users/pablo/OneDrive/Escritorio/TFG/EAMINAR.csv", row.names = TRUE)

epamINAR <- matrix(epamINAR,ncol=2)

epamINAR <- data.frame("EPAM Hombres"=epamINAR[1:10,1],"EPAM
Mujeres"=epamINAR[1:10,2],row.names = unique(dadesBCN3$SectorSanitariDescripcio))

write.csv(epamINAR,"C:/Users/pablo/OneDrive/Escritorio/TFG/EPAMINAR.csv", row.names = TRUE)

cobsINAR <- matrix(cobsINAR,ncol=2)

cobsINAR <- data.frame("Cobertura Hombres"=cobsINAR[1:10,1],"Cobertura
Mujeres"=cobsINAR[1:10,2],row.names = unique(dadesBCN3$SectorSanitariDescripcio))

write.csv(cobsINAR,"C:/Users/pablo/OneDrive/Escritorio/TFG/cobsINAR.csv", row.names = TRUE)

riINAR <- matrix(riINAR,ncol=2)

riINAR <- data.frame("Rang Hombres"=riINAR[1:10,1],"Rang Mujeres"=riINAR[1:10,2],row.names =
unique(dadesBCN3$SectorSanitariDescripcio))

write.csv(riINAR,"C:/Users/pablo/OneDrive/Escritorio/TFG/RangsINAR.csv", row.names = TRUE)

```