

MRLM COVID

Equipo R.E.J.O

24 May, 2023

Contents

1	Introducción	1
1.1	Problema de interés	1
2	Marco teórico	2
2.1	Conceptos básicos	2
2.2	Supuestos del modelo	2
2.3	Método de selección de variables	2
2.4	Limitaciones del modelo	3
3	Análisis exploratorio de datos	4
3.1	Análisis de la base de datos	4
3.2	Selección de la variable explicativa	4
4	Modelo de regresión lineal múltiple	7
4.1	Justificación de la selección del MRLM	7
4.2	Supuestos y validación del modelo	8
5	Predicciones	16
6	Conclusiones	17
6.1	Porqué es útil el modelo	17
6.2	Cómo mejorar el modelo	17
7	Bibliografía	17

1 Introducción

En marzo de 2020 la Organización Mundial de la Salud (OMS) elevó la crisis de salud pública ocasionada por el virus del COVID-19 al grado de pandemia, tan solo cuatro meses después de que el virus fuera identificado en la ciudad de Wuhan, China. Para ese momento había 118 mil casos confirmados en 114 países, y 4291 personas habían perdido la vida. El anuncio incitó a los países de alrededor del mundo a que instauraran medidas para mitigar el contagio y tratar a los enfermos. Algunos países, como Estados Unidos, habían movilizado sus sistemas de respuesta desde enero, no obstante, la pandemia tuvo efectos devastadores.

El Centro para el Control y la Prevención de Enfermedades (CDC por sus siglas en inglés), activó un plan de respuesta desde enero de 2020, priorizando la distribución de información y controles aeroportuarios. En los meses subsecuentes, adoptarían el uso general de pruebas, el aislamiento selectivo y el uso de mascarillas para evitar la propagación del contagio. No obstante, a pesar de las medidas insaturadas, no sería hasta agosto del siguiente año que la pandemia tocaría un punto de quiebre. El 14 de diciembre del 2020, la Administración de Alimentos y Medicamentos de los Estados Unidos (FDA por sus siglas en inglés) aprobaría la vacuna Pfizer-BioNTech para su uso en personas mayores de 18 años. Esta sería la primera vacuna contra el COVID-19 aprobada y representaría un avance fundamental en la estrategia de mitigación del patógeno.

Al 9 de septiembre de 2021, se han administrado 377,6 millones de dosis de vacunas. En general, aproximadamente el 62.7% de la población total de Estados Unidos, han recibido al menos una dosis de la vacuna, mientras que el 53.4% han sido completamente vacunados. Se espera inoculación de la población cambiará el comportamiento de los fallecimientos en los próximos meses. Ahora, al tratarse de una enfermedad estacionaria, los tomadores de decisiones en el sector salud deben hacer uso de la información recabada para anticiparse a las necesidades de los pacientes en temporada de COVID-19. Para ello, nuestra investigación propone aplicar un modelo de regresión lineal múltiple para evaluar la vacunación como un factor de predicción de las olas subsecuentes de COVID-19. Esto permitirá a los hospitales y demás servicios de salud programar costos asociadas a pruebas, tratamientos y medicamentos relacionados al patógeno SARS-Cov-2.

1.1 Problema de interés

Aplicación de un modelo de regresión lineal múltiple para evaluar la vacunación como un factor de predicción de una cuarta ola de muertes por COVID-19 en EUA, incluyendo la estimación a tiempo real de la reproducción efectiva del COVID y el porcentaje de pruebas COVID positivas que hay en la población.

1.1.1 Breve explicación de la base de datos

Glosario de Y y los regresores que utilizamos.

Nuestra variable a analizar [Y] - new_deaths: Son las nuevas muertes atribuidas al COVID-19.

Location: Es la ubicación geográfica, nos ayudó a filtrar nuestra base de datos sólo a USA

Date: Desde la primera vacunación recibida en EUA al final de la 4ta ola

Reproduction_rate: Estimación a tiempo real de la reproducción efectiva del COVID

Positive_rate: El porcentaje de pruebas COVID positivas que hay

New_vaccinations: Nuevas vacunas administradas de COVID 19

Nota: La información está tomando en cuenta los 7 días de la semana.

2 Marco teórico

2.1 Conceptos básicos

El MRLM es una técnica para modelar la relación lineal entre 2 o más variables. De manera general, el MRLM se define en la siguiente ecuación: $Y = B_0 + B_1X_1 + B_2X_2 + \dots + B_nX_n + U$

Donde Y es la variable dependiente y las X_i son las variables independientes, mientras que la variable u , denominada como la perturbación estocástica, son variables aleatorias no observables que representa a todos los factores distintos de X que afectan a Y.

En tanto B_0 , representa el coeficiente del intercepto y las B_i en adelante representan los coeficientes de cada pendiente. Los coeficientes de las pendientes son el interés principal del análisis econométrico, pues mide el efecto de las variables independientes sobre la variable dependiente manteniendo todos los demás factores constantes.

2.2 Supuestos del modelo

Hay algunos requisitos o supuestos que se deben cumplir para que el modelo sea válido y confiable. Estos son los cinco supuestos principales:

Independencia de los residuales: los residuos (las diferencias entre los valores observados y los estimados por el modelo) deben ser independientes entre sí. Esto significa que no hay una relación o patrón entre los errores del modelo. Una forma de asegurar la independencia es seleccionar los datos de forma aleatoria y no sesgada. no obstante esto se puede arreglar aplicandole un Lag a la variable Y.

Homocedasticidad: los residuos deben tener una varianza constante. Esto significa que la variabilidad de los errores del modelo no depende del valor de las variables independientes. $\text{var}(u_i) = \text{var}$

No multicolinealidad: las variables independientes no deben estar relacionadas entre sí o, al menos, su relación debe ser muy débil. Esto significa que no hay redundancia o duplicidad en la información que aportan las variables independientes al modelo. Se puede comprobar que no hay con el Variance Inflation Coefficient (VIF)

Normalidad: los residuos deben seguir una distribución normal, es decir, una distribución simétrica y acampanada alrededor del cero. Esto significa que los errores del modelo son aleatorios y no tienen tendencia o sesgo. Una forma de comprobar la normalidad es mediante un gráfico de probabilidad normal o una prueba estadística.

Valores atípicos: No deben existir datos atípicos ya que eso tiende a sesgar la regresión lineal. Cuando existe una serie de tiempo, como es el caso, se debe aplicar un Cap y un floor para que los datos no se pasen de esos datos.

2.3 Método de selección de variables

Al momento de seleccionar variables, eliminamos variables que explicaban lo mismo pero tenían algún tipo de transformación lineal. Un ejemplo de estas son `new_vaccinations` y `new_vaccinations_per_million`.

Eliminamos variables que eran constantes y variables que fueran de carácter, fuera de location ya que la usamos para filtra nuestra base de datos.

Checamos de las variables restantes con gráficos de `ggpairs`, y vimos las relaciones que tenía `new_deaths` con las variables restantes, dejando en nuestro modelo las variables que tuvieran una relación lineal respecto a nuestra Y

Al meter las variables elegidas al modelo, vimos que quedaban algunas que tenían un Pvalue mayor a .05 por lo que esas variables también fueron eliminadas del modelo.

Ya por último, con la selección restante, utilizamos el Variance Inflation Factor (VIF), su expresión matemática es $(1/(1-R_i^2))$, el cual nos ayuda a cuantificar la intensidad que hay de multicolinealidad entre los regresores. Aceptamos Variables con VIF menores de 10 en el modelo.

2.4 Limitaciones del modelo

Usar un modelo de regresión lineal múltiple como modelo de predicción tiene algunas limitaciones que debemos tener en cuenta. Estas son algunas de ellas:

No siempre hay una relación lineal entre las variables: puede ser que las variables dependientes e independientes no tengan una relación lineal, sino curvilínea o no lineal. En ese caso, el modelo de regresión lineal múltiple no sería adecuado para describir o predecir la variable dependiente.

No se puede establecer causalidad a partir de la correlación: puede ser que las variables independientes estén correlacionadas con la variable dependiente, pero eso no significa que sean la causa de su variación. Puede haber otras variables que no se hayan incluido en el modelo y que sean las verdaderas causas de la variable dependiente.

No se pueden incluir demasiadas variables independientes en el modelo: puede ser que al incluir muchas variables independientes en el modelo se produzca un problema de multicolinealidad, es decir, que las variables independientes estén relacionadas entre sí y aporten información redundante o innecesaria al modelo. Esto puede afectar a la precisión y la estabilidad de los coeficientes del modelo y dificultar su interpretación. De igual forma, puede generarse un sobre ajuste, el cual haría que el modelo no pueda reproducirse en diferentes poblaciones.

Una gran limitación con la que nos encontramos en el MRLM es que aun cuando los datos no se distribuyen normal, asumimos una distribución normal para hacer inferencia. El MRLM es robusto para variables que no se distribuyen normal, no obstante, eso nunca va a ser un modelo exacto apegado a la realidad

3 Análisis exploratorio de datos

3.1 Análisis de la base de datos

De la librería de `tidyr`, utilizamos la función `filter`, para filtrar la base de datos con los datos de USA.

Utilizamos de la librería `dplyr` la función de `select`, para eliminar las variables que la selección backward nos quitó, ya que de esa forma las eliminábamos de la base a modelar.

Debido a las fechas que elegimos no tuvimos que hacer ningún tratamiento de NA's, ya que hubo datos en todas las variables que utilizamos para todas las fechas que analizamos

Seleccionamos `new_deaths` ya que es una variable que nos explica cuantas personas fueron muriendo debido al COVID

La base de datos viene de Data on COVID-19 (coronavirus) por Our World in Data, tiene perfiles de 207 países y monitorea 67 variables de forma diaria a partir del 3 de enero de 2020.

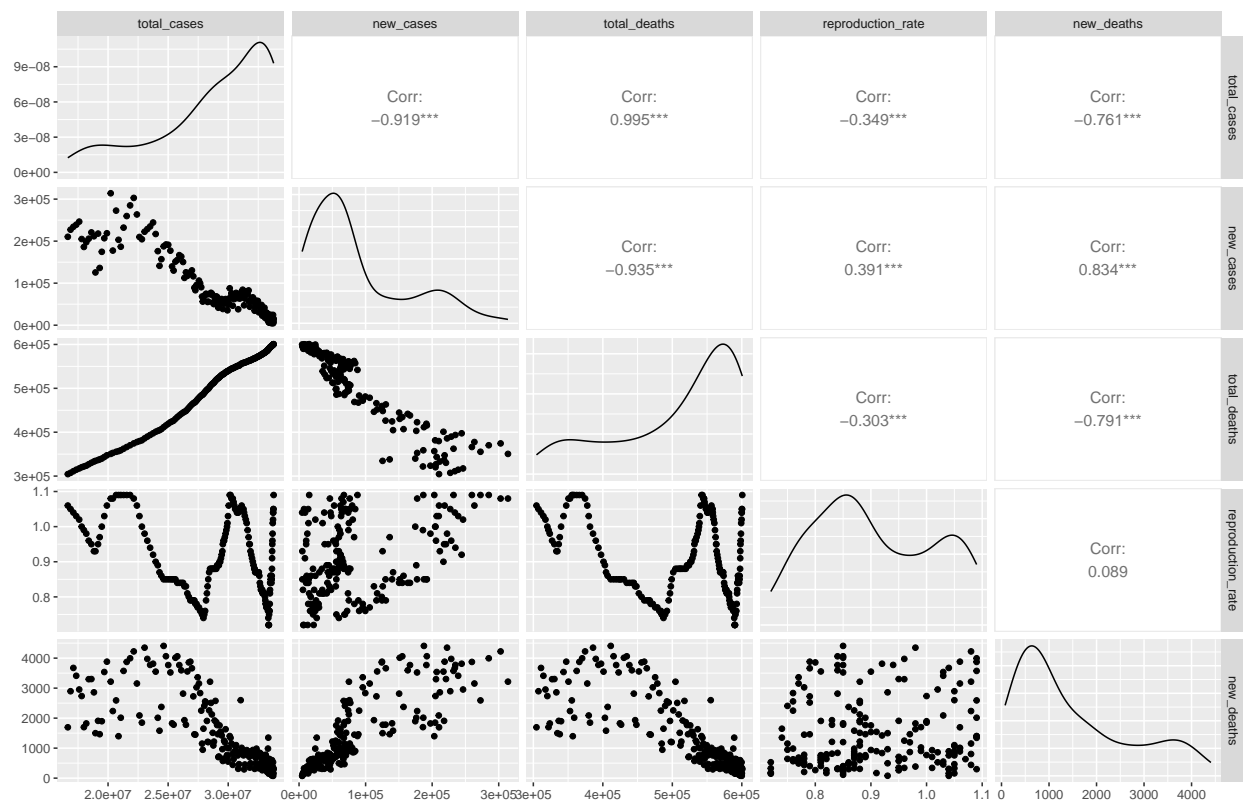
3.2 Selección de la variable explicativa

Dado el enfoque que le quisimos dar al modelo, que es: Situarnos en el punto histórico de las 3 primeras olas de COVID-19, para poder predecir una cuarta ola de COVID y en específico las muertes que esto provocaría. Quisimos observar la influencia de las vacunas en disminuir las muertes por COVID-19. Después de acotar las posibles variables explicativas, con el proceso que anteriormente explicamos en este documento, nos propusimos a usar las que también tuvieran sentido en nuestra realidad.

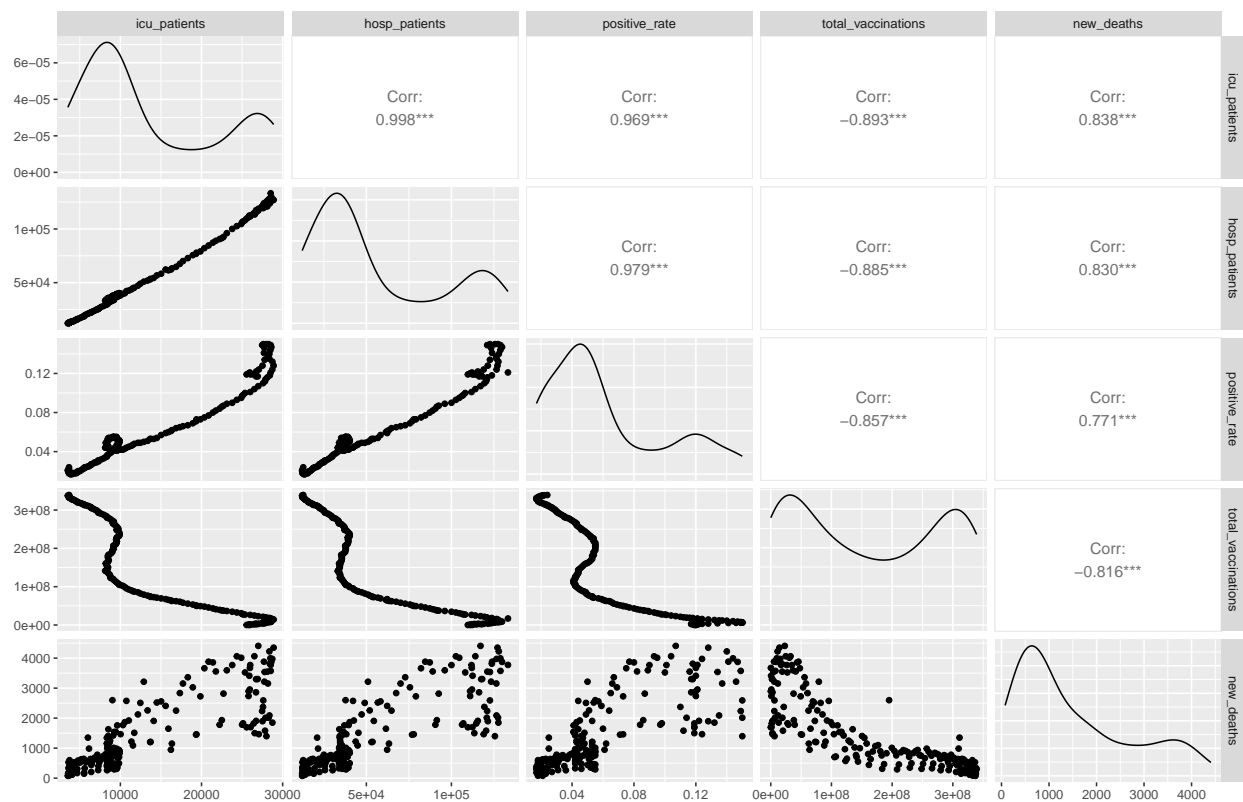
Es por eso, que escogimos usar las nuevas personas vacunadas, la velocidad de propagación de COVID-19 y el porcentaje de la población que dieron positivos. Usando estas variables queremos saber si las políticas relacionadas a las campañas de vacunación tuvieron un impacto positivo en las tasas de muertes

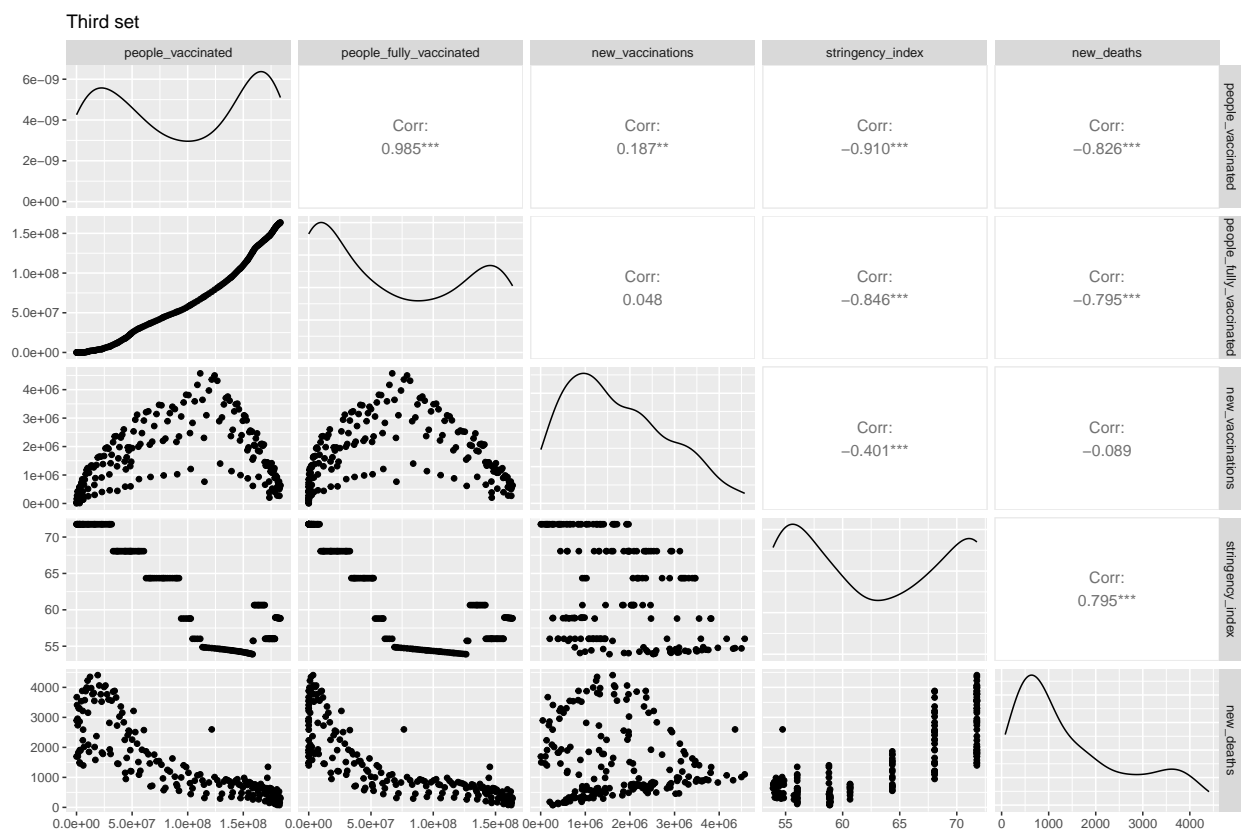
Se pueden observar las variables y su linealidad en los siguientes gráficos de correlación:

First set



Second set





Como se ve en los ggpairs, las variables con relación lineal respecto a las nuevas muertes son: Con relación positiva: icu_patients, hosp_patients, new_cases Relación negativa: total_vaccinations, people_vaccinated

Las nuevas vacunas no tienen una relación establecida con el total de muertes

Call:

```
lm(formula = new_deaths ~ +people_vaccinated + positive_rate +
    reproduction_rate + new_vaccinations + hosp_patients - date,
    data = train_lm_df)
```

Residuals:

Min	1Q	Median	3Q	Max
-1885.52	-275.59	-1.09	284.98	1321.05

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-6.470e+01	6.243e+02	-0.104	0.9176
people_vaccinated	-3.138e-06	1.509e-06	-2.079	0.0389 *
positive_rate	-4.095e+04	8.514e+03	-4.810	3.04e-06 ***
reproduction_rate	5.231e+02	6.248e+02	0.837	0.4035
new_vaccinations	2.472e-04	3.991e-05	6.193	3.48e-09 ***
hosp_patients	6.224e-02	8.884e-03	7.006	3.99e-11 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
Residual standard error: 547.5 on 193 degrees of freedom
Multiple R-squared:  0.7962,    Adjusted R-squared:  0.7909
F-statistic: 150.8 on 5 and 193 DF,  p-value: < 2.2e-16
```

Estas fueron nuestras variables iniciales, no obstante por el modelo anterior, rechazamos aquellas que tuvieran un $p\text{-value} > .05$ ya que no eran significativas para nuestro modelo.

Eso nos dejó con `positive_rate`, `reproduction_rate` y `new_vaccinations`

4 Modelo de regresión lineal múltiple

4.1 Justificación de la selección del MRLM

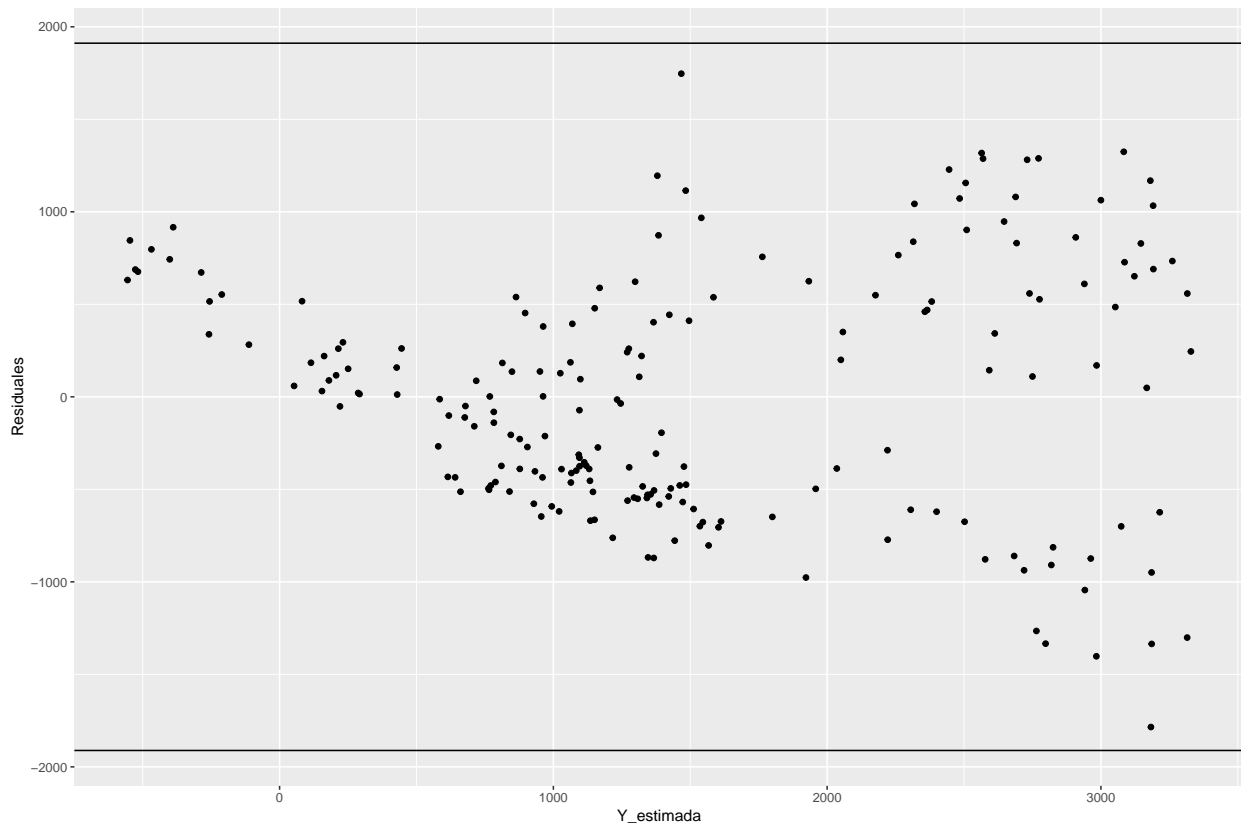
El uso de un MRLM está justificado por las siguientes razones:

1. Control de variables: Al incluir múltiples variables predictoras en el modelo, podemos controlar el efecto individual de cada una de ellas en la variable de respuesta, manteniendo las demás variables constantes. Esto nos ayuda a comprender el impacto específico de cada variable y realizar inferencias más precisas.
2. Mayor precisión: Al considerar múltiples variables predictoras, el MRLM puede proporcionar estimaciones más precisas de la relación entre las variables independientes y la variable dependiente. Al capturar las interacciones entre las variables dependientes, el modelo puede explicar mejor la variabilidad observada en los datos y reducir los errores de predicción.
3. Variables no explicativas eliminables: En situaciones donde existen variables que pueden afectar tanto a la variable de respuesta como a las variables dependientes, el MRLM puede controlar y ajustar los efectos de estas variables. Esto nos permite obtener una interpretación más precisa de las relaciones entre las variables de interés.
4. Predicciones multivariantes: Si nuestro objetivo es realizar predicciones basadas en múltiples variables predictoras, el MRLM es una elección adecuada. El modelo puede combinar la información de varias variables para predecir el valor de la variable de respuesta, lo que resulta útil en escenarios de pronóstico y toma de decisiones.

4.2 Supuestos y validación del modelo

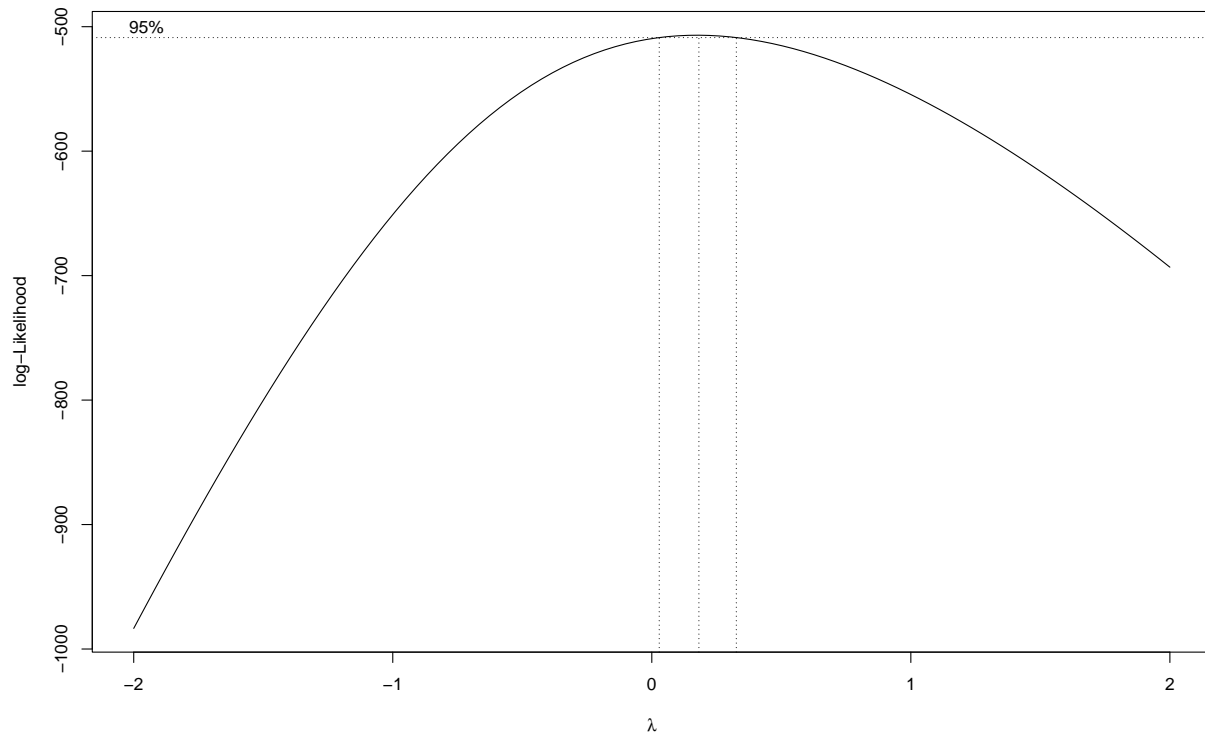
4.2.1 Análisis de residuales

4.2.1.1 Heterocedasticidad Comprobamos homocedasticidad (la varianza de los errores es constante), lo comprobamos con un gráfico comparando los residuales con las Y observadas (\hat{y}), para esto tenemos que hacer un DF con ambos vectores obtenidos de nuestro modelo



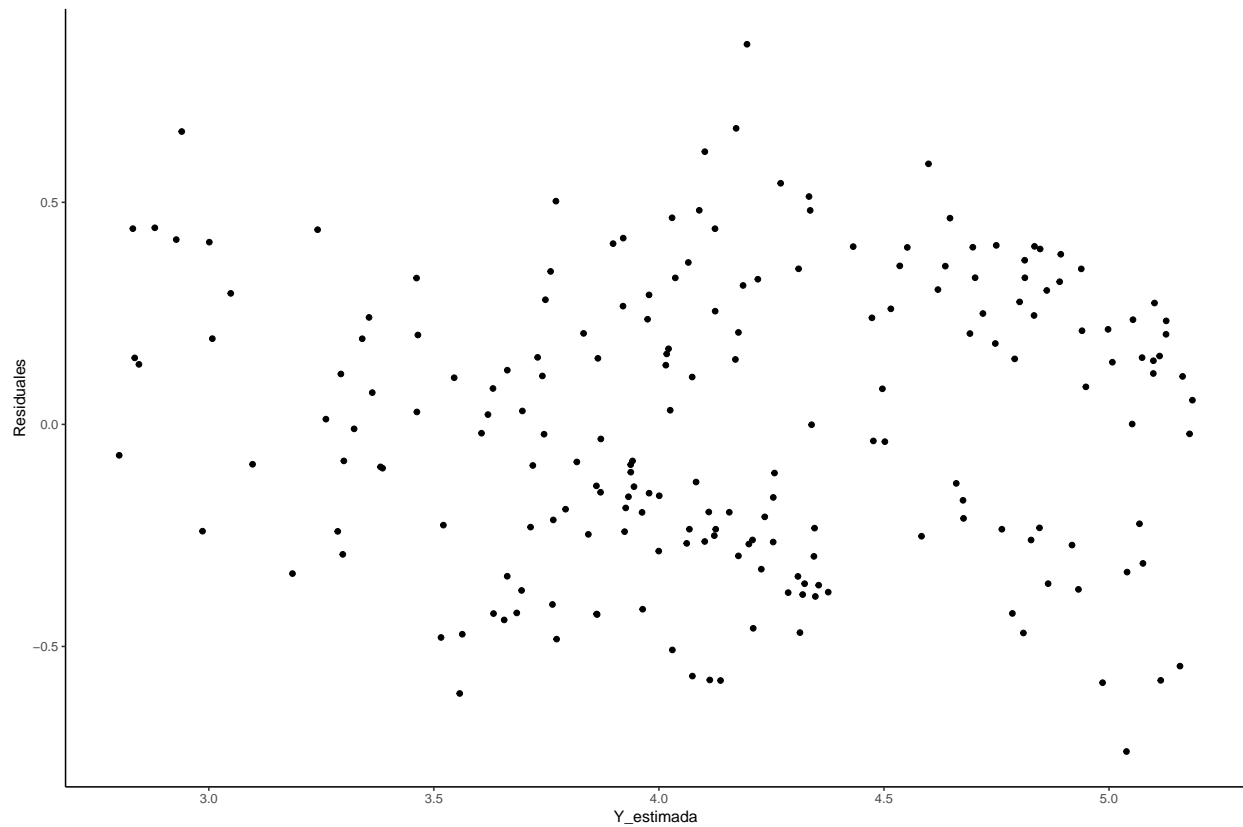
Se puede observar que hay un patrón como un cono de izquierda a derecha, por lo que no podemos afirmar que existe homocedasticidad. Lo que nos lleva a hacer una transformación Box-Cox, la cual es hacer una transformación a Y dependiendo de la lambda obtenida.

Obtenemos el gráfico en el que obtenemos la lambda donde se maximiza la verosimilitud



y se obtiene una lambda maximizadora de verosimilitud de: [1] 0.2

Hacemos la transformación de nuestra Y elevandola a la lambda $[Y^{\wedge}\text{Lambda}]$ y hacemos un modelo nuevo con esta Y, en el que vamos a verificar que se cumpla la homocedasticidad. La cual se observa en el siguiente gráfico.



Ya no se observa un patrón establecido en el gráfico por lo que continuamos con la validación de la significancia de nuestros regresores.

Comprobamos significancia con la Anova, el VIF y el summary de nuestro nuevo modelo:

Analysis of Variance Table

Response: new_deaths_box_cox

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
positive_rate	1	55.156	55.156	529.110	< 2.2e-16 ***
reproduction_rate	1	4.848	4.848	46.508	1.117e-10 ***
new_vaccinations	1	10.820	10.820	103.792	< 2.2e-16 ***
Residuals	195	20.327	0.104		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Call:

```
lm(formula = new_deaths_box_cox ~ positive_rate + reproduction_rate +
    new_vaccinations, data = train_lm_df)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.73657	-0.25117	-0.00078	0.25759	0.85596

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.605e+00	2.014e-01	22.864	<2e-16 ***
positive_rate	1.841e+01	7.108e-01	25.903	<2e-16 ***
reproduction_rate	-2.165e+00	2.408e-01	-8.992	<2e-16 ***
new_vaccinations	2.271e-07	2.229e-08	10.188	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3229 on 195 degrees of freedom

Multiple R-squared: 0.777, Adjusted R-squared: 0.7736

F-statistic: 226.5 on 3 and 195 DF, p-value: < 2.2e-16

positive_rate	reproduction_rate	new_vaccinations
1.377095	1.253100	1.168487

Nuestros regresores siguen siendo significativos, ya que todos siguen dando un p-value menor a .05, tanto en la ANOVA como en el summary, por lo que podemos continuar con las siguientes comprobaciones de supuestos.

4.2.1.2 Independencia en los errores Al ser una serie de tiempo, sí hay que buscar la independencia de los errores, la cual nos dice que $\text{corr}(e_i, e_{i-1}) = 0$. Para esto utilizamos la prueba Durbin-Watson, en la que si está de $[0, dl]$ u $[4-dl, 4]$, se rechaza el coeficiente, cuando está entre $[dl, du]$ u $[4dl, 4du]$, queda inconcluso el dato y podemos continuar con los supuestos, y si está $[du, 4-du]$, se acepta la independencia de los errores.

Durbin-Watson test

```
data: modelo_box_cox
DW = 0.68326, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is not 0
```

Al correr durbin Watson nos da un coeficiente que sí entra en el area de rechazo, por lo que tenemos que no hay independencia de los errores y tenemos que arreglarla.

La forma de arreglar el DW test es aplicando un lag en nuestra Y, pero como ya tenemos la Y transformada, esa es la que se tiene que laggear, no la que no está transformada.

Al momento de laggear una fecha lo que se hace es que el dato de nuestro primer día, se va al segundo, dejando un NA a esa fecha, y después de eso eliminamos el NA, basicamente acortando nuestra base de datos por un día.

Eso nos da un nuevo modelo en el que metemos el lag como un nuevo regresor.

Checamos que se siga cumpliendo la significancia del modelo

Call:

```
lm(formula = new_deaths_box_cox ~ positive_rate + reproduction_rate +
    new_vaccinations + lag_muertes, data = train_lm_df)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.66318	-0.16538	-0.00164	0.14003	0.64670

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.124e+00	2.793e-01	7.604	1.23e-12 ***
positive_rate	9.259e+00	1.016e+00	9.112	< 2e-16 ***
reproduction_rate	-1.094e+00	2.155e-01	-5.078	8.97e-07 ***
new_vaccinations	1.493e-07	1.914e-08	7.801	3.76e-13 ***
lag_muertes	5.293e-01	4.887e-02	10.830	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2559 on 193 degrees of freedom

Multiple R-squared: 0.8611, Adjusted R-squared: 0.8583

F-statistic: 299.2 on 4 and 193 DF, p-value: < 2.2e-16

Analysis of Variance Table

Response: new_deaths_box_cox

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
positive_rate	1	55.307	55.307	844.305	< 2.2e-16 ***
reproduction_rate	1	4.727	4.727	72.164	5.28e-15 ***
new_vaccinations	1	10.693	10.693	163.229	< 2.2e-16 ***
lag_muertes	1	7.683	7.683	117.289	< 2.2e-16 ***
Residuals	193	12.643	0.066		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

positive_rate	reproduction_rate	new_vaccinations	lag_muertes
4.425362	1.580442	1.355002	3.294875

Se sigue cumpliendo la significancia de los regresores, por lo que podemos continuar con el nuevo modelo

Durbin-Watson test

data: modelo_box_cox_updated

DW = 1.4615, p-value = 5.086e-05

alternative hypothesis: true autocorrelation is not 0

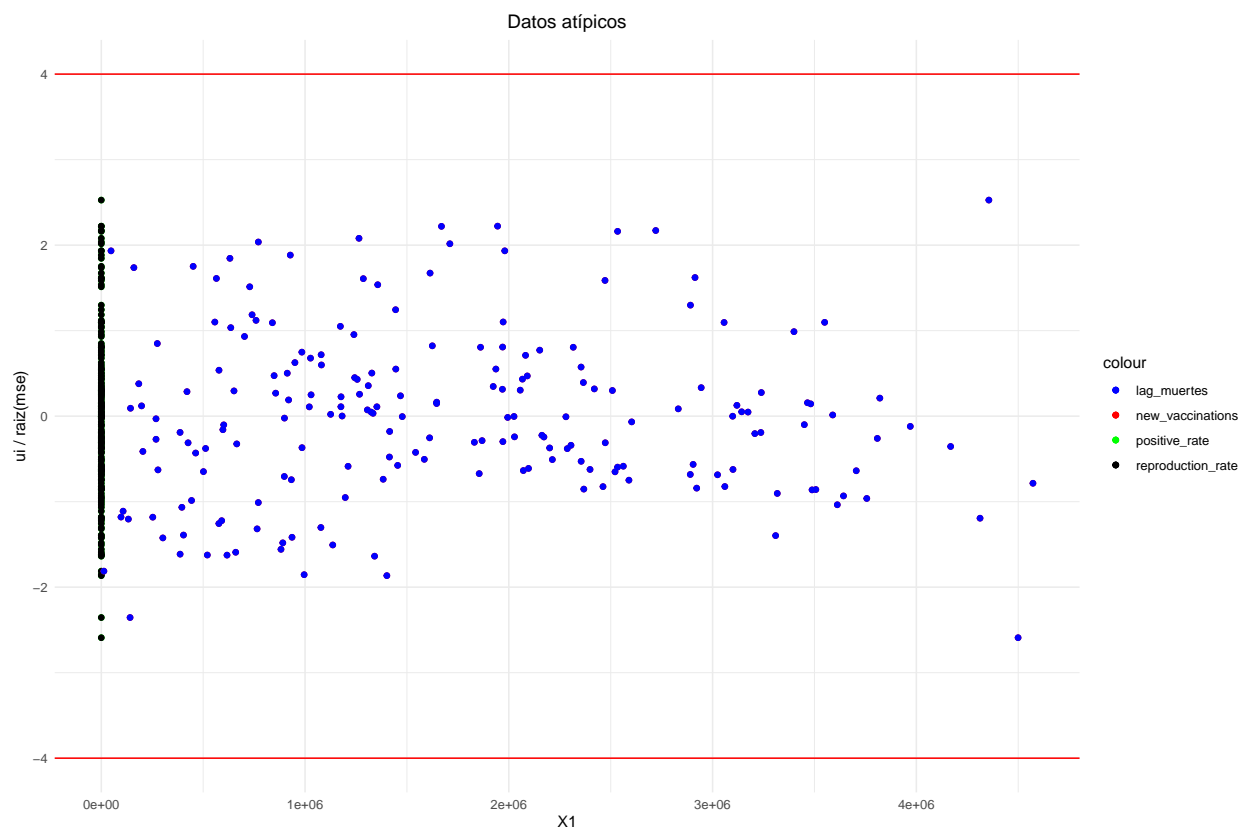
El coeficiente DW, queda en el area inconclusa de la recta de DW, por lo que vamos a continuar con nuestras variables regresoras.

4.2.1.3 Presencia de errores atípicos Por la construcción del modelo de Mínimos Cuadrados Ordinarios (MCO) los datos atípicos pueden desviar la recta estimada de regresión. Esto puede generar una mala interpretación de lo que sucede en un proceso o población.

Por lo tanto, es importante identificar estos errores y tratarlos de forma adecuada. Lo ideal no es ignorarlos, ya que esto puede representar una distorsión del análisis que estamos haciendo.

Para identificarlos, usamos el siguiente gráfico, si alguno de los datos rebasa el valor de 4 o -4 se considera un dato atípico, usamos estos valores porque está aceptado que fuera de ese rango, se determina un dato atípico. Esos datos atípicos los descartamos de nuestro modelo única y exclusivamente cuando representan un error en el registro de datos.

Este grafico está construido de la siguiente forma; en el eje de las Y's se encuentran los valores de cada error divididos por la raíz del MSE; en el eje de las X's se encuentran los datos de cada regresor, es decir el error que se genera por cada valor de los regresores, el -4 y 4 son desviaciones estandar, se considera que un dato que esté 4 desviaciones estandar lejos de la media es un dato muy poco probable de que ocurra, por lo que se puede acotar a un cap&floor, o eliminar.



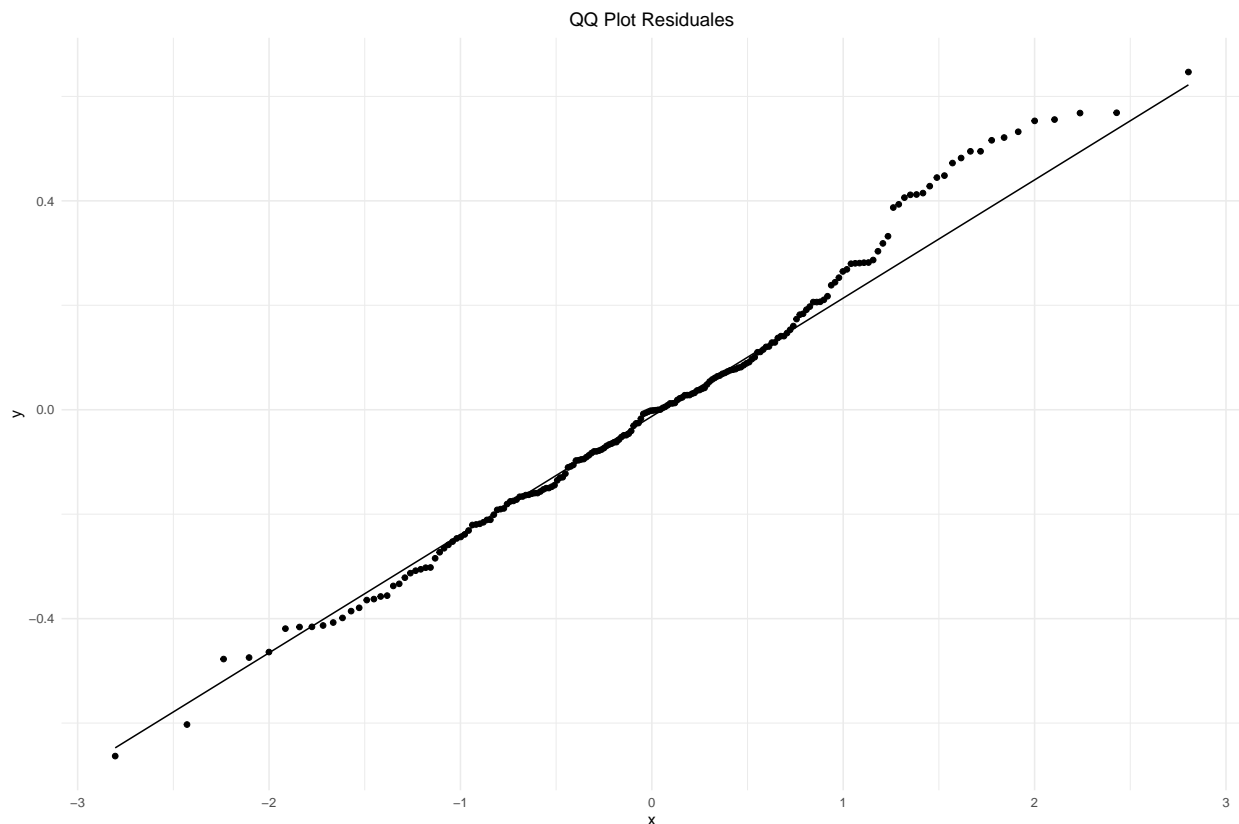
Como podemos observar en este gráfico, todos los errores se encuentran dentro del intervalo $[-4,4]$, con esto sabemos que no contamos con datos atípicos en el modelo.

4.2.1.4 Comprobación de la linealidad de la Fn de regresión Eso lo comprobamos con las R^2 ya que sabemos que matemáticamente es SSE/SST y es el porcentaje de variabilidad que es explicada por el modelo.

Una R^2 ajustada de [1] 0.8611497

nos dice que hay una bondad de ajuste del 86% de nuestro modelo con Y, lo cual nos dice que sí es lineal la fn de regresión.

4.2.1.5 Verificación de la normalidad en los errores La mejor forma de comprobar la normalidad de los errores es con una prueba de QQ plot, o gráfica cuantil cuantil. En este método se comparan los valores que deberían tomar la normal y la distribución que este tomando la muestra.



Si los datos de nuestro modelo se comportasen de manera normal, los puntos se alinearían a lo largo de la línea diagonal. Como se puede apreciar en la siguiente gráfica Q-Q plot, los datos tienen un sesgo a la derecha, por lo que se viola el supuesto de normalidad de los errores.

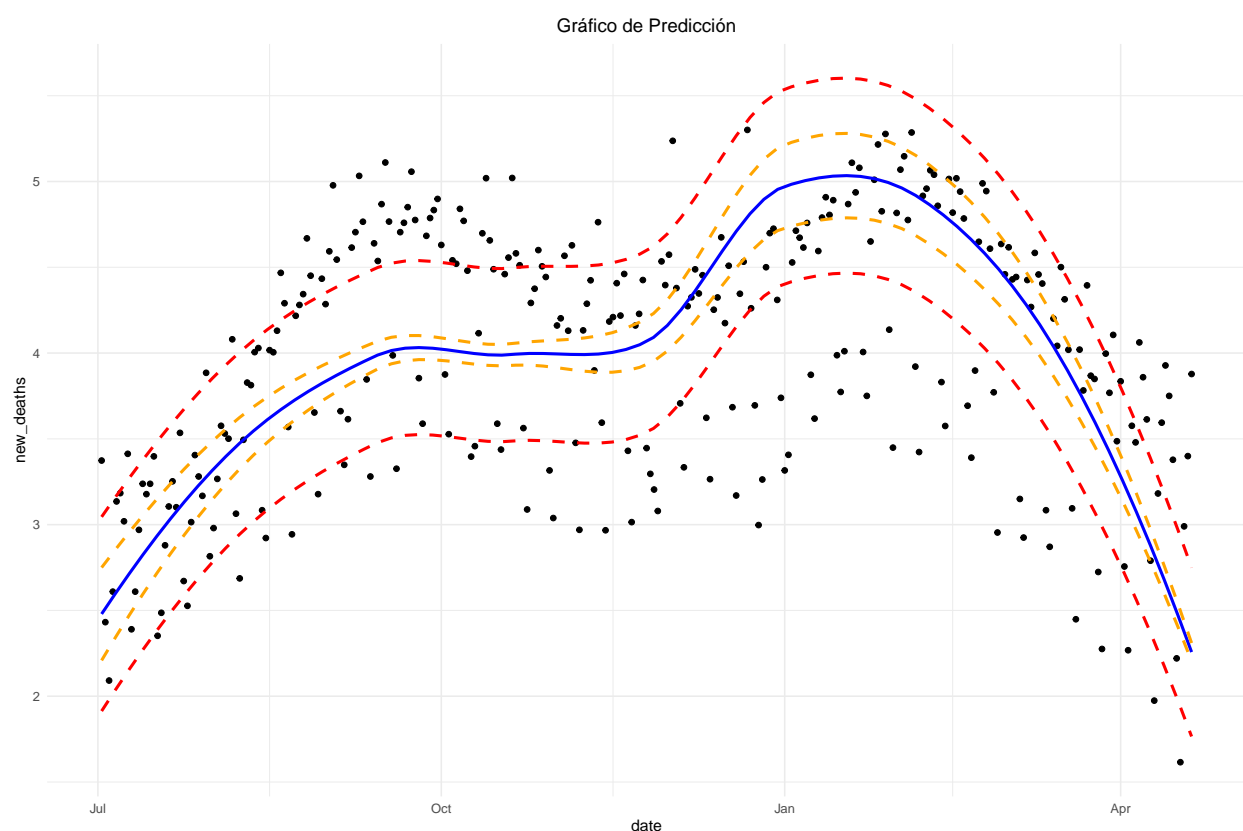
Dado que el MRLM es robusto ante la normalidad podemos utilizar nuestros resultados para hacer inferencia estadística. Esto se debe a que el método empleado para calcular los coeficientes de regresión del modelo, el MCO, busca minimizar la suma de los errores cuadrados entre los valores observados y los valores predichos.

5 Predicciones

Ya que nuestro modelo cumple con los supuestos necesarios, podemos utilizarlo para predecir la base de validación. Nuestra base de validación es de la ola 3, 2021-07-01, al final de las olas en USA, 2022-04-21.

Cuando utilizas la población de validación, necesitas aplicarle la lambda que se utilizó en la transformación Box-Cox y también es necesario ponerle el lag a esta nueva base, ya que si no se podría validar.

Una vez que le hacemos las transformaciones a la base de validación, hacemos los intervalos de predicción y de confianza, el de confianza es para $E[Y/X's]$ y los de predicción buscan meter a todas las observaciones de Y .



Se puede apreciar que nuestra base de entrenamiento logra predecir de forma correcta la subida y bajada de los datos reales, por lo que la predicción es correcta. Se observa que a partir de la fecha en la que inicia la base de validación va a haber un aumento en muertes, y se ve que van a disminuir hasta que termine abril. Esto es muy cercano a lo real en la base de datos que tenemos de validación, lo que se ve con los puntos negros.

6 Conclusiones

6.1 Porqué es útil el modelo

Al producir predicciones basadas en evidencia sobre el comportamiento de una cuarta ola por COVID-19, nuestro modelo ofrece oportunidades al sector salud en general, tanto a los oferentes de los servicios de salud, como a los demandantes. Por el lado de la oferta, el modelo ayudará a hacer más eficiente el uso de recursos, programando correctamente costos asociados al tratamiento del COVID, designando un número adecuado de camas de hospital para pacientes que necesiten ser internados y evitando pérdidas relacionadas a un exceso o deficiencia de insumos.

Esto también ayudará al público en general, pues un ajuste de la demanda de insumos médicos podría desencadenar un reajuste en los precios del mercado. No obstante, la mayor contribución al bienestar de los pacientes radica en el hecho de los hospitales contarán con los recursos necesarios para atender los casos más graves, salvando a la mayor cantidad de personas.

Por último, este modelo busca apoyar al personal de nuestro sistema de salud, asegurándose de que cuenten con los materiales necesarios para cuidar tanto de los pacientes, como de ellos mismos.

6.2 Cómo mejorar el modelo

En el futuro, nos gustaría que nuestro modelo fuese reentrenado con información adicional que refleje el ritmo de vacunación. También, el acceso a información más detallada nos daría la oportunidad de proveer información a nivel local, de manera que pueda realizar predicciones más precisas sobre mercados de salud subnacionales. Así mismo, en el futuro nos gustaría reducir las áreas de oportunidades al ampliar el espectro de información que sirve de base a nuestro modelo. En concreto, nos enfocaremos en dos áreas de oportunidades para hacer las predicciones más precisas, que es ampliar la información a otros mercados nacionales y poder discriminar entre cepas del virus, que se ha demostrado tienen comportamientos distintos.

7 Bibliografía

Edouard Mathieu, Hannah Ritchie, Lucas Rodés-Guirao, Cameron Appel, Charlie Giattino, Joe Hasell, Bobbie Macdonald, Saloni Dattani, Diana Beltekian, Esteban Ortiz-Ospina and Max Roser (2020) - "Coronavirus Pandemic (COVID-19)". Published online at OurWorldInData.org. Retrieved from: '<https://ourworldindata.org/coronavirus>' [Online Resource]

COVID-19 Map. (s.f.). Johns Hopkins Coronavirus Resource Center. Recuperado 24 de mayo de 2023, de <https://coronavirus.jhu.edu/map.html>

Times, T. N. Y. (2020, marzo 3). Coronavirus in the U.S.: Latest Map and Case Count. The New York Times. <https://www.nytimes.com/interactive/2023/us/covid-cases.html>

CDC Museum COVID-19 Timeline. (2023, marzo 15). Centers for Disease Control and Prevention. <https://www.cdc.gov/museum/timeline/covid19.html>