RCG

American Hospital Association™

Advancing Health in America

Claves para la próxima ola.

Aplicación de un modelo de regresión lineal múltiple para evaluar la vacunación, el ratio positivo de pruebas COVID y la estimación real de reproducción del COVID como factores de predicción de la cuarta ola de infecciones por COVID-19.

Julio 2021

The REJO Consulting Group

Contexto del Trabajo

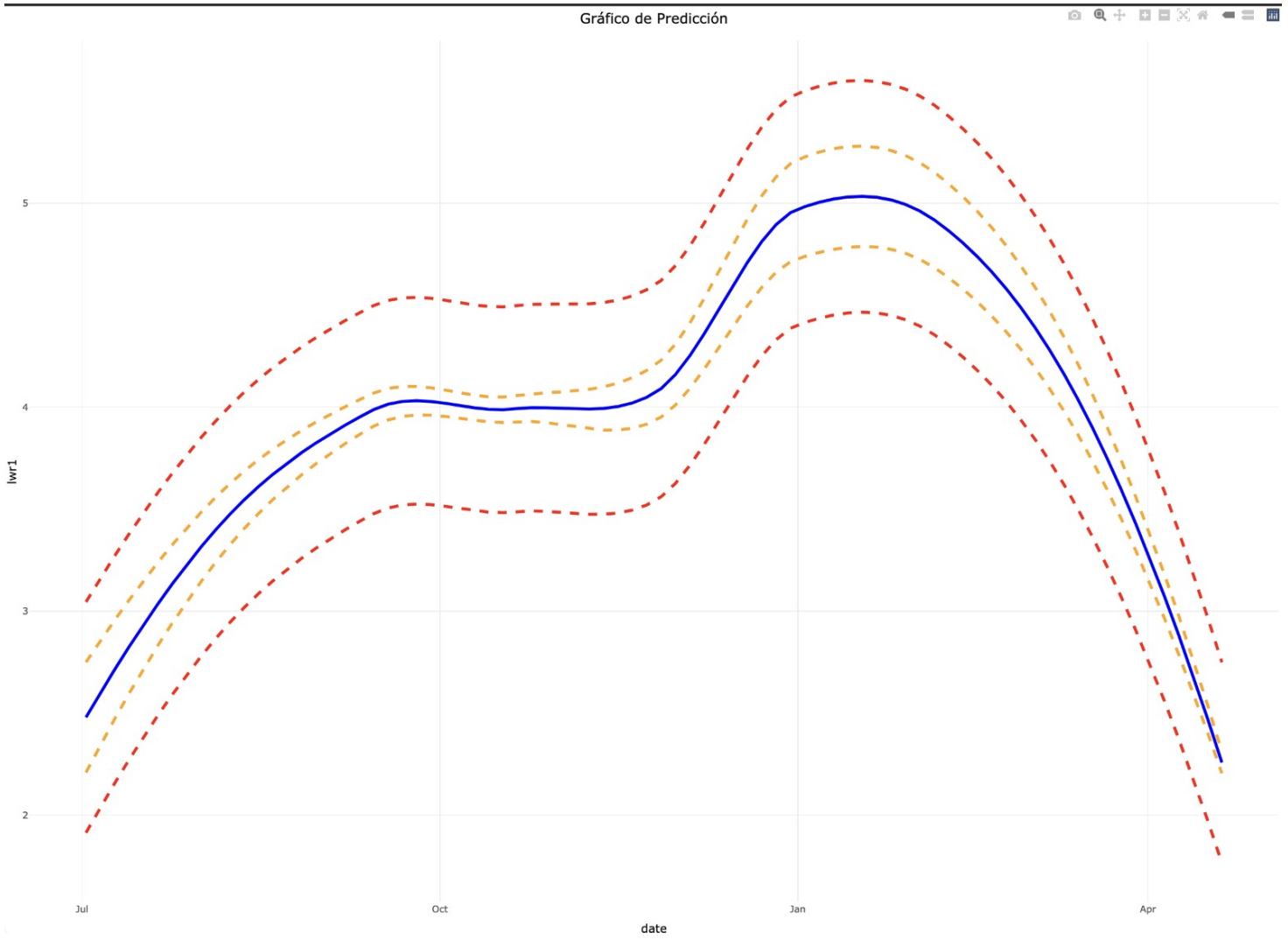
En marzo de 2020, la Organización Mundial de la Salud (OMS) declaró al COVID-19 como una pandemia. Para ese momento había 118 mil casos confirmados en 114 países, y 4291 personas habían perdido la vida. Al día de hoy, sólo en Estados Unidos, hay 40 millones de casos confirmados y más de 650 mil personas han fallecido. No obstante, hay signos de que la situación cambiará.

El 14 de diciembre de 2020, Estado Unidos administró la primera dosis de vacuna contra el COVID-19. Al 01 de junio de 2021, aproximadamente el 66% de la población adulta de Estados Unidos, han recibido al menos una dosis de la vacuna, mientras que el 44.56% han sido completamente vacunados. Se espera inoculación de la población cambiará el comportamiento de los fallecimientos en los próximos meses, pero por ahora el COVID sigue siendo la séptima causa de muerte en los Estados Unidos.

La American Hospital Association (AHA) abordó a RCG en 2021 para evaluar el efecto que tendrá la vacunación en mitigar las muertes en una cuarta ola por COVID-19 y ayudarlos a entender cómo esta información puede ayudarlos a reducir ineficiencias y destinar los recursos necesarios para el ala de COVID.

Este documento es un resumen de los resultados obtenidos por RCG haciendo un análisis de regresión lineal múltiple sobre una base de datos de datos de fuente abierta.

Predicciones



Como base de entrenamiento utilizamos los datos desde el primer vacunado, 2020-12-14, hasta el día de hoy, 2021-07-01.

Dado los datos analizados **predecimos una cuarta ola de muertes entre junio 2021 y abril de 2022,** con un aumento mayor en enero ya que se comportará de manera muy similar a las anteriores.

No será hasta **abril** que los efectos de la vacunación comiencen a mitigar las muertes por COVID-19.

Selección del modelo y análisis

Las ventajas en el uso del MRLM nos permite controlar las variables, obtener una mayor precisión, ajustar y controlar variables que puedan causar problemas con el modelo, y tener un modelo de predicción multivariable.

Los coeficientes de nuestras variables nos ayudan a explicar la relación (positiva o negativa) con nuestra variable número de muertes. La intersección con el eje de las Y es positiva. La relación que existe entre el número de pruebas positivas, la tasa de reproducción efectiva es positiva. El coeficiente de vacunas aplicadas es casi cero, por eso asumimos que en el corto plazo no tendrá efecto para contrarrestar el número de muertes por Covid. Por último, el lag de número de muertes fue utilizado para cumplir con el supuesto de independencia de los errores.

Y = 2.124 + 9.259 (positive_rate) - 1.094 (reproduction_rate) + 1.493X10^-7 (new_vaccinations) + 0.5293 (lag_muertes)

Base de datos

Data on COVID-19 (coronavirus) por Our World in Data

- El objetivo es monitorear el impacto de la pandemia e identificar los países que han tenido el mejor progreso mitigandola.
- Perfiles de **207 países**
- Monitorea **67 variables** de manera diaria desde 3 de enero de 2020

Variables que seleccionamos en el modelo

Nuestra variable a analizar: **new_deaths** (nuevas muertes atribuidas al COVID-19).

- No todos las personas que fallecieron estuvieron internadas, de acuerdo a estudios sólo el 13.4% de los hospitalizados fallecieron, pero los casos fatales nos dan una idea de la demanda real de hospitalización.

Variables explicativas :

Reproduction_rate (estimación a tiempo real de la reproducción efectiva del COVID-19).
Positive_rate (el porcentaje de pruebas COVID positivas).
New_vaccinations (Nuevas vacunas administradas de COVID-19)

Análisis Exploratorio de los datos y selección de variable

Partiendo de la base de datos del COVID-19, filtramos los datos correspondientes a Estados Unidos, que es nuestro país de interés. Después seleccionamos nuestra variable dependiente, new_deaths, es una variable de interés que explica el número de muertes que hubo debido al COVID-19.

Para seleccionar las variables explicativas, eliminamos las que explicaban lo mismo, pero tenían algún tipo de transformación lineal. Por ejemplo, new_vaccinations y new_vaccinations_per_million. De igual forma, eliminamos variables que fueran constantes o de carácter.

Nuestro siguiente paso fue observar que variables tenían una relación lineal con nuestra variable de interes, utilizamos gráficos de ggpairs para esto.
Utilizamos otros dos filtros para mantener las mejores variables explicativas posibles, que contarán con un P-value menor a 0.5 y que el Variance Inflation Factor (VIF) fuera menor a 10.

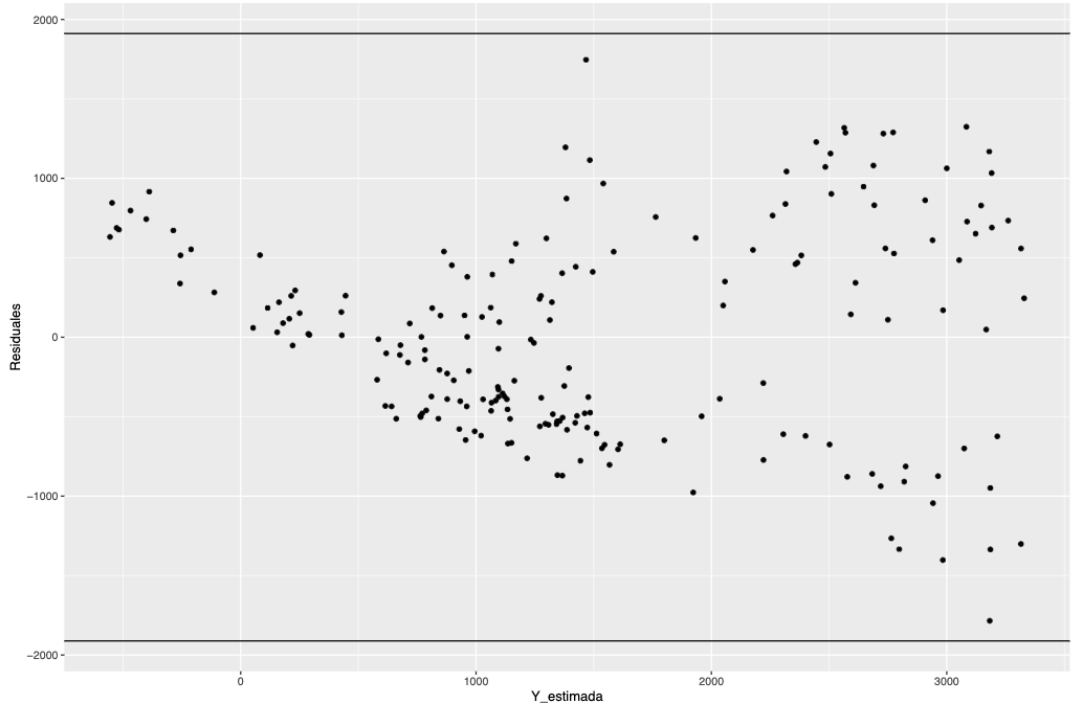
Como resultado, nuestras variables explicativas son las siguientes: positive_rate, reproduction_rate y new_vaccinations.

Validación de supuestos

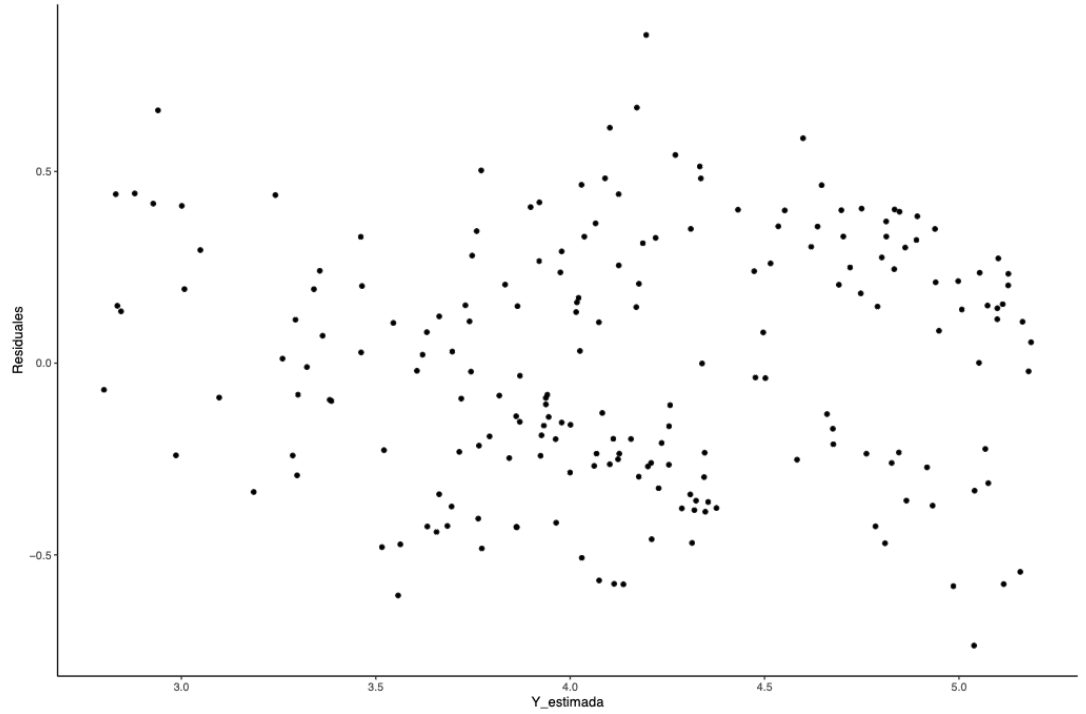
Los supuestos que debe cumplir un MRLM son:

- Homocedasticidad
- Independencia de los errores
- Errores atípicos
- Linealidad de la fn respecto parámetros
- No multicolinealidad
- Normalidad en los residuos
- E[ui] = 0, Esto se cumple teniendo en el modelo a B0
- Numero de observaciones > K+A (parámetros)

Validación de supuestos
Homocedasticidad



Box-Cox
Lambda = .2
Transformación de Y^.2



Para obtener nuestra Y real, elevas la Y transformada a la 5

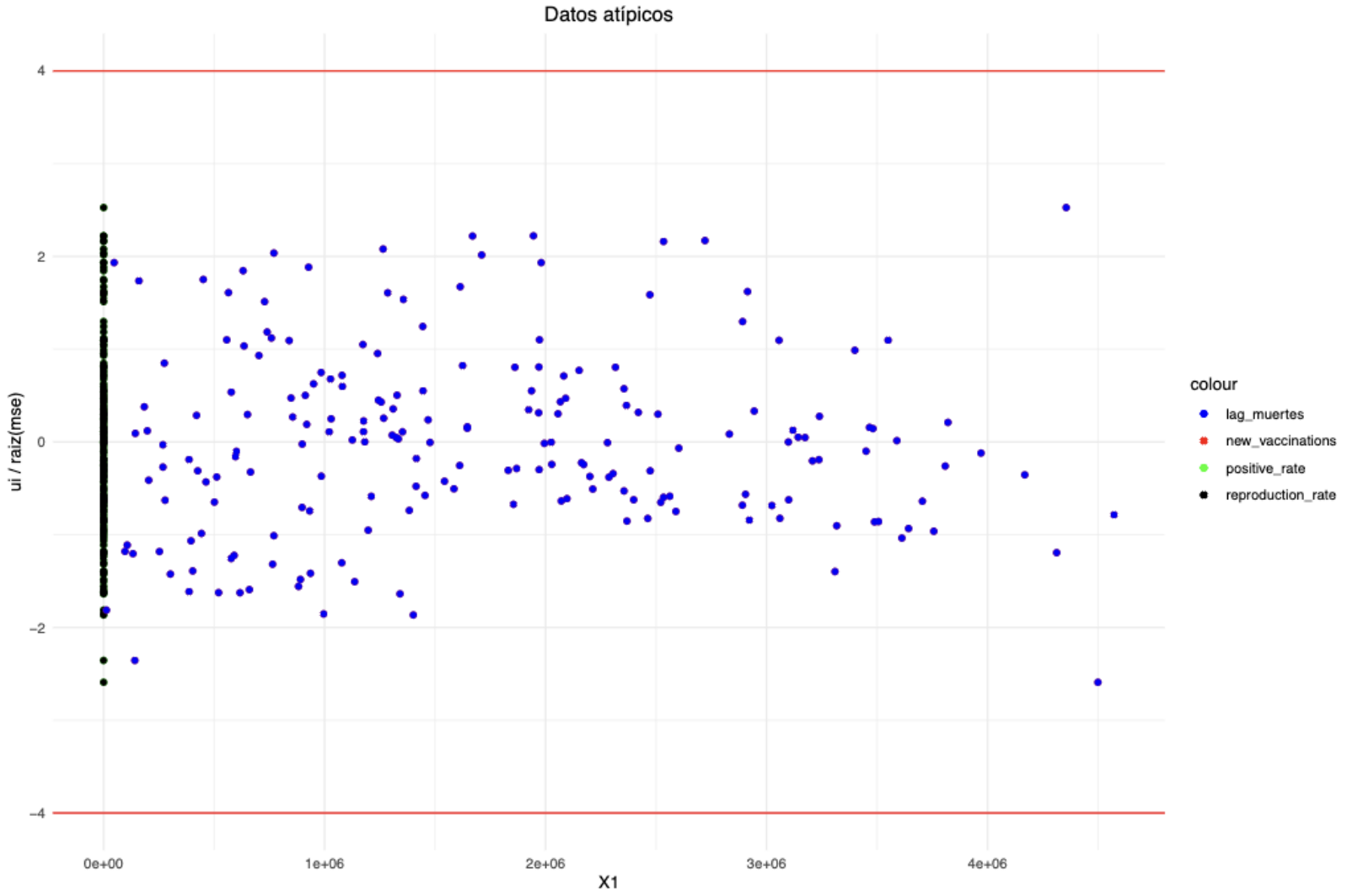
Validación de supuestos
Independencia de errores

Al ser una serie de tiempo, sí hay que buscar la independencia de los errores, la cual nos dice que corr(ei, ei-1) = 0

Esto se puede verificar con la prueba Durwin Watson; la ual nos dice que nuestra base no tiene independencia en los errores.

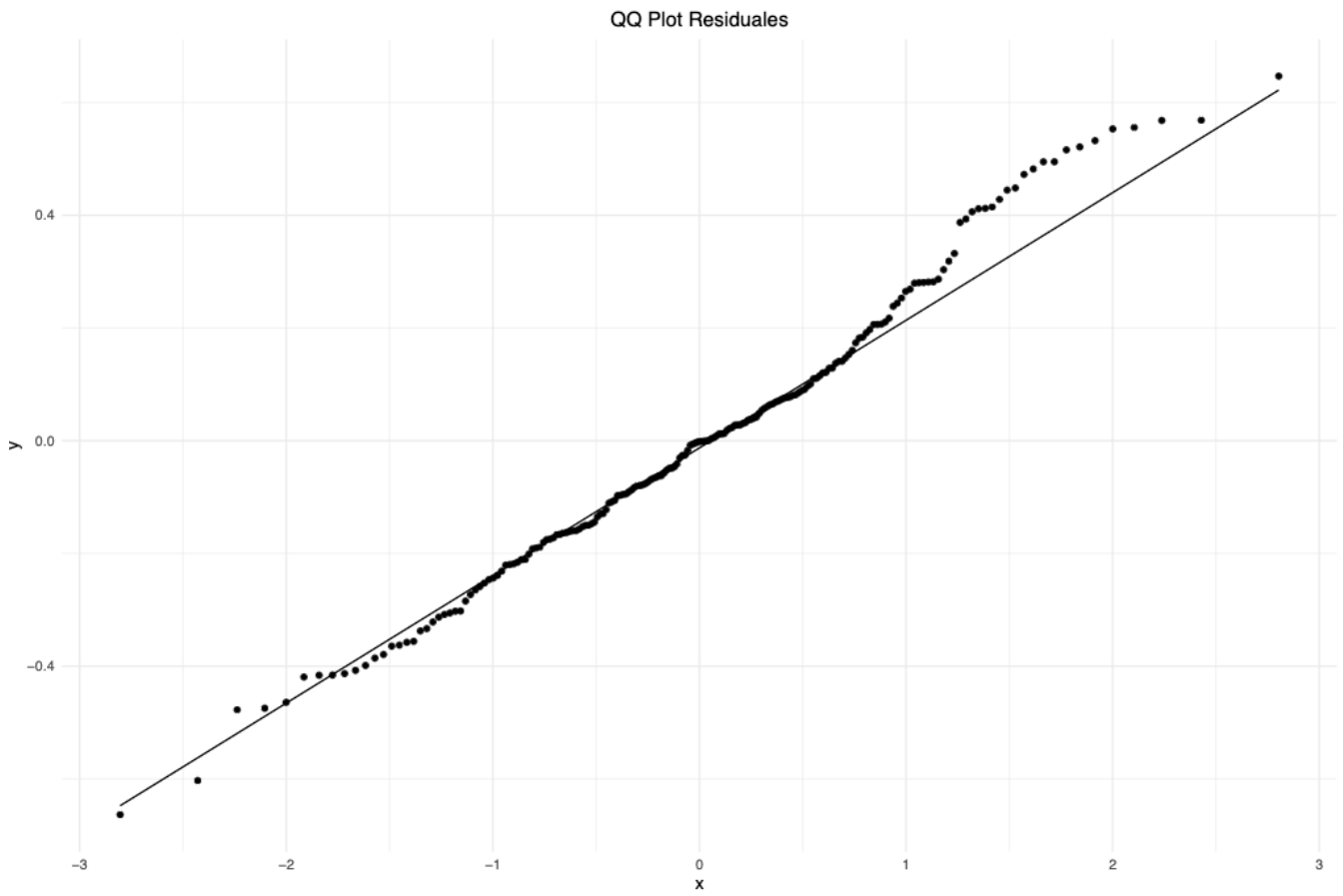
Para arreglar esto se aplica un lag en la Y, y se mete en el modelo de regresión.

Validación de supuestos
Errores atípicos y linealidad de la Fn



Nuestra R^2 es .86, lo cual nos dice que nuestro modelo sí cumple la linealidad a los datos y un ajuste correcto a ellos.

Validación de supuestos
Normalidad de los errores y multicolinealidad



El MRLM es robusto ante normalidad

VIF < 10 por lo que no hay multicolinealidad entre independientes

```
vif(modelo_box_cox_updated)
  positive_rate reproduction_rate  new_vaccinations       lag_muertes
       4.425362          1.580442          1.355002          3.294875
```

Conclusiones

Las muertes causadas por el COVID-19, de acuerdo a nuestro modelo, seguirán en aumento y no esperamos ver una tendencia a la baja hasta el mes de abril. Sin embargo, Al producir predicciones basadas en evidencia sobre el comportamiento de una cuarta ola por COVID-19, nuestro modelo ofrece beneficios a los tres niveles esenciales de sector salud:

1. Hospitales: hacer más eficiente el uso de recursos.

2. Pacientes: proveer servicios suficientes y accesibles a los pacientes más graves.

3. Personal: contar con los materiales necesarios para cuidar tanto de los pacientes, como de ellos mismo.

Nuestro trabajo no ha terminado, queremos que nuestro modelo se fortalezca adquiriendo más y mejor información para que en el futuro pueda discriminar el comportamiento de las cepas de COVID-19, así como ser aplicado a un nivel subnacional, para los aliados estratégicos de la AHA puedan utilizarlo para informar decisiones sobre mercados locales.

RCG

Raúl Blé Rosique I.D. 175992
Eduardo Damián Aoki I.D. 182509
Jorge Araujo Justo I.D. 186767
Óscar Barranca Ventoso I.D. 188242