

Estado de la publicación: El preprint ha sido publicado como artículo en una revista  
DOI del artículo publicado: <https://doi.org/10.1590/1980-549720220001>

# COVID-19 prediction of tendency for 2021 in northwestern Argentina

Eduardo Agustín Mendoza, Octavio Bruzzone, María Julia Dantur Juri

<https://doi.org/10.1590/1980-549720220001>

Enviado en: 2021-12-10

Postado en: 2021-12-15 (versión 2)

(AAAA-MM-DD)

<https://doi.org/10.1590/1980-549720220001>

E220001

## COMUNICACIONES BREVES

### **COVID-19 prediction of tendency for 2021 in northwestern Argentina**

Previsão da tendência COVID-19 para 2021 no noroeste da Argentina

**Short title:** COVID-19 tendency prediction for 2021 in Argentina

**Título corto:** Previsão da tendência COVID-19 para 2021 na Argentina

**Eduardo Agustín Mendoza** ORCID: <https://orcid.org/0000-0003-2216-4332>,  
Fundación Miguel Lillo, Instituto de Ecología, Comportamiento y Conservación,  
Biología Integrativa, San Miguel de Tucumán, Tucumán, Argentina.  
[eamendoza@lillo.org.ar](mailto:eamendoza@lillo.org.ar)

**Octavio Bruzzone** ORCID: <https://orcid.org/0000-0003-2358-1963>

Unidad Ejecutora Instituto de Investigaciones Forestales y Agropecuarias Bariloche,  
(Instituto Nacional de Tecnología Agropecuaria/Consejo Nacional de Investigaciones  
Científicas y Técnicas), San Carlos de Bariloche, Río Negro. Argentina.  
[okktawio@gmail.com](mailto:okktawio@gmail.com)

**María Julia Dantur Juri** ORCID: <https://orcid.org/0000-0002-9219-2577>

Unidad Ejecutora Lillo (Consejo Nacional de Investigaciones Científicas y Técnicas-  
Fundación Miguel Lillo), San Miguel de Tucumán, Tucumán, Argentina. Unidad de  
Microscopía Espectral (Universidad Nacional de Tucumán - Sistema Provincial de  
Salud - Ministerio de Salud de Tucumán), San Miguel de Tucumán, Tucumán,  
Argentina. [jdanturjuri@gmail.com](mailto:jdanturjuri@gmail.com)

**Autor correspondiente:** Eduardo Agustín Mendoza, Instituto de Ecología, Comportamiento y Conservación, Biología Integrativa, Miguel Lillo 251, (CP 4000), San Miguel de Tucumán, Tucumán, Argentina. [eamendoza@lillo.org.ar](mailto:eamendoza@lillo.org.ar)

**Conflicto de intereses:** Los autores declaran la ausencia de conflictos de intereses.

**Financiamiento:** El trabajo no contó con financiamiento.

**Comité de Ética:** Se trabajó con reportes diarios de casos de COVID-19 publicados por el Ministerio de Salud de Tucumán, siendo los mismos de libre acceso, sin necesidad de la aprobación de un Comité de Ética en Investigación.

**Contribución de autor.** Eduardo A. Mendoza trabajó en el diseño, análisis estadístico e interpretación de datos, redacción, revisión final y es el autor correspondiente; Octavio Bruzzone y María J. Dantur Juri, han trabajado en la interpretación de datos para la realización del trabajo, crítica revisión en busca de contenido intelectual importante y corrección preliminar.

## **RESUMEN**

Usando un modelo de regresión polinomial con retraso, que empleó datos de COVID-19 de 2020 con ausencia de vacunas, se realizó la predicción de COVID-19 en un escenario con administración de vacunas para Tucumán en 2021. La modelación incluyó la identificación de un punto de quiebre de contagios entre ambas series con la mejor correlación. Previamente, se indicó por medio de correlación cruzada el lag que sirvió para obtener el menor error entre los valores esperados y los observados. La validación del modelo fue realizada con datos reales. En 21 días fueron predichos 18.640 casos de COVID-19 de 20.400 casos informados. El pico máximo de COVID-19 fue estimado 21 días antes con la intensidad esperada.

**Palabras clave:** Predicción, modelo, COVID-19, vacunas, Tucumán.

## **ABSTRACT**

Using a lagged polynomial regression model, which used COVID-19 data from 2020 with no vaccines, the prediction of COVID-19 was performed in a scenario with vaccine administration for Tucumán in 2021. The modeling included the identification of a contagion breaking point between both series with the best correlation. Previously, the lag that served to obtain the smallest error between the expected and observed values was indicated by means of cross correlation. The validation of the model was carried out with real data. In 21 days, 18,640 COVID-19 cases out of 20,400 reported cases were predicted. The maximum peak of COVID-19 was estimated 21 days in advance with the expected intensity.

**Keywords:** Prediction, model, COVID-19, vaccine, Tucumán.

## INTRODUCCIÓN

En Marzo de 2020, la Organización Mundial de la Salud (OMS) declaró a la enfermedad coronavirus (COVID-19) como pandemia<sup>1</sup>. Instó a activar diversos protocolos para contener la propagación<sup>2</sup>. En Argentina, el primer caso fue detectado en marzo de 2020 en Buenos Aires, declarándose cuarentena obligatoria por Decreto de Necesidad y Urgencia<sup>3</sup>.

A principios de 2021 no se habían administrado vacunas a la población y luego de la reapertura de actividades, la segunda ola de COVID-19 progresó.

El objetivo fue predecir la tendencia de los casos COVID-19 durante 2021 para un escenario con administración de vacunas y su pico máximo, estudiando el comportamiento estadístico de los datos de COVID-19 en 2020 sin la aplicación de vacunas.

## MÉTODOS

El estudio se llevó a cabo en la provincia de Tucumán, en el noroeste de Argentina, resultando elegida debido a la falta de predicción de casos de COVID-19 y porque es la segunda provincia más densamente poblada del país, con un reporte de 1.338.523 habitantes<sup>4</sup>.

La elaboración del modelo de predicción para casos de COVID-19 consistió en identificar en datos de COVID-19 de 2020 un lag de días  $t$  que mejor se correlacione con un lag de días  $t$  de COVID-19 de 2021 empleando como referencia un punto de quiebre de contagios en la primera serie. Con esa identidad lograda se realizó una correlación cruzada entre esos lags, con la finalidad de encontrar el mejor retraso para

ajustar los datos con un modelo polinomial de regresión con retraso y predecir a la tendencia de COVID-19 actual.

Transformación de los datos: para estabilizar el promedio y reducir la tendencia se usó diferenciación de orden uno. El valor de  $p$  se realizó con el estadístico  $t$  con  $n-2$  grados de libertad y con  $n$  basado en el número de muestras que se superponen en las correlaciones cruzadas. El análisis fue realizado con Past 3.22<sup>5</sup>.

<https://github.com/Agustino216/Covid19-prediction.git>

Se emplearon dos conjuntos de datos de COVID-19, los que fueron publicados diariamente por el Ministerio de Salud Pública de la provincia de Tucumán (MSPT)<sup>6</sup>. El primer conjunto desde el 18/03/2020 hasta 27/11/2020. El segundo conjunto desde 19/03/2021 hasta 20/05/2021. Se elaboró una matriz de lags para COVID-19 de 2020 de diferentes cantidades de días de longitud. Previamente, se obtuvieron las fechas de inicio y final de los lags. Se obtuvieron en base a un punto de quiebre de contagios, indicado por un incremento de 50% de la totalidad de casos reportados antes del máximo de COVID-19 en 2020. Se empleó media móvil de 15 días. La longitud en  $t$  días de los lags ( $l$ ) fueron exploradas a 30, 35, 40, 45 días.

La identificación del lag de COVID-19 2020 para elaborar el set de entrenamiento estuvo determinado por correlación de Pearson ( $r_p$ ) con  $p > 0.05$  con el lag de COVID-19 de 2021. Identificado el lag, se realizó correlación cruzada ( $r_d$ ) con  $p > 0.05$  entre ellos. De esta manera se obtuvo la ubicación en la serie predictiva  $y_i$  (COVID-19 2020) para su mejor retraso  $m$ . En el apéndice 1 se indica mediante un diagrama de flujo a la metodología empleada en detalles.

Modelo empleado: el lag de COVID-19 de 2020 identificado en el retraso  $m$  junto al lag de datos de COVID-19 de 2021 fueron ajustados con un modelo de regresión

polinomial con retraso. Se empleó este tipo de modelo debido a que los casos de COVID-19 son aleatorios y no lineales. El modelo de regresión Polinomial<sup>7</sup> usado fue:

$$x_i = a + b y_{i-m} + c y_{i-m}^2 + \dots + e$$

En donde  $x_i$  representa a los casos de COVID-19 diferenciados y predichos para 2021 en el día  $i$ ,  $a$ ,  $b$ ,  $c$  son coeficientes del modelo polinomial,  $y$  es la serie predictiva de COVID-19 de 2020 del día  $i$  que mejor pronostica  $x$  en función de  $y$  para su mejor retraso  $i-m$ , mientras  $e$  representa al error estimado. El proceso fue invertible para la diferenciación realizada. La autocorrelación de los residuos del mejor modelo fue nula. La evaluación del modelo fue realizada con datos reales de COVID-19, empleando el error de pronóstico porcentual promedio (MAPE). Un horizonte de pronóstico posteriormente fue valorado de idéntica forma.

## RESULTADOS

Los resultados indicaron que el punto de quiebre de contagios de COVID-19 en 2020 fue el 25/09/2020 (692 casos), mientras que en los datos de 2021 fue el 13/05/2021 (723 casos).

El lag de datos COVID-19 en 2021 empleado para construir el modelo fue desde 30/03/2021 hasta 13/05/2021 y para el lag de COVID-19 en 2020 fue desde 18/08/2020 el 01/10/2020. Entre las series se obtuvo  $r_p = -0.296$ , con  $p = 0.04$ , mientras para un retraso de  $m = 17$  días fue  $r_d = 0.488$  con  $p = 0.008$  (Figura 1). El modelo obtenido para ese retraso fue:  $x_i = -1.935E-06 y_{i-m}^3 + 0.0006216 y_{i-m}^2 + 0.02296 y_{i-m} + 11.44$ , con  $R^2 = 0.315$ ,  $F = 36.8$ ,  $p = 0.026$  (Figura 2). La autocorrelación de los residuales fue nula.

Los 17 días predichos fueron desde 14/05/2021 hasta 30/05/2021 ( $n = 17$ ) (Figura 2) con 14.042 casos predichos de 15.824 reportados, su MAPE fue de 11.3. En la máxima

variabilidad de los casos de COVID-19 desde 31/05/2021 hasta 03/06/21 se acumularon 4.576 casos predichos sobre 4.598 casos reportados con MAPE de 0.47. El pico máximo de COVID-19 a 21 días fue estimado para el 03/06/2021 con 1200 casos y aconteció el 04/06/202 con 1.485 casos.

## **DISCUSIÓN**

Los resultados mostraron que el modelo infraestimó a la cantidad de sucesos ocurridos antes del 27/05/2021, momento en que se instalaron restricciones sociales estrictas<sup>8</sup> y acompaña a los casos reales hacia el pico máximo cuando se instalan las restricciones. Es posible que la infraestimación se encuentre influenciada por relajación de la sociedad en relación con la administración de vacunas. Antes del inicio del modelado el 22/04/2021, la campaña de vacunación acumuló 230.000 dosis aplicadas<sup>9</sup>. Mientras dos días después del primer pico de COVID-19, el 06/05/2021, acumuló la aplicación de 306.000 dosis<sup>10</sup>. Otra vía de infraestimación del modelo sería la ausencia de restricciones sociales. Comparamos de manera conjunta el incremento de los casos de COVID-19 informados por el MSPT, los predichos y un Índice de Movimiento de Personas en Supermercados y Farmacias<sup>11</sup> y observamos que se comportaron de una manera similar (figura 2).

La precisión del modelo es similar al de otras investigaciones informadas, como el calculado con modelo de supervivencia y convolución, parsimonioso y robusto<sup>12</sup>. La longitud de la predicción obtenida es similar a la lograda con modelo susceptible-expuesto-infeccioso-recuperado extendido<sup>13</sup>.

El modelo que presentamos pudo predecir la tendencia en la dinámica de los casos de COVID-19 esperados hacia el pico máximo. Sin embargo, solamente pudo predecir el

máximo de COVID-19 para el día 03 de junio, ocurriendo realmente dos máximos de COVID-19 en 2021, uno el día 04/06/21 y el otro el 08/06/21.

En conclusión, remarcamos que las tendencias de los casos COVID-19 en 2021 en Tucumán pudieron predecirse analizando el comportamiento estadístico de la primera ola de COVID-19 ocurrida en 2020.

## REFERENCIAS

1. World Health Organization. Coronavirus disease 2019 (COVID-19) Situation Report – 51. Ginebra: World Health Organization; 2020.
2. World Health Organization. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). Ginebra: World Health Organization; 2020.
3. Ministerio de Salud de la Nación. Decreto de Necesidad y Urgencia (DNU) 260/2020 [Internet]. 2020. [Accedido el 5 de junio de 2021]. Disponible en: <https://www.argentina.gob.ar/coronavirus/dnu>
4. Instituto Nacional de Estadística y Censo (INDEC).. República Argentina. Censo Nacional de Población, Hogares y Viviendas 2001 y 2010. [Internet]. 2020 [Accedido el 21 de mayo de, 2021]. Disponible en: <https://www.indec.gob.ar/indec/web/Nivel4-CensoProvincia-3-999-90-000-2010>.
5. Hammer Ø, Harper DAT, Ryan PD. 2001. PAST: Paleontological software package for education and data analysis. Paleontological Electronica 4(1): 9 pp. Acceso en: <https://www.nhm.uio.no/english/research/infrastructure/past/>
6. Ministerio de Salud Pública. Gobierno de Tucumán. [Internet]. 2020. [Accedido el 29 de mayo de 2021]. Disponible en: <https://msptucuman.gov.ar/category/noticias/>
7. Cromwell JB, Hannan MJ, Labys WA, Terraza M. Multivariate Tests For Time Series Models. SAGE University paper. CA. USA. ISBN 0-8039-5440-9.
8. Comité Operativo de Emergencia de Tucumán. Gobierno de Tucumán. 2021. [Internet] [Accedido el 22 de mayo de 2021]. Disponible en: [https://coe.tucuman.gov.ar/recursos/documentos/archivos/archivo\\_333\\_20210522110758.pdf](https://coe.tucuman.gov.ar/recursos/documentos/archivos/archivo_333_20210522110758.pdf)

9. Ministerio de Salud Pública. Gobierno de Tucumán. [Internet]. 2021. [Accedido el 31 de Octubre de 2021]. Disponible en: <https://vacunartuc.gob.ar/llegaron-24600-dosis-de-sputnik-v-a-la-provincia/>
10. Ministerio de Salud Pública. Gobierno de Tucumán. [Internet]. 2021. [Accedido el 31 de Octubre de 2021]. Disponible en: <https://vacunartuc.gob.ar/desde-el-inicio-de-la-campana-tucuman-lleva-aplicadas-306-833-dosis-de-vacunas-contra-el-covid-19/>
11. Google LLC "*Google COVID-19 Community Mobility Reports*". [Internet]. 2021. [Accedido el 01 de Noviembre de 2021]. Disponible en: <https://www.google.com/covid19/mobility/>
12. Qinxia W, Shanghong X, Yuanjia W, Donglin Z. Survival-Convolution Models for Predicting COVID-19 Cases and Assessing Effects of Mitigation Strategies . *Frontiers in Public Health*, 2020; 8:636.  
<https://www.frontiersin.org/article/10.3389/fpubh.2020.00325>  
DOI=10.3389/fpubh.2020.00325
13. Ghostine R, Gharamti M, Hassrouny S, Hoteit I. An Extended SEIR Model with Vaccination for Forecasting the COVID-19 Pandemic in Saudi Arabia Using an Ensemble Kalman Filter. *Mathematics*. 2021; 9(6):636.  
<https://doi.org/10.3390/math9060636>

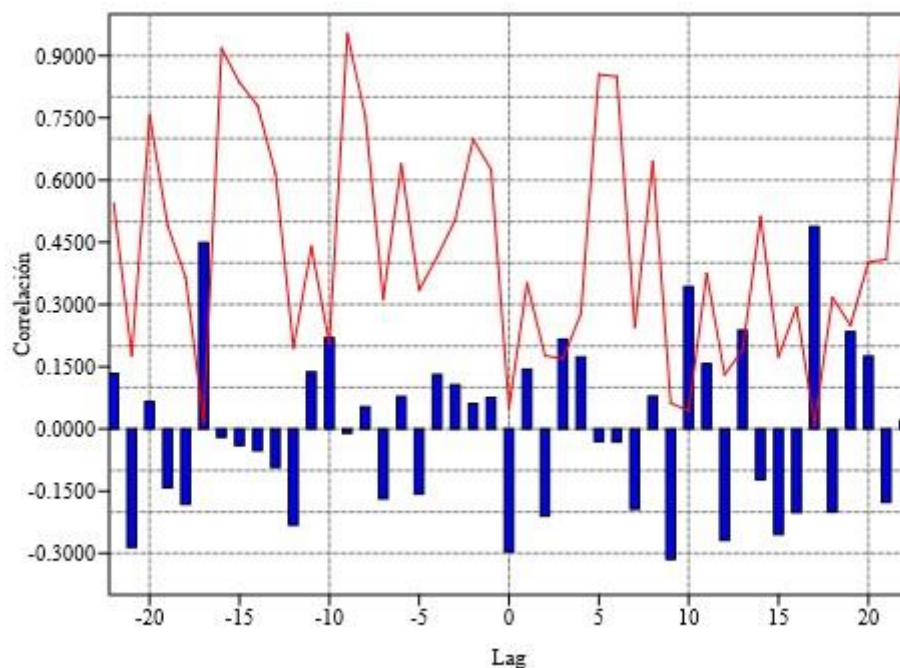
Recibido: 02/09/2021

Revisado: 05/11/2021

Aprobado: 07/12/2021

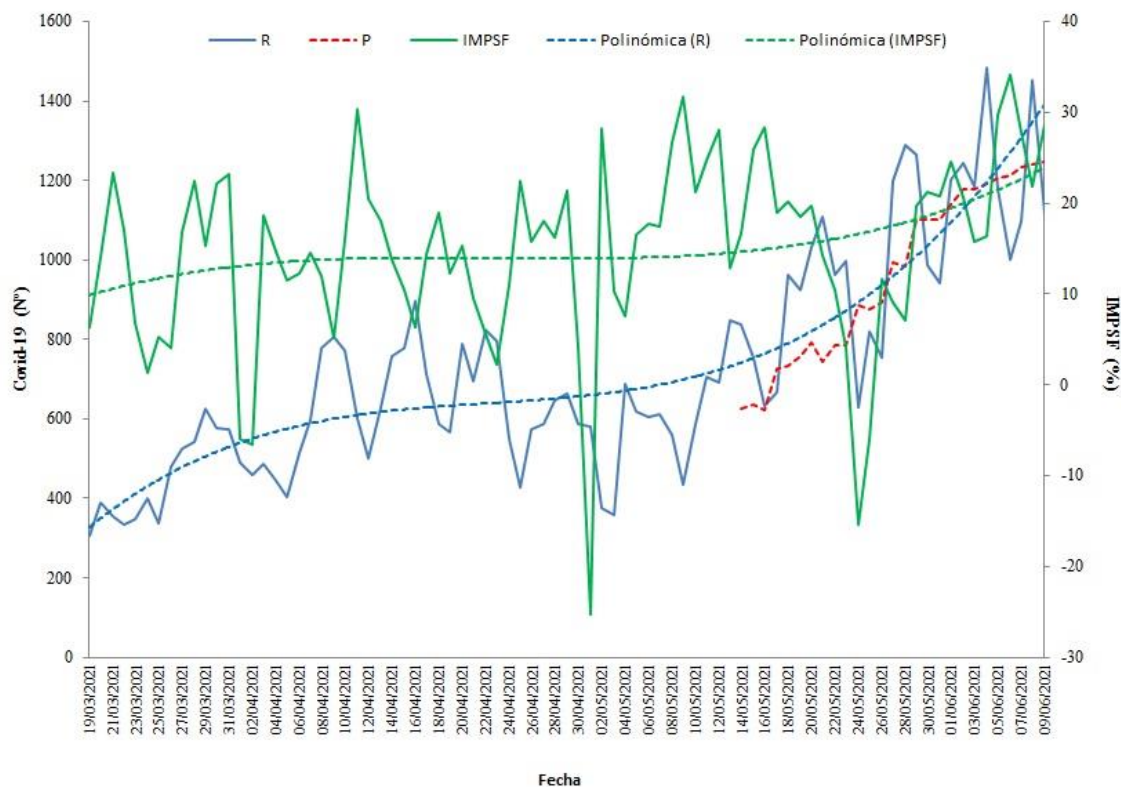
Preprint: 10/12/2021

**Figura 1.** Correlograma entre el lag de entrenamiento de COVID-19 2020 y el lag de COVID-19 de 2021.



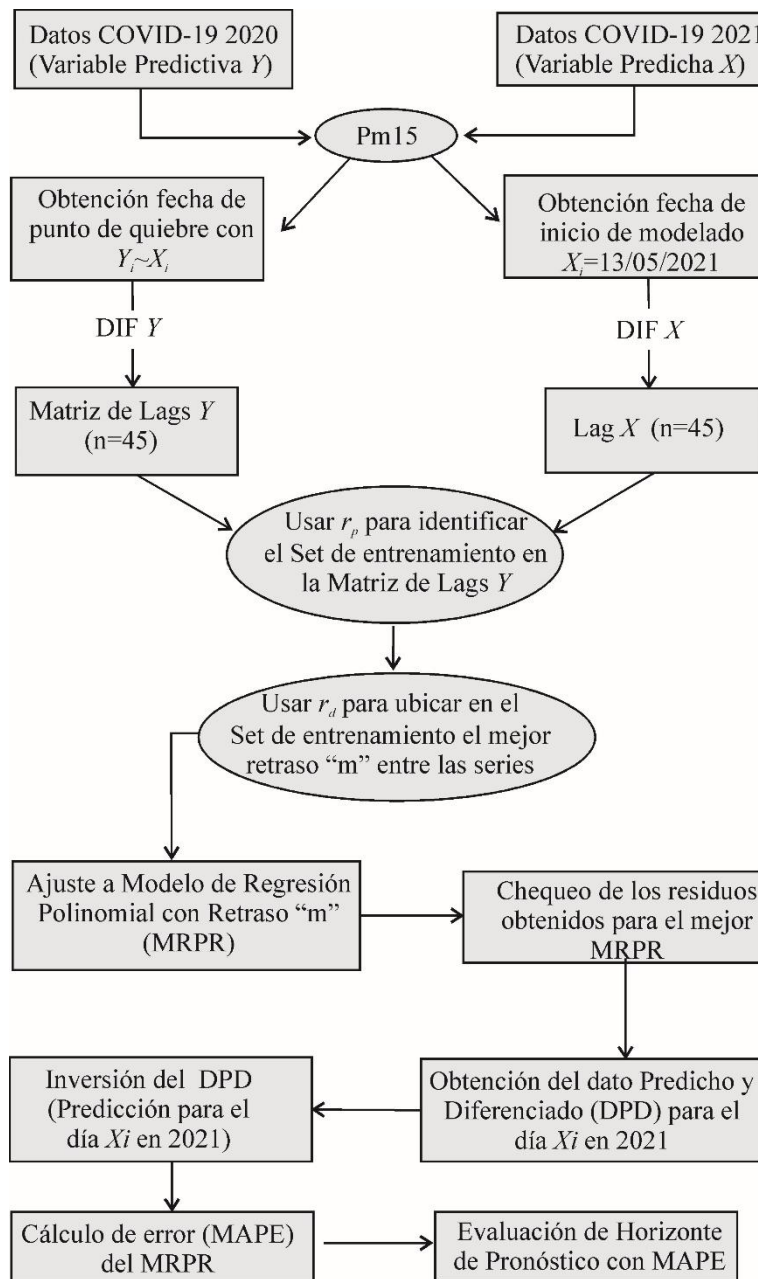
La línea roja indica probabilidad. El lag 10 fue significativo con residuales de modelo no nulo.

**Figura 2.** Datos reales y pronosticados de casos de COVID-19 en 2021 en Tucumán.



Las líneas azul R y Polinómica (R) corresponden a datos reales de COVID-19. La línea roja P corresponde a datos de COVID-19 predichos con MRPR. La línea verde es un Índice de Movimiento de Personas en Supermercados y Farmacias (IMPSF) en porcentaje para Tucumán según su punto de referencia igual acero. Todas las líneas quebradas corresponden a polinomios de grado tres.

**Apéndice 1.** Diagrama de flujo del modelo propuesto para la predicción de COVID-19 en Tucumán.



MAPE=error de pronóstico porcentual promedio, Pm=promedio móvil de 15 días,  
DIF=Diferenciación de datos de orden uno.

## Este preprint fue presentado bajo las siguientes condiciones:

- Los autores declaran que son conscientes de que son los únicos responsables del contenido del preprint y que el depósito en SciELO Preprints no significa ningún compromiso por parte de SciELO, excepto su preservación y difusión.
- Los autores declaran que se obtuvieron los términos necesarios del consentimiento libre e informado de los participantes o pacientes en la investigación y se describen en el manuscrito, cuando corresponde.
- Los autores declaran que la preparación del manuscrito siguió las normas éticas de comunicación científica.
- El manuscrito depositado está en formato PDF.
- Los autores aceptan que el manuscrito aprobado esté disponible bajo licencia [Creative Commons CC-BY](#).
- Los autores declaran que los datos, las aplicaciones y otros contenidos subyacentes al manuscrito están referenciados.
- El autor que presenta el manuscrito declara que las contribuciones de todos los autores y la declaración de conflicto de intereses se incluyen explícitamente y en secciones específicas del manuscrito.
- Los autores declaran que el manuscrito no fue depositado y/o previamente puesto a disposición en otro servidor de preprints o publicado en una revista.
- Si el manuscrito está siendo evaluado o siendo preparando para su publicación pero aún no ha sido publicado por una revista, los autores declaran que han recibido autorización de la revista para hacer este depósito.
- El autor que envía el manuscrito declara que todos los autores del mismo están de acuerdo con el envío a SciELO Preprints.
- Los autores declaran que la investigación que dio origen al manuscrito siguió buenas prácticas éticas y que las aprobaciones necesarias de los comités de ética de investigación, cuando corresponda, se describen en el manuscrito.
- Los autores aceptan que si el manuscrito es aceptado y publicado en el servidor SciELO Preprints, será retirado tras su retractación.