

### **1. Área de investigación**

Estadística e Investigación

### **Duración del proyecto (en meses)**

11 meses

### **2. Tipo de proyecto**

Multidisciplinario

### **3. Datos de los integrantes del proyecto**

Sabino Edgar Mamani Choque

Escuela profesional de Ingeniería Económica

Celular: 987349299

Correo: [semamani@unap.edu.pe](mailto:semamani@unap.edu.pe)

Julio Jesus Espiniza Calsin

Escuela profesional de Ingeniería Económica

Celular: 951896910

Correo: [julioespinozac@yahoo.com](mailto:julioespinozac@yahoo.com)

Roger Cueva Mamani

Ing. Estadístico. Trabajador administrativo UNA Puno

Celular: 950977404

Correo: [cuevanet@gmail.com](mailto:cuevanet@gmail.com)

1. Título del proyecto

Modelos de crecimiento para la predicción del Sars-Cov-2 en Perú

2. Área de Investigación

Área de investigación	Línea de Investigación	Disciplina OCDE
	Estadística e Investigación	

3. Duración del proyecto (meses)

11

4. Tipo de proyecto

Individual	<input type="radio"/>
Multidisciplinario	<input checked="" type="radio"/>
Director de tesis pregrado	<input type="radio"/>

4. Datos de los integrantes del proyecto

<b>Apellidos y Nombres</b>	Mamani Choque Sabino Edgar
<b>Escuela Profesional</b>	Ingeniería Económica
<b>Celular</b>	987349299
<b>Correo Electrónico</b>	semamani@unap.edu.pe

I. Título

**Modelos de crecimiento para la predicción del Sars-Cov-2 en Perú**

II. Resumen del Proyecto de Tesis

En el Perú, así como en todo el mundo, la pandemia del Sars-Cov-2 constituye el problema sanitario más importante cuya letalidad es aproximadamente del 3%, pero además implica un serio problema para la economía por las restricciones que aplican los gobiernos. Actualmente ya se vive una segunda ola donde se muestra una mayor peligrosidad porque el contagio es más rápido, afecta a distintos grupos etarios con mayor letalidad, superando el número de infectados por día y también la mortalidad de la primera ola. En esta situación, es importante contar con información sobre los contagios, la mortalidad y ocupación hospitalaria, pero también es necesario hacer predicciones que contribuyan con la gestión de la crisis por parte del gobierno y las autoridades sanitarias, por esta razón, el objetivo de esta investigación es predecir

los contagios y la mortalidad a través de modelos de crecimiento como el de Gompertz y Logístico que por sus características describen crecimientos sigmoideales con una tasa de crecimiento lento al inicio, luego se acelera hasta alcanzar un punto de inflexión y continúa su curso a tasas de crecimiento menores hasta hacerse asintótico donde el número de infectados sería mínimo. La información para esta investigación está disponible en la página web del Ministerio de Salud. Para la estimación se hará a través del software estadístico SAS mediante el procedimiento NLIN y el método de Marquardt. Los resultados permitirán tomar decisiones al gobierno y las autoridades sanitarias, así como validar la utilidad de estos modelos que ya fueron utilizados en países como Italia y España e incluso en Perú, al inicio de la pandemia.

### III. Palabras claves (Keywords) (Máx. palabras: cinco)

Sars-Cov-2, predicción, Gompertz, Logístico

### IV. Justificación del proyecto

Debido al aumento de casos registrados y de países afectados por el COVID-19, el 30 de enero de 2020 la Organización Mundial de la Salud (OMS) declaró una Emergencia de Salud Pública de Importancia Internacional. De acuerdo con el Centro de Recursos de Coronavirus de la Johns Hopkins University & Medicine, hasta el 18 de marzo de 2020 se habían registrado 214.894 casos confirmados de COVID-19 a nivel mundial, de los cuales 81.102 fueron en China, seguido por Italia con 35.713 casos. En América Latina y el Caribe el mayor número de casos se observó en Brasil (372), Chile (238), Perú (145), Ecuador (111) y México (93). El total de muertes por COVID-19 a nivel mundial fue de 8.732 personas, de las que el 35,7% corresponde a China, seguido por Italia (34%).

En el Perú, el 6 de marzo de 2020 se confirmó el primer caso importado de COVID-19, posteriormente el 5 de mayo, todas las regiones del Perú confirmaron la transmisión del virus. En adelante, los contagios tuvieron una tendencia creciente alcanzando los niveles más altos en el mes de agosto, aunque con algunas diferencias entre regiones. En la última semana de diciembre se observa nuevamente una tendencia creciente y se traduce en una segunda ola de contagios. El Ministerio de Salud, al 7 de febrero del año en curso, reporta 1,191,221 casos positivos y un

total de 42,467 defunciones.

Sin embargo, durante la crisis de salud en el Perú ocasionada por la pandemia COVID-19 se requiere información sobre los contagios, la mortalidad y ocupación hospitalaria, pero también es necesario hacer predicciones que contribuyan con la gestión de la crisis por parte del gobierno y las autoridades sanitarias.

## V. Antecedentes del proyecto

Palacios et al., (2020), en una investigación al inicio de la pandemia reportaron que el coronavirus, denominada 2019-nCoV que surgió el 31 de diciembre de 2019 en Wuhan, China, ha causado conmoción entre la comunidad médica y el resto del mundo por el gran número de casos y fallecimientos en China y en cantidad creciente fuera de ella, convirtiéndose en una emergencia de salud pública a nivel mundial. Su tasa de mortalidad no es tan elevada (2-3%), pero su rápida propagación ha propiciado la activación de protocolos para detener su diseminación. Predijeron una posible pandemia, recomendando seguir cuidado personal dictadas por la Organización Mundial de la Salud.

En el Perú, Maguiña et al., (2020) reportan que el nuevo virus SARS-Cov2, ha causado una severa pandemia a nivel mundial, desatando pánico y alarma universal, además de generar el colapso del sistema sanitario en muchas regiones del planeta, por ser muy contagioso, causando miles de muertes especialmente en adultos mayores con comorbilidades como diabetes mellitus o hipertensión arterial. De los órganos, el sistema respiratorio es el más afectado, pero puede afectar a cualquier órgano del ser humano y aún no existe terapia específica ni vacuna. La mejor forma de prevenirlo es con la cuarentena, higiene con frecuente lavado de manos y la aplicación del distanciamiento social.

Arias-Reyes et al., (2020), al analizar los datos epidemiológicos de COVID-19 del Tíbet y las regiones de gran altitud de Bolivia y Ecuador, compararon con datos de las tierras bajas, para probar la hipótesis de que los habitantes de las zonas altas (+2,500 m.s.n.m) son menos susceptible a desarrollar efectos adversos graves en la infección aguda por el virus del SARS-CoV-2. Sus hallazgos indican que la aclimatación y adaptación fisiológica que contrarresta el entorno hipóxico a gran altitud puede proteger del impacto severo de la infección aguda por el virus del SARS-CoV-2.

Martelloni & Martelloni, (2020) analizaron la evolución de Sars-Cov-2 en Italia con datos de Protección Civil mediante el modelo Logístico, encontrando que predice adecuadamente al inicio del contagio. Encontraron que los asintomáticos constituyen las fuentes de contagio más importantes en las casas unifamiliares de los italianos. Respecto a las diferencias entre regiones, reportan un retraso en la aparición del virus, por lo que sugieren analizarlos individualmente. Dado que la investigación continúa, consideran la posibilidad de que los datos recientes cambien el escenario, tal vez para mejor, luego de las medidas más restrictivas del gobierno.

Sánchez-Villegas & Daponte Codina, (2020) muestran una metodología utilizada para modelizar la epidemia de COVID-19 en España y en sus comunidades autónomas en el que resaltan la fortaleza del modelo Gompertz basado en su sencillez ya que se utilizan exclusivamente los datos sobre contagios y defunciones que después se predicen, sin contar con ningún tipo de covariable externa para su cálculo. Sin embargo, la ausencia de más información hace que la calidad de las predicciones dependa de la calidad de los datos. Aunque las predicciones de los modelos de contagios pueden sufrir grandes variaciones con respecto a los valores observados debido a los cambios en la definición de caso, consideran que las predicciones de las defunciones son más robustas a lo largo del tiempo porque su definición no ha sufrido cambios.

Villalobos-Arias, (2020) concluye que se puede ajustar curvas de crecimiento de poblaciones, utilizando las funciones Logística Generalizada y Gompertz, inclusive en el caso más general como en los datos de Corea del Sur, donde se está haciendo de la logística con recta. Considera que estos métodos podrían ser utilizados para predecir el crecimiento de poblaciones, en este caso de personas infectadas por el Covid-19, y podrían ayudar a los expertos en pandemias para tomar las medidas necesarias.

Wilson, (1994), reporta que las funciones de supervivencia de Gompertz, Weibull o logística se ajustan a la supervivencia de más del 95% de una población, y los extremos de las curvas de supervivencia parecen caer entre los valores predichos por las tres funciones. Sin embargo, para algunas poblaciones, dichos extremos parecen ser demasiado complejas para adaptarse bien a una función simple. Los datos de supervivencia de hombres y mujeres en algunas poblaciones se ajustan mejor a diferentes funciones. Además, se necesitan poblaciones de 100 o más para distinguir

entre las funciones.

Nguimkeu, (2014) indica que la comparación de los modelos Gompertz y Logístico a través de las medidas tradicionales como el coeficiente de determinación ( $R^2$ ) y el error cuadrático medio (RMSE) son adecuadas, pero se deben usar siempre que dichos modelos sean los correctos. Aunque estos dos modelos son los más utilizados para describir y pronosticar la tendencia de una amplia variedad de tales datos, otros modelos podrían ajustarse mejor a los datos en la práctica, en cuyo caso se requiere un método para evaluar la adecuación del modelo seleccionado.

## **VI. Hipótesis del trabajo**

Los modelos de crecimiento Gompertz y Logístico predicen adecuadamente el número de contagios y defunciones ocasionados por la pandemia COVID-19.

## **VII. Objetivo general**

Predecir los contagios y defunciones debidos a la pandemia COVID-19 mediante los modelos de crecimiento de Gompertz y Logístico a partir de la información abierta del Ministerio de Salud, para el periodo marzo de 2020 a mayo de 2021.

## **VIII. Objetivos específicos**

- Predecir los contagios y defunciones provocados por el COVID-19 en el Perú mediante modelos de crecimiento.
- Seleccionar el mejor modelo que describa la tendencia y predicción según criterios estadísticos.

## **IX. Metodología de investigación**

Los modelos de crecimiento se han utilizado para describir y pronosticar la tendencia de una amplia variedad de datos cuya característica es un crecimiento sigmoideo, lento al inicio, luego un crecimiento acelerado hasta alcanzar un punto de inflexión y finalmente un crecimiento que tiene hacia una asíntota. En ese caso se usará la información abierta del Ministerio de Salud que se encuentra en Internet, para el periodo que comprende el inicio de la pandemia hasta el mes de marzo donde la curva se ve afectada por el proceso de vacunación que inició en el país.

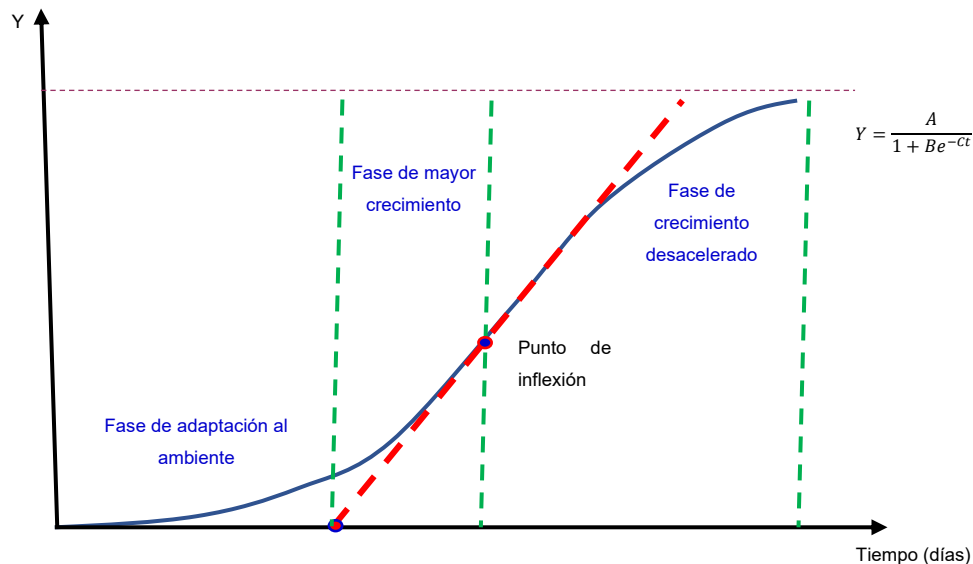
Los modelos no lineales dependen directamente de los parámetros específicos a estimar, por ello se utilizan procesos iterativos que definen una secuencia de

operaciones, cuyo límite es la solución o estimación de los parámetros (Bard, 1970); el proceso de estimación de los parámetros se realizará mediante el procedimiento NLIN del software estadístico SAS que estima los modelos mediante diferentes métodos, en este caso se usará el método de Marquardt (Pérez, 2001; SAS Institute Inc., 2015); es el método más utilizado y depende de varios factores que pueden influir en su convergencia o no, tales como el número de datos, el tipo de ecuación a ajustar y estimaciones iniciales de los coeficientes (WU, 2000). El procedimiento requiere de valores iniciales de los parámetros, luego evalúa los residuos y la suma de cuadrados para cada combinación de los valores, de este modo determina los mejores valores mediante una interacción algorítmica (Agudelo et al., 2007)

Como indicadores de ajuste de los modelos se usará el coeficiente de determinación ( $R^2$ ) como  $R^2 = 1 - \frac{SCE}{SCT}$ , donde SCE es la suma de cuadrados del error y SCT es la suma de cuadrados total (SAS Institute Inc., 2015). Una medida de criterio de convergencia que utiliza el procedimiento NLIN de SAS es la medida de compensación relativa R de Bates & Watts, (1981) que indica convergencia cuando esta medida es menor a  $10^{-5}$ . Finalmente, otro criterio de ajuste es el número de iteraciones que son necesarias para lograr convergencia, pero depende del valor inicial de los parámetros a estimar y dado que éstos se pueden reajustar, el número de iteraciones solo indica cuán rápido disminuye la suma de cuadros de los residuales.

**Las Predicciones basadas en el modelo Logístico tiene tres parámetros (A, B, C)** que describe una función sigmoidea donde la velocidad de crecimiento aumenta conforme transcurre el tiempo, alcanza la máxima tasa de crecimiento en el punto de inflexión, luego disminuye la velocidad de crecimiento hasta alcanzar su valor máximo (A). Se representa como:  $Y = \frac{A}{1+Be^{-Ct}}$

Los puntos de interés se muestran en la siguiente figura: i) fin de la fase de adaptación al ambiente, definido como el intercepto con el eje de la abscisa (t), de la ecuación tangente a la curva de crecimiento en el punto de inflexión, ii) el punto de inflexión que define la culminación de la fase de aceleración e inicia la fase de desaceleración del crecimiento y iii) el valor asintótico que representa el valor potencial a la madurez.



El modelo de Gompertz fue propuesto por Winsor, (1932), surge a partir de modelos de crecimiento autorregulados, donde la tasa de crecimiento decrece exponencialmente con el tiempo después de alcanzar el punto de inflexión, usualmente se expresa como:  $Y_t = A \left[ e^{-e^{B-Ct}} \right]$ , Donde  $Y_t$  es el la variable dependiente en el tiempo  $t$ ,  $A > 0$  es la asíntota,  $B > 0$  controla la diferencia entre el valor inicial y el valor final a un tiempo  $t$ , y  $C > 0$  describe el índice de madurez o la tasa específica de crecimiento. El punto de inflexión es el punto en el tiempo donde  $y = \frac{A}{e}$  con  $t = \frac{B}{C}$ , muestra crecimientos tempranos rápidos, pero valores más lentos en la medida que se aproxima a la asíntota, con un largo periodo de crecimiento lineal alrededor del punto de inflexión (Casas et al., 2010).

#### X. Referencias bibliográficas

- Agudelo, D., Cerón, M., & Restrepo, M. (2007). Modelación de funciones de crecimiento aplicadas a la producción animal. In *Revista Colombiana de Ciencias Pecuarias*. Universidad de Antioquia. <http://aprendeenlinea.udea.edu.co/revistas/>
- Arias-Reyes, C., Zubieta-DeUrioste, N., Poma-Machicao, L., Aliaga-Raduan, F., Carvajal-Rodriguez, F., Dutschmann, M., Schneider-Gasser, E. M., Zubieta-Calleja, G., & Soliz, J. (2020). Does the pathogenesis of SARS-CoV-2 virus decrease at high-altitude? *Respiratory Physiology and Neurobiology*, 277(April). <https://doi.org/10.1016/j.resp.2020.103443>



- Bard, Y. (1970). *Comparison of gradient methods for the solution of nonlinear parameter estimation problems*. 1, 157–186. <https://doi.org/10.1137/0707011>
- Bates, D. M., & Watts, D. G. (1981). Relative Curvature Measures of Nonlinearity (with Discussion). *Journal of the Royal Statistical Society, Series B* 42:1–25.
- Casas, G., Rodriguez, D., & Afanador, G. (2010). Propiedades matemáticas del modelo de Gompertz y su aplicación al crecimiento de los cerdos. *Revista Colombiana de Ciencias Pecuarias*. <http://www.redalyc.org/articulo.oa?id=295023477010>
- Maguiña, C., Gastelo, R., & Tequen, A. (2020). El nuevo Coronavirus y la pandemia del Cov. *Rev Med Hered*, 9(2), 125–131.
- Martelloni, G., & Martelloni, G. (2020). Analysis of the evolution of the Sars-Cov-2 in Italy, the role of the asymptomatics and the success of Logistic model. *Chaos, Solitons and Fractals*, 140, 110150. <https://doi.org/10.1016/j.chaos.2020.110150>
- Nguimkeu, P. (2014). A simple selection test between the Gompertz and Logistic growth models. *Technological Forecasting and Social Change*, 88, 98–105. <https://doi.org/10.1016/j.techfore.2014.06.017>
- Palacios, M., Santos, E., Velásquez, M., & León, M. (2020). *COVID-19, una emergencia de salud pública mundial M. xx*, 1–7.
- Pérez, C. (2001). *El sistema estadístico SAS, The SAS System for Windows*. Prentice Hall.
- Sánchez-Villegas, P., & Daponte Codina, A. (2020). Modelos predictivos de la epidemia de COVID-19 en España con curvas de Gompertz. *Gaceta Sanitaria*, xx. <https://doi.org/10.1016/j.gaceta.2020.05.005>
- SAS Institute Inc. (2015). *SAS Stat(R) 14.1 User's Guide*.
- Villalobos-Arias, M. (2020). *Estimación de población contagiada por Covid-19 usando regresión Logística generalizada y heurísticas de optimización*. 66818, 1–16. <http://arxiv.org/abs/2004.01207>
- Wilson, D. L. (1994). The analysis of survival (mortality) data: Fitting Gompertz, Weibull, and logistic functions. *Mechanisms of Ageing and Development*, 74(1–2), 15–33. [https://doi.org/10.1016/0047-6374\(94\)90095-7](https://doi.org/10.1016/0047-6374(94)90095-7)
- Winsor, C. (1932). The gompertz curve as a growth curve. *Proceedings of the National Academy of Sciences*.

## **XI. Uso de los resultados y contribuciones del proyecto**

Los resultados se podrán usar para la gestión de la crisis sanitaria en términos de previsión hospitalaria, así como decisiones del gobierno expresadas como políticas

públicas.

La metodología también se podrá aplicar para hacer predicciones por departamentos ya que desde el inicio hubo diferencias sustanciales en tiempo de inicio y evolución.

## **XII. Impactos esperados**

### **i. Impactos en Ciencia y Tecnología**

Los modelos de crecimiento ya se utilizan en una amplia gama de casos, también puede aplicarse en la predicción de pandemias ya que la característica de éstas es que inicialmente crece de manera lenta y posteriormente se acelera hasta llegar a un límite definido por una asíntota.

## **XIII. Recursos necesarios**

La ejecución del proyecto requiere de información secundaria, equipos de informática para procesarlos, Internet, materiales de escritorio y material bibliográfico, principalmente.

## **XIV. Localización del proyecto (indicar donde se llevará a cabo el proyecto)**

La investigación se realizará se realizará en la ciudad de Puno utilizando información secundaria del Ministerio de Salud.

## **XV. Cronograma de actividades**

Actividades	Meses											
	Feb	Mar	Abr	May	Jun	Jul	Ago	Set	Oct	Nov	Dic	
Planteamiento del Proyecto de Investigación	X											
Estructuración de la investigación	X	X										
Revisión de bibliografía relacionada con el proyecto.	X	X	X									
Planteamiento de antecedentes y marco teórico		X	X	X								
Planteamiento definitivo de la metodología de investigación				X	X	X						
Obtención y sistematización de información secundaria					X	X	X					
Procesamiento de la información en programas computacionales						X	X	X				
Análisis y discusión de resultados								X	X			

Redacción del informe de investigación					X		X	X		X	X
Revisión y correcciones de la investigación								X	X	X	
Puesta en estructura formal el trabajo									X	X	
Presentación final											X

## XVI. Presupuesto

Concepto	Unidad de medida	Precio unitario (S/ )	Cantidad	Total (S/ )
Papel bond	Millar	25.00	2	50.00
Toner para impresora	Unidad	100.00	3	300.00
Recopilación de información	Varios	-	-	1000.00
Acopio de literatura pertinente	Varios	-	-	3000.00
Movilidad local	Soles por mes	300.00	11	3300.00
Procesamiento de la información	Varios	-	-	1000.00
Capacitación en herramientas de apoyo para la investigación	Unidad	700.00	3	2100.00
Redacción de informe final y publicación	Unidad	500.00	1	1000.00
<b>Total</b>				<b>11,750.00</b>