

Tiempo transcurrido desde los primeros síntomas del COVID-19 hasta la defunción en México

Barco Gabriel Missael, Cabal Páramo Laura del Carmen, Pedroza Segoviano David,
Rodríguez Rivera Gil Estéfano, Segura Gómez Guillermo
Universidad de Guanajuato - División de Ciencias e Ingenierías (DCI)
Lomas del Bosque 103, Lomas del Campestre – León, Gto., México

Resumen

En el presente trabajo, se realiza un análisis estadístico para una variable aleatoria (los días transcurridos desde los primeros síntomas de COVID-19 de un paciente hasta su defunción). Se obtuvieron las medidas de tendencia central, se hizo el ajuste de distintas distribuciones y, para cada una, se contrastaron algunas hipótesis sobre el tiempo medio que tarda un paciente en morir.

1. Introducción

COVID-19, es el nombre de la enfermedad que al año de 2020 ha paralizado al mundo por completo. El padecimiento, detectado por primera vez en diciembre de 2019 en Wuhan China, rápidamente se ha esparcido por el mundo, ocasionando una pandemia mundial sin precedentes en los últimos años. El virus causante de dicha pandemia se conoce como SARS-CoV-2 y tiene la capacidad de transmitirse de persona a persona a través del aire y por contacto directo. La característica principal de la enfermedad, es una infección a nivel tracto respiratorio, y sus principales manifestaciones clínicas son: tos, fiebre y cambios radiográficos típicos con la posibilidad de desarrollar neumonía en distintos grados.[2]

Lo contagiosa que es esta enfermedad, además del hecho de que no todas las personas que portan el virus desarrollan la enfermedad inmediatamente, son motivos mas que suficientes para extremar precauciones, especialmente en los llamados "grupos de riesgo". Aunque todas las personas se pueden contagiar, existen grupos mas vulnerables que otros, ejemplo de ellos son las personas de edad avanzada o las personas que tienen alguna morbilidad asociada a complicaciones de COVID-19 (hipertensión, diabetes, obesidad, neumonía y bronconeumonía). [3]

En los casos graves, la duración desde el inicio de la enfermedad hasta presentar dificultad respiratoria es de 5 días, para precisar hospitalización 7 días y, desde el inicio de la enfermedad hasta presentar síndrome de distrés respiratorio agudo, 8 días. [4] Si bien, la enfermedad no tiene un alto índice letalidad, es interesante saber de qué manera evoluciona la enfermedad en el tiempo -particularmente en los enfermos que mueren-.

2. Metodología

Luego de descargar el archivo con datos (actualizados el 26 de mayo del 2020) desde la página de los datos abiertos la Dirección General de Epidemiología (Nota: el archivo con los datos tiene un código numérico, para entenderlo, es necesario revisar también los catálogos que se incluyen en otro archivo en la misma página) [1], se importaron todos los requisitos para realizar el análisis en Python en una versión mayor o igual a la 3.0 (numpy, pandas y matplotlib) y se almacenaron los datos en forma DataFrame de pandas. Se le dio una estructura a los datos (para tener una visión completa de los mismos) y se creó un diccionario para relacionar cada columna con su respectivo catálogo. Para analizar la evolución de la enfermedad, sólo son relevantes los casos positivos donde el paciente murió, por ende, se filtraron a una variable llamada "muertos"; luego, se contó la cantidad de casos en el arreglo, o sea, el total de muertes. Luego, se obtuvo la fecha de aparición de los síntomas y la fecha de muerte y se convirtieron a un objeto reconocible por Python, obteniendo una lista de los días que tardó cada persona en fallecer (Nota: hubo algunos casos anómalos donde la fecha de los primeros síntomas era posterior a la fecha de fallecimiento (se descubrió porque en el conteo de los días aparecieron números negativos y luego se verificó manualmente en la base de datos para descartar que fuera un error en el manejo de datos), esos no fueron considerados al colocar otro filtro). Se calculó el promedio y la desviación estándar de los días que las personas tardaron en fallecer y se graficó la lista. Se rescataron las frecuencias de la gráfica anterior y con ellas se obtuvo la probabilidad -por día transcurrido- de morir (*muertos en x día / muertos totales*); se graficó la lista; se ajustó una densidad normal con Scipy (Nota: pero, como el dominio de la función no es todo el conjunto de los reales, hubo que normalizarla (calculando el área bajo la curva entre cero e infinito y dividiendo la función entre ese valor) para que la suma de probabilidades sea igual a uno), también con la

densidad previa pero normalizada y se ajustó una distribución log-normal con Scipy. El archivo se encuentra en un repositorio de GitHub en [5].

3. Resultados

En la primera etapa del análisis, el total de muertos de la base de datos es 12477 mientras que los casos considerados (los no-anómalos) en las gráficas siguientes son 12409. El promedio de los días que las personas tardaron en fallecer es de 10,0279 días con una desviación estándar de 6,4821 días. Las distribuciones que hay de la cantidad de defunciones de pacientes con COVID-19 en función de los días que transcurrieron desde la aparición de los primeros síntomas hasta el deceso (tales aparecen en las figuras 1 y 2). Además, en la figura 2 se presenta la comparación de la curva de probabilidades con una distribución normal:

$$N(X = x) = \frac{1}{6,4821\sqrt{2\pi}} e^{-\frac{(x-10,0279)^2}{2(6,4821)^2}} \quad (1)$$

La misma pero normalizada:

$$N_n(X = x) = \frac{1}{(0,9391)(6,4821)\sqrt{2\pi}} e^{-\frac{(x-10,0279)^2}{2(6,4821)^2}} \quad (2)$$

y una distribución log-normal:

$$L(X = x) = \frac{1}{x(0,4369)\sqrt{2\pi}} e^{-\frac{(\ln(x)-13,2294)^2}{2(0,4369)^2}} \quad (3)$$

Note que la distribución normal ajustada con Scipy tiene el mismo promedio y la misma desviación que los calculados a partir de los datos.



Figura 1: Cantidad de decesos en función de los días transcurridos entre la aparición de los primeros síntomas y la defunción.

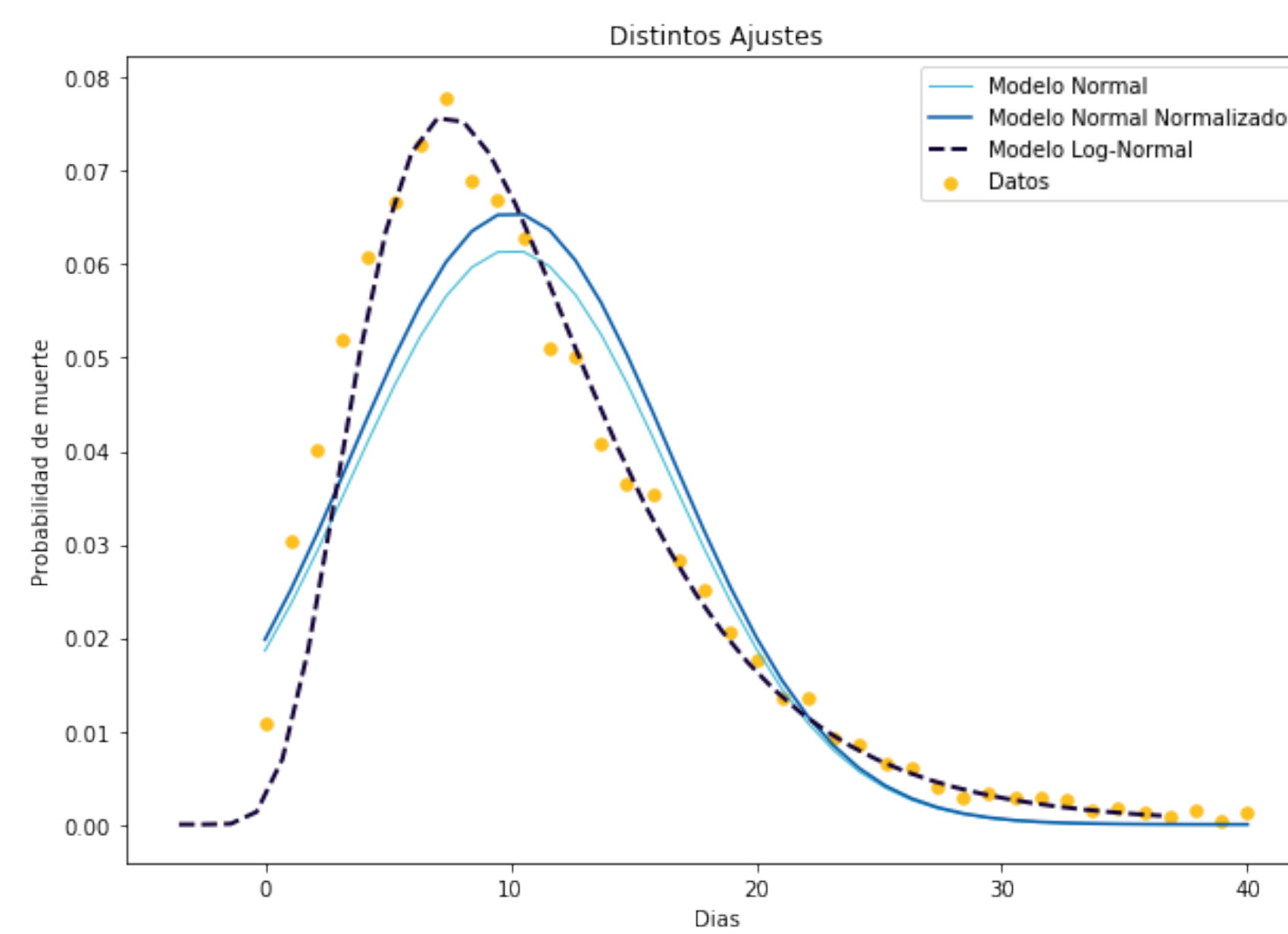


Figura 2: Probabilidad de morir en función de los días transcurridos entre la aparición de los primeros síntomas y la defunción (variable aleatoria X) en amarillo. Los distintos ajustes en diferentes tonos de azul.

En la segunda etapa del análisis, hay que notar que la contrastación de las hipótesis se realizó, para cada distribución, con las medidas de dispersión de cada una ($\bar{x} = 10,0279$ y $S = 6,482 \approx \sigma$ para la normal, $\bar{x} = 10,0279$ y $S = 6,9024 \approx \sigma$ para la normalizada y $\bar{x} = 13,2294$ y $S = 0,4369 \approx \sigma$ para la log-normal) y, como medida adicional, se aprovechó la propiedad de la log-normal de que

su logaritmo sigue una distribución normal para, con esa nueva función, evaluar las hipótesis. El procedimiento fue calcular los valores críticos de distribución normal estándar como $z_\alpha = (\bar{x}_\alpha - \mu)/(\sigma/\sqrt{N})$, donde $N = 12409$, y luego se evaluó el área de la normal estándar entre los valores críticos para determinar si se cumplen las hipótesis o no. Se empieza contrastando para la distribución normal que $P_N(z < 0,4795) = 0,6480$, para la normalizada, $P_{N_n}(-161,3870 < z < 0,4503) = 0,7173$ y para el logaritmo de la log-normal, $P_{ln(L)}(-\infty < X < 71,3546) = 0,9999$, entonces, con una significancia de hasta 0,6480 para la primera, de 0,7173 para la segunda y 0,9999 para la tercera, la gente con COVID-19 tarda 10 días en morir (contrastada con la hipótesis de que la gente muere en menos tiempo); entonces, con un 99,99 % de confianza podemos asegurar eso para los tres ajustes anteriores. Trabajando para otra hipótesis, $P_N(z < -85,4462) = 0,0000$, $P_{N_n}(-242,0803 < z < -80,2432) = 0,0000$ y $P_{ln(L)}(-\infty < X < -32,0262) = 0,0000$ entonces, con una significancia de hasta 0,0000 para la primera, de 0,0000 para la segunda y de 0,0000 para la tercera, la gente con COVID-19 tarda 15 días en morir (contrastada con la hipótesis de que la gente muere en menos tiempo); entonces, con un 99,99 % de confianza podemos asegurar, para los ajustes anteriores, que la gente muere en menos de 15 días.

4. Discusión y Conclusiones

Como primer punto a discutir, resulta muy interesante analizar como no cambian los resultados de las hipótesis según la distribución que se use. El resultado obtenido para la distribución normal refuta la misma hipótesis del resultado para la distribución Log-Normal; en la introducción se dijo que en los casos mas graves, se precisa la hospitalización alrededor del 7 día, entonces, hace sentido nuestra hipótesis de que la media de las personas que fallecen de COVID-19 este en los días menores a 15. Resulta evidente al ver la figura 1 que hay mas personas que mueren después del día 15, pero el resultado tiene sentido ya que la desviación estándar es bastante pequeña.

Los resultados de la inferencia implican que en una persona con COVID-19, según nuestra distribución normal, después de 15 días el riesgo de morir disminuye considerablemente, lo mismo ocurre con la distribución log-normal, siendo esta la que mejor se ajusta a los datos, véase la figura 2. Por lo que podemos concluir que en una persona infectada con COVID-19 el riesgo de morir es muy poco a partir del décimo quinto de haber presentado los primeros síntomas.

Referencias

- [1] Secretaría de Salud. (Actualizado: 2020, Mayo 26). *Datos Abiertos - Dirección General de Epidemiología*. Recuperado en Mayo 27, 2020, de <https://www.gob.mx/salud/documentos/datos-abiertos-152127>
- [2] Organización Mundial de la Salud. (Actualizado: 2020, Mayo 29) *Preguntas y respuestas sobre la enfermedad por coronavirus (COVID-19)*, Recuperado en Mayo 29, 2020 de <https://www.who.int/es/emergencias/diseases/novel-coronavirus-2019/advice-for-public/q-a-coronaviruses>
- [3] Lastra, M. S., González, C. M. V., Pérez, M. C. G., Guzmán, L. E. S., Ruiz-Rivera, N., Alcántara-Ayala, I., ... & Gutiérrez, M. D. C. J. (2020). Índice de vulnerabilidad ante el COVID-19 en México.
- [4] Centro de coordinación de alertas y emergencias sanitarias, (Actualización: 2020, Mayo 18), *Información científica-técnica, enfermedad por coronavirus, COVID-19*. Recuperado en Mayo 29 de <https://www.msbs.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov-China/documentos/ITCoronavirus.pdf>
- [5] BARCO GABRIEL MISSAEL, PEDROZA SEGOVIANO DAVID, SEGURA GÓMEZ GUILLERMO, *Análisis de datos sobre COVID-19 en México*, archivo en: https://github.com/GabrielMissael/Covid_probabilidad/blob/master/Analisis_de_datos.ipynb