



Un modelo de datos panel para predicción de la fatalidad del COVID-19 en el Perú

A panel data model for predicting fatality from COVID-19 in Peru

Carlos Minchón-Medina^{1*}, Daphne Timaná-palacios¹, Leonardo Sanes-Berrú²

¹ *Departamento Académico de Estadística de la UNT, Trujillo-La Libertad, Perú*

² *Escuela Académico Profesional de Estadística de la UNT, Trujillo-La Libertad, Perú*

Resumen

El presente estudio tiene el propósito de evaluar modelos de data panel de Poisson y binomial negativa para predicción de la fatalidad del COVID-19 en el Perú, con la finalidad explicar la variación de la fatalidad del COVID-19 en los departamentos, que permita la implementación de políticas sanitarias y socioeconómicas por parte del gobierno central, gobiernos locales y Ministerio de Salud. La investigación explicativa de datos panel, es de carácter transversal con componentes espacio-temporal, considerando la fatalidad del COVID-19 como variable dependiente, los contagios confirmados como variable de exposición, y la edad en grupos etarios como variable predictora. Los departamentos y periodos mensuales (junio 2020 - mayo 2021), constituyen componentes espacio-temporales. Los modelos de panel data, Poisson y binomial negativa, consideran como respuesta a una variable de conteo, El modelo de data panel binomial negativa fue mejor que el de Poisson; y entre los modelos binomial negativa, los de efectos aleatorios muy similares a los de efecto fijo. Se concluye que el modelo de datapanel binomial negativa a efectos aleatorios con respuesta de conteo es el mejor para la predicción de la fatalidad de COVID-19 en el Perú, mostrando diferencias de la fatalidad entre departamentos y grupos etarios.

Palabras clave: Fatalidad, data panel, conteo, Poisson, binomial negativa, efectos fijos, efectos aleatorios.

Abstract

The purpose of this study was to evaluate Poisson and negative binomial data panel models for predicting the fatality from COVID-19 in Peru, with the purpose of explaining the variation of the fatality of COVID-19 in the departments, which allows the implementation of health and socioeconomic policies by the central government, local governments and the Ministry of Health. The panel data explanatory research is cross-sectional with space-time components, considering the fatality of COVID-19 as a dependent variable, confirmed infections as an exposure variable, and age in age groups as predictor variable. The departments and monthly periods (June 2020 - May 2021), constitute space-time components. The data panel models, Poisson and negative binomial, consider a count variable as a response. The negative binomial data panel model was better than the Poisson model; and among the negative binomial models, those with random effects are very similar to those with fixed effects. It is concluded that the random effects negative binomial panel data model with count response is the best for predicting COVID-19 fatality in Peru, showing differences in fatality between departments and age groups.

Keywords: Fatality, data panel, count, Poisson, negative binomial, fixed effects, random effects.

1. Introducción

La COVID-19 es la enfermedad infecciosa causada por el virus SAR-Cov-2, cuyo primer caso fue reportado el 31 de diciembre de 2019 en Wuhan (China), propagándose rápidamente, primero dentro de China, y posteriormente en el resto del mundo, convirtiéndose en una pandemia (Organización Mundial de la Salud [OMS], 2020a), , alcanzando al 27 de diciembre de 2020 un total de 79,062,802 casos confirmados y 1,751,311 muertes en el mundo (Organización Mundial de la Salud [OMS], 2020b), con 208,518 casos adicionales y 3,285 muertes en las últimas 24 horas (Organización Mundial de la Salud [OMS], 2020c). El primer caso reportado en el Perú fue el 6 de marzo (Gobierno del Perú, 2020), y a la fecha se tienen 1,005,546 casos confirmados y 37,368 fallecidos (Ministerio de Salud [MINSAL], 2020).

Una de las principales características de una enfermedad infecciosa es su severidad o capacidad para causar la muerte, se mide a través de la razón de fatalidad de la infección (IFR) y la razón de la fatalidad de los casos (CFR) (World Health Organization, 2020). La CFR, es definida como la razón de muertes entre los casos confirmados, diferenciándose de la tasa de mortalidad en que la primera se determina tomando como razón al número de personas con una enfermedad específica (O'Toole, 2016), en el caso del COVID-19 a los contagios confirmados (World Health Organization, 2020).

Diversos estudios han empleado la CFR para estudio de la fatalidad por COVID-19. Cuando Italia presentaba el mayor brote, fue comparada con China, encontrándose CFR muy similares, 2.29 en Italia y 2.36 en China, por cada mil casos confirmados (Porcheddu et al., 2020). En China, se empleó un modelo de regresión lineal simple considerando diariamente el número de casos confirmados como predictor y el acumulativo de los fallecimientos como resultado, en el entendido de que la pendiente de la regresión podía ser utilizada como estimador de la CFR (Yang et al., 2020); asimismo, se utilizó un modelo de regresión no lineal considerando la CFR diaria como la variable resultado y la edad como entrada en la función logística, y comparó Italia con China, y hombres frente a mujeres (Gao y Dong, 2020).

Es usual abordar el tema de la mortalidad por COVID-19 empleando la regresión logística. Sin embargo, se requiere de información tanto de los fallecidos como de los supervivientes. Mediante una revisión de datos clínicos, se analizó factores de riesgo de la mortalidad en ancianos en China (Leung, 2020).

Las tasas caso fatalidad varían espacialmente entre países y temporalmente conforme la epidemia ha progresado, lo cual se deduce de estudios realizados. El modelo lineal mixto (LMM, por sus siglas en inglés) fue empleado para analizar la variación de la tasa de letalidad (CFR) de COVID-19 en función del tiempo (efecto lineal y cuadrático) y países (143 países, factor aleatorio), así como varios descriptores de comorbilidades, demográficos, económicos, y políticos de cada país (Sorci et al., 2020). Asimismo, modelos de datos panel fueron empleados para determinar la relación entre los factores y los efectos espaciales, considerándola tasa de muertes como variable dependiente, y comprendiendo 31 regiones de

Mailand China (Guliyev, 2020). Similarmente, se empleó un modelo de datos panel en 52 estados de África (factor a efectos fijos), relacionando el número de muertes atribuibles a COVID-19 y los casos confirmados, con los estados como factor a efectos fijos, y considerando un rezago de un periodo, transformándose las observaciones diarias de conformidad con el convencional procedimiento de estimación (Ayoade et al., 2020).

Los estudios anteriores, basados en datos secundarios, no consideran el hecho de que las muertes son variables discretas y que deben ser tratadas como datos de conteo, excepto un estudio para pronosticar las muertes diarias y acumuladas en los Estados Unidos, bajo el modelo de regresión de Poisson, usual, y un modelo de regresión de Poisson demasiado disperso, considerando las muertes diarias como variable dependiente y la cronología temporal como independiente (Kim et al., 2020); en este estudio el modelo de regresión binomial negativa es considerado un caso particular.

Las tasas de fatalidad de las enfermedades infecciosas, como el COVID-19, ayudan a comprender la severidad de la enfermedad, a identificar las poblaciones en riesgo y a evaluar la calidad de la atención sanitaria (World Health Organization, 2020). La variación de las tasas de fatalidad CFR de un país a otro, se explica debido a que la epidemia se propagó en unos países antes que otros (Sorci et al., 2020), pero tal variación también puede darse entre regiones dentro de un país por razones similares.

Actualmente, se ha desarrollado el análisis de datos de conteo (Hilbe, 2014), empleando modelos de Poisson o binomial negativa (Hilbe, 2011), modelos lineales generalizados (GLM) (Hardin, 2018) o modelos de datos panel para datos de conteo (Wooldridge, 2010; Winkelman, 2008; Sun y Zaho, 2013), que muy bien pueden ser aplicados para predecir los casos fatales por COVID-19 en Perú, mostrando diferencias regionales.

El presente estudio tiene el propósito de evaluar modelos de data panel de Poisson y binomial negativa para predicción de la fatalidad del COVID-19 en el Perú, con la finalidad explicar la variación de la fatalidad del COVID-19 en los departamentos del Perú, que permita la implementación de políticas sanitarias y socioeconómicas por parte del gobierno central, gobiernos locales y Ministerio de Salud.

2. Material y Método

La investigación explicativa de datos panel, es transversal y longitudinal, con componentes espacio-temporal. Las variables en estudio son la fatalidad del COVID-19 como variable dependiente, los contagios confirmados como variable de exposición, y la edad en grupos etarios es la variable predictora (niño, adolescente, joven, adulto y adulto mayor). La componente espacial está conformada por los 24 departamentos del Perú y la Provincia Constitucional del Callao, considerándose para efectos del estudio en 25 departamentos. La componente temporal está dada por periodos mensuales junio 2020 a mayo 2021.

La investigación se realizó en base a datos secundarios obtenidos de la Sala Situacional

COVID-19 Perú del Ministerio de Salud (MINSA, 2020). Estos datos constituyen el material de estudio.

La sección Datos Abiertos COVID-19 presenta dos reportes, que se actualizan diariamente, uno corresponde a los Casos positivos por COVID-19, que son casos confirmados por pruebas rápidas o pruebas moleculares (PCR), y el otro corresponde a los Fallecidos por COVID-19, reportes independientes unos de otros, en los que se consigna la fecha del suceso, edad, sexo, departamento y provincia, en donde ocurrió el hecho. En resumen, se tiene dos poblaciones que conforman los casos y los fallecidos por COVID-19. No es posible identificarnos a los contagiados ni a los fallecidos, por lo que no es posible identificar quienes de los contagiados fueron los que fallecieron, lo cual determinó que se trabajara con el numerador de la razón de fatalidad de los casos (CFR).

El estudio incluye a todos los departamentos del Perú. El periodo junio 2020 a mayo 2021 es considerado una muestra longitudinal. Y debido a la sobre dispersión del número de fallecidos, bajo las consideraciones de Hausman, Hall y Griliches (Stata Press, 2019), se ha propuesto modelos de regresión binomial negativa de datos panel.

En el procesamiento y análisis de datos se combinando los programas EXCEL, IBM SPSS Statistics 26 y Stata 16. El primero para convertir los archivos de extensión *cvs* a *xlsx* y así importar los datos desde SPSS, el cual permitió elaborar una base de datos necesarios para el modelo de datos panel, uniendo los dos archivos (casos positivos y fallecidos). Stata 16 se empleó en el análisis de datos, debido a que presenta un menú apropiado. En la terminología de Stata, se debe referir a la razón de fatalidad de los casos (CFR) como razón de las tasas de incidencia (RTI).

El análisis de datos comprendió la evaluación de modelos de data panel con respuestas de conteo, mediante modelos de Poisson y binomial negativa, a efectos fijos y a efectos aleatorios, reportándose las razones de tasas de incidencia (RTI) o simplemente tasas de fatalidad, en lugar de los coeficientes de regresión estimados. La elección del modelo para la predicción de la fatalidad del COVID-19 se realizó empleando los criterios de información de Akaike (AIC) y bayesiano (BIC), y las propiedades de los estimadores (Hilbe, 2011). La significancia se consideró si $p < 0.05$.

3. Resultados

En el periodo de estudio se reportaron en el Perú un total de 1798508 casos positivos de COVID-19 y 164040 fallecidos, alcanzando una tasa de 91.2 fallecimientos por 1000 contagios (91.2). Las tasas mensuales de la fatalidad se muestran en la tabla 1 por departamentos, las tasas más altas de fallecimientos ocurrieron en junio (142.1) y julio (126.9) del 2020, disminuyendo paulatinamente hasta octubre del mismo año, pero llegando a crecer nuevamente hasta mayo del 2021 (115.1). Los departamentos con tasa de fallecidos superior a los 100 por cada mil positivos fueron Ica (142.1), Lambayeque (135.2), La Libertad (120.9), Piura (127.3) y Puno (104.0). Amazonas presentó el más bajo índice de fatalidad (38.1)

Tabla 1

Perú: Tasas de fatalidad del COVID-19 por departamentos, junio 2020 – mayo 2021 (x 1000 positivos)

Departamento	2020							2021					Total
	Junio	Julio	Agosto	Setiembre	Octubre	Noviembre	Diciembre	Enero	Febrero	Marzo	Abril	Mayo	
Amazonas	53.9	48.7	17.1	20.0	19.6	24.3	53.4	42.1	53.1	42.1	46.4	73.5	38.1
Ancash	166.3	140.5	71.4	55.8	32.2	50.2	68.1	68.9	71.7	68.0	95.3	121.2	85.6
Apurímac	63.3	78.9	43.7	40.4	47.3	55.1	45.1	69.8	68.6	56.9	47.1	84.8	57.5
Arequipa	157.2	187.2	62.5	44.8	66.2	83.6	82.2	84.9	95.5	85.6	78.7	82.7	88.1
Ayacucho	59.9	74.1	66.1	40.5	36.8	33.9	43.0	45.5	74.2	80.1	73.8	92.1	64.9
Cajamarca	77.1	101.9	48.4	43.7	52.3	52.6	73.5	46.5	58.0	51.6	69.0	82.7	63.5
Callao*	199.8	216.2	110.9	63.8	55.0	47.6	65.5	90.5	111.9	70.7	59.1	178.1	95.3
Cusco	41.2	85.3	70.8	51.6	53.8	53.3	56.7	61.7	72.4	63.8	65.0	77.6	66.9
Huancavelica	31.2	85.6	53.6	28.6	56.7	61.7	53.8	94.7	130.0	103.9	80.6	91.3	72.9
Huánuco	87.4	60.9	39.6	33.8	59.1	50.2	59.8	148.9	177.4	90.2	70.4	113.2	75.6
Ica	193.5	143.2	75.8	57.1	115.0	184.6	168.2	121.2	191.6	192.9	193.0	228.8	142.1
Junín	102.3	176.8	88.3	62.5	40.7	40.2	36.0	77.4	99.4	81.2	70.8	92.3	84.1
La Libertad	259.0	223.5	98.6	50.5	53.8	82.8	129.1	140.4	112.6	100.1	119.4	98.2	120.9
Lambayeque	137.5	135.6	101.5	93.9	129.4	138.2	139.9	166.4	160.4	125.9	132.7	187.6	135.2
Lima	141.7	127.8	79.2	62.8	53.8	61.4	70.3	73.2	90.4	96.9	109.3	124.3	94.8
Loreto	99.0	48.9	27.6	14.1	22.8	20.2	24.6	72.6	95.0	99.1	68.9	62.9	60.7
Madre de Dios	92.7	83.0	31.4	14.0	31.3	24.7	46.2	185.0	51.8	57.7	108.3	56.2	56.5
Moquegua	70.2	118.0	67.0	16.9	39.8	25.8	57.7	83.9	33.5	24.2	42.2	81.1	54.8
Pasco	50.7	70.9	48.3	52.5	23.7	49.4	90.7	89.7	68.8	58.9	63.2	115.4	66.4
Piura	145.8	135.0	84.1	67.5	109.5	154.8	227.1	186.7	133.6	92.7	136.9	174.7	127.3
Puno	73.8	145.6	108.6	42.5	53.7	47.6	84.7	132.0	157.7	139.4	117.2	142.3	104.0
San Martín	139.1	105.3	45.6	27.8	17.6	22.5	21.6	33.9	46.5	53.1	67.3	95.5	61.8
Tacna	45.4	81.4	65.7	39.1	41.1	27.3	28.9	74.9	101.8	99.3	68.1	76.6	68.4
Tumbes	119.0	71.5	44.3	35.4	45.5	58.1	102.4	70.1	74.3	100.1	114.1	103.0	81.5
Ucayali	79.8	73.0	36.1	35.0	24.9	41.0	33.7	89.7	132.3	88.4	169.9	211.8	80.8
Total	142.1	126.9	71.9	52.6	53.0	62.1	77.2	83.1	95.8	89.7	97.7	115.1	91.2

*El Callao es considerado como departamento.

En la tabla 2 se muestran los casos positivos por COVID-19 y los fallecidos, permitiendo obtener las tasas de fatalidad y las razones de tasas de fatalidad por este virus. En cuanto a los contagiados, el mayor número de presenta en los adultos (1 008 779 contagiados), los jóvenes (359 585 contagiados) y adultos mayores (321 145); pero el número de fallecidos fue muy superior en los adultos mayores (114 567 fallecidos) y adultos (46 672 fallecidos), con respectivas tasas de fatalidad de 355.6 y 46.3 fallecidos por cada mil contagiados, respectivamente. En promedio la tasa de fallecimiento fue de 91.2 por cada mil contagiados. Además, tomando como referencia a los jóvenes, se encontró que la tasa de fatalidad en los adolescentes disminuyó en comparación con ellos, pero se multiplicó por 2.241 en los niños, por 8.642 en los adultos y por 66.432 en los adultos mayores.

Tabla 2

Perú: Tasas de fatalidad del COVID-19 por grupos etarios, junio 2020 – mayo 2021.
(x 1000 positivos)

Grupo etario	Positivos	Fallecidos	Tasa	RTI [#]
Niño	52016	624	12.0	2.241
Adolescente	55983	252	4.5	0.841
Joven	359585	1925	5.4	1.000
Adulto	1008779	46672	46.3	8.642
Adulto mayor	322145	114567	355.6	66.432
Total	1798508	164040	91.2	

[#] Razones de tasas de fatalidad, con referencia a los jóvenes.

En la tabla 3 se presentan modelos de data panel para respuestas de conteo a efectos fijos de Poisson y binomial negativa, en los cuales se presentan las razones de tasas de fatalidaden lugar de los coeficientes de regresión. La variable de respuesta de conteo fue “fallecidos”, la variable de exposición fue “positivos” y la de agrupamiento fue “departamento” donde ocurrieron los contagios y fallecimientos. En todos los modelos evaluados, se encontródiferencias en las tasas de fallecimientos de los grupos etarios en conjunto según el test de Wald ($p < 0.05$, en los tres modelos), y también diferencias de cada grupo etario respecto a los jóvenes ($p < 0.05$, en todos los casos) según la prueba Z.

En los modelos de data panel a efectos fijos de la tabla 3, el binomial negativa es el que presenta mejores criterios de información (AIC=10118.09 y BIC=10144.65), en comparación con los modelos de Poisson, aun cuando se haya considerado la varianza robusta. Un aspecto relevante, es que, ajustado por los departamentos, el modelo binomial negativa estima una mayor tasa de fatalidad a los adolescentes en comparación con los jóvenes (RTI=1.628), diferente al mostrado en los modelos de Poisson (RTI=0.848). Las tasas de fatalidad de los adultos mayores se multiplican por 33.630 en comparación con los jóvenes, y la de los adultos por 4.415, según este modelo.

Tabla 3

Modelos de data panel a efectos fijos para la fatalidad del COVID-19, junio 2020 – mayo 2021

Grupo etario	Poisson		Poisson robusto		Binomial negativa	
	RTI [#]	P&	RTI [#]	p&	RTI [#]	p&
Niño	2.270	0.000	2.270	0.000	3.218	0.000
Adolescente	0.848	0.014	0.848	0.018	1.628	0.000
Adulto	8.649	0.000	8.649	0.000	4.415	0.000
Adulto mayor	66.171	0.000	66.171	0.000	33.630	0.000
Constante					0.001	0.000
Wald: X ²	169888.65		20997.55			11146.81
p	0.0000		0.0000			0.0000
AIC	23498.19		23498.19			10118.09
BIC	23519.45		23519.45			10144.65

[#] Razones de tasas de fatalidad, con referencia a los jóvenes

& Prueba de Wald

Los modelos a efectos aleatorios, mostrados en la tabla 4, comparan nuevamente los modelos de Poisson, Poisson robusto y binomial negativa. Los criterios de información (AIC=10118.09 y BIC=10144.65), permiten deducir nuevamente que la binomial negativa es mejor que los modelos de Poisson, aunque ahora los efectos de los departamentos son aleatorios. Además, el reporte de STATA revela que la binomial negativa tiene efectos aleatorios con distribución Beta (15.536, 216.406).

Los modelos a efectos aleatorios de data panel estimados, a través de las pruebas Chi-cuadrado para α , indican que son mejores que los modelos agrupados de Poisson y Binomial negativa, lo cual justifica el ajuste en la estimación de las tasas de fatalidad del COVID-19 por departamento, ajuste que STATA lo hace internamente.

Tabla 4

Modelos de data panel a efectos aleatorios para la fatalidad del COVID-19, junio 2020 – mayo 2021.

Grupo etario	Poisson		Poisson robusto		Binomial negativa	
	RTI [#]	p ^{&}	RTI [#]	p ^{&}	RTI [#]	p ^{&}
Niño	2.269	0.000	2.269	0.000	3.216	0.000
Adolescente	0.848	0.014	0.848	0.018	1.628	0.000
Adulto	8.649	0.000	8.649	0.000	4.416	0.000
Adulto mayor	66.172	0.000	66.172	0.000	33.649	0.000
Constante	0.005	0.000	0.005	0.000	0.001	0.000
Alfa (α)	0.076		0.076			
Wald: X ²	169893.3		21842.71		11162.25	
p	0.000		0.000		0.000	
AIC	23912.68		23912.68		10532.07	
BIC	23944.56		23944.56		10569.27	
Test RL vs. agrupado: (alfa=0)						
chibar2	6959.44		6959.44		272.58	
p	0.000		0.000		0.000	

[#] Razones de tasas de fatalidad, con referencia a los jóvenes

[&] Prueba Z

Los modelos a efectos aleatorios estimados, mostrados en la tabla 4, evidencian nuevamente las diferencias en la razón de las tasas de fatalidad de los grupos etarios mediante la prueba Z, y que las tasas de fatalidad de cada grupo etario difieren de la tasa de fatalidad correspondiente a los jóvenes.

La binomial negativa a efectos aleatorios encontró diferencias de las tasas de fatalidad de cada grupo etario con respecto a los jóvenes, multiplicándose la tasa por 3.216, 1.628, 4.416 y 33.649 en el caso de niños, adolescentes, adultos y adultos mayores, respectivamente; muy similares a las del modelo a efectos fijos, pero con criterios de información 4% más grandes. La similitud de las tasas de fatalidad es el resultado de la similitud de coeficientes de regresión de los modelos y en sus correspondientes errores estándar, modificando el valor de la prueba Z en céntimas, como se muestra en las figuras A1 y A2 del anexo, reportados por STATA.

4. Discusión

En el presente estudio, se tomó como variable dependiente a la fatalidad de los casos de COVID-19, los contagios confirmados como la variable de exposición y los grupos etarios como factor de estudio. El interés del modelamiento fue predecir la tasa de fatalidad, la variable de exposición fue construir la razón de fatalidad de los casos (CFR), que para efectos de empleo de Stata 16, se denomina razón de las tasas de incidencia (RTI). El empleo de esta variable dependiente se justifica como un indicador de la capacidad del COVID-19 para causar la muerte (O'Toole, 2016), estudiada con propósitos similares, pero empleando modelos descriptivos (Porcheddu et al, 2020) o modelo de regresión logística (Gao y Dong, 2020), pero sin considerar la variación espacial y temporal que ofrece el modelamiento de data panel.

El análisis de modelos de data panel realizado para respuestas de conteo para predicción de la fatalidad del COVID-19 se determinó que los modelos de data panel binomial negativa eran mejores que los respectivos modelos de Poisson, robustos o no, basados en los criterios de información de Akaike y bayesiano. El criterio de Akaike fue empleado en un estudio sobre COVID-19, para predecir el número de muertes diarias y muertes acumuladas durante la pandemia comparando modelo de Poisson y de Poisson sobredispersos (Kim et al, 2020), pero sin considerar la metodología de panel data.

En cuanto a los modelos binomial negativa, a efectos fijos o aleatorios, no se dispone en Stata (Stata Press, 2019) de pruebas específicas para respuestas de conteo, aun cuando ya existe la prueba de Hausman (Winkelmann, 2008), la cual es ampliamente utilizada en modelos lineales para decidir entre los modelos fijos y aleatorios (Hardin y Hilbe, 2018; Hilbe, 2011; Wooldridge, 2010).

Al no tener una prueba formal en modelos de data panel para respuestas de conteos y decidir entre modelos a efectos fijos y a efectos aleatorios, seguimos la misma idea de la prueba de Hausman y comparamos numéricamente los coeficientes de regresión, encontrando una mínima diferencia entre ambos modelos. Por esta razón, en el presente trabajo se considera el modelo de datos panel a efectos aleatorios como el apropiado para las tasas de fatalidad por COVID-19, que se suponen varían de un departamento a otro, de modo que el inverso de uno más la dispersión sigue una distribución Beta (Stata Press, 2019).

La elección del modelo a efectos fijos, también se decidió debido a que, en modelos de Poisson, los estimadores son inconsistentes en presencia de un gran número de efectos fijos, cuya forma de consistencia está aún en debate la binomial negativa (Hilbe, 2011). Los 25 departamentos en el presente estudio dan lugar a un número similar de efectos fijos. Asimismo, el número de periodos mensuales considerados (junio 2020 – mayo 2021), 12 meses, constituye un número mayor de paneles que periodos, al respecto Hilbe (2011) indica que los estimadores de efectos aleatorios son más eficientes que los estimadores a efectos fijos.

Es posible emplear otros modelos alternativos a los planteados. En 52 estados de África se modeló el número de muertes atribuibles al COVID-19, explicado por el número de casos confirmados en un periodo de 31 días (Ayoade et al, 2020), empleándose el método generalizado de momentos (GMM) para modelos de data panel dinámicos, que incluía al

número de muertes en el periodo anterior como explicativa. Los autores no consideraron que la respuesta constituye una variable de conteo, y solo emplearon modelos a efectos fijos, tampoco consideraron algún factor de riesgo para orientar las políticas de cuidado en la población.

En la línea de modelos de data panel, para determinar el efecto espacial de COVID-19, se midió los casos diarios confirmados, muertes de los mismos y casos recuperados en 31 regiones de China, analizaron una serie de 7 modelos de data panel espaciales (Guliyev, 2020). El seleccionado fue el modelo espacial-rezagado X (SLX), basado en coeficientes de determinación y criterios de información AIC y BIC. Sin embargo, ninguno de los modelos propuestos correspondió a una respuesta de conteo; pero encontró diferencias espaciales, como en el presente estudio, y temporales.

Un estudio más interesado por letalidad que induce la infección del COVID-19, analizan factores de comorbilidad y asociados a variables demográficas, económicas y políticas, empleando modelo lineal mixto (LMM), considerando espacio y tiempo (Guliyev, 2020). Los factores que explican la variación de la letalidad fueron seleccionados la significancia de las pruebas estadísticas básicas. Igualmente, no consideró la respuesta de conteo y la edad como factor de riesgo. Un estudio muy similar (Sorci, Faivre y Morand, 2020), empleó modelos lineales mixtos, encontrando resultados similares al anterior.

Los adultos mayores fueron considerados el grupo etario de mayor riesgo de muerte en el presente estudio, se incorpora en la discusión para hacer notar la importancia de considerar políticas sanitarias especiales para los adultos mayores (Leung, 2020).

En resumen, los estudios realizados para las muertes por COVID-19 muestran una amplia variedad de modelos estadísticos, incluyendo los modelos de datos panel, no es común la propuesta seleccionada, a pesar de la literatura y softwares especializados. Por lo que, los modelos de datos panel binomial negativa con respuesta de conteo, con efectos aleatorios, constituye una buena alternativa para pronosticar la fatalidad por COVID-19 en el Perú.

5. Conclusiones y recomendaciones

Conclusiones

Se concluye que el modelo de data panel binomial negativa a efectos aleatorios con respuesta de conteo es el mejor para la predicción de la fatalidad de COVID-19 en el Perú, el cual muestra diferencias de la tasa de fatalidad entre departamentos, y encuentra diferencia de las tasas de fatalidad de los niños, adolescentes, adultos, y adultos mayores con respecto a los jóvenes, tomados como referencia.

Recomendaciones

Las autoridades políticas y sanitarias pueden tomar medidas sanitarias de protección a los pacientes de mayor riesgo, los adultos y adultos mayores, en cada uno de los departamentos, según diferencias de la fatalidad por COVID-19 o pandemias similares en el futuro, demostrada con el empleo de modelos de regresión binomial negativa con respuesta de conteo, con efectos aleatorios.

6. Referencias

- Ayoade I, Toyin A, Olugbenga O, et al. (2020). Modelling spatial variations of coronavirus disease (COVID-19) in Africa.
<https://doi.org/10.1016/j.scitotenv.2020.138998>
- Gao X, Dong Q. (2020). A logistic model for age-specific COVID-19 case-fatality rates. *JAMIA Open*, 3(2), 151–153. doi: 10.1093/jamiaopen/ooaa025
- Gobierno del Perú (2020). Coronavirus en el Perú: casos confirmados.
<https://www.gob.pe/8662-coronavirus-en-el-peru-casos-confirmados>
- Guliyev H. (2020). Determining the spatial effects of COVID-19 using the spatial panel datamodel. <https://doi.org/10.1016/j.spasta.2020.100443>
- Hardin J, Hilbe J. (2018). *Generalized Linear Models and Extensions*. 4th ed. Texas: Stata Press
- Hilbe J. (2014). *Modeling Count Data*. New York: Cambridge University Press.
- Hilbe J. (2011). *Negative Binomial Regression*. 2nd ed. New York: Cambridge University Press.
- Kim T, Luta G, Lieberman B, et al. (2020). Prediction Regions for Poisson and Over-Dispersed Poisson Regression Models with Applications to Forecasting Number of Deaths during the COVID-19 Pandemic.
<https://www.researchgate.net/publication/342733886>
- Leung Ch. (2020). Risk factors for predicting mortality in elderly patients with COVID-19: A review of clinical data in China. <https://doi.org/10.1016/j.mad.2020.111255>
- Ministerio de Salud (MINSA). Sala Situacional COVID-19 Perú 2020.
https://covid19.minsa.gob.pe/sala_situacional.asp
- Organización Mundial de la Salud. (2020a). Preguntas y respuestas por la enfermedad del coronavirus (COVID-19). <https://www.who.int/es/emergencias/diseases/novel-coronavirus-2019/advice-for-public/q-a-coronaviruses>
- Organización Mundial de la Salud. (2020b). WHO Coronavirus Disease (COVID-19) Dashboard. <https://covid19.who.int/>
- Organización Mundial de la Salud (2020c). Situación del COVID-19 en la Región de las Américas. <https://www.paho.org/es/temas/coronavirus/brote-enfermedad-por-coronavirus-covid-19>
- O’Toole M, editor. (2016). *MOSBY’S MEDICAL DICTIONARY*. 10th ed. Canada: Elsevier.
- Porcheddu R, Serra C, Kelvin V, et al. (2020). Similarity in Case Fatality Rates (CFR) of COVID-19/SARS-COV-2 in Italy and China. *J Infect Dev Ctries*; 14(2):125-128. doi:10.3855/jidc.12600
- Sorci G, Faivre B, Morand S. (2020). Explaining among-country variation in COVID-19 case fatality rate. <https://doi.org/10.1038/s41598-020-75848-2>
- Stata Press. (2019). *STATA. LONGITUDINAL-DATA/PANEL: REFERENCE MANUAL. RELEASE 16*. Texas: StataCorp LLC.
- Sun J, Zhao X. (2013). *Statistical Analysis of Panel Count Data*. New York: Springer; 2013.
- Winkelmann R. (2008). *Econometric Analysis of Count Data*. 5th edic. Switzerland: Springer.
- Wooldridge J. (2010). *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. USA:

Massachusetts Institute of Technology.

World Health Organization (2020). Estimating mortality from COVID-19. <https://www.who.int/publications/i/item/WHO-2019-nCoV-Sci-Brief-Mortality-2020.1>

Yang S, Cao P, Du p, et al. (2020). Early estimation of the case fatality rate of COVI-19 in mailand China: a data-driven análisis. *Ann Transl Med.*;8(4):128. <http://dx.doi.org/10.21037/atm.2020.02.66>