

Modelos ARIMA para el análisis de datos de sismos en el Perú en 2017

Carlos Risco Franco

Resumen

Este proyecto busca explorar el fenómeno de los sismos mediante el uso de herramientas estadísticas, en este caso para caracterizar y buscar patrones de comportamiento de los mismos, que se han presentado en el Perú en el 2017; para ello se usan técnicas como el Cluster Análisis y Modelos de Series de Tiempo. Primero se ha identificado las zonas de mayor actividad sísmica y luego se ha hallado sus características principales, así como la interrelación entre magnitud y profundidad. Aparte de identificar las zonas sísmicas, se observa una pequeña tendencia a la disminución en la magnitud de los sismos en Lima en los últimos tres meses. Los datos se han ajustado a un modelo de series de tiempo ARIMA (1,1,0), el cual ha resultado significativo, tanto a nivel nacional, como para Lima-Ica y Arequipa. Se ha utilizado datos del Instituto Geofísico del Perú en su página web. La zona de mayor actividad es la zona de Arequipa, seguida de la de Lima.

Palabras clave: sismo, Cluster Analysis, Modelos de Series de tiempo ARIMA.

Introducción

Es conocido en el mundo, de los daños materiales y pérdidas humanas que ocasionan los sismos, por tanto, el problema de buscar predecir cuándo y dónde se va a producir constituye la preocupación de los diversos Institutos de investigación en muchos países en el mundo. Es por ello que diversas investigaciones apuntan contribuir en esta dirección. Hasta ahora se han ensayado diversas metodologías, diversos caminos, utilizando información de un tipo u otro, que van desde modelos físico matemáticos, modelos probabilísticos, hasta información de ondas electromagnéticas que se producen en

momentos previos a los sismos, etc. En el Perú, que forma parte de una de las zonas más sísmicas del Pacífico, esta preocupación señalada arriba también forma parte del IGP y de diversos investigadores. En el presente trabajo, que consideramos como un inicio en este campo, exploramos con diversas herramientas estadísticas patrones, en forma similar al que ya se aplicado en otras partes del mundo, en particular con series de tiempo. Previamente se han identificado las zonas sísmicas con mayor actividad.

En esta investigación, sobre el fenómeno de los sismos en el Perú, nos ha guiado la idea que podrían existir ciertos patrones de comportamiento aleatorio, en la ocurrencia de este fenómeno.

Sobre la búsqueda de estos patrones y dada su complejidad en el mundo existen diversas propuestas y teorías, las cuales vamos a presentar a continuación, extraído de El Committee on the Economic Benefits of improved seismic monitoring (7), que resume los principales métodos de pronóstico de sismos.

“El método de brecha sísmica es una aplicación de uso frecuente de este concepto de predicción sísmica a largo plazo y la identificación de potencial sísmico A lo largo de los límites de placas simples, como la falla de San Andreas, la mayor parte del movimiento de la placa, a largo plazo ocurre durante infrecuentes grandes terremotos”.

“Este enfoque fue exitoso y aplicado para varios terremotos grandes ($M_w > 7.5$) a lo largo de las zonas de subducción y límites de placas de deslizamiento durante los años 60 y 70. Aplicación de este método para eventos con $M_w < 7.5$, o en áreas con configuraciones tectónicas complejas, es más controvertido”.

“El patrón repetitivo de acumulación y liberación de tensión también es expresado en patrones de sismicidad, donde los bajos niveles de sismicidad en la primera parte del ciclo (una vez que las réplicas del último evento disminuyan) son seguidos por un aumento en la actividad regional a medida que la placa reacumula y libera en última instancia, por la ocurrencia de otro terremoto con sus respectivos sismos previos y réplicas”.

“Los métodos de reconocimiento de patrones, se basan en estadísticas de cambios, tales como la tasa de ocurrencia de terremotos y la proporción de sismos grandes a pequeños, dentro de una región. Específicamente, el fenómeno que es monitoreado y analizado incluye pequeños sismos que se convierten más frecuentes, en un área que no es necesariamente donde el inminente terremoto ocurrirá; terremotos cada vez más agrupados en el tiempo y espacio; terremotos que ocurren casi simultáneamente a grandes distancias dentro de la región sísmica; y una proporción creciente de magnitud media a terremotos de pequeña magnitud. Estos métodos son utilizados para predecir la ocurrencia de un gran terremoto dentro de un periodo de meses a años, en una región circular grande especificada teniendo un radio de varios cientos de kilómetros”.

“El periodo de tiempo de sucesivas predicciones ha ido disminuyendo a medida que el desarrollo de los métodos han progresado. Los desarrolladores afirman haber predicho un número de terremotos que ocurrieron dentro de grandes regiones circulares: el 1989 $M_w 6.9$ Terremoto de Loma Prieta dentro de un periodo de 5 años, el

1994 Mw 6.7 terremoto de Northridge dentro de un periodo de 18 meses (aunque en realidad ocurrió 21 días después del final de la ventana), y el 2004 Mw 6.5 Terremoto de San Simón dentro de un periodo de 9 meses. Los desarrolladores predijeron que un terremoto de magnitud 6,4 o mayor ocurriría en el sudeste de California antes del 5 de septiembre de 2004, pero no ocurrió tal terremoto".(7)

La presente investigación se realiza, dado el interés público del tema sísmico, el mismo que puede servir para entender mejor el fenómeno y sus tendencias.

Marco teórico y conceptual

Desde los años 60, China ha estado conduciendo sistemáticas observaciones, monitoreando anomalías de los sismos e intentando predecir los terremotos y la aplicación de evaluaciones, dependiendo del tiempo para disminuir los daños que éstos ocasionen. En tales esfuerzos, en algunos casos con éxitos y fracasos, sirven como lecciones en la ciencia de los terremotos, así como en la sostenibilidad social.

Para Pozos, Gómez y Hong (2014), el uso de redes neuronales artificiales es explorado para predecir aceleraciones máximas del terreno y pseudo aceleraciones para sismos de tipo intraslab e interplaca. Un total de 277 y 418 registros sísmicos de dos componentes para sismos de intraslab.

Hung y Nan (2003) utilizan el método bayesiano y usan un modelo epidémico para analizar datos de sismos en Taywan, siguiendo la idea de Ogata (1988). Han realizado estimaciones de riesgo de sismo por zonas con características especiales, y han realizado simulación mediante el método de Montecarlo. El ha usado el estimador de Máxima Verosimilitud (MLE), para analizar y predecir la actividad sísmica a lo largo de Nor-Este de la costa de Japón. La elección de la función de intensidad condicional ha sido descrito ampliamente en Ogata (1988) y varias referencias allí mencionadas.

También Peruggia y Santner (1996) (7) aplicó el método Bayesiano a datos sísmicos en el área de Sannio Matese en Italia, basado en la función de intensidad

$$\lambda(t | \underline{\theta}) = \mu + \sum_{t_i < t} e^{\beta(m_i - M_T)} \alpha e^{-\alpha(t-t_i)}$$

Donde $\theta = (\mu, \alpha, \beta, a)$ son todos parámetros desconocidos.

Un método similar ha sido aplicado en el área de Hualien en el Nor-este de Taiwan por Fan y Li en el año 2002.

Alawadhi y Alhulail (2016) analizan datos de sismos en el El cinturón Zagros el cual es uno de los más sísmicamente activos del mundo. Debido a la ubicación de Kuwait en el suroeste del cinturón de Zagros, se ve afectada por movimientos tectónicos relativos en la región vecina. En su trabajo consideran fundamental evaluar la sismicidad de Zagros riesgos en Kuwait utilizando datos recientes y coordinar con los organismos respectivos para reducir dichos riesgos. Utilizando la magnitud de la onda del cuerpo (Mb) recogidos en Kuwait, evalúan los cambios recien-

tes en la magnitud de los terremotos y sus variaciones en la vecindad de Kuwait. Han construido un modelo de punto de cambio para detectar los cambios significativos en sus parámetros.

Volpe y Piersanti (2016) analizan la posible relación entre los dos sismos ocurridos en Emilia Romagna, Italia, el 20 y 29 de mayo de 2012. Realizan simulaciones en 3D, con datos satelitales de las ondas de los sismos. Asimismo, investigan las posibles conexiones entre las principales Acontecimientos de la secuencia sísmica Emilia del 2012 que ocurrieron los días 20 y 29 de mayo, con el fin de aclarar el papel que la Transferencia de estrés y migración fluida profunda en la ocurrencia del terremoto del 29 de mayo después del primer sismo.

Algunos conceptos

Sismo, también conocido como terremoto, se denomina una sacudida de la tierra que consiste en una serie de vibraciones en la superficie como consecuencia del movimiento de las capas interiores de la tierra.

Localidad más cercana al sismo. Es la localidad que según el IGP esta cerca al epicentro del sismo.

La escala sismológica de Richter (13) o escala de magnitud local (ML), es una escala logarítmica arbitraria que asigna un número para cuantificar la energía liberada en un terremoto, denominada así en honor del sismólogo estadounidense Charles Richter (1900-1985). Escala de Richter.

Fecha del sismo, se refiere a la fecha de la ocurrencia del sismo

La latitud es la distancia angular entre la línea ecuatorial (el ecuador), y un punto determinado de la Tierra, medida a lo largo del meridiano en el que se encuentra dicho punto. Según el hemisferio en el que se sitúe el punto, puede ser latitud norte o sur.

La altitud

Magnitud del sismo, La magnitud de un sismo es un número que busca caracterizar el tamaño de un sismo y la energía sísmica liberada. Se mide en una escala logarítmica, de tal forma que cada unidad de magnitud corresponde a un incremento de raíz cuadrada de 1000, o bien, de aproximadamente 32 veces la energía liberada

Intensidad local, es la intensidad del sismo en referencia a la localidad o localidades mas cercanas.

Metodología

El presente estudio es de tipo exploratorio en la medida en que el fenómeno sísmico es muy complejo y aleatorio y a través de los resultados se pretende conocer sus características y tendencias que serán expresadas en el modelo o los modelos estadísticos.

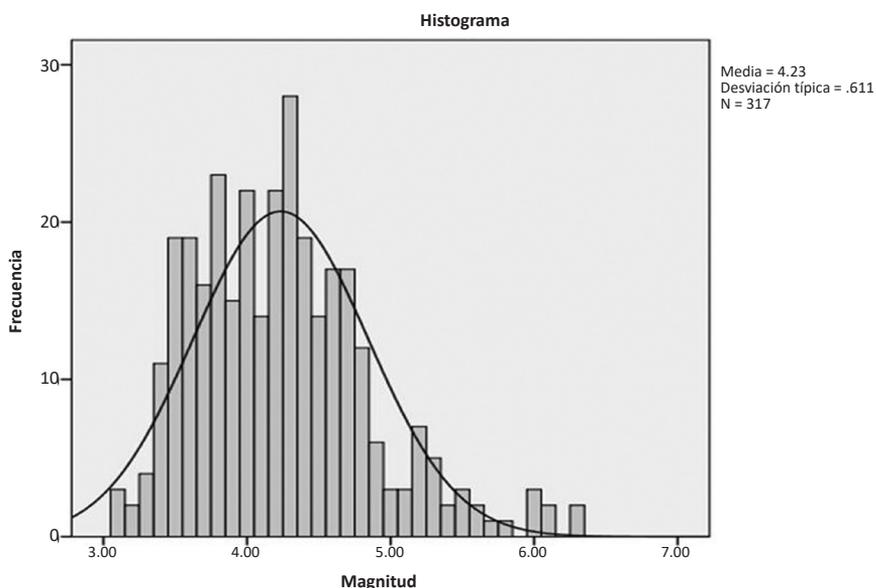
Por otro lado se utilizó información secundaria, los que proporciona el Instituto Geofísico del Perú y otras fuentes, disponibles.

Para realizar el análisis se emplearon diversa herramientas estadísticas, tales como el Cluster Analysis, modelos de Series de Tiempo ARIMA.

Se utilizó el software estadístico SPSS y R.

Resultados y aportaciones

Grafico 1
Distribución de las magnitudes de los sismos 2017



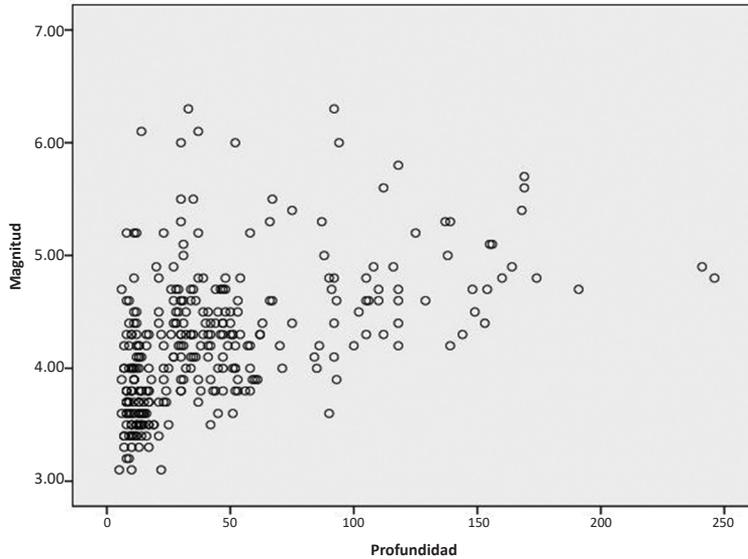
Fuente: Elaboración propia (Datos del IGP magnitud de los sismos de Enero a Noviembre 2017).

Se puede observar que los sismos en el Perú, se distribuyen aproximadamente como una normal, según el gráfico y el cuadro 1-A del Anexo, que muestra la prueba de Kolmogorov – Smirnov para la distribución normal.

Relación entre la magnitud y la profundidad del sismo

Gráfico 2

Relación entre la magnitud y la profundidad



Fuente: Elaboración propia

Tabla 1

		Profundidad
magnitud	Correlación de Pearson	0.492
	Sig. (bilateral)	0
	N	317

Fuente: Elaboración propia

La correlación entre la magnitud de los sismos y la profundidad no es alta, como se muestra en la tabla 1. Tampoco esta correlación lineal no es tan buena (tabla 2).

Tabla 2

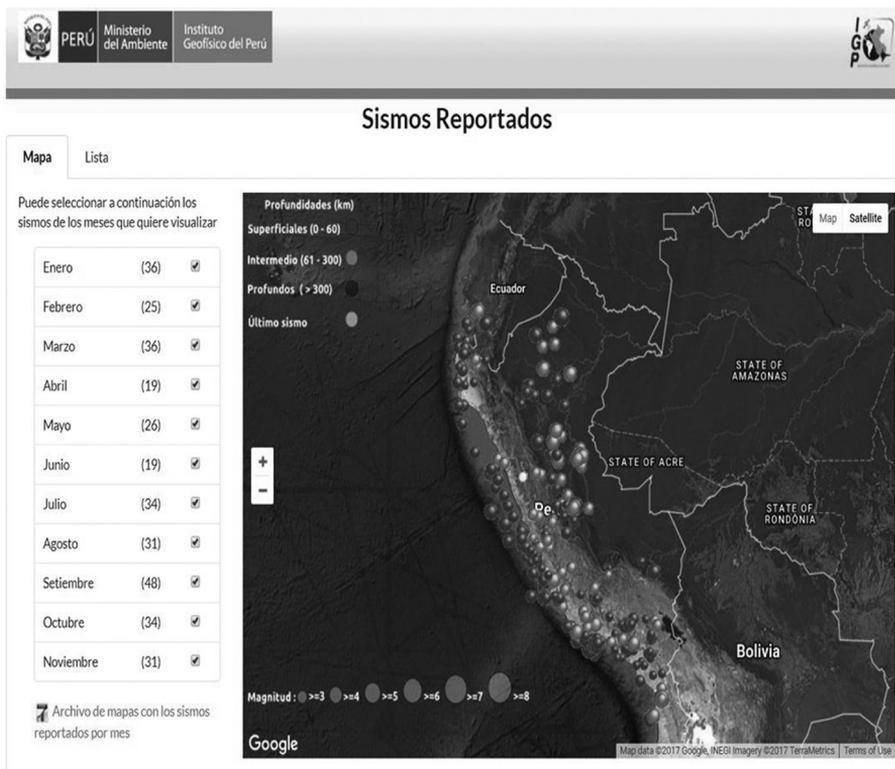
Descriptivos para la serie magnitud						
	N	Rango	Mínimo	Máximo	Media	Desv. típ.
magnitud	317	3.20	3.10	6.30	4.2341	.61137
Profundidad	317	241	5	246	44.13	42.954
N	317					

Fuente: Elaboración propia

A nivel nacional el promedio de la magnitud de los sismos de enero a noviembre del 2017 fue de 4.23 en la escala de Richter y la profundidad fue de 44.13 Km.

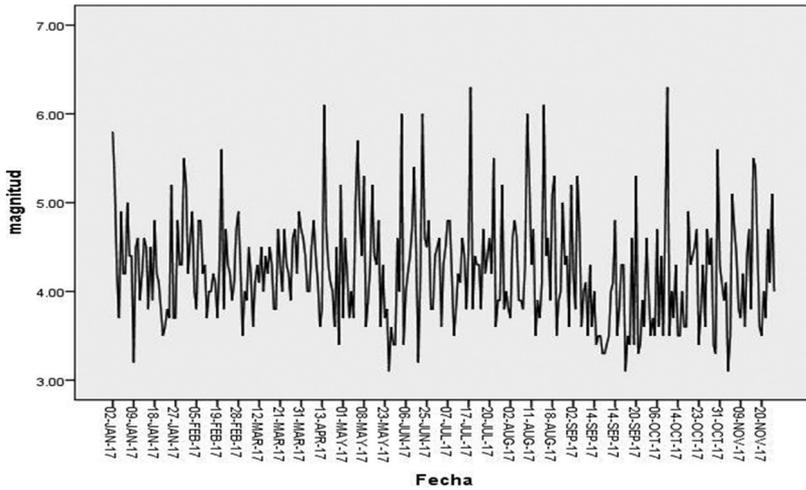
Gráfico 3

**Reporte de los sismos del 2017 por parte del Instituto Geofísico del Perú (IGP)
30-11-17**



En este mapa de los sismos reportados por el IGP, se puede observar la mayor concentración en la zona de Arequipa, seguido de Lima-Ica.

Gráfico 4
Serie de la magnitud de los sismos del 2017



Fuente: Elaboración propia.

En este gráfico se puede observar la oscilación de la magnitud de los sismos, de enero a noviembre de 2017.

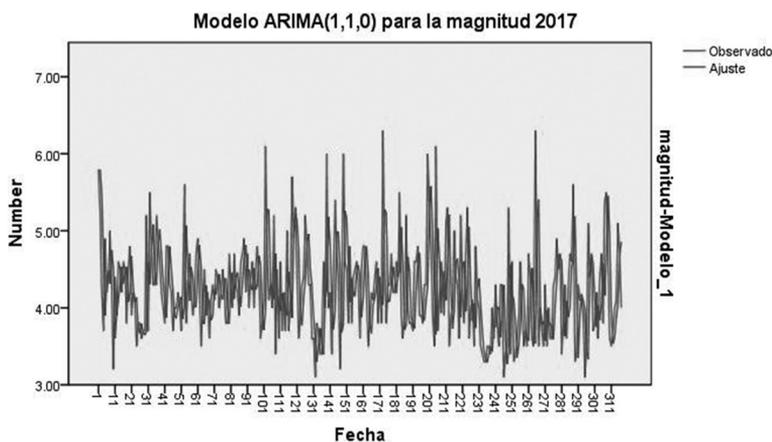
Luego de ensayar con varios modelos de series de tiempo, hemos encontrado el modelo ARIMA(1,1,0), el cual se ajusta mejor a los datos de la magnitud de los sismos en el Perú.

Tabla 3

Estadísticos del modelo ARIMA (1,1,0)					
Modelo	Estadísticos de ajuste del modelo		Ljung-BoxQ(18)		Número de valores atípicos
	R-cuadrado estacionaria	Estadísticos	GL	Sig	
magnitud	.166	54.076	17	.000	0

Fuente: Elaboración propia

Gráfico 5
La magnitud observada y la ajustada



Fuente: Elaboración propia

Este gráfico se puede observar un cierto buen ajuste entre la serie observada y la serie ajustada. Lo anterior nos plantea la siguiente interrogante, existe un patrón de comportamiento de la serie de los sismos en el Perú en el 2017?

Análisis de grupos y sus características

Para hallar los conglomerados, se ha utilizado la herramienta Cluster Analisis, con distancia euclidea entre los grupos, y con la ayuda del grafico Dendograma, detectamos los 8 grupos.

Tabla 4
Características de los sismos por grupo (2017)

Grupo	Nº de sismos	Ciudad	Magnit prom	Prof prom	Desv. Tip Mag	Desv. Tip Prof
1	14	Moyobamba	4.81	72.79	0.66432	63.011
2	7	Tacna	4.8	62.14	0.92556	38.321
3	91	Lima - Ica	4.07	41.87	0.45088	28.933
4	127	Arequipa	4.04	32.9	0.59831	41.694
5	19	Chimbote	4.59	45.05	0.34302	15.629
6	32	Pucallpa	4.66	80.25	0.57283	67.431
7	25	Sullana	4.47	43.8	0.52144	19.17
8	2	San Gaban	4.5	12	0.98995	5.657
Total	317					

Fuente: Elaboración propia.

Respecto a la magnitud de los sismos, Moyobamba y Tacna presentan los mayores promedios, sin embargo se puede observar que no difieren de los demás grandemente. También son los mismos quienes presentan mayor varianza.

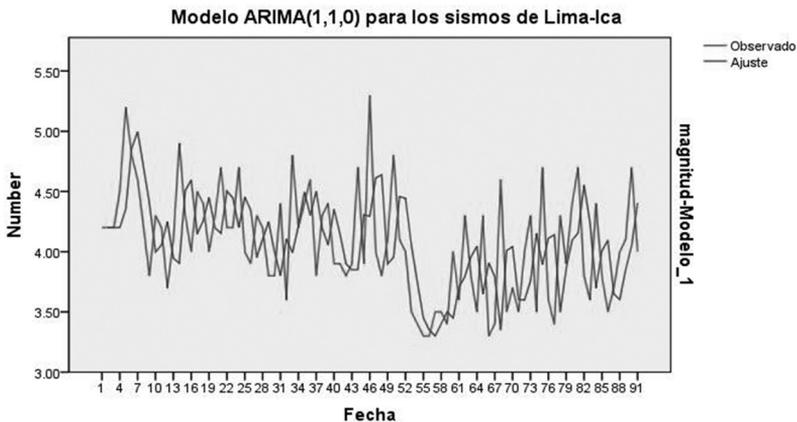
1. Un análisis para el grupo de Lima - Ica

Tabla 5

Estadísticos del modelo ARIMA (1,1,0)					
Modelo	Estadísticos de ajuste del modelo		Ljung-BoxQ(18)		Número de valores atípicos
	R-cuadrado estacionaria	Estadísticos	GL	Sig	
magnitud	.242	34.557	17	.007	0

Fuente: Elaboración propia

Gráfico 6



Fuente: Elaboración propia

En el caso de los sismos del grupo de Lima-Ica, se observa que a partir de la mitad de año, (septiembre) la tendencia es a una disminución de la media de la serie. El ajuste es relativamente bueno, sin embargo que se ajuste a un modelo requiere de un análisis más especializado desde el punto de vista del sismólogo?

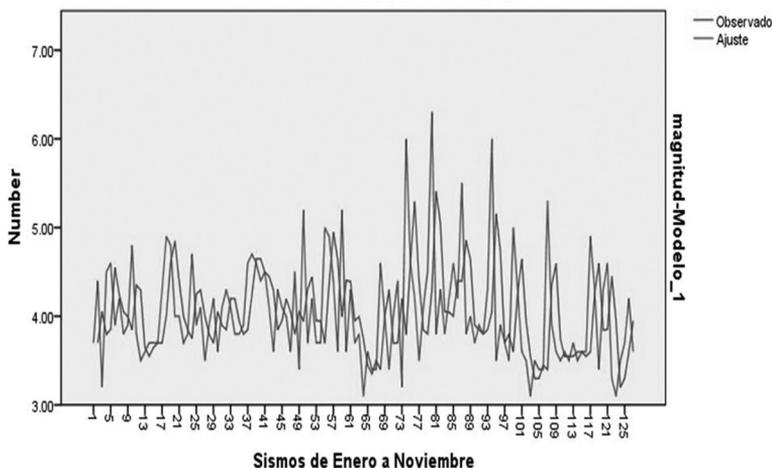
2. Un análisis para el grupo de Arequipa

Tabla 6

Estadísticos del modelo ARIMA (1,1,0)					
Modelo	Estadísticos de ajuste del modelo		Ljung-BoxQ(18)		Número de valores atípicos
	R-cuadrado estacionaria		Estadísticos	GL Sig	
magnitud	.429		39.471	17 .002	0

Fuente: Elaboración propia

Gráfico 7
Modelo ARIMA(1,1,0) Para Arequipa



Fuente: Elaboración propia

En el caso de Arequipa, no se presenta la tendencia a disminuir la media como es el caso de Lima-Ica, pero es parecido en el sentido del relativo buen ajuste con el modelo ARIMA(1,1,0).

Discusión de resultados

Dada la complejidad de los datos de los sismos en el Peru, sin embargo nos ha guiado la búsqueda de patrones de comportamiento, de los resultado obtenidos en la sección anterior se puede observar lo siguiente:

- Para el conjunto de datos de los sismos del 2017, se ha constatado una relación lineal significativa entre la profundidad del sismo y la magnitud.

- Asimismo, se ha ajustado a los datos de la magnitud de los sismos, un modelo ARIMA(1,1,0), el cual ha resultado significativo, tanto a nivel nacional, así como para el caso de Lima-Ica y Arequipa.
- También se ha encontrado ocho grupos de zonas sísmicas, evaluadas por las distancias. Siendo el grupo de Arequipa, seguido de Lima-Ica, como los que presentan mayor actividad sísmica.
- Para el conjunto de datos de los sismos del grupo de Lima-Ica
- Se ha encontrado una tendencia lineal negativa de la magnitud de los sismos.
- Y respecto a la profundidad, son Pucallpa y Moyobamba, quienes presentan la mayor profundidad promedio 80,25 km y 72,29 km, respectivamente, y también la mayor varianza, 67 km y 63 km.
- Es la selva la que presenta la mayor profundidad y la mayor magnitud promedio.
- Hemos encontrado que la magnitud promedio de Arequipa y de Lima-Ica son bastantes parecidos 4,04 y 4,07 y su profundidad promedio no difieren grandemente, 32,9 km y 41,87 km, respectivamente.

Conclusiones

1. Hemos encontrado ocho zonas de actividad sísmica diferenciados, de las cuales Arequipa y Lima son las más activas.
2. El modelo de series de tiempo ARIMA(1,1,0) de las magnitudes de los sismos de Enero a Noviembre del 2017, ha resultado significativo a nivel nacional, y también ha sido significativo para Lima-Ica y Arequipa.
3. En el grupo de Lima, existe una tendencia hacia una pequeña disminución en el nivel de la media de la serie, lo que podría estar indicando la cercanía de un sismo de mayor magnitud en los próximos meses.

Referencias bibliográficas

- Tsay Hung Fan and Eng Nan Kuo " A Bayesian analysis for seismic data on Taywan. 2003.
- Pozos-Estrada, A, Gómez, R., y Hong, H.P (2014). "Use of Neural network to predict the peak ground accelerations and pseudo spectral accelerations for Mexican Inslab and Interplate Earthquakes".
- Fuqiong Huang*, Mei Li, Yuchuan Ma, Yanyan Han, Lei Tian, Wei Yan, Xiaofan Li. "Studies on Earthquake precursors in China", 2016.
- Pecker, A., Fassioli, E. An overview on the research SIGMA Proyect. Edit Springer 2017.

Ogata 2005, Anomalies of Seismic Activity and Transient Crustal Deformations Preceding the 2005 M 7.0 Earthquake West of Fukuoka.

Committee on the Economic Benefits of Improved Seismic Monitoring Improved seismic monitoring decision making. Assessing the Value of Reduced Uncertainty The National Academies press. Washington, D.C. 2006

“Escala de Richter”. En: Significados.com. Disponible en: <https://www.significados.com/escala-de-richter/>

Douglas C. Montgomery , Cheryl L. Jennings , Murat Kulanci, 2015, Introduction To Time Series Analysis and Forecasting.

John K. Kruschke, Doing Bayesian Data Analysis: A Tutorial with R and BUGS, 2010

Mike West, Jeff Harrison, Bayesian Forecasting and Dinamic Models. 1989.

Volpe, M. y Piersanti, A. (2016)The role of fluid migration and static stress transfer in searching connections between the May 2012 Emilia earthquakes through a fully 3D finite element modeling. DOI 10.1007/s40808-016-0087-9

ANEXO 1

Este tipo de gráficos, nos sirve para identificar que tipo de modelo de series de tiempo podemos ensayar ajustar a la serie de la magnitud de los sismos.

Tabla 1-A

Prueba de Kolmogorov-Smirnov para una muestra		
		magnitud
N		317
Parámetros	Media	4.2341
normales ^{a,b}	Desviación típica	0.61137
	Absoluta	0.082
Diferencias más extremas	Positiva. Positiva más extremas	0.082
	Negativa	-0.058
Zde Kolmogorov-Smirnov		1.454
Sig. asintót. (bilateral)		0.029

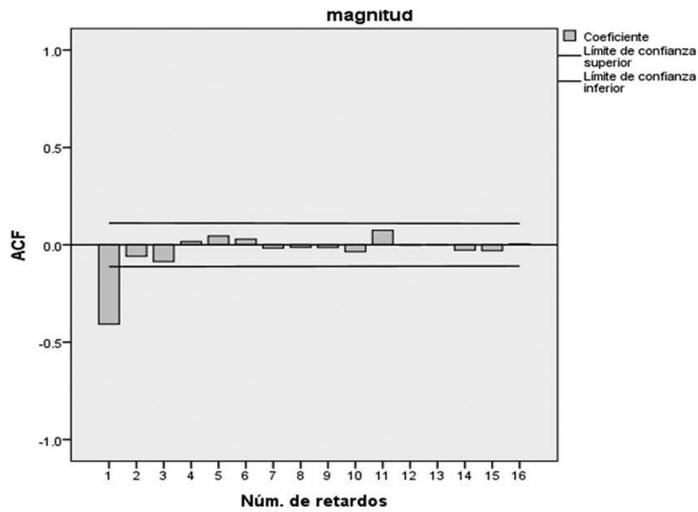
a. La distribución de contraste es la Normal.
b. Se han calculado a partir de los datos.

Fuente: Elaboración propia

Un análisis de series de tiempo para la magnitud a nivel nacional

Gráfico A-1

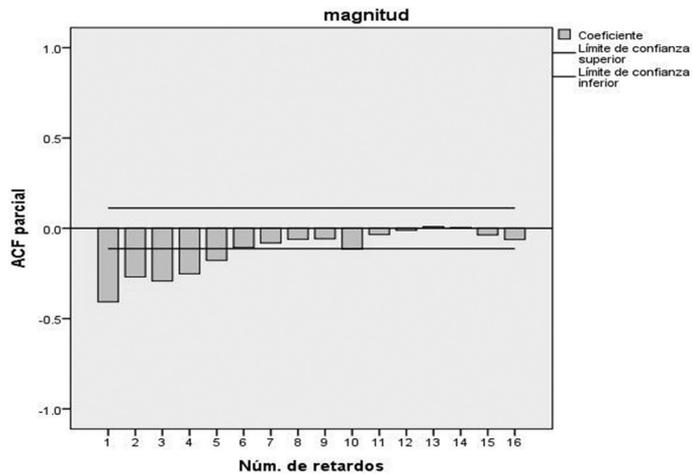
La función de autocorrelación de la magnitud



Fuente: Elaboración propia

Gráfico A-2

La función de autocorrelación parcial de la magnitud a nivel nacional



Fuente: Elaboración propia

Tabla A- 2
Descriptivos por grupos

Grupos	Variable	N	Mínimo	Máximo	Media	Desv. típ.
Moyobamba	magnitud	14	3.60	6.10	4.8143	.66432
	Profundidad	14	13	174	72.79	63.011
Tacna	magnitud	7	3.60	6.30	4.8000	.92556
	Profundidad	7	9	110	62.14	38.321
Lima-Ica	magnitud	91	3.30	5.30	4.0736	.45088
	Profundidad	91	6	129	41.87	28.933
Arequipa	magnitud	127	3.10	6.30	4.0402	.59831
	Profundidad	127	5	246	32.92	41.694
Chimbote	magnitud	19	4.10	5.50	4.5895	.34302
	Profundidad	19	27	91	45.05	15.629
Pucallpa	magnitud	32	3.90	6.10	4.6656	.57283
	Profundidad	32	8	191	80.25	67.431
Sullana	magnitud	25	3.80	6.00	4.4760	.52144
	Profundidad	25	11	93	43.80	19.170
San Gaban	magnitud	2	3.80	5.20	4.5000	.98995
	Profundidad	2	8	16	12.00	5.657

Fuente: Elaboración propia.