

Modelamiento prospectivo de la tendencia de adquirir casa propia o usada, aplicando series de tiempo como método de pronóstico a partir de datos suministrados por la herramienta Google Trends

Prospective modeling of the tendency to buy the house or used with the application of the time series as a forecast method based on data provided by the Google Trends tool

Elkin Orlando Vélez Sánchez*

Tania Paola Ortiz González**

Alejandra Marín Betancur***

Ángela Mayellis Melo Hidalgo****

Nelson Octavio Casas Atehortua*****

Guillermo Giraldo Vargas*****

Nelson Marín Latorre Arias*****

Equipo de trabajo ECACEN Zona Occidente

* Ingeniero industrial, Especialista en Administración, Magíster en Gerencia de la Universidad Nacional Abierta y a Distancia (UNAD). Número ORCID <https://orcid.org/0000-0003-2903-7061>

** Administradora de Empresas, Magíster en Administración de Organizaciones de la Universidad Nacional Abierta y a Distancia (UNAD). Número ORCID <https://orcid.org/0000-0001-9321-6908>

*** Administradora de Empresas, Especialista en Gerencia y Prospectiva, Magister en Gestión de la Innovación. Número ORCID <https://orcid.org/0000-0002-9608-2359>

**** Administradora de Empresas, Magister en Administración de Organizaciones de la Universidad Nacional Abierta y a Distancia (UNAD).

***** Administrador de Empresas. Número ORCID <https://orcid.org/0000-0002-1172-5475>

***** Administrador de Empresas, Especialista en Gestión de Costos, Magíster en Administrador de Organizaciones de la Universidad Nacional Abierta y a Distancia (UNAD). Número ORCID <https://orcid.org/0000-0003-2695-4745>

***** Economista, Especialista en Gerencia Empresarial, Magister en Administración de Negocios. Universidad Católica Luis Amigó

Introducción

La tradicional frase “quien tiene la información tiene el poder”, en estos tiempos de auge de la Internet parece sufrir un cambio a “quien entiende la información tiene el poder”. Actualmente todo ser humano con acceso a la Web puede consultar, aprender y visitar en tiempo real los sitios a los cuales no ha podido ir; así, por ejemplo, a través de la herramienta libre Google Trends se puede acceder a una diversidad de variables de búsquedas que utilizamos los habitantes de la tierra, sin ser de uso exclusivo de un país, organismo, edad o estrato social.

Al acceder a Google Trends, la herramienta genera un historial a partir de las búsquedas en la Web que realiza la población mundial o un determinado país. Esto da origen a una mega cantera de datos (Big Data), a través de la cual se pueden formular análisis en torno a aspectos como, por ejemplo, el interés de la población en momentos o temas específicos, con lo cual se pueden identificar líneas de actuación masivas. Estas últimas pueden ser puntos de partida para abordar cuestiones más profundas gracias a la correlación de variables, en especial cuando se construyen patrones específicos en un modelo planteado, como pueden ser terapias de medicina alternativa y complementaria, la tendencia de comprar carro nuevo o usado, la dispersión de una enfermedad, entre otros (Rocha y Correia, 2013).

La herramienta Google Trends se ha utilizado ampliamente en los últimos años, durante los cuales se ha demostrado su aplicación en modelos de pronóstico en diversas áreas como la medicina, propagación de enfermedades y compra de carros. No obstante, su mayor aplicación se encuentra en el marketing y la vigilancia tecnológica, donde pueden identificarse tendencias de compra y de búsqueda de tecnología o productos a nivel mundial, nacional, regional o local.

Las series de tiempo, como herramienta de pronóstico, son básicamente la sucesión de los eventos convertidos en variables, ocurridos aleatoriamente con el transcurrir del tiempo; esto da como resultado un conjunto de datos que se pueden tratar estadísticamente arrojando tendencias que se pueden

modelar matemáticamente, entregando así modelos que se pueden utilizar para pronosticar los posibles valores que tomará la variable en periodos posteriores a los relacionados en los datos iniciales.

Como caso de aplicación, se desea identificar la tendencia de compra de casa nueva o usada en el periodo 2004 a agosto de 2017 para Colombia, con el fin de analizar patrones de búsqueda y posibles modelos de pronóstico propuestos para cada conjunto de datos “casa nueva” y “casa usada”, consideradas como las frases de búsqueda del interés en la investigación, utilizando las series de tiempo como herramienta generadora de modelos de pronóstico que, pese a ser complejas, presentan la posibilidad de evaluar la pertinencia o no de su aplicación. Los cálculos se realizan en Microsoft Excel, debido a la facilidad para construir el modelo y las funcionalidades del Solver, como lo son: la posibilidad de seleccionar entre dos algoritmos de optimización (GRG con reinicios aleatorios múltiples y algoritmos evolutivos) y el manejo de restricciones (Sánchez y Martínez, 2009).

Marco teórico

La tecnología, en especial las TIC, son una herramienta fundamental para la sociedad actual, en la que hay una ola de información que los investigadores pueden analizar periódicamente. De esta forma, Google Trends cuantifica la búsqueda con base en las consultas que los internautas realizan día a día, a través de un computador, smartphone o cualquier otro dispositivo que utilice como ruta de comunicación la Web.

La búsqueda de información en la Web involucra una gran cantidad de fuentes, tales como libros, diarios, revistas, documentales, grandes pensadores, etc.; siendo la Internet una gran biblioteca que agrupa dicha información. Es claro que para poder acceder a estos contenidos, además del acceso a equipos y una conexión a la Web, también se deben adquirir estrategias que permitan hallar información adecuada, es decir, tener la capacidad de discernir cuáles son los datos pertinentes y confiables que arroja la búsqueda (Ibáñez y Ponce, 2016).

En el estudio del comportamiento de la evolución de las variables con respecto al tiempo, es esencial tener en cuenta los procesos estocásticos, que son, en esencia, fenómenos aleatorios que dependen solo del tiempo. Según Rincón, “un proceso estocástico es una colección de variables aleatorias parametrizada por un conjunto T , llamado espacio parametral, en donde las variables toman valores en un conjunto S llamado espacio de estados” (Rincón, 2012). En el presente estudio se identifica que los datos extraídos de Google Trends tienen un comportamiento estocástico, toda vez que provienen de búsquedas que se dieron en un tiempo determinado conformando un espacio muestral de estados (Rincón, 2012).

La correlación de variables es una técnica estadística que busca determinar la posible relación entre dos o más variables, dando origen a modelos predictivos óptimos que parten de estadísticos propios que sirven como parámetros de evaluación de la pertinencia o no del modelo matemáticos para predecir el fenómeno de estudio. Uno de los más utilizados es el estadístico R^2 o de Pearson, que en regresión lineal presenta el grado que el modelo explica la realidad, se mide entre 0 a 1, entre más se acerque a cero más mal es el modelo, mientras que, si se acerca a 1, el modelo es apropiado y por ende se detecta una mayor correlación entre las variables (Ramón, 2015).

Las series de tiempo son colecciones de datos con comportamientos estocásticos, que se obtienen de mediciones de un evento aleatorio de cualquier tipo, tratados a lo largo del tiempo a intervalos temporales idénticos. Con el estudio de estas se busca construir un modelamiento matemático que pronostique los comportamientos futuros que pueden ser utilizados en la toma de decisiones, como es el caso de compra de casa nueva o usada. El estudio de las series puede mostrar patrones característicos de la misma como estacionalidad, ergódica, con tendencia definida o no. Hay diversas series de tiempo que tienen aplicaciones en la economía, física, geofísica, marketing, demografía, etc. Se analiza en una serie de tiempo la tendencia, que consiste en identificar en la serie si el valor de la variable tiende a disminuir o aumentar en un periodo largo. La fluctuación puede presentar

ciclos, estacionalidades o recurrencias, a partir de las cuales se identifican los patrones específicos de la serie (Patiño Jiménez y Silva Salcedo, 2016).

Metodología

Para la toma de datos se realizó un estudio de las tendencias de búsqueda mediante la herramienta Google Trends en español, utilizando como punto de partida las palabras “casa nueva”, “casa usada”; en primera instancia, la búsqueda se hizo de manera global y luego delimitada a Colombia. Se evaluó la asociación entre las dos tendencias de datos mostrados por la herramienta de búsqueda.

El análisis sectorial de la actividad de compra de vivienda se realizó bajo la óptica de la proyección de ventas o incidencias que esta puede tener en el impacto sectorial de la construcción, con base en datos extraídos, en su mayoría, de las intenciones de ventas de las constructoras o del Departamento Nacional de Estadísticas de Colombia (DANE).

La metodología que aquí se propone contribuye a la creación de un modelo exploratorio de la actividad de búsqueda de casa nueva o usada, por parte de los internautas. Lo anterior, suponiendo que cuando el usuario realiza este tipo de búsqueda, lo hace por un interés particular en comprar vivienda o visualizar ofertas, lo cual genera indicadores sociales y sectoriales con los que se puede visualizar, de forma global, la intención de compra o venta en periodos determinados del año. A su vez, estos pueden ser un indicador para reducir la incertidumbre que genera la ausencia de datos al momento de comprar o vender un bien raíz de este tipo (Mejía y Monsalve, 2013).

Se procede a realizar un análisis de los datos obtenidos mediante series de tiempo que incluyen cuatro componentes específicos: 1) análisis de la tendencia, 2) estacionalidad o recurrencia, 3) heterocedasticidad y 4) búsqueda de la variación alrededor de la línea de tendencia, para luego seleccionar el mejor modelo de pronóstico, según sea el caso (Patiño Jiménez y Silva, 2016).

Resultados

Se observa una tendencia descendente en el volumen de búsquedas de casa nueva, casa usada, arrojada por la página Google Trends, estos serán los insumos para los análisis de series de tiempo siguientes (Méndez, 2013).

Se analizan los datos tomados de Google Trends para las frases “casa nueva” y “casa usada”, primero a nivel mundial y luego para Colombia.

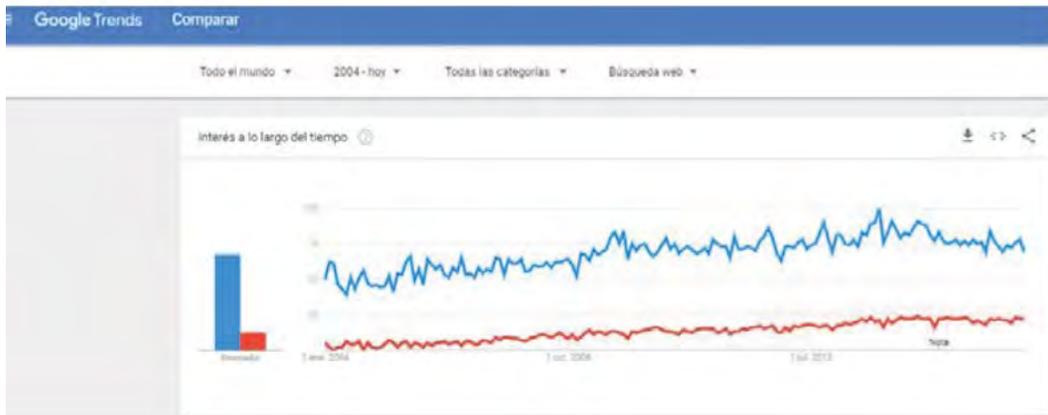


Figura 1. Censo para las frases “casa nueva” Vs “casa usada”

Fuente: Google Trends (10 de septiembre de 2017)

En la figura 1 se observa cómo en el mundo las búsquedas de casa nueva (línea azul) son superiores a las de casa usada (línea roja), con una tendencia creciente aproximadamente lineal cada una. Las dos series carecen de estacionalidad, pero se evidencia que en algunos periodos pasa de picos muy altos a muy bajos, presentando la serie de casa nueva un mayor índice en este aspecto. El histograma de la izquierda muestra que casa nueva tiende a ser muy superior a casa usada.

Resultados de la búsqueda para Colombia entre 2004 y septiembre de 2017 (figura 2).

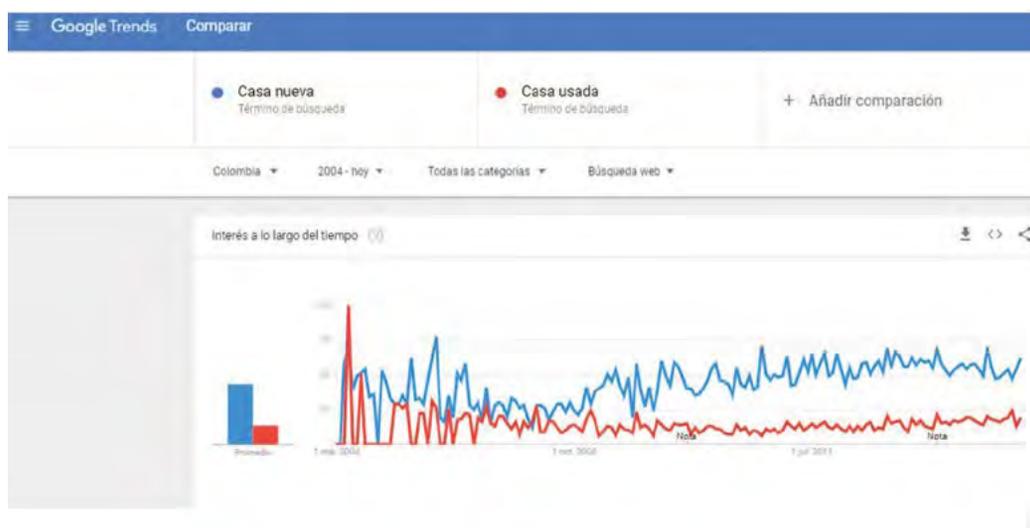


Figura 2. Censo para las frases “casa nueva” Vs “casa usada” en Colombia

Fuente: Google Trends (10 de septiembre de 2017)

Para Colombia, la tendencia de las series tiende a ser idéntica a la de búsquedas a nivel mundial, salvo que se nota más la estacionalidad de aproximadamente un semestre. Debido a la expansión paulatina del acceso a la Internet, es notorio que para el periodo comprendido entre 2004 y 2008, se carece de tendencia y la estacionalidad es mucho más marcada, aunque, sin contar con un patrón específico de regularidad, las series se acercan bastante y en algunos periodos tienden a no alejarse tanto una de la otra. Pese a esto, a lo largo de los periodos se conserva la tendencia mundial de estar muy por encima la búsqueda del término “casa nueva”, comparada con el de “casa usada”.

A continuación se expone el análisis de los datos, según la serie de tiempo para los datos de Colombia, utilizando las herramientas de Excel y Statgraphics (figura 3).

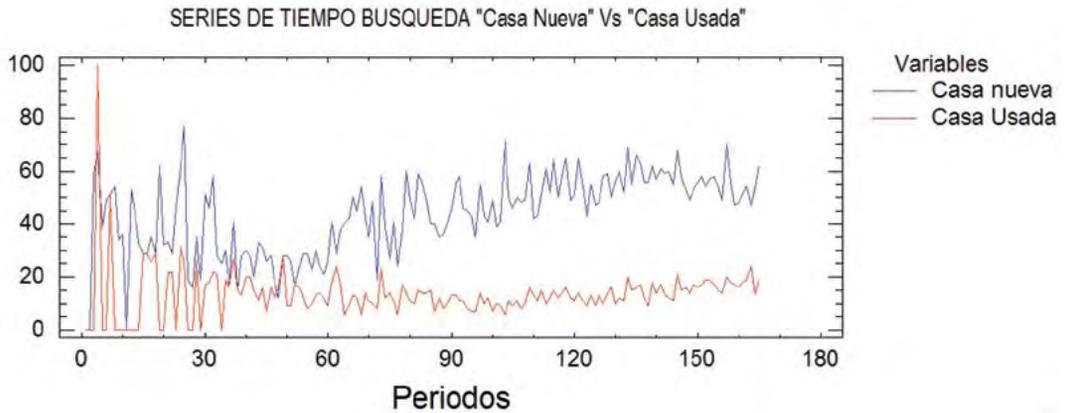


Figura 3. Series de tiempo tratadas con Statgraphics

Fuente: Google Trends (10 de septiembre de 2017)

En primera instancia se debe hacer el ajuste estacional de la serie, que consiste en aislar los efectos que pueden tener los datos por factores exógenos, con lo cual se minimizan las tendencias de desarrollo propias de la Internet en Colombia. Esto se puede hacer mediante un análisis de regresión que consiste en observar la gráfica y ver que no es posible hallar una correlación entre las variables, por lo que dicho modelo no es aplicable. En consecuencia se puede utilizar el método de media móvil, también conocido como de medias móviles. Este último utiliza las medidas para hallar el promedio de búsquedas del subgrupo de un periodo en particular como patrones de cálculo del periodo estacional, contrastándolo con el promedio total (Ariza, 2013).

Con los datos extraídos directamente de la página Google Trends se procede a modelar la serie en Excel. En este programa se pueden realizar algunos ajustes y tratamientos estadísticos que nos permitirán aplicar uno de los modelamientos o suavisamientos propios de las series de tiempo que permitan generar pronósticos.

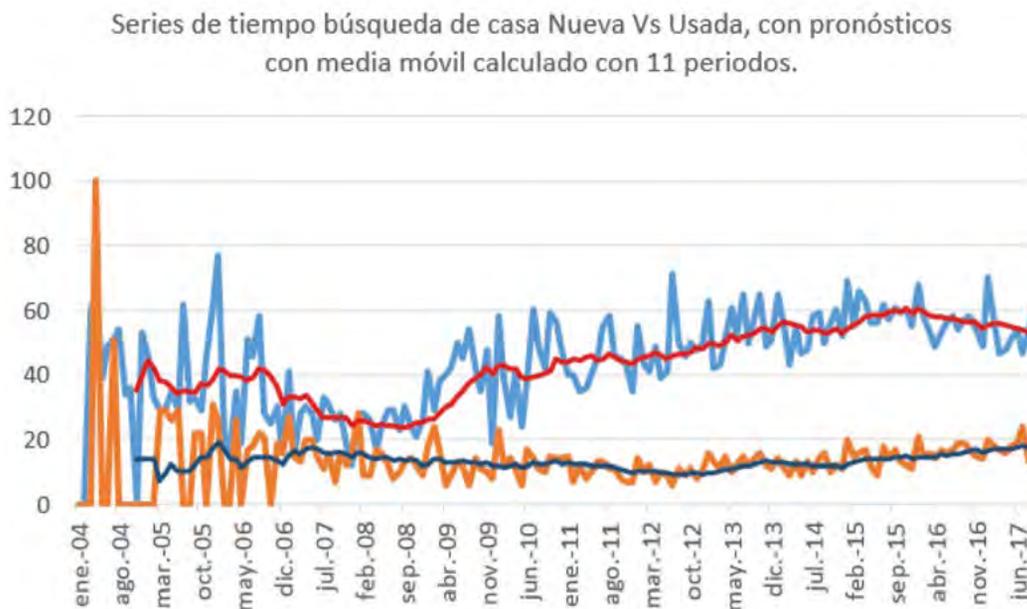


Figura 4. Serie de datos del censo para las palabras “casa nueva” Vs “casa usada” para Colombia, con líneas de pronóstico suavizadas con el método de media móvil

Fuente: Google Trends (10 de septiembre de 2017)

La figura 4 muestra las líneas suavizadas de tendencia del modelo de pronóstico, tanto para la búsqueda de casa nueva como para la de casa usada, con un rango de rezago de 11 periodos. Allí se puede evidenciar que la tendencia a buscar casa nueva es mucho mayor que la de casa usada. En las siguientes figuras se observa más de cerca esta tendencia y cómo los pronósticos siguen el patrón, generando un resultado aceptable según la confiabilidad del método media móvil.



Figura 5. Serie de datos del censo para las palabras “casa nueva” (en Colombia), con líneas de pronóstico suavizadas con el método de media móvil y pronósticos hasta mayo 2018

Fuente: Google Trends (10 de septiembre de 2017)

En esta gráfica aislada se puede observar como la tendencia de búsqueda de casa nueva va en aumento. La línea roja describe los datos pronosticados, suavizados y desestacionalizados, utilizando el método de media móvil con 11 periodos de rezago, entregando unos pronósticos ajustados para los periodos hasta el mes de mayo de 2018 (Méndez, 2013).



Figura 6. Serie de datos del censo para las palabras “casa usada” (en Colombia), con líneas de pronóstico suavizado con el método de media móvil y pronósticos hasta junio de 2018

Fuente: Google Trends (10 de septiembre de 2017)

En la gráfica se puede observar cómo la tendencia de búsqueda de casa usada va en aumento en Colombia. La línea roja describe los datos pronosticados, suavizados y desestacionalizados, utilizando el método de media móvil con 11 periodos de rezago, entregando unos pronósticos ajustados para los periodos hasta el mes de junio de 2018.

Conclusiones

La realización de esta investigación permitió demostrar que se pueden aplicar los datos de búsqueda de Google Trends, complementados con el modelo de estudio de las series de tiempo, como una herramienta exploratoria viable para profundizar en patrones más específicos, como es el caso de la compra de vivienda (teniendo en cuenta el estrato, costo, o sector). Además, al correlacionar las variables de búsqueda se puede determinar hasta qué punto una

variable, por ejemplo el costo, puede inferir en el interés de encontrar casa nueva. Con este análisis se pueden obtener predicciones confiables y determinar prospectivas para el año siguiente, con respecto a la tendencia de adquirir casa nueva para el 2018 por parte de la población colombiana. En este sentido, se cumple con el objetivo de la presente investigación, dando como respuesta la aplicación de los modelos de series de tiempo a los datos temporales que arroja Google Trends.

El impacto tecnológico y social sobre la información que arroja Google Trends es fundamental al momento de realizar una prospectiva con base en el comportamiento de las diferentes series de tiempo que se pueden extraer. De esta forma se puede profundizar, aún más, en los posibles patrones que la pueden rodear, al poder comparar el comportamiento de éstas bajo la observación directa de sus variables altamente correlacionadas.

La teoría que aporta los modelos de series de tiempo es apropiada para la descripción y realización de pronósticos (tendencias) con base en los datos que arroja Google Trends. Esta, además, puede ser una herramienta de gran valor para diversos campos de la investigación en ciencias sociales, económicas, de mercadeo u otras áreas.

Para esta investigación se tomó como método de suavizamiento y pronóstico las medias móviles, por ser de fácil aplicación; sin embargo, este método tiene una gran limitante: sus predicciones solo son buenas a corto plazo. En consecuencia, otras investigaciones podrán incorporar modelos de mayor envergadura, tales como los ARIMA.

Los datos de búsqueda que nos entrega de forma gratuita Google Trends se transforman en una herramienta fundamental al momento de analizar, de forma exploratoria, las diversas búsquedas de variables que la humanidad tiende a consultar con respecto a determinados temas. A partir de allí se pueden establecer tendencias que, a su vez, permiten construir modelos matemáticos o estadísticos para hacer proyecciones sobre temáticas específicas. Para el caso de la compra de casa propia o usada, el desarrollo de la investigación

mostró como el modelo de series de tiempo, generado a partir de las medias móviles, plantea unos pronósticos coherentes con el desarrollo históricos de los datos de búsqueda, los cuales pueden ser el punto de partida para investigaciones más profundas de marketing, entre otras.

Referencias

Ariza, R. (2013). *Métodos utilizados para el pronóstico de demanda de energía eléctrica en sistemas de distribución*. Pereira, Colombia: Universidad Tecnológica de Pereira.

Gómez, R. (2002). *Aprendizaje organizativo y sistemas complejos con capacidad de adaptación*. Mexico: Universitat Jaume.

Gorgas, J., Cardiel, N. y Zamorano, J. (2011). *Estadística básica para estudiantes de ciencias básicas*. Madrid, España: Universidad de Complutense.

Ibáñez, L. y Ponce, E. (2016). Uso de la Web e Internet como herramientas para la búsqueda de información médica científica. En *Archivos en Medicina Familiar – UNAM*, vol. 18, pp. 95-106.

Mejía, L. y Monsalve, D. (2013). Indicadores ISAAC: Siguiendo la actividad sectorial a partir de Google Trends. En *Notas Fiscales Ministerio de Hacienda y Crédito Público*, 2 - 25.

Méndez, B. (2013). *Series de tiempo: una aplicación a registros hidrométricos en una cuenca del Estado de Oaxaca*. Mexico: UNAM.

Patiño, L. y Silva, D. (2016). *Análisis de series de tiempo*. Cali, Colombia: Univalle.

Ramón, S. (2015). *Correlación entre variables. Apuntes de clase del curso Seminario Investigativo VI*. Medellín, Colombia: UdeA.

Rincón, L. (2012). *Introducción a los procesos estocásticos*. Mexico: UNAM.

- Rocha, B. y Correia, S. (2013). Google Trends como herramienta de investigación: un estudio aplicado a la Copa del Mundo Brasil 2014. *Gestão estratégica da inovação, criatividade e competitividade* [en línea]. Disponible en https://www.researchgate.net/profile/ari_mariano/publication/276410342_google_trends_como_herramienta_de_investigacion_un_estudio_aplicado_a_copa_del_mundo_brasil_2014/links/5558c1aa08ae6943a876acc8/google-trends-como-herramienta-de-investigacion-un-estud
- Rodríguez, J. (2012). *Trabajo de grado para obtener el título de Profesional en Finanzas y Comercio Internacional*. Bogotá: Universidad Colegio Mayor de Nuestra Señora del Rosario.
- Russell, B. (2003). *Conocimientos actuales sobre nutrición*. Washington, D.C., EUA: Organización Panamericana de la Salud.
- Sánchez, M. & Martínez, R. (2009). Aplicación del modelo de series de tiempo a la proyección del servicio de referencia en línea de la Biblioteca del IIMAS-UNAM. En *Ibersid*, vol. 17, pp. 1-13.