



PLANTILLA DE PRESENTACIÓN DE ABSTRACTS PARA COMUNICACIÓN ORAL

III ReD-N – 20 JULIO 2021

TÍTULO DE LA COMUNICACIÓN

Análisis de series temporales para la detección precoz de brotes de Salmonella.

Time series analysis for the early detection of Salmonella outbreaks.

NOMBRE, APELLIDOS Y FILIACIÓN DE LOS AUTORES

Izan Hoyuela Iglesias¹ (izanhoyuela97@gmail.com), Julio Álvarez Sánchez^{2,3}, Miguel Ángel Moreno Romo².

¹Facultad de Estudios Estadísticos, Universidad Complutense de Madrid.

²Facultad de Veterinaria, Universidad Complutense de Madrid.

³Centro de Vigilancia Sanitaria Veterinaria (VISAVET).

TIPO DE TRABAJO

TFG TFM Tesis Doctoral Otros

Si otros, especificar:



ABSTRACT

Background: *Salmonella* es la causante del 11% de los 9,4 millones (intervalo de confianza 90%: 6,6 – 12.7) de infecciones de origen alimentario notificadas en EE. UU cada año. Las infecciones por este patógeno son de obligada notificación a las autoridades sanitarias, por lo que una gran cantidad de información sobre la frecuencia y distribución de la enfermedad está disponible.

Objetivo: Este trabajo busca explorar el uso de modelos autorregresivos integrados de media móvil (ARIMA) para la detección precoz de brotes de *Salmonella* mediante la identificación de aumentos inesperados en la notificación usando datos de un estado de EE. UU. (Minnesota).

Métodos: Una serie temporal, comprendida entre los años 2005-2017, del número de casos semanales definidos como esporádicos por el departamento de salud de Minnesota fue analizada para la elaboración del modelo. La capacidad del modelo para detectar brotes se evaluó a través de la generación de predicciones sobre cada semana del año 2018 y una comparación de estas con los casos totales reales observados, incluyendo los procedentes de brotes conocidos. Adicionalmente se generaron curvas ROC para la fijación de un punto de corte que optimizase la sensibilidad y la especificidad de la diferencia entre casos totales observados (Esporádicos y pertenecientes a brote) y casos esporádicos estimados a la hora de definir brotes activos.

Resultados: El mejor modelo fue un ARIMA con componente estacional $(5,1,2) \times (1,1,1)$ con la temperatura media máxima de cada semana como variable regresora. El punto de corte para definir una semana como con brote activo se fijó en 1.57 casos esporádicos predichos más que totales observados. Con este método se logró clasificar correctamente al 93.3% (Intervalo de confianza 95%: 0.844 – 1) de las semanas de 2018 con brote activo y al 63.6% (Intervalo de confianza 95%: 38.6% - 79.7%) de las carentes de brote activo.

Conclusiones: Dado que la finalidad es maximizar la sensibilidad para detectar precozmente brotes, los resultados sugieren que esta aproximación podría ser útil para la detección precoz de brotes de *Salmonella*.



Background: *Salmonella* is responsible for 11% of the 9.4 million (90% confidence interval: 6.6 - 12.7) reported foodborne infections in the US each year. Infections by this pathogen are mandatory notification to health authorities, so a large amount of information on the frequency and distribution of the disease is available.

Objective: The aim of this work is to explore the use of Autoregressive Integrated Moving Average (ARIMA) models for the early detection of *Salmonella* outbreaks by identifying unexpected increases in reported cases using data from a US state (Minnesota).

Methods: A time series, between the years 2005-2017, of the number of weekly cases defined as sporadic by the Minnesota Department of Health was analyzed for the elaboration of the model. The ability of the model to detect outbreaks was evaluated through the generation of predictions for each week of the year 2018 and a comparison of these with the total real cases observed, including those from known outbreaks. Additionally, ROC curves were generated to set a cut-off point that would optimize the sensitivity and specificity of the difference between total observed cases (sporadic and belonging to an outbreak) and estimated sporadic cases to define active outbreaks.

Results: The best model was an ARIMA with a seasonal component $(5,1,2) \times (1,1,1)$ with the maximum mean temperature of each week as regressor variable. The cut-off point for defining a week as an active outbreak was set at 1.57 predicted sporadic cases more than total observed cases. With this method, it was possible to correctly classify 93.3% (95% confidence interval: 0.844 - 1) of the weeks of 2018 with an active outbreak and 63.6% (95% confidence interval: 38.6% - 79.7%) of those lacking active outbreak.

Conclusions: Since the aim is to maximize the sensitivity to an early detection of outbreaks, the results suggest that this approach could be useful for the early detection of *Salmonella* outbreaks.



BREVE BIBLIOGRAFÍA

1. Ward MP, Iglesias RM, Brookes VJ. Autoregressive Models Applied to Time-Series Data in Veterinary Science. *Front Vet Sci.* 2020;7(1):604-619.
2. Unkel S, Farrington CP, Garthwaite PH, Robertson C, Andrews N. Statistical methods for the prospective detection of infectious disease outbreaks: a review. *J.R. Statistics Soc a.* 2012;175(1):49-82.
3. Rounds JM, Taylor AJ, Eikmeier D, Nichols NM, Lappi V, Wirth SE, et al. Prospective Salmonella Enteritidis surveillance and outbreak detection using whole genome sequencing, Minnesota 2015-2017. *Epidemiol Infect.* 2020;148(5):1-8.
4. Hyndman RJ, Athanasopoulos G. *Forecasting: principles and practice.* 3rd ed. Melbourne: OTexts; 2021.
5. Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, M Widdowson, Roy SL et al. Foodborne Illness Acquired in the United States - Major Pathogens. *Emerg Infect Dis.* 2011; 17(1): 7-15.
6. Elmore JG, Wang P, Kerr KF, Schriger DL, Morrison DE, Brookmeyer R, et al. Excess Patient Visits for Cough and Pulmonary Disease at a Large US Health System in the Months Prior to the COVID-19 Pandemic: Time-Series Analysis. *J Med Internet Res.* 2020;22(9):30-41.
7. Nielsen A. *Practical Time Series Analysis.* 1st ed. Canada: O'Reilly; 2020.

CONFLICTOS DE INTERÉS

N/D