

MORBILIDAD POR IRA EN ANTIOQUIA

Verónica Mendoza Iguarán¹, Pablo A. Osorio Marulanda², Mariana Uribe Orrego³

¹vmendozai@eafit.edu.co, ²paosorion@eafit.edu.co, ³muribeo3@eafit.edu.co

Universidad EAFIT, Medellín, Antioquia, Colombia

15 de mayo de 2019

Resumen

El siguiente trabajo presenta la realización de un modelo epidemiológico probabilístico aplicado a la infección respiratoria aguda (IRA) en el departamento de Antioquia, con observaciones desde enero de 2018 hasta marzo de 2019, evaluando tal fenómeno a través de los datos recopilados por el Instituto Nacional de Salud. Adicionalmente se plantea un posible modelo de predicción aplicado a las observaciones semanales de esta epidemia. Para la ejecución y desarrollo de dicho modelo se hace uso del software MatLab R2019a.

Palabras clave: Infección respiratoria aguda, IRA, Distribución weibull, Morbilidad, Modelo probabilístico, Distribución log-normal.

Abstract

The following work presents the realization of a probabilistic epidemiological model applied to the acute respiratory infection (ARI) in the department of Antioquia, with observation data from January 2018 to March 2019. This phenomenon has been evaluated through the data collected by the National Health Institute. Additionally a possible prediction model is applied to the weekly observations made to this epidemic. The MatLab R2019a software is used for the execution and development of the aforementioned model.

Keywords: Acute respiratory infection, ARI, Weibull distribution, Morbidity, Probabilistic model, Log-normal distribution.

1. Introducción

La infección respiratoria aguda (IRA) es una enfermedad que se produce en el aparato respiratorio y es causada por diferentes microorganismos como virus y bacterias. La mayoría de estas infecciones, como el resfriado común, son leves, pero dependiendo del estado general de la persona puede llegar complicarse y podría amenazar la vida, como en el caso de la neumonía.

La IRA es una de las primeras causas de atención médica en todo el mundo, tanto en la consulta ambulatoria como en la hospitalización, además se encuentra entre las primeras causas de mortalidad en niños y niñas menores de 5 años y en adultos

mayores de 60 años [1]. Es un evento de interés en salud pública dado su gran impacto de morbilidad, por esta razón, en este trabajo, se abordará un estudio de dicho problema en uno de los departamentos colombianos donde se presentan más casos de la IRA como lo es Antioquia, mediante la determinación del comportamiento de esta infección respecto a variables de tiempo.

Los resultados de este trabajo nos permite evidenciar la relación que ocurre entre el modelo teórico de casos observados de la IRA en Antioquia y los datos reales que se tienen. Probablemente puede contribuir al conocimiento de las características epidemiológicas de las personas afectadas por esta enfermedad y el comportamiento que se genera con

respecto al tiempo.

2. Materiales y método

Para el desarrollo del presente documento se realizó una recopilación de boletines epidemiológicos semanales colombianos correspondientes al año 2018 y lo corrido de 2019, generados por el Instituto Nacional de Salud (INS) quienes tienen como fuente de información a las diferentes entidades territoriales (ET) a través del Sistema de vigilancia en salud pública (Sivigila). Estos boletines presentan datos de tipo cuantitativo que corresponden a la cantidad de casos presentados de un tipo de enfermedad en un departamento colombiano determinado en cada semana epidemiológica, la cual consiste en una semana de año que siempre empieza en domingo y termina en sábado. Los datos que se extraen de estos boletines para trabajar son datos que representan la morbilidad de la IRA en Antioquia. Estos datos se dividen en tres, los casos que se generaron para consulta externa, casos de hospitalizaciones generales y casos en UCI, es decir hospitalizaciones en la unidad de cuidados intensivos. La suma de estos tres tipos de datos nos permite conocer los datos totales que fueron observados de la IRA en cada semana epidemiológica en Antioquia.

A modo de desarrollar un modelo de predicción se generaron datos nuevos con respecto a la distribución de aproximación que se obtuvo, con el propósito de predecir comportamientos de esta infección con la comparación entre los datos simulados y los reales.

Para el análisis de datos se utilizó el software Matlab, el cual nos permitió verificar si la distribución de probabilidad a la que se aproximaron los datos era apropiada. Se generó la función de densidad y la función de distribución teórica; de esta última se observó la variación con respecto la distribución empírica, logrando así extraer diferentes conclusiones. Adicionalmente se incluyó un análisis respectivo a el modelo sin los datos que se salían del rango, esto es, posibles datos que podrían llegar a ser errores.

Resultados

Con los datos obtenidos de las observaciones de la IRA se realizó un histograma que muestra la frecuencia con que ocurre un intervalo de números de casos. (Ver figura 1)

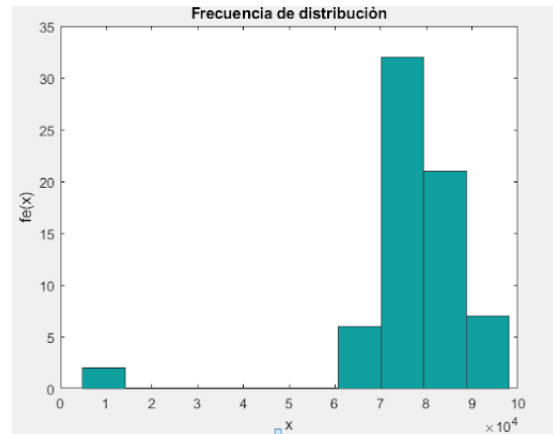


Figura 1: Histograma de los casos observados de IRA en Antioquia en el tiempo estipulado.

Del histograma se deduce que la distribución weibull puede ser una buena aproximación, dada la asimetría que tienen los datos.

Sabiendo a cuál distribución se aproximan mejor los datos, se utiliza la función `fitdist` de matlab que a través de la estimación por máxima verosimilitud encuentra los parámetros de la función de distribución supuesta que mejor se adapta al volumen de datos que se tienen.

```
pd = fitdist(x,'Weibull')
```

Con `x` como el vector de observaciones

La estimación por máxima verosimilitud determina los valores para los cuales la derivada de la probabilidad de todos los eventos dada la distribución a la que se va aproximar, se anula, es decir, los máximos de los parámetros de la distribución; para que este proceso funcione, los datos deben ser independientes.

Dado que los datos son independientes, la probabilidad del conjunto de datos dada la distribución supuesta, es igual a la multiplicación de las probabilidades individuales. Para calcular la derivada de forma simple se utiliza la función logaritmo pues es una función monótona creciente al igual

que la función de probabilidad, lo que asegura que el máximo de tal función ocurra en el mismo punto que en la función original. Todo lo anterior dicho es matemáticamente visto de la siguiente manera: Considere una función de distribución genérica, con parámetros

$$\theta_1, \theta_2 \dots \theta_k$$

$$f(x; \theta_1, \theta_2, \dots, \theta_k)$$

Por lo dicho en el enunciado anterior, la función de verosimilitud se calcula a partir de la función de distribución de los datos x , de esta manera, y como condición dijimos que eran individuales, entonces la función esta dada finalmente por:

$$L = \prod_{i=1}^n f(x; \theta_1, \theta_2, \dots, \theta_k)$$

Ahora, para manipular mejor las derivadas sacamos logaritmo natural a ambos lados

$$\Lambda = \ln(L) = \sum_{i=1}^n \ln(f(x; \theta_1, \theta_2, \dots, \theta_k))$$

Y para calcular los valores respectivos de los parámetros derivamos total o parcialmente e igualamos a 0, buscando, como se dijo, los máximos de la función

$$\frac{\partial \Lambda}{\partial \theta_j} = 0, j = 1, 2, \dots, k$$

Para ciertos casos no es posible resolver las ecuaciones de manera analítica, allí es donde se debe usar ciertos métodos numéricos. [5]

Para nuestro caso de aplicación, tenemos la función de distribución *Weibull*, que cuenta con parámetros λ y α , los que representan parámetros de escala y forma respectivamente. El cálculo de la máxima verosimilitud de la función *Weibull* se calculará a continuación:

Considere

$$f(x_i, \lambda, \alpha) = \lambda \alpha (\lambda x_i)^{\alpha-1} e^{-(\lambda x_i)^\alpha}$$

Donde cada dato x_i es generado independientemente de los demás. Entonces, la función de verosimilitud se expresaría así:

$$L(\lambda, \alpha) = \prod_{i=1}^n \lambda \alpha (\lambda x_i)^{\alpha-1} e^{-(\lambda x_i)^\alpha}$$

Ahora generamos la función de *log - verosimilitud*

$$\begin{aligned} \ln(L(\lambda, \alpha)) &= \ln\left(\prod_{i=1}^n \lambda \alpha (\lambda x_i)^{\alpha-1} e^{-(\lambda x_i)^\alpha}\right) \\ &= \ln((\lambda^\alpha \alpha)^n \prod_{i=1}^n (\lambda x_i)^{\alpha-1} e^{-(\lambda x_i)^\alpha}) \\ &= \ln(\lambda^\alpha \alpha)^n + \ln\left(\prod_{i=1}^n (\lambda x_i)^{\alpha-1} e^{-(\lambda x_i)^\alpha}\right) \\ &= \ln(\lambda^\alpha \alpha)^n + \ln\left(\prod_{i=1}^n (\lambda x_i)^{\alpha-1}\right) + \ln\left(\prod_{i=1}^n e^{-(\lambda x_i)^\alpha}\right) \\ &= \ln(\lambda^\alpha \alpha)^n + \ln[(x_1)^{\alpha-1} \cdot (x_2)^{\alpha-1} \cdot \dots \cdot (x_n)^{\alpha-1}] + \\ &\quad \ln[e^{-(\lambda x_1)^\alpha} \cdot e^{-(\lambda x_2)^\alpha} \cdot \dots \cdot e^{-(\lambda x_n)^\alpha}] \\ &= n \ln(\lambda^\alpha) + n \ln(\alpha) + [(\alpha-1) \ln(x_1) + (\alpha-1) \ln(x_2) + \\ &\quad \dots + (\alpha-1) \ln(x_n)] + [-(\lambda x_1)^\alpha - (\lambda x_2)^\alpha - \dots - (\lambda x_n)^\alpha] \\ &= n \ln(\lambda^\alpha) + n \ln(\alpha) + (\alpha-1) [\ln(x_1) + \ln(x_2) + \\ &\quad \dots + \ln(x_n)] - \sum_{i=1}^n (\lambda x_i)^\alpha \\ &= n \alpha \ln(\lambda) + n \ln(\alpha) + (\alpha-1) \sum_{i=1}^n \ln(x_i) - \sum_{i=1}^n (\lambda x_i)^\alpha \\ \ln(L(\lambda, \alpha)) &= n \alpha \ln(\lambda) + n \ln(\alpha) + (\alpha-1) \sum_{i=1}^n \ln(x_i) - \sum_{i=1}^n (\lambda x_i)^\alpha \end{aligned}$$

Ahora derivamos parcialmente respecto a λ

$$\begin{aligned} \frac{\partial \ln(L(\lambda, \alpha))}{\partial \lambda} &= \alpha n \frac{1}{\lambda} - \frac{\partial \sum_{i=1}^n (\lambda x_i)^\alpha}{\partial \lambda} \\ &= \alpha n \frac{1}{\lambda} - \left[\frac{\partial (\lambda x_1)^\alpha}{\partial \lambda} + \frac{\partial (\lambda x_2)^\alpha}{\partial \lambda} + \dots + \frac{\partial (\lambda x_n)^\alpha}{\partial \lambda} \right] \\ &= \alpha n \frac{1}{\lambda} - [\alpha (\lambda x_1)^{\alpha-1} (x_1) + \alpha (\lambda x_2)^{\alpha-1} (x_2) + \dots + \alpha (\lambda x_n)^{\alpha-1} (x_n)] \\ &= \alpha n \frac{1}{\lambda} - [\alpha \lambda^{\alpha-1} (x_1 + x_2 + \dots + x_n)] \\ &= \alpha n \frac{1}{\lambda} - [\alpha \lambda^{\alpha-1} \sum_{i=1}^n x_i] \end{aligned}$$

Ahora igualamos a 0 la ecuación para encontrar su máximo

$$0 = \alpha n \frac{1}{\lambda} - [\alpha \lambda^{\alpha-1} \sum_{i=1}^n x_n]$$

Despejando λ obtenemos que

$$\lambda = \left(\frac{n}{\sum_{i=1}^n x_n} \right)^{\frac{1}{\alpha}}$$

Utilizando el mismo proceso anterior, pero ahora derivando parcialmente respecto a α , obtenemos algo derivado de la siguiente ecuación

$$0 = n \ln(\lambda) + \frac{n}{\alpha} + \sum_{i=1}^n \ln(x_i) - \lambda \alpha [\ln(\lambda) \sum_{i=1}^n (x_i^\alpha) + \dots + \sum_{i=1}^n (x_i^\alpha) \ln(x_i)]$$

Ecuación que no puede ser resuelta mediante método analítico, solo a través de métodos numéricos como el de Newton-Raphson. A continuación se presenta la gráfica de la aproximación a la distribución weibull (Ver figura 2):

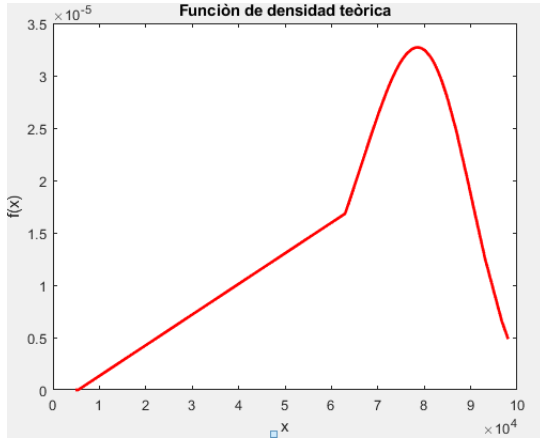


Figura 2: Aproximación a la función de distribución weibull de los casos observados de IRA en Antioquia en el tiempo estipulado.

Cabe recalcar que para nuestro caso los parámetros de forma y escala tendrían un valor respectivo de: $\lambda = 80275,5$ $\alpha = 7,06047$. En una weibull de la forma

$$f(x_i, \lambda, \alpha) = \lambda \alpha (\lambda x_i)^{\alpha-1} e^{-(\lambda x_i)^\alpha}$$

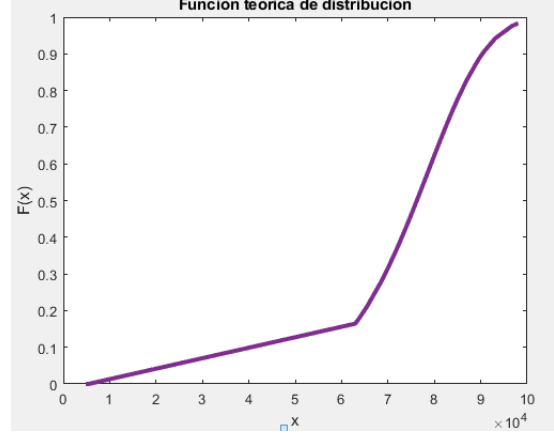
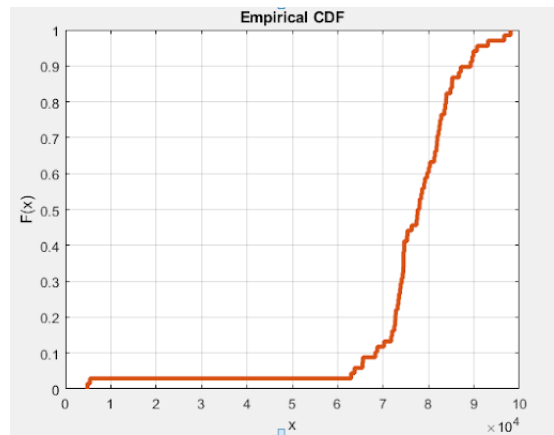


Figura 3: Aproximación a la función de densidad weibull de los casos observados de IRA en Antioquia en el tiempo estipulado.

Para conocer la función de distribución acumulativa empírica se utiliza la función **cdfplot** definida en matlab que crea un gráfico de esta. Esta función consiste en asignarle un valor de proporción a cada dato que se estudia. El valor obtenido para este es la proporción de los datos estudiados (Ver figura 4). Matemáticamente la distribución empírica tiene la siguiente forma: Sea una muestra de X_1, \dots, X_n

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n I[X_i \leq x]$$

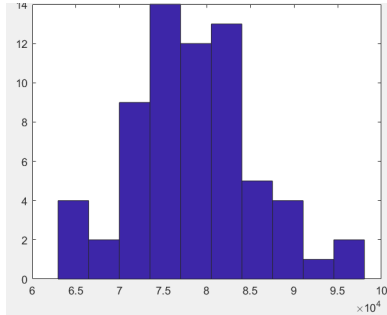


Como se dijo anteriormente, se hizo adicionalmente una simulación sin los datos outlayer, esto

es, datos que podrían llegar a representar un error al momento de realizar el modelo (2 en específico).

A través de la simulación sin estos datos se obtuvieron los siguientes resultados.

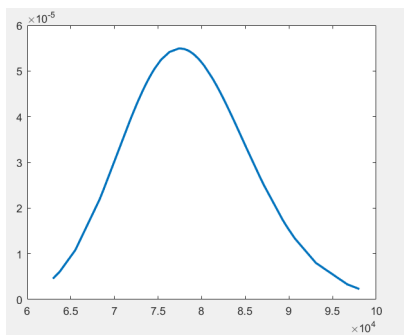
Figura 5: Histograma generado a partir de los la muestra sin los datos fuera del rango



En la figura 5 se puede observar la inmensa diferencia entre el histograma y los intervalos generados por esta comparados con el histograma con la muestra completa.

A partir entonces de esta nueva muestra se uso igualmente el comando `fitdist` anteriormente usado para calcular la weibull y bajo el mismo método se obtuvo o siguiente (Ver figura 6)

Figura 6: Distribución generada por la muestra de datos sin los datos fuera de rango (Distribución lognormal)



Asi entonces se generó una distribución lognormal con parámetros

$$\mu = 11,2668$$

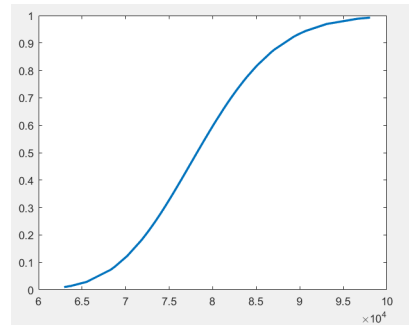
$$\sigma = 0,0933053$$

Con la log-normal de la forma:

$$\frac{1}{x\sigma(2\pi)^{\frac{1}{2}}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)$$

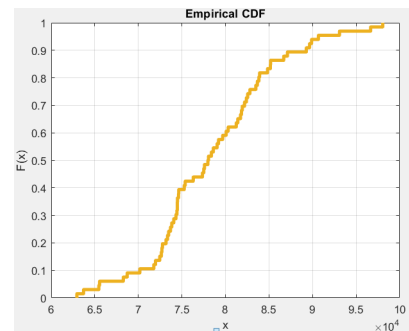
Asi mismo, y generando la función de densidad respectiva a la distribución respectiva obtenemos (Ver figura 7)

Figura 7: Función de densidad relacionada con la lognormal generada a partir de la parametrización vista)



E igualmente, dado que es un volumen diferente de datos se calcula la distribución empírica bajo el mismo fundamento teórico que lo dicho anteriormente(Ver figura 8).

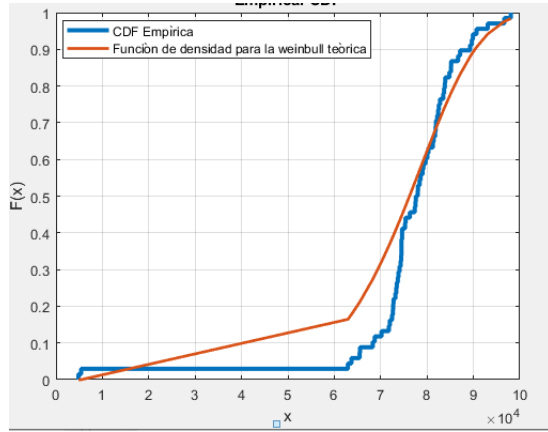
Figura 8: Función empírica de los datos relacionados a la nueva muestra)



3. Análisis de resultados

Al poner la gráfica de distribución teórica y la distribución empírica acumuladas juntas podemos visualizar su diferencia gráficamente(Ver figura 9)

Figura 9: Función de densidad teórica y función de distribución empírica en Antioquia en el tiempo estipulado



Siendo la curva de color azul la distribución empírica y la curva de color naranja la distribución teórica.

Para la validación del modelo es necesario hacer una validación de comportamiento a través del cálculo de algunas estadísticas considerables en esta evaluación y aplicables para este caso.

Se obtuvo una media experimental de $7,6363e + 04$ y una media teórica de $7,5126e + 04$, es evidente la similitud de estos valores lo que permite inferir que la distribución *Weibull* es una buena aproximación. También se obtuvo una varianza experimental de $2,0867e + 08$ y una varianza teórica de $1,1445e + 10$, en este caso percibimos mayor diferencia entre los valores pues existen dos datos muy alejados, estos se presentan en las semanas 4 y 56 con un número de casos de 5421 y 4831 respectivamente. Esto se puede generar por reportes tardíos presentados por las entidades territoriales al Sistema de vigilancia en salud pública.

Adicionalmente se tomaron los datos de la función de distribución (acumulada) empírica y los datos de la teórica; para cada par de puntos (uno de cada gráfica) se halló la diferencia.

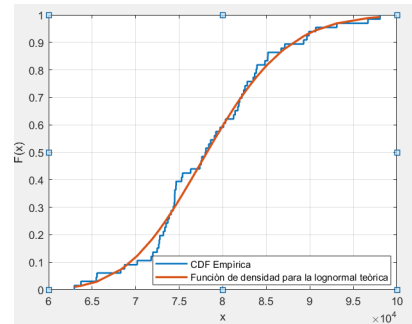
```
VH=ecdf(xo);
d=abs(Ya-VH);
promedioerror=mean(d);
Max=max(d)
```

Donde Ya representa los datos acumulados teóricos y VH los experimentales

Posteriormente se obtuvo el promedio de las diferencias de los puntos el cual es igual a 0,0805. Este valor nos indica que en promedio las dos gráficas tienden al mismo valor y por tanto el modelo teórico elegido se adapta en gran medida a los datos experimentales observados. Sin embargo el mayor error encontrado es igual a 0,1085

Igualmente, haciendo el análisis de la distribución lognormal encontrada obtenemos lo siguiente (Ver figura 10)

Figura 10: Función de densidad teórica y función de distribución empírica en Antioquia en el tiempo estipulado relacionadas a la nueva muestra



En esta nueva muestra el error máximo de distancia es de 0,0845. Además de esto, si comparamos la varianza y la media logarítmica muestral y experimental podemos ver una gran diferencia respecto al anterior modelo, dado que este, por supuesto es más aproximado:

Valores teóricos de la media y varianza logarítmica

$$\mu = 11,2668$$

$$\sigma = 0,0933053$$

Valores experimentales de la media y varianza logarítmica

$$\mu = 11,2668$$

$$\sigma = 0,0087$$

De hecho, la media logarítmica en ambos casos es la misma.

Ahora bien, a través de la observación y análisis del modelo teórico dado, y observando los resultados y el nivel de adaptación que este tiene en un contexto real es preciso analizar la validación estructural del modelo a través de su evaluación en

condiciones extremas, esto es, evaluar, por ejemplo el comportamiento del modelo para un número de semanas total a 1000, y 10000.

Para poder hacer esto se utilizará el hecho de que la inversa de la función de distribución con los parámetros específicos dados evaluada en una distribución uniforme de n datos (donde, en este caso n sería igual al número de semanas) es igual a n datos que cumplen la distribución de F, esto es

$$F(U)^{-1} = X$$

donde, para el caso aplicado, con la función de distribución weibull

$$F(X) = 1 - e^{-(\frac{x}{\lambda})^k}$$

Obteniendo así

$$F(U)^{-1} = -\lambda(-\log(1 - U))^{\frac{1}{k}}$$

con parámetro de escala y k parámetro de forma para un X Weibull(λ, k)

De esta manera, usando la función anteriormente dicha podemos modelar a futuro. Un ejemplo de ello sería el modelo aplicado a 1000 semanas(Ver figura 11).

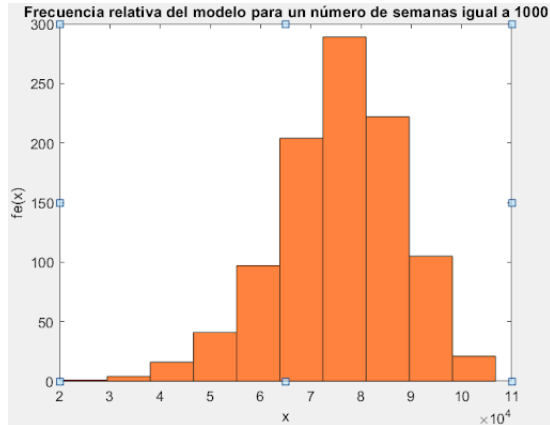


Figura 11: Histograma de observaciones para un número de semanas igual a 1000

De la misma manera, podemos observar, cómo, a través de la distribución usada para generar los datos la probabilidad alusiva a los mismos se graficaría de la siguiente manera(Ver figura 12).

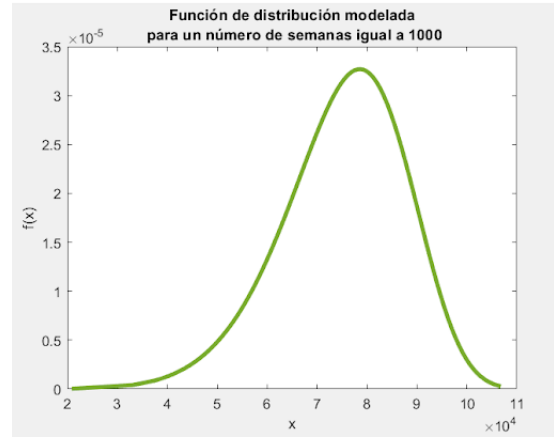


Figura 12: Distribución de observaciones para un número de semanas igual a 1000

De la misma manera, si ejecutamos el modelo con la nueva muestra y la nueva distribución encontrada obtenemos lo siguiente(Ver figura 13):

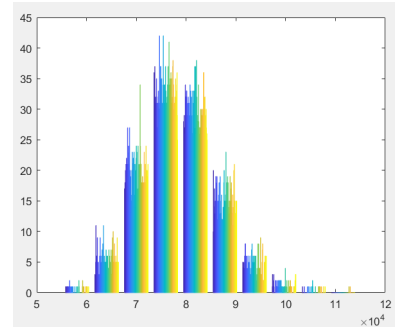


Figura 13: Frecuencia de observaciones para un número de semanas igual a 10000

Y de la misma manera, la distribución relacionada a la misma (Ver figura 14)

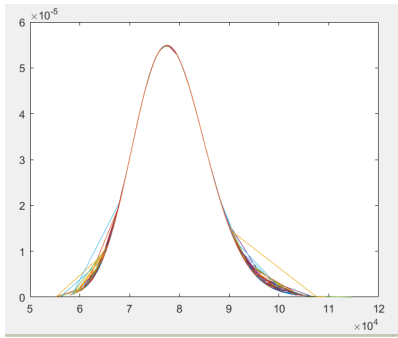


Figura 14: Distribución de observaciones para un número de semanas igual a 10000

Lo anterior mostrado y analizado puede ser un indicio de que el modelo entonces no es sensible a los cambios extremos abruptos.

Así entonces, la función de distribución y la resultante de la misma se podría interpretar como la probabilidad de que durante una semana cualquiera suceda una x cantidad de observaciones de una magnitud específica.

Cabe recalcar que, dado que el modelo es una versión simplificada del mundo real, este no considera la diferencia entre una semana a otra, es decir, considera que cada una de las semanas están bajo las mismas condiciones, tanto externas como de observación.

4. Conclusiones

A través del uso de un modelo probabilístico se puede evaluar el comportamiento general de los datos, esto es, qué cantidad de observaciones pueden suceder en un periodo de tiempo específico, en este caso, semanas.

El análisis de los datos también nos permitió comprobar que se presentan épocas del año en las que los casos aumentan notablemente, por lo que es necesario generar mayor cantidad de controles en estas épocas. Igualmente se puede verificar a través del modelo que la mayor cantidad de semanas presentan observaciones en un intervalo entre 70000 y 80000.

También, en los últimos años se ha identificado que la mayoría de los casos de IRA se presentan en los periodos altas temperaturas y mayor precipitación; la humedad atmosférica es un factor que tiene

gran relación con las enfermedades respiratorias de origen viral, pues cuando la humedad relativa es alta, los virus suelen ser menos estables y se transportan en el aire con mayor facilidad, generando mayor propagación de enfermedades en diferentes personas. (2)

Del mismo modo, el modelo se puede interpolar a futuro, sin embargo estas consideraciones tendrían un considerable porcentaje de error, dado que el modelo no considera cambios climáticos y cambios medioambientales abruptos.

Es claro igualmente, que, el modelo encontrado que se realizó sin los datos fuera de rango es mucho mas aproximado a la realidad que el otro, pues, como se dijo, todo esta reñido por las observaciones que notifiquen los entes encargados.

Referencias

- [1] MINISTERIO DE SALUD Y PROTECCION SOCIAL. (14 DE JULIO DE 2017). MINISTERIO DE SALUD. *Obtenido de* <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/DIJ/circular-23-de-2017.pdf>.
- [2] INSTITUTO NACIONAL DE SALUD. (20 DE ABRIL DE 2018). INSTITUTO NACIONAL DE SALUD. *Obtenido de* <https://www.ins.gov.co/buscador-eventos/Informesdeevento/Informe>
- [3] WEATHER SPARK. (S.F.) *Obtenido de* <https://es.weatherspark.com/y/22512/Clima-promedio-en-Antioquia-Colombia-durante-todo-el-a>
- [4] INSTITUTO NACIONAL DE SALUD. (S.F.). BOLETÍN EPIDEMIOLÓGICO SEMANAL (SEMANAS 1-52 2018 Y SEMANAS 1-16 2019) *Obtenido de* <https://www.ins.gov.co/buscador-eventos/BoletinEpidemiologico/2018>
- [5] RELIABILITY ENGINEERING RESOURCE WEBSITE (NOVIEMBRE DE 2003) *Obtenido de:* <https://www.weibull.com/hotwire/issue33/relbasics33.htm>
- [6] APROXIMACIÓN EMPÍRICA DE LA FUNCIÓN DE DISTRIBUCIÓN (2016) *Obtenido de:* <https://slideplayer.es/slide/9190788/>

5. Anexos

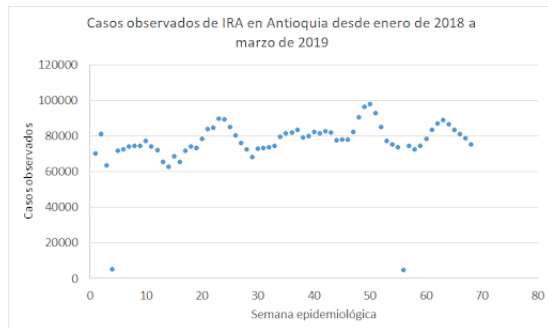


Figura 15: Casos observados en Antioquia desde enero de 2018 a marzo de 2019.

A través de la figura 15 se han detectado comportamientos inusuales: Se percibió que durante las

semanas 11, 12, 13 y 14 (del 11 de marzo al 7 de abril del 2018) los casos observados de IRA disminuyeron notoriamente, esto coincide con los climas constantes que se presentaron durante las fechas. Se observó que hasta la semana 23 (del 3 al 9 de Junio) del 2018 se presentó un aumento significativo de los casos observados correspondientes a 89889; esto se debe en gran medida a los altos cambios climáticos; en este mes la precipitación es la más baja de todo el año, las temperaturas generalmente varían de 23 °C a 32 °C, siendo las más altas de todo el año. Otro comportamiento inusual se presenta durante los meses de noviembre y diciembre, durante estos meses se presentan grandes cambios climáticos por lo que la propagación de enfermedades aumenta.