A lo largo de la materia estudiamos distintos métodos para tratar una muestra de datos, enfocándonos en los valores centrales de la distribución de la misma, ignorando los valores extremos que esta pudiera tener y en muchos casos desechándolos. No obstante, en muchas disciplinas como la economía, la ingeniería o las ciencias ambientales, resulta importantísimo analizar estos valores extremos y los eventos que ellos producen por su alto impacto. Encontrar un análisis que lejos de ignorarlos se centre en describirlos, entonces, resulta vital.

Nos propusimos presentar la Teoría de Valores Extremos y trabajar un poco la problemática de su análisis. En particular, como en meteorología unas de las grandes divas para el estudio de eventos extremos son las olas de calor, decidimos visualizar este tema utilizando datos contenidos en un paquete de R llamado heatwaveR. Este dataset contiene las temperaturas mínimas y máximas diarias en grados Celsius registradas en Argel durante el período 01/01/1961 - 31/12/2005.

La ciudad de Argel es la capital de Argelia y se encuentra situada en el litoral mediterraneo del continente africano, en 36°46′35″N 3°03′31″E. Como esta ciudad se ubica en el hemisferio norte, particularmente en una zona subtropical y a orillas del mar Mediterráneo, su climatología es mediterránea marítima, con lo cual las temperaturas cumplen con un régimen normal de temperaturas más frías en los meses invernales de DEF y mas cálidas en el verano de JJA, aunque moderadas por la influencia de este gran cuerpo de agua. Es por eso que en verano, gracias a las brisas que vienen desde el mar, las máximas no tiende a superar los 30°C; sin embargo, en esta ciudad se dan eventos de olas de calor, especialmente cuando soplan los vientos del Sur provenientes del desierto del Sahara, que son muy cálidos.

MODELO 1:

Primero se establece una serie de bloques máximos o mínimos, según el análisis que se quiera hacer. Por ello, de una serie temporal se extraen los máximos o mínimos valores para un determinado intervalo de tiempo generando así una serie de máximos o mínimos.

Según el Teorema de Fisher-Tippett-Gnedenko, si y sólo si el máximo de una muestra de variables aleatorias independientes e idénticamente distribuidas converge, entonces lo hará hacia uno de tres tipos posibles de distribución: la distribución de Gumbel, la distribución de Fréchet o la distribución de Weibull.

Formalmente, el teorema dice:

Sea una sucesión de variables aleatorias independientes e idénticamente distribuidas provenientes de una función de distribución en común y sea donde cada representa un valor de un proceso medido en una escala temporal fija- por ejemplo temperaturas máximas o mínimas diarias- de modo que representa al máximo a lo largo de observaciones- en el ejemplo, si fuera igual a la cantidad de días que tiene un año, entonces representa la temperatura máxima anual. Si existe una sucesión de pares de números reales tal que para cada y donde es una función de distribución no degenerada[[1]](#footnote-1), entonces la distribución límite pertenece a alguna de estas tres familias:

1. Gumbel
2. Fréchet
3. Weibull

con los parámetros de ubicación b y de escala y en el caso de las familias 2 y 3, también el parámetro de forma.

Estas tres familias se agrupan bajo el nombre de Distribuciones para valores extremos, y tienen comportamientos distintos debidos a las diferentes conductas que toman sus colas: por ejemplo, en el caso de la distribución Weibull, tiene límite finito, mientras que para las distribuciones de Fréchet y Gumbel o mismo la densidad de F decae exponencialmente para la distribución de Gumbel, pero de forma polinómica en la de Fréchet. Primariamente para un set de datos se podría adoptar alguna de estas tres familias y encontrar los parámetros correspondientes, pero esto no resulta útil porque habría que saber de antemano qué distribución es la apropiada para los datos y la realidad es que eso implica una suposición muy grande que a veces puede no ser acertada.

En cambio, se pueden reformular estos tres modelos, combinándolos en una sola familia de modelos llamada Generalized extreme value (GEV) tal que su función de distribución cumpla con la forma

definida para donde los parámetros satisfacen que y Los parámetros son de ubicación, escala y forma para y respectivamente. En particular, las distribuciones de las familias de Fréchet y Weibull se corresponden con para la primera y para la segunda, mientras que la de Gumbel se corresponde con el caso en que que se interpreta como el límite en que conduciendo a la distribución de la familia de Gumbel: , con .

La unificación de las tres familias de distribuciones de valores extremos en una sola gran familia simplifica la implementación estadística. A través de realizar inferencias sobre los mismísimos datos determinan el comportamiento más apropiado para la cola de la distribución, sin realizar juicios previos sobre cuál familia de distribuciones tomar por oportuna. Incluso la incertidumbre en el valor inferido para da una medida de la incertidumbre a la hora de elegir un tipo de familia por sobre los demás para un set de datos.

Teniendo esto en cuenta entonces el Teorema de Fisher-Tippett-Gnedenko podemos reformularlo de la siguiente manera: si existe una sucesión de pares de números reales tal que para cada y donde es una función de distribución no degenerada[[2]](#footnote-2), entonces la distribución límite pertenece a la familia GEV:

definida para donde los parámetros satisfacen que y

Para modelar extremos de una serie de observaciones independientes los datos son separados en bloques de secuencias de observaciones de longitud , con valores grandes de generando así una serie de bloques de máximos para la cual se ajusta la distribución GEV. Los bloques de datos se eligen tal que se correspondan a un período de tiempos, por ejemplo un año, en cuyo caso el es el número de observaciones en un año y los máximos se corresponden con máximos anuales.

Los estimadores de los cuantiles para extremos en una distribución anual de máximos se obtienen invirtiendo las ecuaciones

donde , siendo el nivel de retorno asociado al período de retorno tal que éste sea superado en promedio una vez cada años. En otras palabras, es excedido por el máximo anual de cualquier año particular con una probabilidad

Como los cuantiles permiten expresar a los modelos de probabilidad acorde a los datos en uso, la relación entre el modelo GEV y sus parámetros se interpreta más fácilmente en términos de sus cuantiles. En particular, definiendo tal que

se tiene que si se realizara un gráfico vs tomando una escala logarítmica, el gráfico resulta linear cuando Si el gráfico es convexo con un límite asintótico cuando en ; en cambio, si el gráfico es cóncavo y tiende a infinito. Este gráfico es un gráfico de nivel de retorno.

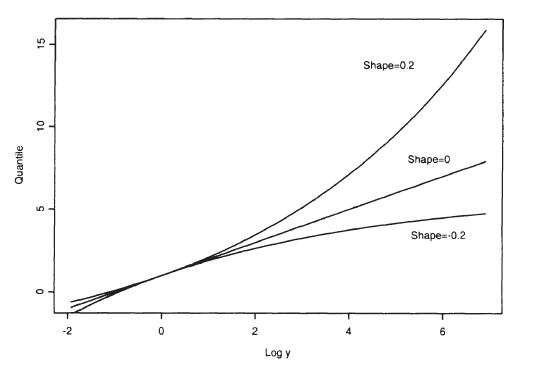


Figura 1: Gráfico de nivel de retorno para una distribución GEV con parámetros de forma ξ=-0.2, ξ=0 y ξ=0.2 respectivamente[[3]](#footnote-3).

1. [↑](#footnote-ref-1)
2. [↑](#footnote-ref-2)
3. Hoja 50 Coles&Tenner An introduction to statistical modeling of extreme values [↑](#footnote-ref-3)