



REVISTA FIGEMPA

Análisis estadístico de datos de las precipitaciones usando métodos robustos y bootstrap.

Statistical analysis of rainfall data using robust and bootstrap methods

Cristian Paliz | [iD](#) Instituto Nacional de Meteorología e Hidrología, Ecuador

Nelly Perugachi | [iD](#) Escuela Superior Politécnica de Chimborazo, Ecuador

Juan Martínez | [iD](#) Escuela Superior Politécnica de Chimborazo, Ecuador

Mariela Moreno | [iD](#) Proyecto Monitoreo de precipitación hídrica de la provincia de Chimborazo, Ecuador.

Carina Yaucán | [iD](#) Proyecto Monitoreo de precipitación hídrica de la provincia de Chimborazo, Ecuador

Romel Palaguachi | [iD](#) Proyecto Monitoreo de precipitación hídrica de la provincia de Chimborazo, Ecuador

HISTORIAL DEL ARTÍCULO

Recepción: 02/07/2021

Aceptación: 08/11/2021

PALABRAS CLAVE

Estadística robusta, modelación, precipitación hídrica, meteorología.

RESUMEN

Con el objetivo de identificar los periodos de mayor o menor precipitación y sus respectivas variaciones, se analizan las precipitaciones hídricas en la ciudad de Riobamba, usando datos de precipitación diaria y mensual entre 1978 y 2019. Empleando métodos estadísticos no convencionales, para establecer la normalidad de los datos, se aplica el método de Kolmogorov Smirnov, que demuestra la necesidad de usar estadística robusta en el análisis, la cual se complementa con el método bootstrap, lo que permitió consolidar los resultados estableciendo las tendencias de los datos y sus rangos de variación; el análisis con el método robusto nos proporciona la tendencia de los datos a lo largo del año que coincide con los resultados hallados con los métodos bootstrap. Los meses de febrero, marzo, abril y mayo presentan más regularidad en sus precipitaciones, con mayor intensidad en el mes de abril y menor intensidad en el mes de julio, identificando dos estaciones con mayores precipitaciones, entre enero y mayo, así como entre octubre y diciembre, mientras que de junio a septiembre se tiene menores precipitaciones, épocas conocidas en los primeros casos como temporadas húmedas y, en la última, como temporada seca.

ARTICLE HISTORY

Received: 02/07/2021

Accepted: 08/11/2021

KEY WORDS

Potential, hydraulic fracturing, methodology for selection, candidate wells, economic analysis.

ABSTRACT

In order to identify the periods of higher or lower precipitation and their respective variations, the water precipitation in the city of Riobamba is analyzed, using daily and monthly precipitation data from 1978 to 2019. Using non-conventional statistical methods, the Kolmogorov Smirnov method is applied to establish the normality of the data, which demonstrates the need to use robust statistics in the analysis, which is complemented with the Bootstrap method, which allowed consolidating the results by establishing the trends of the data and their ranges of variation, the analysis with the robust method provides us with the trend of the data throughout the year that coincides with the results found with the Bootstrap methods. The months of February, March, April and May present more regular precipitation, with higher precipitation in April and lower precipitation in July, identifying two seasons with higher precipitation, between January and May, as well as between October and December, while from June to September there is less precipitation, known in the first case as wet season and in the last case as dry season.

INTRODUCCIÓN

La línea ecuatorial posee características climáticas que implican alta temperatura, abundante lluvia, viento apacible y baja presión, los cuales son indicadores del clima ecuatorial lluvioso, siendo afectado por diferentes sistemas sinópticos que ocurren durante todo el año, cuya

principal manifestación es evidenciada en la variabilidad de precipitación (López Calvajar *et al.*, 2017); la región donde se ubica la ciudad de Riobamba es uno de los lugares de alta producción agrícola, por ello es muy importante realizar el estudio estadístico de precipitación (López Calvajar *et al.*, 2017).

La lluvia es una variable meteorológica que interviene sobre el clima tanto espacial como temporalmente, e incide sobre las actividades humanas vitales como la agricultura (Priyan, 2015). La provincia de Chimborazo es una de las zonas de mayor producción agrícola, pero tiene zonas con déficit hídrico (precipitación < 400 mm durante el ciclo) que ha obligado a implementar sistemas de riego (Fay *et al.*, 2017).

Los cambios extremos en la precipitación constituyen un factor limitante para el desarrollo de los cultivos agrícolas, las actividades sociales y en la mayor parte de las actividades económicas. La sequía es un fenómeno acumulativo y ese efecto se refleja en el agotamiento de los recursos hídricos, de difícil recuperación en un corto espacio de tiempo (Hidalgo, 2017). Las sequías son un fenómeno que han afectado a distintas zonas del Ecuador, y Chimborazo no ha estado exenta de este fenómeno, provocado en gran parte por el cambio climático (Balmaseda Espinosa y Mederos Machado, 2019).

En la actualidad se realizan estudios estáticos por series de tiempo para el estudio de sequías, dando paso incluso al desarrollo de modelos de pronóstico; determinando las fluctuaciones climáticas (Saquina, 2020), los estudios por series de tiempo son claves para encontrar diferencias y anomalías en las variaciones de las precipitaciones en determinados tiempos (Zhiña *et al.*, 2017).

Un problema que se afronta en la práctica al trabajar con series temporales es que es muy común encontrar defectos, como la existencia de lagunas (falta de datos) y/o falta de homogeneidad (derivadas de cambios en los parámetros de observación, o por errores en las mismas) (Hidalgo, 2017).

En el trabajo estadístico es habitual encontrarse con observaciones que parecen proceder de alguna distribución distinta al resto; ya que las precipitaciones hídricas de la lluvia se presentan esporádicamente a través del tiempo, y es muy necesario conocerlas para analizar el ciclo de los cultivos ya que afectan la humedad del suelo o de otros fenómenos naturales como la prevención de riesgos y para tener un registro de ésta.

Las series de precipitaciones presentan un alto grado de irregularidad, mostrando valores atípicos debido a la naturaleza misma de dicha variable, haciendo que las estimaciones habituales se vean muy afectadas. Por esta razón, la media muestral, dada por los datos medidos en general, no representa la media real que se observa en ciertos periodos, acarreando fenómenos naturales de difícil control como inundaciones y que pueden conllevar riesgos en las actividades productivas o a las personas mismas (Méndez *et al.*, 2020).

Usando métodos estadísticos adecuados para estos datos, conocidos como análisis de datos robustos, se podría mejorar las previsiones y análisis de la precipitación, lo cual puede resultar más acorde con la realidad, en el análisis de datos, no solo la media resulta afectada,

también la estimación de la desviación típica está seriamente distorsionada por la existencia de anomalías, que puede llevar a un análisis erróneo de los datos ((Méndez *et al.*, 2020).

Por lo tanto, ante la existencia de valores atípicos que pueden distorsionar las estimaciones obtenidas caben tres opciones:

- Detectar los valores atípicos, y separarlos de la muestra para, posteriormente, aplicar los métodos tradicionales de estimación.
- Aplicar métodos robustos de estimación que no se vean afectados por la existencia de estos valores atípicos, aunque, si la muestra no tuviera valores anómalos, las estimaciones fueran algo menos eficientes que las tradicionales.
- Aplicar el método Bootstrap para el análisis de datos de la precipitación mensual media acumulada.

Un indicador importante para calcular el grado de susceptibilidad de una estadística a la distorsión es la estimación de su punto de ruptura. El punto finito de ruptura de la muestra para las medidas más comunes suele ser muy bajo y el origen de la distorsión puede provenir de datos denominados «valores atípicos». Los valores atípicos son puntuaciones muy diferentes al resto del conjunto de datos, cuya naturaleza extrema y de alta influencia llega a afectar de modo sustancial el cálculo. Una solución a estos problemas es el uso de la estadística robusta para la detección de valores atípicos (Vianez, 2019).

MATERIALES Y MÉTODOS

La zona para el análisis de este estudio es la ciudad de Riobamba, provincia de Chimborazo, con las series de datos de precipitación en el período comprendido entre los años 1978 y 2019 generados por la estación meteorológica del INAMHI en la latitud -1.650789, longitud -78.685213. Dichos datos se organizaron por sumas en meses y fueron desarrolladas en Excel y RStudio.

Las predicciones de estimaciones puntuales y sus intervalos de confianza basados en estadísticas clásicas pueden verse estropeados por valores atípicos, mientras que los modelos predictivos ajustados utilizando estadísticas robustas no sufren esta desventaja (Belío, 2020).

Los métodos de la estadística robusta están diseñados para verse poco afectados cuando se registran anomalías en las muestras. Si se usa la variación de los datos la mediana muestral (definida como el valor central que deja el 50% de la distribución a ambos lados) en lugar de la media muestral, el resultado obtenido es más ajustado para la muestra sin la observación atípica (Belío, 2020).

Si X_i proceden de una distribución normal $N(\mu; \sigma^2)$, la varianza de la mediana muestral sería un 57% mayor

que la de la media muestral, por tanto, sería menos eficiente utilizar la mediana que la media cuando todos los datos muestrales proceden del modelo supuesto. Un valor atípico afecta en un alto grado al estimador y define una función de influencia, no está acotada para la media muestral (es decir, si el valor anómalo crece sin límite, lo mismo ocurriría para la media) y, por otro lado, está acotada para la mediana muestral.

ESTIMADORES ROBUSTOS DE LOCALIZACIÓN

Existen otros estimadores alternativos a la media que son robustos frente a la existencia de valores atípicos. Entre ellos, quizá los más intuitivos son las α -medias recortadas, las cuales eliminan un porcentaje α de las observaciones muestrales en cada extremo. Por ejemplo, la 0,2-media sería la media una vez que eliminamos el 20% de las observaciones en cada cola.

Una alternativa para recortar la muestra consiste en winsorizar la muestra. Una muestra α -winsorizada sustituye el porcentaje α de observaciones extremas en cada cola de la distribución con las observaciones más cercanas (Gil, 2014).

A partir de aquí la estimación winsorizada de la media consiste, simplemente, en calcular la media muestral sobre dicha muestra winsorizada.

Como ya se ha comentado anteriormente, en general, las estimaciones robustas (mediana, media recortada y media winsorizada) son menos eficientes que los estimadores clásicos, como la media muestral, cuando las observaciones muestrales proceden de distribuciones normales. Sin embargo, como hemos visto en el ejemplo, son mucho más robustos cuando hay valores anómalos (su función de influencia está acotada mientras que la de la media muestral no lo está) (Kluwer, 2018).

Para diseñar estimadores que sigan siendo robustos, y además ganen eficiencia en el caso de distribuciones normales, se han propuesto los denominados M-estimadores que constituyen una familia de estimadores robustos basados en una generalización de los estimadores máximo-verosímiles. Aunque su cálculo resulta a veces complejo, en gran parte del software estadístico especializado se pueden obtener sin esfuerzo, especialmente el estimador de Huber lo cual está permitiendo su difusión en los análisis estadísticos habituales (Kluwer, 2018).

En general, un modo intuitivo de comprobar si la muestra puede contener outliers que distorsionen los resultados, es calcular estimadores clásicos y robustos analizando las diferencias entre ellos. Si ambos tipos de estimadores difieren poco, probablemente no haya grandes anomalías en la muestra y los estimadores clásicos sean adecuados, en otro caso, puede ser más conveniente utilizar estimadores robustos (Gil, 2014).

ESTIMADORES ROBUSTOS DE ESCALA

Estimadores robustos en el contexto no paramétrico pueden definirse como no sensibles a un valor atípico vertical aislado. En este contexto, se definieron distintos procedimientos robustos para estimar la función η cuando los errores no necesariamente tienen primer momento (Orozco, 1992).

Al igual que existen alternativas robustas a la media (el cual es un estimador de localización) también existen alternativas robustas a la desviación típica como estimador de escala. Entre las más habituales se encuentra la desviación absoluta mediana, definida como la mediana de las desviaciones en valor absoluto respecto a la mediana, donde X_i es un elemento dentro de un conjunto de datos, como se muestra a continuación:

$$MAD = \text{Mediana } \{|X_i - Me|\}^{n_i = 1}$$

En general, para comparar esta medida con la desviación típica, es necesario estandarizarla dividiendo por 0,6745 por lo cual se obtiene el valor del MAD estandarizado, o NMAD, el cual es una medida de dispersión comparable con la desviación típica. Si se compara este valor con el de la cuasi-desviación típica muestral de la muestra original puede comprobarse hasta qué punto hemos disminuido la influencia de un valor anómalo (de hecho, la función de influencia del NMAD está acotada mientras que la de la cuasi-desviación típica muestral no lo está) (Orozco, 1992).

Otro estimador de dispersión robusto, como el rango intercuartílico que comprende $RIQ = \text{percentil } 0,75 = \text{percentil } 0,25$ entre las cuales está el 50% central de los valores muestrales, cómo es el caso de estimadores de localización; los estimadores robustos de escala son menos eficientes que la cuasi-desviación típica muestral para el caso de observaciones procedentes de la normal y son más adecuados cuando existen outliers en la muestra.

En general, conocer el comportamiento de los estimadores robustos en muestras pequeñas suele ser un problema complejo y, en muchas ocasiones, hay que recurrir a técnicas de remuestreo como el de Bootstrap y Jackknife (Orozco, 1992).

MUESTREO ESTRATIFICADO

Se determina los estratos que conforman la población para seleccionar y extraer de ellos la muestra, este paso se define como estrato a los subgrupos de unidades de análisis que difieren en las características que van a ser analizadas (Velázquez, 2017).

Como parte de este muestreo, se toma una submuestra a partir de cada estrato mediante un procedimiento aleatorio simple, de modo que para obtener la muestra general se combinan las submuestras de todos los estratos (Velázquez, 2017).

La estratificación es aplicable en las variables estadísticas de tipo cuantitativo, normales, entre otras, donde se divide la población compuesta por «N» individuos en «x» subpoblaciones o estratos, el estudio se basa en el manejo por extractos con tamaños respectivos N₁, N₂, N₃, ... N_k de las variables importantes, además se realiza muestreos aleatorios simples de tamaños n_i; obteniendo un cierto número de muestras seleccionadas en cada uno de los extractos. Para lo cual se los clasifica con base en los siguientes parámetros, asignación proporcional (el tamaño de la muestra de cada estrato es proporcional al tamaño del estrato que le dio origen, respecto a la población total) y asignación óptima al tamaño de la muestra de cada estrato, son definidos por quien realiza el muestreo (Guijarro *et al.*, 2018).

MUESTREO ESTRATIFICADO PROPORCIONAL

El muestreo estratificado proporcional se basa en la toma de la muestra de cada estrato a partir del tamaño de cada uno de éstos. En este tipo de muestreo la población se divide en L clases, estratos o grupos, y todos ellos son examinados parcialmente (González-Hidalgo *et al.*, 2020).

MÉTODO BOOTSTRAP

La metodología bootstrap fue iniciada por Efron (1976) e introducida en la estimación del riesgo de reserva de compañías aseguradoras por Lowe (1994). Entre otras, esta metodología fue utilizada por England y Verrall (1999) quienes proponen un método de estimación del riesgo de reserva con bootstrap como una forma fácil de obtener el riesgo de reserva (Villegas *et al.*, 2019).

El bootstrap se basa en la analogía y en la aproximación finita (como opuesto a cálculo exacto). Es un método simple y un principio general que surge de una observación: es posible aproximar bien la distribución de los datos, en un sentido preciso, sin asumir que la distribución tiene una forma particular (Irigoyen *et al.*, 1997).

Los métodos bootstrap son una clase de métodos Monte Carlo no paramétricos que pretenden estimar la distribución de una población mediante remuestreo. Los métodos de remuestreo tratan una muestra observada como una población finita, y generan muestras aleatorias a partir de ella para estimar características poblacionales y hacer inferencia sobre la población muestreada. Una de las problemáticas que ataca este tipo de método es la estimación de intervalos de confianza (Bello, 2019).

INTERVALO DE CONFIANZA BOOTSTRAP PERCENTIL

Usando bootstrap podemos obtener intervalos precisos sin tener que hacer suposiciones normales (Gil Marín *et*

al., 2012). Utiliza la distribución empírica de las réplicas Bootstrap como distribución de referencia. Los cuantiles de la distribución empírica son estimadores de los cuantiles de la distribución muestral del parámetro a estimar, con lo que estos cuantiles aleatorios reproducirán mejor la verdadera distribución cuando la distribución del parámetro a estimar no sea normal (Bello, 2019).

PRUEBA DE NORMALIDAD DE KOLMOGOROV-SMIRNOV

Conocida como prueba K-S, es una prueba de significación estadística para verificar si los datos de la muestra proceden de una distribución normal. Se emplea para variables cuantitativas continuas y cuando el tamaño muestral es mayor de 50 (Romero Saldaña, 2016).

Cuando la prueba Kolmogorov-Smirnov se aplica para contrastar la hipótesis de normalidad de la población, el estadístico de prueba es la máxima diferencia:

$$D = \max|Fn(x) - Fo(x)|$$

siendo Fn(x) la función de distribución muestral y Fo(x) la función teórica o correspondiente a la población normal especificada en la hipótesis nula.

RESULTADOS Y DISCUSIÓN

Las series de datos de precipitación son proporcionadas por el INAMHI, sede Riobamba, en el período 1978 a 2019, las cuales son usadas en el estudio de su variabilidad, para detectar cambios y tendencias en la precipitación.

El análisis de una muestra permite realizar inferencias, extrapolar o generalizar conclusiones a la población blanco con un alto grado de certeza; de tal modo que una muestra se considera representativa de la población blanco, cuando la distribución y valor de las diversas variables se pueden reproducir con márgenes de error calculables. Entonces, el muestreo tiene por objetivo estudiar las relaciones existentes entre la distribución de una variable en la población blanco y la distribución de esta variable en la muestra de estudio (Otzen, y Manterola, 2017).

Los datos de la figura 1 muestran la variación de la precipitación promedio mensual, que oscilan entre 0 y 201,8 mm siendo éste el máximo observado, con fuertes oscilaciones en los datos medidos en el transcurso del tiempo en los que fueron registrados (ver Figura 1).

En la figura 2, el diagrama de caja muestra los datos atípicos que se presentan en los diferentes meses del año y que la mediana en general no coincide con la media (ver Figura 2), lo cual nos indica que los datos no se distribuyen simétricamente lo que implica una falta de normalidad en la distribución de los datos conllevando a la

conveniencia de realizar un análisis utilizando la estadística robusta (Ruano, 2018).

En los histogramas se observa que en el primer cuatrimestre, en general, la mayor parte de días no existe precipitación en los diferentes meses notándose un incremento de la precipitación a medida que avanzamos al mes de abril (ver Figura 3).

Según la figura 4 se mantiene la predominancia de los reportes con valores de precipitación 0 con una frecuencia mayor en el mes de mayo y, sin embargo, muestra un mayor nivel de intensidad, la cual empieza a disminuir hasta finales del segundo cuatrimestre (ver Figura 4).

En la figura 5 se observa que la intensidad de precipitación respecto al cuatrimestre anterior aumenta, particularmente los meses de septiembre y octubre y luego empieza a disminuir desde noviembre y diciembre (ver Figura 5).

Los resultados de la figura 6, muestran el comportamiento de la precipitación media mensual de los diferentes años de estudio, notándose que los meses de abril con un valor de 3360,6 mm presenta el mayor valor de precipitación y el mes de agosto con 653,3 mm presenta menor precipitación. Notándose que entre los meses de julio a septiembre existe menor precipitación (ver Figura 6).

Los resultados de la tabla 1, prueba de normalidad de datos de precipitación hídrica, muestran datos normales entre febrero a mayo, mientras que los demás meses no serían normales, mostrando en el primer caso mayor regularidad en las precipitaciones con mayor intensidad, sin embargo, a pesar de no mostrar normalidad los meses de octubre a diciembre también son de alta precipitación y más esporádicas (ver Tabla 1).

Como lo presentado en algunos trabajos para definir métodos de inferencia confiables cuando la muestra puede contener datos atípicos tanto para modelos de regresión como para el análisis de datos funcionales con diferentes estimadores potencialmente relevantes dentro del proceso de estimación (Boente, 2021; Calvo, 2017) (ver Tabla 2).

Los resultados de la tabla 2, muestran la variación de los parámetros entre los datos de la media y la mediana, la media acotada y la mediana winzorizada, que permiten observar la presencia de datos anómalos en los registros medidos, debido a la diferencia en sus valores numéricos. En el mes de enero se observa que el valor de la media es mayor que la mediana, la media acotada y la mediana winzorizada, mientras que se observa una gran similitud entre estas últimas variables consideradas como magnitudes robustas. En los meses de febrero, marzo, abril y mayo se puede observar una mayor coincidencia entre la media y las magnitudes robustas, confirmando la tendencia a la normalidad en los datos en estos meses, finalmente los meses de junio a diciembre tienden a variar las magnitudes robustas con respecto a la media, indicándonos anomalías en los datos medidos, que prevén irregularidades

de precipitaciones, presentando posibles eventos anómalos con mayor probabilidad. La desviación absoluta mediana estandarizada N_{MAD} mantiene la tendencia de la desviación típica con mayor dispersión en los meses de abril y menor dispersión en el mes de julio con valores numéricos más altos que evidencian estas referencias.

En la figura 7 se presenta un consolidado de los análisis realizados, observándose una gran similitud entre los resultados del método Bootstrap y el de la estadística descriptiva, a excepción del mes de diciembre, mientras que los métodos robustos presentan mayores diferencias con éstos, sin embargo, en todos los casos se mantiene las tendencias de precipitación en los diferentes meses del año, con mayor precipitación en el mes de abril y menor en el mes de julio (ver Figura 7).

La figura 8, muestra con mayor claridad la variación de la precipitación a lo largo del año usando el método Bootstrap, coincidiendo la tendencia con los estimadores del método usado con estadística robusta, para el análisis se usa la distribución sobre la mediana, cuyos resultados tienen una alta similitud con la media, que indicaría que los datos anómalos afectan menos usando este método, adicionalmente se gráfica los máximos y mínimos de los datos de precipitación, encontrándose el mayor valor en noviembre de 1999 con 201,8 mm de precipitación.

Comparando con el trabajo realizado en el que se co-teja dos tratamientos de datos usando métodos paramétrico y bootstrap (propuesto) a la variable meteorológica promedios de radiación solar en cada hora-día, para probar la diferencia en medias de radiación en las épocas lluviosa y seca en la ciudad de Riobamba, nos indica que el método Bootstrap propuesto es una nueva alternativa científica para hipótesis concernientes a la diferencia de medias en variables independientes en la cual se observa de mejor manera la variación de la radiación (Meneses-Freire, 2017).

CONCLUSIONES

- Se identificaron los valores atípicos, los que separan la muestra, aplicando los métodos robustos, para estimar el comportamiento de los datos en la ciudad de Riobamba, disminuyendo los sesgos que se puedan dar por la presencia de éstos.
- El análisis robusto muestra las variaciones de la precipitación, disminuye los sesgos que se dan con las medias tradicionales, identificando efectos de la variabilidad de los datos en el que el mayor valor fue en noviembre de 1999 con 201,8 mm de precipitación.
- Con el método bootstrap se determinó una similitud entre los valores de la media y la media winzorizada en algunos meses de febrero, marzo y abril. Determinando así que en estos meses los datos no tienen muchos valores atípicos durante los treinta años de monitoreo

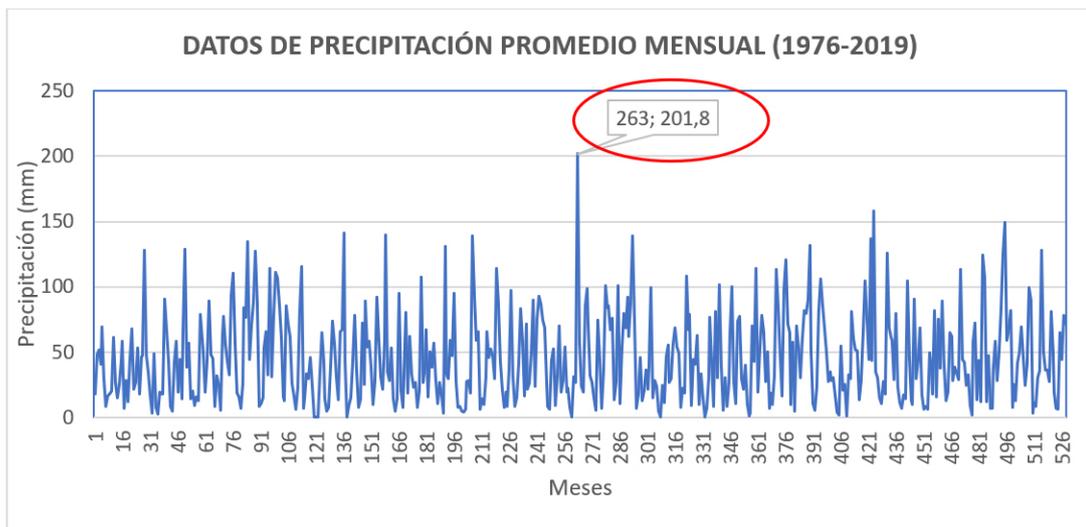


Figura 1. Precipitación promedio mensual registrada en el periodo 1978 a 2020

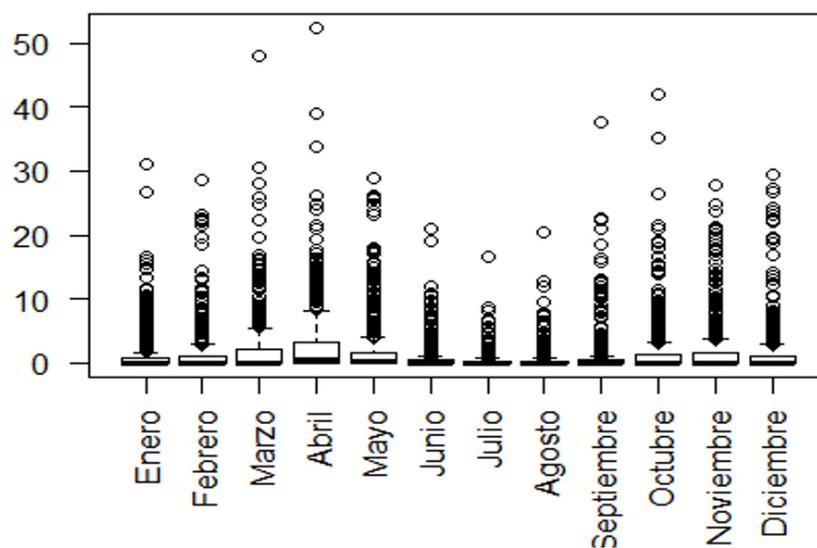


Figura 2. Diagrama de caja de los datos de precipitación promedio mensual en el periodo 1978 a 2019

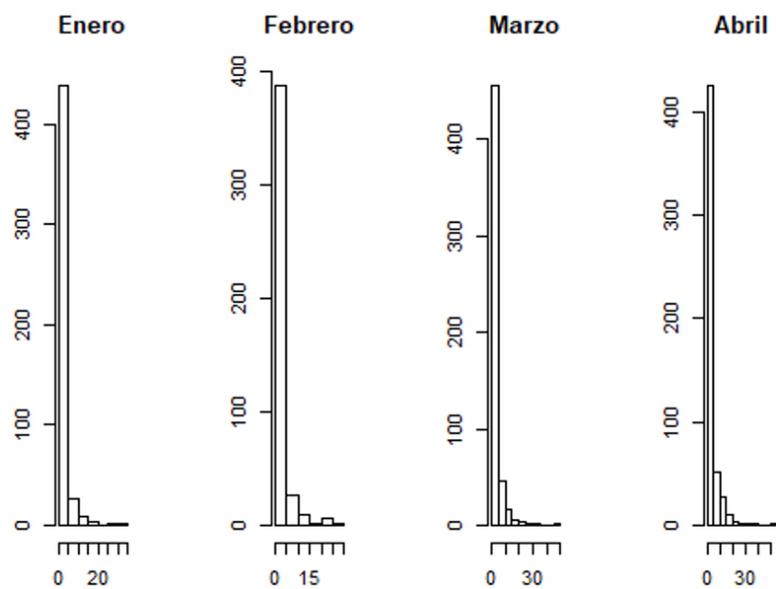


Figura 3. Histograma de precipitación diaria, primer cuatrimestre

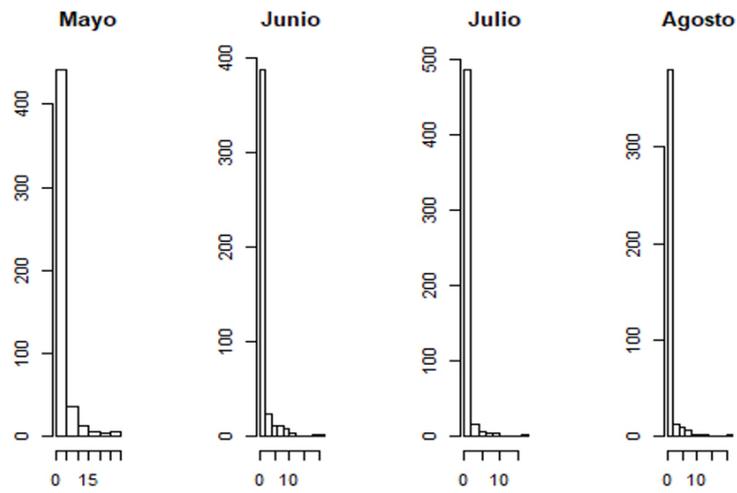


Figura 4. Distribución precipitación, segundo cuatrimestre

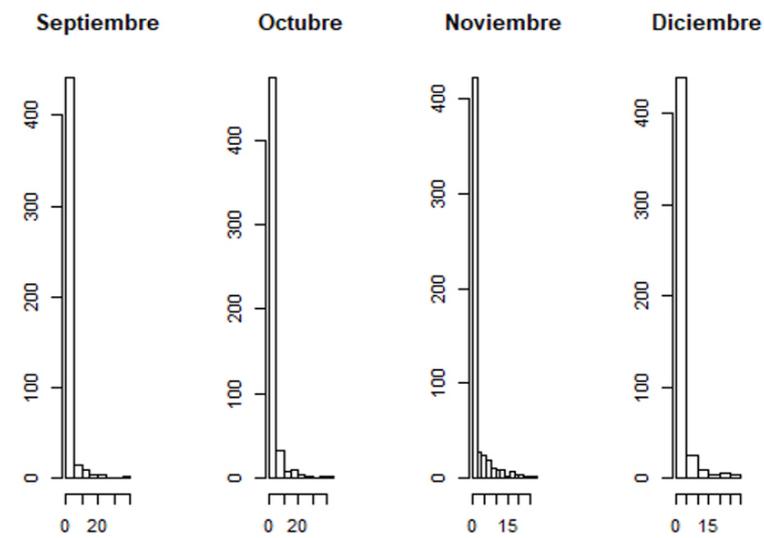


Figura 5. Distribución precipitación, tercer cuatrimestre

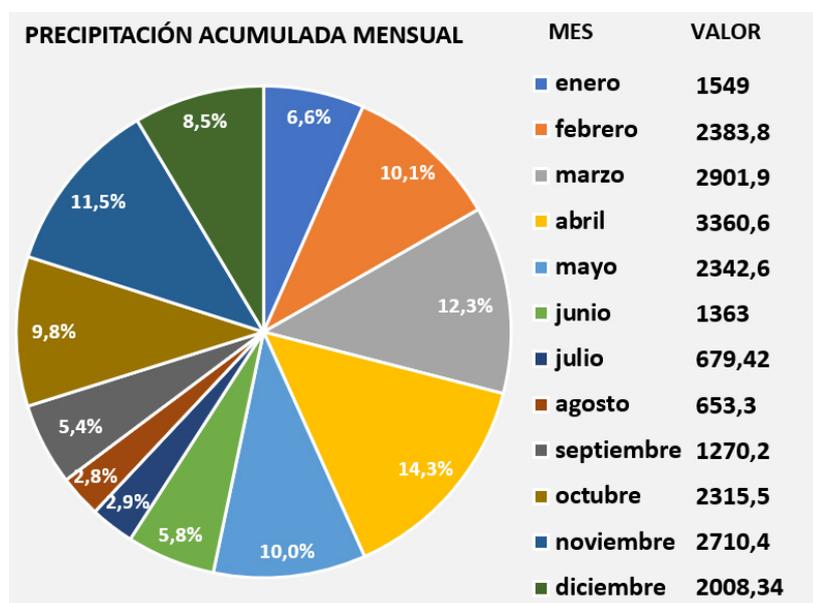


Figura 6. Precipitación en (mm) promedio mensual período 1978 a 2019

Tabla 1. Prueba de normalidad de Kolmogorov-Smirnov de los datos de precipitación por meses

Variables	Parámetros normales	Máximas diferencias extremas	Positivo	Negativo	Estadístico de prueba	Sig. asintótica (bilateral)
	Media	Absoluta				
Enero	1,2925	0,162	0,162	-0,080	0,162	0,032
Febrero	1,6517	0,143	0,143	-0,112	0,143	0,119
Marzo	2,1575	0,076	0,076	-0,056	0,076	0,200
Abril	2,9558	0,100	0,100	-0,064	0,100	0,200
Mayo	2,0784	0,116	0,116	-0,073	0,116	0,200
Junio	1,1090	0,207	0,207	-0,169	0,207	0,002
Julio	0,6572	0,327	0,327	-0,308	0,327	0,000
Agosto	0,8672	0,347	0,347	-0,288	0,347	0,000
Septiembre	1,3735	0,244	0,244	-0,236	0,244	0,000
Octubre	2,0147	0,206	0,206	-0,173	0,206	0,001
Noviembre	2,3187	0,247	0,247	-0,147	0,247	0,000
Diciembre	1,9175	0,257	0,257	-0,196	0,257	0,000

Tabla 2. Análisis robusto de datos

	Media	Mediana	Media acotada 20%	Media winzorizada	Desviación estándar	NMAD
Enero	36,98	31,10	34,63	35,93	22,96	23,72
Febrero	56,23	48,10	51,60	52,13	35,79	37,66
Marzo	66,18	58,80	63,37	62,19	34,40	42,48
Abril	75,73	76,95	74,74	74,24	31,63	30,84
Mayo	52,84	49,20	50,65	49,37	30,30	29,13
Junio	33,17	28,10	29,77	31,04	21,02	20,61
Julio	15,77	10,90	12,94	11,84	13,12	5,86
Agosto	14,05	13,25	13,18	13,08	12,11	9,56
Septiembre	27,24	25,55	26,11	25,17	23,01	21,42
Octubre	53,13	56,20	52,32	51,48	29,29	31,50
Noviembre	63,58	53,25	57,55	56,58	39,37	36,32
Diciembre	45,92	38,80	43,53	45,21	30,28	24,46

ESTIMADORES DE LOCALIZACIÓN

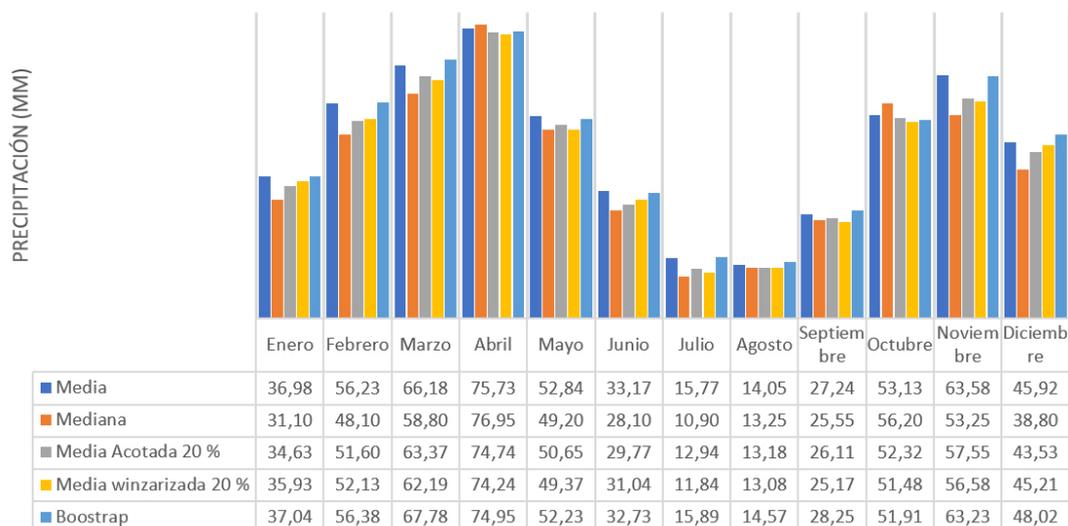


Figura 7. Estimadores de localización media, mediana, media acotada y media winzorizada de la precipitación media diaria

DISTRIBUCIÓN NORMAL CALCULADA USANDO EL MÉTODO BOOTSTRAP

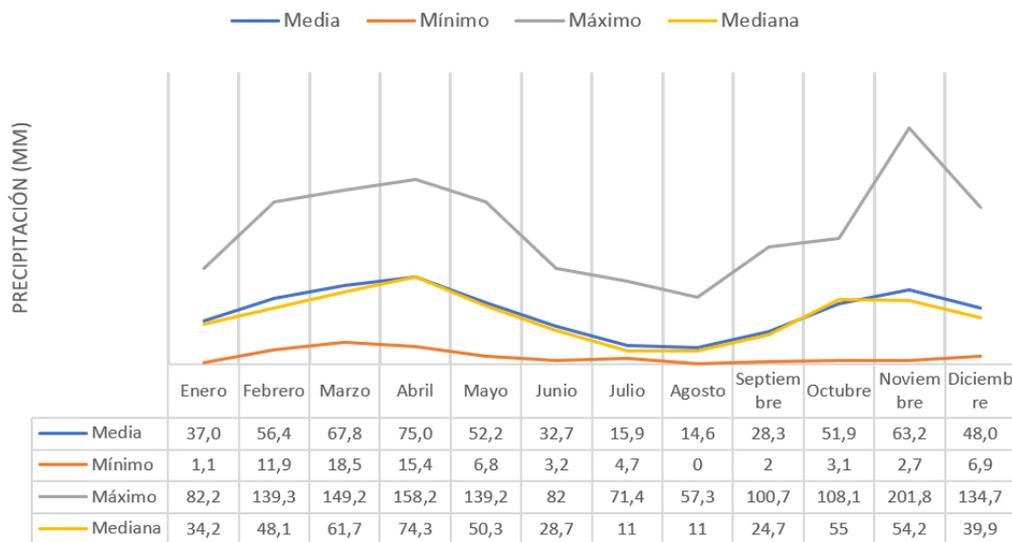


Figura 8. Precipitación estimada con el método bootstrap en los diferentes meses del año

- En el caso del mes de noviembre, con el método bootstrap se tiene más datos atípicos con un valor de media de 63,58 y la media winsorizada de 56,58, una de las razones de esta anomalía es el dato más alto de la figura 1, teniendo en cuenta que en los demás años el mes de noviembre no tiene valores de precipitación muy significativos.
- Los meses de febrero, marzo, abril y mayo presentan mayor regularidad en sus precipitaciones, la mayor precipitación es abril con 3360,6 mm y el de menor precipitación agosto con 653,3 mm, con dos temporadas húmedas a inicios y fin de año con menor intensidad en la última parte.

AGRADECIMIENTOS

Agradecemos el Grupo de Energías Alternativas y Ambiente (GEAA) de la ESPOCH por los datos de precipitación que fueron analizados para la presente investigación.

REFERENCIAS

- Balmaseda Espinosa, C. E. y Mederos Machado, M. C. (2019). Visión del cambio climático desde la perspectiva del riego y drenaje en Ecuador. *Revista Iberoamericana Ambiente & Sustentabilidad*, 2(1), 68-73. DOI: 10.46380/rias.v2i1.41

- Belío, J. (2020). *Métodos bootstrap y sus aplicaciones* (grado en matemáticas). Universidad de Zaragoza.
- Bello, A. M. (2019). Técnicas de imputación para datos de precipitación máxima mensual en la zona central de Boyacá. *Ingeniería, Investigación y Desarrollo*, 19(1), 64-79.
- Boente, G. (2021). Inferencia robusta: de lo finito a lo infinito-dimensional, 150-78. Available from: <https://www.ancefn.org.ar/user/CONTINUACION-ANALES/TOMO72/Boente-Inferenciारobusta.pdf>
- Calvo, F. (2017). *Estimadores robustos de localización en Fase I basados en esquemas CUSUM* (trabajo de grado). Universidad Nacional de Colombia. <https://repositorio.unal.edu.co/bitstream/handle/unal/62275/1024520821-2017.pdf?sequence=1&isAllowed=y>
- Fay, D. L., Akhavan, S. y Goldberg, V. M. (2017). INIAP-Josefina: Nueva variedad de papa con tolerancia a la sequía. *Rev. Latinoamericana de la Papa*, 20(465), 39-54.
- Gil Marín, J. A., Lozada García, B. I., López Pérez, N. y Grozdanac Berardelli, L. M. (2012). Régimen pluviométrico de Maturín, estado Monagas, Venezuela. *Revista Científica UDO Agrícola*, 12(1), 188-97. <https://dialnet.unirioja.es/servlet/articulo?codigo=4277146>
- Gil, S. (2014). *Bootstrap en poblaciones finitas* (trabajo de grado). Universidad de Granada. https://masteres.ugr.es/moea/pages/tfm1314/tfm_samuel_gil_abreu/
- González-Hidalgo J. C., Peña-Angulo, D., Beguería, S. y Brunetti, M. (2020). MOTEDAS century: a new high-resolution secular monthly maximum and minimum temperature grid for the Spanish mainland (1916-2015). *International Journal of Climatology*, 40(12), 5308-5328. DOI: 10.1002/joc.6520
- Guijarro, J. A., Azorín Molina, C., González Hidalgo, J. C., Sánchez Lorenzo, A., Herrera García, S. y López Díaz, J. A. (2017). Observaciones climáticas atmosféricas y reconstrucciones instrumentales sobre la península Ibérica I. obtención de series climáticas de alta calidad. Volumen especial sobre el clima en la península Ibérica: una visión científica global. Agencia Estatal de Meteorología. *CLIVAR Exchange*, (73), 11-14. <http://hdl.handle.net/20.500.11765/10216>
- Hidalgo, M. (2017). Variabilidad climática interanual sobre el Ecuador asociada a ENOS. *CienciaAmérica*, 6(2), 32-7.
- Irigoyen, A., Suero, E. y Gardiol, J. (1997). Tendencia y estacionalidad de la precipitación en Balcarce, (1), 5.
- Kluwer, W. (2018). *Diccionario empresarial*. https://guias-juridicas.wolterskluwer.es/Content/Documento.aspx?params=H4sIAAAAAAEAMtMSbF1jTAAAS-NjM2MztlbLUouLM_DxbIwMDS0NDA1OQQGZapUt-ckhlQaptWmJOCSoAvvhbpzUAAAA=WKE
- López Calvajar, M. G. A., López Fernández, D. C. R. y León González, D. C. J. L. (2017). Análisis de la influencia de factores climatológicos en la pérdida de superficie sembrada de cultivos transitorios en el Ecuador. *Revista Científica Agroecosistemas*, 5(1), 176-183. <https://aes.ucf.edu.cu/index.php/aes/article/view/155>
- Méndez, D. R., Elizabeth, M., Zaira, C. y Valero, V. (2020). Estadística robusta aplicada a las medidas de localización y escala. *Nota Técnica*, (6), 499-517.
- Meneses-Freire, A., Zúñiga-Lema, L. y Haro, A. (2017). Método bootstrap propuesto para hipótesis concernientes a la diferencia de medias en variables independientes. *Perfiles*, 1(17), 67-73. <http://ceaa.esPOCH.edu.ec:8080/revista/perfiles/Articulos/Perfiles17Art8.pdf>
- Orozco, M. (1992). *Estudio comparativo del bootstrap clásico y suavizado* (tesis de grado). Universidad de los Andes. <https://repositorio.uniandes.edu.co/flexpaper/handle/1992/39311/u806819.pdf?sequence=1&isAllowed=y#page=1>
- Otzen, T. y Manterola, C. (2017). Técnicas de muestreo sobre una población a estudio. *International Journal of Morphology*, 35(1), 227-32.
- Priyan, K. (2015). Spatial and temporal variability of rainfall in anand district of Gujarat State. *Aquatic Procedia*, (4), 713-20. <http://dx.doi.org/10.1016/j.aqpro.2015.02.092>
- Romero Saldaña, M. (2016). Pruebas de bondad de ajuste a una distribución normal. *Enfermería del Trabajo*, 6(3), 105-14. <https://dialnet.unirioja.es/servlet/articulo?codigo=5633043>
- Ruano, E. (2018). *Manual del buen uso del muestreo estadístico*. Universidad de Cantabria Miguel Hernández. <https://sdei.unican.es/Paginas/servicios/correo/Manual-del-buen-uso-del-Correo-Electronico.aspx>
- Saquina, C. (2020). *Correlación de variables climáticas (temperatura y precipitación) con el índice de cobertura vegetal con el método NDVI en la zona noroccidental de la provincia de Cotopaxi en el piso bioclimático, bosque siempreverde montano de la cordillera occidental* (bachelor thesis). Universidad Técnica de Cotopaxi. <http://repositorio.utc.edu.ec/bitstream/27000/4501/1/PI-000727.pdf>
- Velázquez, A. P. (2017). *Tipo de muestreo* (diplomado en análisis de información geoespacial). Conacyt, 14. <http://centro-geo.repositorioinstitucional.mx/jspui/handle/1012/163>
- Vianez, J. P. (2019). Estimación del riesgo de reserva con solvencia II: Distribución libre de Mack vs. Bootstrap con simulación. *Espacios*, 40(18), 10.
- Villegas, D., Rivas, Y., Pérez, Y., Villegas, S. y Milla, M. (2019). Evaluación de observaciones atípicas en datos climatológicos y en modelos lineales simulados. *Revista Climatología*, (19), 67-73.
- Zhiña, D., Pacheco, J. y Avilés, A. (2017). Modelos estocásticos para el pronóstico de sequías en la microcuenca del río Chulco en Ecuador. *Revista de la Facultad de Ciencias Químicas*, 1(16), 44-61. <https://publicaciones.ucuenca.edu.ec/ojs/index.php/quimica/article/view/1585>