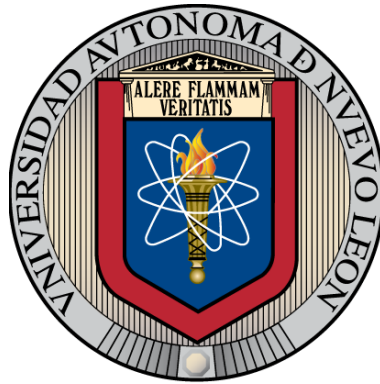


UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS



**FACTORES METEOROLÓGICOS DETERMINANTES DE SEQUÍA EN
CONDADOS DE ESTADOS UNIDOS**

Por

IRVING DANIEL ESTRADA LÓPEZ

Como requisito parcial para obtener el Grado de
MAESTRÍA EN CIENCIA DE DATOS

Octubre, 2023

Universidad Autónoma de Nuevo León

Facultad de Ciencias Físico Matemáticas

Los miembros del Comité de Tesis recomendamos que la Tesina “Factores Meteorológicos Determinantes De Sequía En Condados De Estados Unidos”, realizada por el alumno Irving Daniel Estrada López, con número de matrícula 1739907, sea aceptada para su defensa como opción al grado de Maestría en Ciencia de Datos.

El Comité de Tesis

Dr. José de Jesús Rocha Salazar
Director

Dra. Elena Cristina Villanueva Guerra
Revisor

Dr. Álvaro Eduardo Cordero Franco
Revisor

Vo.Bo.

M.C. Azucena Yoloxóchitl Ríos Mercado
Coordinadora de la Maestría en Ciencia de Datos

San Nicolás de los Garza, N.L.

DEDICATORIA

Para mis papás.

ÍNDICE GENERAL

Capítulo	Página
1. Introducción	10
2. Delimitación y Planteamiento del Problema de Investigación	12
3. Justificación	14
4. Formulación de Objetivos	16
4.1 Objetivo General	
4.2 Objetivos Específicos	
5. Marco Teórico	17
5.1 Sequía	
5.1.1 Sequía Meteorológica	
5.1.2 Sequía Agrícola	
5.1.3 Sequía Hidrológica	
5.1.4 Sequía Socioeconómica	
5.2 Indicadores de Sequía	
5.2.1 <i>Palmer Drought Severity Index (PDSI)</i>	
5.2.2 <i>Climate Prediction Center Soil Moisture Model (CPC-SMM)</i>	
5.2.3 <i>United States Geological Services Weekly Streamflow (USGS-WS)</i>	
5.2.4 <i>Standardized Precipitation Index (SPI)</i>	
5.2.5 <i>Objective Drought Indicator Blends (ODIB)</i>	
5.3 Series de Tiempo	
5.3.1 Media, Varianza y Covarianza	

5.3.2 Estacionalidad	
5.3.3 Panel de Datos	
5.3.3.1 Características	
5.4 Regresión	
5.4.1 Regresión Simple	
5.4.2 <i>Ordinary Least Squares (OLS)</i>	
5.4.3 Efectos Ajustados	
5.4.4 Efectos Aleatorios	
5.4.5 Modelo Lineal Mixto	
5.5 Antecedentes	
6. Metodología	35
6.1 Datos	
6.1.1 <i>U.S. Drought Monitor (USDM)</i>	
6.1.2 Conjunto de Datos y Variables	
6.1.2.1 Códigos de Condados	
6.1.2.2 Precipitación Meteorológica	
6.1.2.3 Presión Superficial	
6.1.2.4 Temperatura	
6.1.2.5 Punto de Rocío	
6.1.2.6 Temperatura de Bulbo Húmedo	
6.1.2.7 Temperatura de la Superficie de la Tierra	
6.1.2.8 Velocidad de Viento	
6.1.2.9 Índice de Severidad y Cobertura de Sequía	
6.2 Python	
6.2.1 Librería <i>statsmodels</i>	
6.2.2 Librería <i>linearmodels</i>	
6.3 Limpieza de Datos	
6.3.1 Resumen Semanal	
6.3.2 Multicolinealidad	
6.3.2.1 Coeficiente de Correlación de Pearson	

6.3.2.2 *Variance Inflation Factor (VIF)*

6.4 Modelos

6.4.1 Efectos Ajustados: *OLS*

6.4.2 Modelo Lineal Mixto

7. Resultados y Conclusiones 51

7.1 Modelo de Efectos Ajustados

7.2 Modelo Lineal Mixto

7.3 Comparativa de Modelos

7.4 Discusión

7.5 Conclusión

Referencias62

ÍNDICE DE FIGURAS

Figura	Página
6.1 Severidad de Sequía USDM	36
6.2 Ciclo del Agua	38
6.3 Porcentaje de Área en Sequía	41
6.4 Muestra de Datos (5 variables de 20)	44
6.5 Resumen Semanal en Python	45
6.6 Matriz de Correlación de Pearson	46
6.7 VIFs Finales	48
6.8 Efectos Ajustados en Python	49
6.9 Modelo Lineal Mixto en Python	49

ÍNDICE DE TABLAS

Tabla	Página
6.1 Rangos de Índices con su Categoría	41
6.2 Resumen de Variables	42
7.1 Resultados de Modelo de Efectos Ajustados	53
7.2 Resultados del Modelo Lineal Mixto	56
7.3 Comparativa Entre Modelos	58

AGRADECIMIENTOS

Agradezco a mis papás por siempre apoyarme de forma incondicional, sin importar la situación profesional o económica en la que me encuentre. Gracias por siempre creer en mí.

A Andrea Morón por siempre estar a mi lado sin importar las circunstancias. Gracias por siempre creer en mí y permitirme caminar juntos en la vida. Eres mi persona favorita.

A Cynthia Estrada, Diego Narváez, Erick Estrada, Karen Capistran, Rodrigo Ríos, Carlos Ibáñez y Rodrigo Torres por siempre darme ánimos y su comprensión.

A mi asesor el Dr. José de Jesús Rocha Salazar por guiarme, darme buen material de lectura y brindarme la confianza de siempre preguntar. Lo respeto y lo admiro enormemente.

A mis revisores la Dra. Elena Cristina Villanueva Guerra y el Dr. Álvaro Eduardo Cordero Franco por ayudarme a complementar de una forma excepcional este documento.

A la Facultad de Ciencias Físico Matemáticas por ser la institución que es y tener excelentes docentes. Agradezco a la M.C. Azucena Ríos por la excelente coordinación de la Maestría en Ciencia de Datos.

CAPÍTULO 1

INTRODUCCIÓN

La sequía se define como la reducción drástica de precipitación en un largo periodo de tiempo (Malik, 2020). Según la ONU (2022), la población mundial llegó a los 8 mil millones de personas con algunas comunidades sin agua potable. En el planeta cerca de 2.1 billones de personas no cuentan con acceso al agua potable, suceso que aunado a las sequías afectan el crecimiento económico de grandes áreas o países enteros (Sundararajan, 2021). Según el *U.S. Drought Monitor* (2022) en Estados Unidos más del 50% de los estados enfrentan sequías de distintas severidades.

La sequía, un complicado fenómeno natural causado por un gran número de variables, es un problema global que no conoce de fronteras (será descrito más a profundidad en el capítulo 3). El entendimiento de sus factores determinantes es un tema de máxima relevancia en la actualidad que permite definir medidas mitigantes.

El objetivo general de esta investigación es generar un modelo explicativo capaz de mostrar las variables meteorológicas determinantes de la sequía. Además del objetivo general, se van a cubrir dos objetivos específicos:

1. Elaborar el modelo explicativo en Python.
2. Comparar el comportamiento que tiene un modelo lineal mixto contra un modelo de efectos ajustados.

Los modelos serán aplicados al mismo panel de datos que contiene registros meteorológicos diarios de ciertas regiones de Estados Unidos.

Después de esta introducción, en el capítulo 2 se delimitan y plantean los problemas de investigación explicando cómo las sequías afectan los condados de Estados Unidos. En el capítulo 3 se expone la motivación de la investigación, exponiendo las diferentes afectaciones globales y en Estados Unidos de las sequías. La intención del capítulo 4 es definir el objetivo general y los objetivos específicos de la investigación. En el capítulo 5 se revisa la literatura relacionada al tema de estudio. La metodología de investigación, expuesta en el capítulo 6, busca documentar la fase experimental. El capítulo expone la preparación de los datos y la construcción de los modelos en Python. El capítulo 7 finaliza con los principales resultados y conclusiones. En esta sección se muestran las salidas de los códigos implementados y se compararán los dos modelos. Después, se comparan los coeficientes del modelo de efectos ajustados y los coeficientes del modelo lineal mixto.

CAPÍTULO 2

DELIMITACIÓN Y PLANTEAMIENTO DEL PROBLEMA DE INVESTIGACIÓN

La sequía se define como la reducción drástica de precipitación en un largo periodo de tiempo (Malik, 2020). En la literatura se definen 4 tipos de sequías según el mecanismo de déficit del agua: la sequía meteorológica, agrícola, hidrológica y socioeconómica (Hao, 2018). La sequía meteorológica se relaciona con la falta de precipitación en un largo periodo de tiempo. La sequía agrícola se define en función de la humedad del suelo. La sequía hidrológica involucra la falta de escurrimientos superficiales y subterráneos de agua. Finalmente, la sequía socioeconómica se relaciona con la oferta demanda de los bienes económicos. Esta última, tiene una fuerte relación con la interacción de las actividades humanas. Sundararajan (2021) menciona que las tres variaciones críticas que causan una sequía son el aumento de evaporación, transpiración y el déficit de precipitación. El aumento de evaporación y transpiración es debido a altas temperaturas, fuertes vientos, pocas nubes y poca humedad. Por otro lado, la deficiencia de precipitación reduce los escurrimientos de agua.

En Estados Unidos más del 50% de los estados enfrentan sequías de distintas severidades afectando a 378.5 millones de acres y cerca de 139.7 millones de personas (*National Integrated Drought Information System*, 2022). En California durante el periodo 2011 a 2017 se registraron pérdidas de \$5.5 mil millones de dólares representadas en la pérdida de 130 millones de árboles, debido a una sequía extrema (Han, 2020). Un fenómeno similar se observó en Texas en el año 2022, afectando más del 93% del estado (*U.S Drought Monitor*, 2022).

Existen una gran cantidad de índices de sequía. El índice utilizado en esta tesina está dado por el *U.S. Drought Monitor* que se basa en 5 indicadores para su clasificación: *Palmer Drought Severity Index (PDSI)*, *Climate Prediction Center Soil Moisture Model (CPC-SMM)*, *United States Geological Services Weekly Streamflow (USGS-WS)*, *Standardized Precipitation Index (SPI)* y el *Objective Drought Indicator Blends (ODIB)*. El monitor lleva registros diarios de indicadores meteorológicos. Estos registros no corresponden a pronósticos de sequía, más bien son evaluaciones manuales de sus condiciones periféricas. La evaluación se lleva a cabo de forma semanal y es utilizada por organizaciones y el gobierno de Estados Unidos. El presente estudio utiliza registros meteorológicos de distintos condados de Estados Unidos para identificar los factores determinantes de la sequía. Los datos se estudian a detalle en el capítulo 6, en el apartado de Datos.

CAPÍTULO 3

JUSTIFICACIÓN

La escasez de agua cada vez es más común y es uno de los desastres naturales más costosos. Esta afecta a más personas que cualquier desastre natural, sin importar su nivel socioeconómico. Las sequías afectan desde la producción de alimentos hasta su consumo.

Según Salvador (2020) las sequías causan afectaciones a corto y largo plazo en diversos sectores que se pueden resumir en los siguientes puntos:

- El déficit de agua causa la pérdida de la biodiversidad.
- La calidad del aire es afectada por condiciones de sequía grave, y la materia suspendida puede irritar los conductores branquiales y los pulmones.
- En casos extremos, las sequías prolongadas se han asociado con un aumento de la tasa de suicidio, principalmente en poblaciones rurales.
- El riesgo de incendios forestales se ve incrementado.
- La sequía contribuye al desabasto de alimentos.
- Las sequías prolongadas se asocian con pérdidas socioeconómicas y desplazamientos forzados de comunidades.

La población mundial llegó a los 8 billones en noviembre del 2022 y en el mismo año cerca de 2.1 billones de personas no contaban con acceso a agua potable (ONU, 2022). Distintas comunidades como Etiopía, Kenia y Somalia experimentaron fuertes sequías impactando negativamente entre 9.5 y 16.2 millones de personas (UNICEF, 2022). En España las productoras de aceite de oliva se vieron afectadas por una sequía agrícola (BBC, 2022) y en México el

75% de los municipios enfrentaron un recorte de agua por sequía extrema en el mismo año (The New York Times, 2022).

De manera particular, Estados Unidos experimentó grandes pérdidas económicas en el periodo de 1980 a 2017 debido a la sequía. En el periodo de 1988-1989 una sequía impactó a los cultivos estadounidenses causando una pérdida de \$15 billones de dólares (Lu, 2020). En California durante el periodo de 2011 a 2017 una sequía extrema impactó 130 millones de árboles equivalente a una pérdida de \$5.5 mil millones de dólares (Han, 2020). Además de los impactos económicos, Christensen (2019) muestra que la sequía tiene un efecto en la salud, siendo un factor determinante de las enfermedades hemorrágicas en los Estados Unidos.

En resumen, el entendimiento de factores meteorológicos determinantes de la sequía es un tema de máxima relevancia en la actualidad que permite la definición de los planes de mitigación. La falta de agua trae consigo problemas en cadena, causados inicialmente por la falta de precipitación. Por lo tanto, es crucial comprender y abordar estos efectos en cadena, para mitigar los impactos de la sequía en diferentes sectores y garantizar la sustentabilidad hídrica a largo plazo.

CAPÍTULO 4

FORMULACIÓN DE OBJETIVOS

4.1 OBJETIVO GENERAL

El objetivo de esta investigación es identificar las variables meteorológicas determinantes en la sequía de Estados Unidos. Se encontrarán indicadores de su significancia a través de los p – *valores* de un modelo explicativo, generado con datos longitudinales, o un panel de datos, de registros meteorológicos, en el cual el “elemento” son los distintos condados de Estados Unidos. El modelo se ajustará a la serie de tiempo del respectivo condado.

4.2 OBJETIVOS ESPECÍFICOS

Se cubrirán 2 objetivos específicos:

1. Elaborar el modelo explicativo en Python.
2. Comparar el comportamiento que tiene un modelo lineal mixto contra un modelo de efectos ajustados.

Ambos modelos son aplicados a un panel de datos que contiene registros meteorológicos diarios de las regiones de Estados Unidos.

CAPÍTULO 5

MARCO TEÓRICO

En este capítulo, se aborda la revisión bibliográfica necesaria para fundamentar la metodología propuesta. En primer lugar, se presenta una definición precisa de los conceptos climatológicos vinculados a la sequía, sentando las bases teóricas para el análisis. Asimismo, se menciona la fundamentación matemática subyacente a las librerías de Python empleadas en el estudio, con el objetivo de garantizar una implementación robusta y confiable. Por último, se exploran los antecedentes relacionados con la sequía, principalmente en aquellos que han utilizado un enfoque de efectos ajustados o modelos lineales mixtos en diversas regiones. Esta revisión de la literatura científica asegura la validez de la metodología presentada en esta tesina, ofreciendo una sólida contribución al conocimiento en el campo de estudio y facilitando la replicabilidad de los métodos propuestos para futuras investigaciones.

5.1 SEQUÍA

Malik (2020) define la sequía como la reducción drástica de precipitación en un largo periodo de tiempo. Richardson (2019) define la sequía como un periodo de clima anormalmente seco con una duración larga que causa una afectación al balance hidrológico. La sequía afecta a grandes regiones por meses o incluso años. Tiene un gran impacto en la producción de comida que reduce las expectativas de vida.

Según Sundararajan (2021) la sequía afecta el crecimiento económico de grandes áreas o países enteros. La sequía es un complicado fenómeno natural,

ya que es de los menos entendidos debido a sus múltiples mecanismos causantes. La sequía a diferencia de otros desastres naturales se desarrolla de forma lenta. Esta es causada por distintos factores que contribuyen en diferente escala temporal y espacial, principalmente por la falta de precipitación, altas temperaturas y la evapotranspiración. Las altas temperaturas son causantes de la evaporación y reducción de la humedad del suelo, afectando al proceso hidrológico. La sequía evoluciona debido a las anomalías meteorológicas, los procesos de la superficie y la actividad humana. Según Hao (2018) existen cuatro tipos de sequías: meteorológica, agrícola, hidrológica y socioeconómica. La categoría depende del impacto causado por la sequía.

5.1.1 Sequía Meteorológica

La sequía meteorológica se relaciona con la falta de precipitación (Han, 2020) y es causada generalmente por anomalías persistentes en los patrones de circulación atmosférica a gran escala. Las anomalías de circulación se deben a las anomalías temperaturas superficiales del mar.

Las tres variaciones del clima críticas son el aumento de evaporación, transpiración y la falta de precipitación (Sundararajan, 2021). El aumento de evaporación y transpiración es debido a las altas temperaturas, fuertes vientos, menos nubosidad y poca humedad. La deficiencia de precipitación reduce los escurrimientos de agua, propagándose por todo el ciclo hidrológico. Este tipo de sequía usualmente es la causante de las demás. Los índices más utilizados para este tipo de sequía es el *SPI*, *PNP* (*Percent of Normal Precipitation*) y el *PDSI*.

5.1.2 Sequía Agrícola

La sequía agrícola se relaciona con la falta de humedad del suelo y es causada por el aumento de evaporación, que reduce la humedad del suelo (Hao, 2018). Esta prolonga los efectos de la sequía meteorológica. Los indicadores más utilizados para la sequía agrícola son: *SMP* (*Soil Moisture Percentile*), *CMI* (*Crop Moisture Index*), *SMDI* (*Soil Moisture Deficit Index*), *SMDI* (*Soil Moisture Deficit Index*) y el *SSI* (*Standardized Soil Moisture Index*).

5.1.3 Sequía Hidrológica

La sequía hidrológica está relacionada con la falta de escurrimientos superficiales y subterráneos de agua (Hao, 2018). Las sequías meteorológicas usualmente evolucionan a sequías hidrológicas, sin embargo, esto no es instantáneo. Esa evolución está dada por complejos mecanismos físicos. Las posibles causas, además de los antecedentes de falta de precipitación, son: el bajo almacenamiento de agua, las bajas temperaturas y la acumulación de nieve. Los índices comúnmente utilizados para la sequía hidrológica son: *PHDI (Palmer Hydrologic Drought Index)* y el *SRI (Standardized Runoff Index)*.

5.1.4 Sequía Socioeconómica

Según Hao (2018) la sequía socioeconómica se relaciona con la oferta y demanda de los bienes económicos. A diferencia de los anteriores tipos de sequías, para la sequía socioeconómica no se han desarrollado tantos indicadores. Los indicadores que se utilizan son el rendimiento de cultivos y la escasez de suministro de agua. Su definición resalta la fuerte interacción entre la sequía y las actividades humanas. El *Objective Drought Indicator Blends (ODIB)* es un indicador de sequía socioeconómica.

5.2 INDICADORES DE SEQUÍA

Según Zargar (2011) los indicadores de sequía son medidas cuantitativas que caracterizan los niveles de sequía de una o varias variables en un solo valor numérico. Los indicadores pueden reflejar la sequedad anómala en el clima, pérdida de humedad del suelo o incluso la disminución de un embalse. Debido a que los índices de sequías pueden estar basados en múltiples factores, se han creado múltiples índices de sequía. Sundararajan (2021) nos dice que existen más de 100 distintos indicadores de sequía. Nuestra variable objetivo está basada en 5 indicadores que se van a presentar a continuación. La justificación del uso de estos indicadores se encuentra en el capítulo 6.

5.2.1 *Palmer Drought Severity Index (PDSI)*

El Indicador de Severidad de Sequía de Palmer o *PDSI*, está basado en el concepto de oferta demanda del agua. Este indicador hace mayor énfasis en las anomalías de la humedad del suelo, que en las anomalías del clima (Yihdego, 2019). El *PDSI* utiliza datos de la precipitación, la temperatura y la disponibilidad local del agua. Con estos datos calcula tres términos en el balance del agua: la evapotranspiración, escurrimientos y la humedad. El *PDSI* no considera el impacto humano en el balance del agua. Este indicador usualmente es utilizado como indicador de sequía hidrológica.

5.2.2 *Climate Prediction Center Soil Moisture Model (CPC-SMM)*

El Centro de Predicción de Clima tiene como propósito la mitigación de desastres naturales. Este centro realiza monitoreo en tiempo real de variables meteorológicas. El Modelo de Humedad del Suelo del Centro de Predicción del Clima o *CPC-SMM*, se estima mediante un modelo hidrológico de cubeta con fugas de una sola capa (*One-Layer Leaky Bucket*). El modelo utiliza como datos de entrada la precipitación y las temperaturas observadas, después calcula la humedad del suelo, la evaporación, el escurrimiento y el espesor de la capa de nieve.

5.2.3 *United States Geological Services Weekly Streamflow (USGS-WS)*

El registro diario de los flujos de corrientes alrededor de Estados Unidos lo monitorea los Servicios Geológicos Semanales de Escurrimientos de los Estados Unidos o *USGS-WS*. Este organismo provee monitoreo, análisis y predicciones de los desastres naturales, recursos naturales, ecosistemas y los efectos del cambio climático.

5.2.4 *Standardized Precipitation Index (SPI)*

El Indicador Estandarizado de Precipitación o *SPI*, es un popular indicador de sequía que solamente se basa en la precipitación. El *SPI* compara la precipitación con el promedio de precipitación de múltiples años (Yihdego, 2019). Según Zagar (2011) al igual que el porcentaje de precipitación normal o *PPN*, el

SPI compara la cantidad de precipitación con su promedio a lo largo de los años. Los registros de precipitación son ajustados a una distribución gamma y luego se convierten en una distribución normal. La media se establece en cero, de esta manera los valores positivos indican periodos húmedos y los valores negativos indican periodos secos.

El *SPI* representa cuantas desviaciones estándar respecto a su déficit de precipitación acumulada, se aleja del promedio normalizado. Si se observa consistentemente un valor menor que cero y alcanza un valor de -1 o menos, se considera que ha ocurrido una sequía.

5.2.5 *Objective Drought Indicator Blends (ODIB)*

La Mezcla de Indicadores Objetivos de Sequía o *ODIB* es un indicador que aproxima los impactos relacionados con la sequía que responden a la precipitación en escalas de tiempo. Estos van desde unos pocos días hasta unos pocos meses. Algunos de los impactos son: incendios forestales, la humedad en la capa superior del suelo, las condiciones de los pastizales y los flujos de agua no regulados.

5.3 SERIES DE TIEMPO

Según Wooldridge (2009) las series de tiempo consisten en observaciones de una o más variables a través del tiempo. La frecuencia es una característica importante de las series de tiempo que se relaciona con la forma de recolección de los datos. Esta recolección puede ser diaria, semanal, mensual o anual.

El análisis de una sola secuencia de datos es llamado análisis de serie de tiempo univariado. El análisis de distintas secuencias de periodos de tiempo es llamado análisis de series de tiempo multivariado. La secuencia de una variable aleatoria $\{Y_t : t = 0, \pm 1, \pm 2, \pm 3, \dots\}$ es llamada un proceso estocástico y sirve como un modelo para las observaciones de una serie de tiempo (Cryer, 2008).

5.3.1 Media, Varianza y Covarianza

Para un proceso estocástico $\{Y_t: t = 0, \pm 1, \pm 2, \pm 3, \dots\}$ la función de media está definida por:

$$\mu_t = E(Y_t) \quad \text{para } t = 0, \pm 1, \pm 2, \dots$$

Donde μ_t , es el valor esperado del proceso al tiempo t . En general μ_t puede ser diferente en cada tiempo t . La función de autocovarianza, $\gamma_{t,s}$, se define como:

$$\gamma_{t,s} = Cov(Y_t, Y_s) \quad \text{para } t, s = 0, \pm 1, \pm 2, \dots$$

Donde $Cov(Y_t, Y_s) = E[(Y_t - \mu_t)(Y_s - \mu_s)] = E(Y_t Y_s) - \mu_t \mu_s$.

La función de autocorrelación, $\rho_{t,s}$, esta dada por

$$\rho_{t,s} = Corr(Y_t, Y_s) \quad \text{para } t, s = 0, \pm 1, \pm 2, \dots$$

Donde:

$$Corr(Y_t, Y_s) = \frac{Cov(Y_t, Y_s)}{\sqrt{Var(Y_t)Var(Y_s)}} = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t}\gamma_{s,s}}}$$

La covarianza y correlación son medidas de la dependencia lineal entre variables aleatorias (Cryer, 2008). Las siguientes propiedades son importantes y salen de sus definiciones:

$$\gamma_{t,t} = Var(Y_t) \quad \rho_{t,t} = 1$$

$$\gamma_{t,s} = \gamma_{s,t} \quad \rho_{t,s} = \rho_{s,t}$$

$$|\gamma_{t,s}| \leq \sqrt{\gamma_{t,t}\gamma_{s,s}} \quad |\rho_{t,s}| \leq 1$$

Los valores de $\rho_{t,s}$ cercanos a ± 1 indican fuerte dependencia lineal, y los cercanos a cero representan una dependencia lineal débil. Si $\rho_{t,s} = 0$ se dice que Y_t y Y_s no están correlacionadas.

5.3.2 Estacionalidad

Para hacer inferencia estadística sobre la estructura de un proceso estocástico, se tienen que cumplir algunos supuestos. El supuesto más importante es la estacionalidad (Cryer, 2008). La estacionalidad implica que las leyes de probabilidad que gobiernan el comportamiento del proceso no cambian con el tiempo. En cierto sentido, el proceso está en equilibrio estadístico. El proceso $\{Y_t\}$ es estrictamente estacional si la distribución conjunta de $Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}$ es la misma que el conjunto de distribución de $Y_{t_1-k}, Y_{t_2-k}, \dots, Y_{t_n-k}$ para todos los puntos de tiempo t_1, t_2, \dots, t_n y para todos los lapsos de tiempo k . Por lo tanto:

$$\begin{aligned}\gamma_{t,s} &= Cov(Y_{t-s}, Y_0) \\ &= Cov(Y_0, Y_{s-t}) \\ &= Cov(Y_0, Y_{|t-s|}) \\ &= \gamma_{0,|t-s|}\end{aligned}$$

Entonces la covarianza entre Y_t y Y_s depende solo del tiempo a través de las diferencias del tiempo $|t - s|$ y no de los actuales tiempos t y s . Para un proceso estacionario podemos decir que:

$$\gamma_k = Cov(Y_t, Y_{t-k}) \quad y \quad \rho_k = Corr(Y_t, Y_{t-k})$$

5.3.3 Panel de Datos

Según Hsiao (2005), los términos panel de datos y datos longitudinales, hacen referencia a datos que contienen series de tiempos de un número de elementos. Las observaciones en un panel de datos tienen por lo menos dos dimensiones: la dimensión transversal y la dimensión de la serie de tiempo. La

dimensión transversal es una colección de información, recolectada para múltiples variables de interés en una población o muestra específica. En este tipo de conjunto de datos, se observan y registran las características de los elementos de estudio en un momento determinado, sin seguimiento a lo largo del tiempo.

Según Wooldridge (2009) un panel de datos consiste en una serie de tiempo por cada elemento transversal de un conjunto de datos. Su característica principal es que utiliza las mismas unidades transversales a través de los periodos de tiempo. Hsiao (2022) define el panel de datos como aquellos que siguen una muestra dada de elementos a lo largo del tiempo y, por lo tanto, proporcionan múltiples observaciones sobre cada individuo de la muestra.

Según Baltagi (2021) existen dos tipos de panel de datos:

- Panel de datos micro: Gran conjunto de unidades (N) y corto número de periodos (T).
- Panel de datos macro: Pequeño conjunto de unidades (N) y largo número de periodos (T).

En un panel balanceado, todos los miembros del panel tienen mediciones en todos los periodos. En una tabla transversa y de series de tiempo, cada uno de los registros debe tener una frecuencia. Si un panel balanceado tiene N miembros en el panel y T periodos, el número de observaciones n en el conjunto de datos es $n = N * T$. En un panel no balanceado, uno o más miembros del panel tiene un número diferente de observaciones. Si un panel no balanceado tiene N miembros y T periodos, entonces el número de observaciones es $n < N * T$.

5.3.3.1 Características de un Panel de Datos

El panel de datos tiene ciertas ventajas sobre los datos transversales y las series de tiempo (Hsiao, 2005); tiene inferencias más precisas de los parámetros del modelo, el panel de datos usualmente tiene más grados de libertad y menos

multicolinealidad que los datos transversales. También tiene una capacidad de capturar la complejidad del comportamiento humano. Por lo tanto, es posible construir y probar hipótesis de comportamiento más complejas. Además, considera el impacto de las variables omitidas.

El panel de datos permite descubrir las relaciones dinámicas, ya que el comportamiento económico es inherentemente dinámico, por lo que la mayoría de las relaciones económicamente interesantes son explícita o implícitamente dinámicas (Nerlove, 1996). Este genera predicciones más precisas para los resultados individuales al agrupar los datos en lugar de generar predicciones de resultados individuales usando los datos del individuo en cuestión. Si los comportamientos individuales son similares condicionados por ciertas variables, el panel de datos brinda la posibilidad de aprender sobre el comportamiento de un individuo al observar el comportamiento de otro.

5.4 REGRESIÓN

5.4.1 Regresión Simple

Los análisis de regresión son una de las herramientas más utilizadas en la econometría. Maddala (1992) define el análisis de regresión como la evaluación y descripción de la relación entre una variable, llamada variable dependiente, y una o más variables, llamadas variables independientes o explicativas. La regresión lineal simple está dada por la ecuación (1). También llamada regresión lineal bivariada, ya que implica dos variables.

$$y = \beta_0 + \beta_1 x + u \quad (1)$$

La variable y es llamada variable dependiente. La variable x es la variable independiente. El término u es llamado el término error o ruido en la relación, esta representa todos los factores que afectan a y que no son x , aquellos que no han sido observados. En caso de que $\Delta u = 0$ entonces x tiene un efecto lineal en y . El término β_1 es el parámetro de la pendiente en la relación entre x y y . Por último, el término β_0 es el intercepto, algunas veces llamado el término constante.

Siempre que se incluya β_0 en la ecuación, no hay ninguna pérdida asumiendo que el promedio del valor de u en la población es 0.

$$E(u) = 0 \quad (2)$$

El supuesto, en la ecuación (2), no nos dice nada acerca de la relación entre x y u , solamente hace una aclaración sobre la distribución de lo que no observamos en la población. El supuesto crucial es que el valor promedio de u no depende del valor de x . La ecuación (3) sostiene que u tiene media independiente de x .

$$E(u|x) = E(u) \quad (3)$$

Cuando combinamos la media independiente (Ecuación 3) con el valor esperado del término error (Ecuación 2), obtenemos la ecuación (4) que representa el supuesto de media condicional.

$$E(u|x) = 0 \quad (4)$$

5.4.2 Ordinary Least Squares (OLS)

Una vez establecido la ecuación del modelo de la regresión simple (Ecuación 1), se deben estimar los parámetros β_0 y β_1 . Teniendo una muestra aleatoria denotada por $\{(x_i, y_i): i = 1, \dots, n\}$ de tamaño n , se obtiene la ecuación (5). Donde, el término u_i es el error de la observación i ya que contiene todos los factores que influyen en la y_i que no están en x_i .

$$y_i = \beta_0 + \beta_1 x_i + u_i \quad (5)$$

La ecuación (6) se obtiene partiendo del supuesto que, en la población, la variable u no está correlacionada con x .

$$Cov(x, u) = E(xu) = 0 \quad (6)$$

La ecuación (7) se obtuvo a partir de la ecuación (6), se realizan operaciones algebraicas y utilizando el método de momentos para estimar los parámetros. Donde \bar{x}, \bar{y} son el promedio aritmético de x_i, y_i respectivamente.

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \quad (7)$$

La ecuación (8) representa la covarianza muestral entre x y y dividida entre la varianza de x . Si la correlación entre x y y es positiva, entonces $\hat{\beta}_1$ es positiva y viceversa.

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (8)$$

Una vez estimada la pendiente β_1 podemos llevar a cabo la ecuación (9), estimando el intercepto. La ecuación (9), $\hat{\beta}_0$ o intercepto estimado, está escrita en términos de $\bar{x}, \bar{y}, \hat{\beta}_1$ (Ecuación 7).

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (9)$$

El método utilizado para estimar $\hat{\beta}_0$ y $\hat{\beta}_1$ es usualmente el de mínimos cuadrados ordinarios o *OLS*. En la ecuación (10) para cada $\hat{\beta}_0$ y $\hat{\beta}_1$ se define un valor ajustado por y cuando $x = x_i$.

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \quad (10)$$

5.4.3 Efectos Ajustados

La ecuación (11) representa la ecuación de efectos ajustados donde en y_{it} , i denota el elemento y t denota el periodo de tiempo. Por otro lado, a_i representa todo efecto no observado, lo cual es constante al tiempo, que afecta a y_{it} . Por último, u_{it} es llamado error variable al tiempo, debido a que representa los factores no observados que cambian alrededor del tiempo y afectan a y_{it} .

$$y_{it} = \beta_1 x_{it} + a_i + u_{it}, \quad t = 1, 2, \dots, T \quad (11)$$

La ecuación (12) representa cada i promediada respecto al tiempo. Donde $\bar{y}_i = T^{-1} \sum_{t=1}^T y_{it}$. Debido a que a_i esta ajustada sobre el tiempo aparece en ambas ecuaciones.

$$\bar{y}_i = \beta_1 \bar{x}_i + a_i + \bar{u}_i \quad (12)$$

La ecuación (13) es obtenida de la resta de la ecuación de efectos ajustados (Ecuación 11) y el promedio respecto al tiempo para cada t (Ecuación 12).

$$y_{it} - \bar{y}_i = \beta_1 (x_{it} - \bar{x}_i) + u_{it} - \bar{u}_i, \quad t = 1, 2, \dots, T \quad (13)$$

O

$$\ddot{y}_{it} = \beta_1 \ddot{x}_{it} + \ddot{u}_{it}, \quad t = 1, 2, \dots, T \quad (14)$$

Donde $\ddot{y}_{it} = y_{it} - \bar{y}_i$ (el valor y_t menos la media al tiempo i) son los datos desestacionalizados (*time-demeaned data*) en y , de la misma forma pasa con \ddot{x} y \ddot{u} . Lo relevante de esta ecuación es que han desaparecido el efecto no observado, a_i . De esta forma podemos estimar por mínimos cuadrados ordinarios agrupados. Un estimador de mínimos cuadrados ordinarios agrupados que está basado en variables desestacionalizadas, es conocido como estimador de efectos ajustados.

El nombre viene del hecho que los mínimos cuadrados ordinarios en la ecuación (14), utiliza la variación a través del tiempo en x y y dentro de cada observación transversal.

Al tener múltiples variables explicativas, consideremos la ecuación (15).

$$y_{it} = \beta_1 x_{it1} + \beta_2 x_{it2} + \dots + \beta_k x_{itk} + a_i + u_{it}, \quad t = 1, 2, \dots, T \quad (15)$$

Para cada variable explicativa desestacionalizada estimaremos, igual que en el caso univariado, por mínimos cuadrados ordinarios agrupados, los parámetros β_i . Obtenemos la siguiente ecuación:

$$\ddot{y}_{it} = \beta_1 \ddot{x}_{it1} + \beta_2 \ddot{x}_{it2} + \cdots + \beta_k \ddot{x}_{itk} + \ddot{u}_{it}, \quad t = 1, 2, \dots, T \quad (16)$$

5.4.4 Efectos Aleatorios

Considerando la ecuación (17).

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \cdots + \beta_k x_{itk} + a_i + u_{it} \quad (17)$$

Al incluir el intercepto consideramos el supuesto que los efectos no observados (a_i), tienen media cero. El objetivo en el estimado de los efectos ajustados es eliminar los efectos no observados, debido a que se piensa que estos efectos están correlacionados con una o más x_{itj} .

La ecuación (18) representa el principal supuesto de los efectos aleatorios, en donde se supone que los efectos no observados (a_i), no están correlacionados a ninguna variable explicativa en todos los periodos de tiempo. Entonces la ecuación (17) se vuelve un modelo de efectos aleatorios si asumimos que los efectos no observados (a_i) no están correlacionados a las variables explicativas.

$$Cov(X_{itj}, a_i) = 0, \quad t = 1, 2, \dots, T; j = 1, 2, \dots, k \quad (18)$$

5.4.5 Modelo Lineal Mixto

La regresión por efectos mixtos es una extensión del modelo lineal general que toma en cuenta la jerarquía de la estructura de los datos. Los efectos mixtos también son conocidos como modelos jerárquicos, modelos multinivel o modelos lineales mixtos. Este modelo se puede utilizar con distintos tipos de datos como los datos longitudinales. Los efectos mixtos son llamados de esta manera porque combinan los efectos ajustados y los efectos aleatorios (Brown, 2020). Nos

referiremos a elemento como el grupo de interés o las observaciones divididas en grupos independientes, perteneciente a los datos longitudinales.

Brown (2020) resalta que el análisis ANOVA y la regresión múltiple enfrentan inconvenientes, los cuales son resueltos al utilizar modelos lineales mixtos. Por ejemplo:

- Permiten examinar la condición de interés, considerando la variabilidad dentro y entre los elementos simultáneamente.
- A diferencia del ANOVA, manejan datos faltantes y no balanceados, con un efecto negativo menor.
- El análisis de ANOVA considera todas las respuestas dentro de los elementos como una sola observación. Por otro lado, los elementos con un número mayor de datos faltantes, en los modelos de efectos mixtos, tendrán una significancia menor en la estimación de parámetros.
- Utilizan predictores continuos. También, proporcionan coeficientes estimados para medir la magnitud y dirección de los efectos de interés.
- Son aplicables en situaciones donde el análisis ANOVA y la regresión múltiple no son adecuados.

$$y_i = X_i\beta + Z_i\alpha_i + \epsilon_i \quad i = 1, \dots, m \quad (19)$$

La ecuación (19) pertenece al modelo de efectos mixtos con datos longitudinales. Donde la variable y representa un vector de observaciones del i –ésimo individuo, β es el vector de coeficientes de regresión desconocidos que son llamados efectos ajustados. X_i y Z_i son, cada uno, una matriz conocida. El término α_i es el vector de efectos aleatorios, por último ϵ_i es el vector de error. Se asume que α_i y ϵ_i son independientes. Los efectos mixtos en los datos longitudinales contienen observaciones divididas en grupos independientes, con un efecto aleatorio correspondiente a cada grupo. Los modelos longitudinales de

efectos mixtos tienen covariables dependientes del tiempo que aparecen en X , Z o en ambos. La información que proporcionan los efectos aleatorios indica la variabilidad alrededor del intercepto. La variabilidad se basa en los diferentes niveles en la variable clúster.

$$y_{ij} = \xi_i + \eta_i x_{ij} + \zeta_{ij} + \epsilon_{ij} \quad (20)$$

En la ecuación (20), el término ξ_i representa el intercepto aleatorio, η_i representa la pendiente aleatoria, x_{ij} representa la covariable, ζ_{ij} representa la correlación serial, por último, ϵ_{ij} representa el error.

Los efectos ajustados, en los efectos mixtos, representan el nivel de la población o los efectos promedio que se utilizarán durante el experimento. Los efectos ajustados, pueden ser efectos de condición debido a que, se espera que operen de manera predecible en varias muestras de elementos.

Los efectos aleatorios varían dependiendo del grupo o elemento, estos son grupos de puntos de datos dependientes, en los cuales las observaciones componentes provienen del mismo nivel. Los efectos aleatorios son incluidos en los modelos lineales mixtos, para tomar en cuenta el comportamiento de ciertos elementos que puedan diferenciarse de la tendencia promedio. Los efectos aleatorios no pueden ser una variable aleatoria continua, de ser así debemos tomar esa variable como un efecto ajustado.

El problema de la no dependencia se resuelve al incluir efectos aleatorios. El efecto usualmente impacta a la regresión múltiple, al tener en cuenta el hecho de que los elementos se comportan de forma distinta.

El intercepto aleatorio es la desviación aleatoria. El intercepto aleatorio de cada elemento permite al modelo estimar la desviación de cada participante, en relación con los efectos ajustados. En la regresión múltiple utiliza, en todos los elementos, la misma línea de regresión con el mismo intercepto y pendiente. Las predicciones que utilizan la misma línea de regresión provocan menor precisión y grandes residuales. En los modelos lineales mixtos el intercepto ajustado

representa el intercepto promedio y el intercepto aleatorio permite a cada elemento desviarse del promedio.

La pendiente aleatoria modela la desviación individual. Las pendientes aleatorias permiten al modelo, estimar la desviación de cada participante de la tendencia general.

5.5 ANTECEDENTES

Jiménez (2020) evaluó los cambios de crecimiento facial en la población mestiza de Colombia utilizando un modelo lineal mixto. El estudio consistió en registrar por 18 años a 49 mestizos con características faciales normales. Los sujetos no debían tener historial de tratamiento de ortodoncia. El modelo lineal mixto se utilizó para comparar las medidas de crecimiento craneofaciales a través del tiempo. La variable objetivo fue la altura posterior facial. Los efectos ajustados fueron la edad y el sexo de la persona. Se agregó un intercepto para los sujetos, el cual modifica de forma aleatoria el intercepto poblacional. Los resultados indicaron que los cambios en las medidas cefalométricas, que fueron detectadas durante 18 años de seguimiento, eran significativamente afectados por la edad y el sexo.

Anshuka (2019) utilizó un modelo lineal mixto para identificar el modelo más efectivo para pronosticar el *SPI*. El análisis con el modelo lineal mixto fue elaborado con los paquetes en R: NLME, lme4, lmerTest y lsmeans. El modelo consta de la escala de tiempo del *SPI* y el plazo de meses como las variables independientes. El efecto ajustado fue el tipo de modelo que fue utilizado, las categorías de modelos son: WANN (*Wavelet Artificial Neural Network*), SVM (*Support Vector Machine*), MM (*Marcov Model*), ANN (*Artificial Neural Network*) y ARIMA/SARIMA. Los efectos aleatorios son representados por el identificador del artículo y los autores que lo componen en la ubicación geográfica. Los efectos aleatorios son para tener en cuenta la variación regional y de investigación en los resultados del estudio.

Christensen (2020) evaluó el rol que tiene la severidad de sequía como determinante de la enfermedad hemorrágica en el este de Estados Unidos. En el estudio se utilizó un modelo lineal mixto para explicar la presencia de la enfermedad hemorrágica, también se evaluó los predictores espaciotemporales en toda la región. La investigación se centra en condados dentro de los 23 estados al este. Los estados habían presentado reportes de enfermedad hemorrágica, por lo menos una vez dentro del periodo del 2000 al 2014. El modelo lineal mixto fue implementado utilizando el paquete “lme4” de R. El modelo tiene como efectos aleatorios el estado y el año. Los efectos ajustados se implementaron como: la severidad de sequía, latitud, longitud, región fisiográfica y el porcentaje de cobertura de humedales como efectos aleatorios.

Kuwayama (2018) estimó el impacto de la sequía en rendimientos de cultivos, en granjas de Estados Unidos, durante el periodo del 2001 al 2013. El análisis se realizó utilizando el *U.S. Drought Monitor*. Se utilizó un panel de datos, con efectos ajustados, para explotar la variabilidad espacial y temporal de las condiciones de sequías. Los efectos fijos fueron representados como los condados. Los coeficientes fueron estimados utilizando la relación entre las sequías y los resultados agrícolas.

Bittencourt (2020) midió como los árboles del Amazonas tienen una capacidad limitada, para aclimatar las propiedades hidráulicas de las plantas en respuesta a sequías prolongadas. La medición fue realizada utilizando un modelo lineal mixto con parcela, diámetro y su interacción como efectos fijos. Los efectos aleatorios fueron los genes, representados como intercepto en el modelo. La pendiente de la variable independiente fue representada con los términos de los efectos ajustados. El modelo fue desarrollado en R, utilizando el paquete “lme”.

Lu (2020) realizó un mapeo de la sensibilidad de la agricultura hacia la sequía. Además, estimó el efecto del riego en Estados Unidos en un periodo de 1950 a 2016. El modelo de efectos ajustados se utilizó para medir el efecto del riego, sobre la sensibilidad de la agricultura a la sequía para cada cultivo. Los efectos ajustados fueron utilizados, para incluir un efecto fijo específico del

condado. Esto para tener en cuenta las diferencias entre condados en las variables omitidas. La tendencia lineal específica del condado también fue incluida, para dar cuenta de las diferencias entre condados en los cambios científicos y tecnológicos.

CAPÍTULO 6

METODOLOGÍA

El estudio realizado en esta investigación es de tipo cuantitativo. Fernández (2002) define la investigación cuantitativa como aquella en la que se recogen y analizan datos cuantitativos sobre variables. La investigación cuantitativa busca determinar la fuerza de asociación de los resultados de una muestra para realizar inferencias de la población. Jiménez (2020) estudió el impacto de la investigación cuantitativa en la actualidad, concluyendo que *“La investigación cuantitativa mantiene su impacto y relevancia en los tipos de estudios que requieren de un orden secuencial, riguroso y demostrativo, con una amplia gama de criterios que permita su valoración y aportes a la comunidad científica.”*

En este capítulo se explica el proceso llevado a cabo para la identificación de factores determinantes de sequía en condados de Estados Unidos. El estudio abarca dos modelos: un modelo de efectos ajustados y un modelo lineal mixto. El capítulo contiene desde la explicación de los datos, hasta la construcción del modelo.

La población objetivo de este estudio son todos los condados de Estados Unidos y se consideran sus diferentes métricas meteorológicas.

6.1 DATOS

6.1.1 U.S. Drought Monitor (USDM)

Los datos utilizados en este estudio son del *U.S. Drought Monitor*. Este monitor existe desde 1999, es producto de *National Drought Mitigation Center*

(NDMC) de la Universidad de Nebraska-Lincoln, del *National Oceanic and Atmospheric Administration (NOAA)* y el *U.S. Department of Agriculture (USDA)*. Cada que hagamos referencia a los datos en esta investigación nos referiremos a los registros meteorológicos diarios de cada condado.

Según el sitio oficial del *U.S. Drought Monitor* “La Agencia del Servicio Agrícola (FSA) utiliza el monitor para determinar la elegibilidad para el programa *Livestock Forage Program (LFP)*. El *Internal Revenue Service (Servicio de Impuestos Internos)* lo utiliza para el aplazamiento de impuestos para las ventas de ganado forzadas debido a la sequía. Los encargados del estado y los responsables locales, tribales y a nivel de cuencas utilizan el monitor para tomar decisiones sobre las acciones en respuesta a la sequía, idealmente junto a otros indicadores locales de sequía.”

Este monitor de sequía es un mapa publicado cada miércoles, en donde muestra la situación de Estados Unidos con relación a la sequía. En la figura 6.1, tenemos la severidad actual (2023) de sequía de Estado Unidos. Una gran parte del territorio presenta algún grado de sequía. El monitor utiliza cinco categorías de sequía, sin embargo, la severidad está dada por una métrica continua.

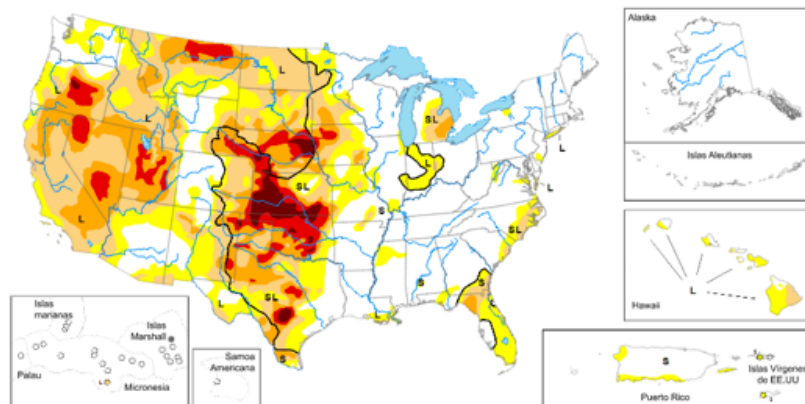


Figura 6.1: Severidad de Sequía USDN Fuente: *U.S. Drought Monitor*

6.1.2 Conjunto de Datos y Variables

El conjunto de datos utilizado cuenta con 23,841,468 registros meteorológicos que abarcan desde el año 2000 hasta el año 2020. Los datos se obtuvieron en febrero del 2023. El historial pertenece a los 3,108 diferentes condados de Estados Unidos. La razón de utilizar solamente datos meteorológicos es debido a que cualquier tipo de sequía tiene que pasar por una sequía meteorológica. Por otro lado, la evaluación de sequía se lleva de forma semanal, publicada cada miércoles. El monitor no realiza un pronóstico, más bien es una evaluación semanal manual de las condiciones de sequía. Se cuenta con 20 variables de las cuales la primera variable (“fips”) representa el condado al que pertenece dicho registro, o nuestro elemento. Por otro lado, la última (“score”) representa el grado de sequía, nuestra variable objetivo. Las 18 variables restantes representan datos meteorológicos recolectados por el *U.S. Drought Monitor*.

6.1.2.1 Códigos de Condados

Los códigos de sistema federal de procesamiento de información, o mejor conocidos como FIPS por sus siglas en inglés (*Federal Information Processing System*), son identificadores geográficos únicos. El número de dígitos en estos códigos varía según el nivel geográfico. Los códigos a nivel estatal tienen dos dígitos y a nivel condado tienen 5 dígitos. Los primeros dos dígitos, en los códigos FIPS, son del estado al que pertenece el respectivo condado.

6.1.2.2 Precipitación Meteorológica

La precipitación meteorológica se define como “*Agua procedente de la atmósfera, y que en forma sólida o líquida se deposita sobre la superficie de la tierra.*” (RAE, 2023). La precipitación es un hidrometeoro que cae desde una nube o un grupo de nubes y toca el suelo, puede estar formada por partículas líquidas o sólidas.

Según el Instituto de Hidrología, Meteorología y Estudios Ambientales (2019) las formas de precipitación son: lluvia, llovizna, nieve, cinarra,

nieve granulada, polvo diamante, granizo y gránulos de hielo. En la figura 6.2 tenemos ilustrado la precipitación dentro del ciclo del agua. Las sequías comienzan con la falta de precipitación.

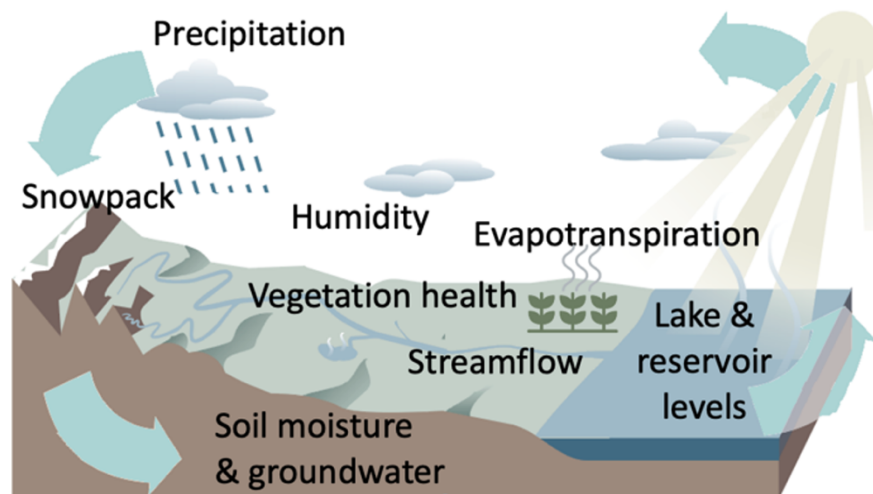


Figura 6.2: Ciclo del Agua

Fuente: *U.S. Drought Monitor*

6.1.2.3 Presión Superficial

La presión Superficial se define como la presión atmosférica dada en una ubicación particular de la superficie de la tierra (*American Meteorological Society*, 2022). Según National Geographic (2022) cuando un sistema de baja presión se traslada a un área, generalmente provoca nubosidad, viento y precipitaciones. Los sistemas de alta presión generalmente conducen a un clima tranquilo y despejado.

6.1.2.4 Temperatura

Según Rodríguez (2004) la temperatura es una magnitud relacionada con la rapidez del movimiento de las partículas que constituyen la materia. Cuanta mayor agitación haya en las partículas, mayor será la temperatura. Esta es un índice del calentamiento o enfriamiento del aire, que resulta del intercambio de calor entre la atmósfera y la tierra. En resumen, la temperatura indica en valores

numéricos, el nivel de energía interna que se encuentra en ese lugar en ese momento.

6.1.2.5 Punto de rocío

Según el *National Weather Service* (2023), el punto de rocío es la temperatura que necesita el aire para alcanzar la humedad relativa del 100%. El aire en ese momento no se puede mantener en estado gaseoso. Al incrementar el punto de rocío también se ve incrementada la humedad en el aire.

6.1.2.6 Temperatura de Bulbo Húmedo

La temperatura del bulbo húmedo es la temperatura más baja que se puede alcanzar solo evaporando el agua, esta es un indicador de la cantidad de humedad en el aire. Cuanto menor sea la humedad relativa del aire, mayor será el enfriamiento. La temperatura del bulbo húmedo se mide con un termómetro especializado. La evaporación y la reducción de la temperatura del bulbo húmedo, es mayor cuanto más seco está el aire atmosférico y es nula cuando la atmósfera está saturada de vapor de agua.

6.1.2.7 Temperatura de la Superficie de la Tierra

La temperatura de superficie de la tierra es una medida del calentamiento directo de la superficie terrestre, donde los rayos del sol son absorbidos y reemitidos. Según *EARTHDATA NASA* (2023) la temperatura de superficie de la tierra se refiere a la temperatura de radiación efectiva del suelo, más la superficie del dosel. La temperatura de la superficie de la tierra muestra una variación diurna, mayor a la temperatura del aire en la superficie.

6.1.2.8 Velocidad del Viento

El viento consiste en el movimiento del aire de una zona a otra. Las causas por las que se produce el viento son diversas, pero normalmente se origina cuando hay diferencias de presión o de temperatura. En la atmósfera existe una relación directa de la presión y el viento. El viento usualmente es medido con el anemómetro de cazoleta, el cual sus giros nos indican la velocidad del viento.

Existen vientos especiales, como son: las brisas de mar, brisa de montaña y valles, vientos generados por huracanes y tornados, huracanes y tornados.

6.1.2.9 Índice de Severidad y Cobertura de Sequías

La variable *score* representa el Índice de Severidad y Cobertura de Sequía, o *DSCI* por sus siglas en inglés. Este índice representa la evaluación de sequía semanal de cada condado. La métrica es evaluada de forma manual, no hay un modelo que la calcule. En el *U.S. Drought Monitor* trabajan expertos que sintetizan los datos de distintas fuentes, se colabora con gente local para proveer información más específica. Los indicadores obtenidos para calcular este índice son: *Palmer Drought Severity Index (PDSI)*, *Climate Prediction Center Soil Moisture Model (CPC-SMM)*, *United States Geological Services Weekly Streamflow (USGS-WS)*, *Standardized Precipitation Index (SPI)* y el *Objective Drought Indicator Blends (ODIB)*.

En la tabla 6.1 tenemos las cinco categorías de sequía que puede tener cada condado. La categoría de sequía, ubicada en la primera columna, va de D0 a D4. El número especifica la severidad de sequía siendo 4 la categoría más grave. La segunda columna contiene la descripción de la categoría, desde “Inusualmente Seco” hasta “Sequía Excepcional”. La columna de “Impacto” contiene algunas de las repercusiones causadas por categoría. Por último, en la parte de “Rango”, tenemos los valores de los indicadores para que se genere dicha categoría de sequía. El *PDSI* y el *SPI* van de -5 a -1 o menos, donde entre menor sea el número es mayor el grado de sequía. Por otro lado, el *CPC-SMM*, *USGS-WS* y el *ODIB* representan los percentiles, donde entre menor sea el número es mayor el grado de sequía. Debido a que los rangos de los diversos indicadores a menudo no coinciden, la categoría final de sequía tiende a basarse en lo que muestran la mayoría de los indicadores. Además, se apoyan de las observaciones de los locales.

Categorías	Descripción	Impacto	Rangos				
			PDSI	CPC-SMM	USGS-WS	SPI	ODIB
D0	Inusualmente Seco	<ul style="list-style-type: none"> Sequedad a corto plazo. Pequeña deficiencia de Agua. 	-1.0 a -1.9	21 a 30	21 a 30	-0.5 a -0.7	21 a 30
D1	Sequía Moderada	<ul style="list-style-type: none"> Algunos Daños a los cultivos Solicitud de restricción de agua. 	-2.0 a -2.9	11 a 20	11 a 20	-0.8 a -1.2	11 a 20
D2	Sequía Severa	<ul style="list-style-type: none"> Perdida parcial de sembradíos Cortes de Agua. Restricciones de uso de agua impuestas. 	-3.0 a 3.9	6 a 10	6 a 10	-1.3 a -1.5	6 a 10
D3	Sequía Extrema	<ul style="list-style-type: none"> Perdida mayoritaria de sembradíos. Escasez de agua generalizada. 	-4.0 a -4.9	3 a 5	3 a 5	-1.6 a -1.9	3 a 5
D4	Sequía Excepcional	<ul style="list-style-type: none"> Perdida total de sembradíos. Escasez de los almacenes de agua. Emergencia severa de agua. 	-5 o menos	0 a 2	0 a 2	-2.0 o menos	0 a 2

Tabla 6.1: Rangos de Índices con su Categoría Fuente: *U.S. Drought Monitor*

El Índice de Severidad y Cobertura de Sequías (*DSCI*) es un método experimental, el cual convierte los niveles de sequía del mapa del *U.S. Drought Monitor* en un único valor para un área específica. Los valores del *DSCI* se encuentran en las tablas de datos del *U.S. Drought Monitor*, estos varían entre 0 y 5. Un valor de cero indica que no hay áreas anormalmente secas o en sequía, mientras que un valor de 5 significa que toda el área está experimentando una sequía excepcional (D4). En la figura 6.3 tenemos el porcentaje del área en sequía del estado de Dakota del norte donde la categoría de Inusualmente Seco predomina en esta región. Este índice se obtiene con la suma de la multiplicación del porcentaje de cobertura por la categoría de sequía, finalmente dividiéndolo entre 100. La categoría D0 tiene el valor de 1 y D4 el valor de 5. La ecuación (21) representa el cálculo del *DSCI*, unificando estas cuatro categorías en un solo índice.

Week	None	D0	D1	D2	D3	D4
2017-08-29	0.41	33.75	14.50	29.26	21.69	0.39

Figura 6.3: Porcentaje de Área en Sequía Fuente: *U.S. Drought Monitor*

$$DSCI = \frac{1(33.75)+2(14.50)+3(29.26)+4(21.69)+5(0.39)}{100} = 2.39 \quad (21)$$

La Tabla 6.2 representa el resumen de cada una de las variables, con sus respectivas unidades de medida. La primera columna contiene la enumeración de las 20 variables utilizadas. Después, tenemos los nombres originales de las variables en relación con el conjunto de datos. Los nombres representan, en su mayoría, las siglas en inglés de la variable. Por ejemplo, la variable *T2M* representa la temperatura en 2 metros. Por último, la tercera columna muestra las unidades de medida de cada variable. La variable *fip* y la de *score*, no tienen unidades de medida. Por otro lado, las variables relacionadas con temperatura utilizan los grados Celsius. Las variables relacionadas a la velocidad del viento, como *WS10M*, utilizan metros por segundo. Otras variables como la precipitación, la presión superficial y la humedad específica en 2 metros, utilizan milímetros, kilopascales y gramos por kilo respectivamente. Se añade un filtro por multicolinealidad para remover las variables fuertemente correlacionadas.

No.	Variable	Significado	Unidad
1	fips	Códigos de sistema federal de procesamiento de información.	-
2	PRECTOT	Precipitación.	mm
3	PS	Presión superficial.	kPa
4	QV2M	Humedad específica en 2 metros.	g/Kg
5	T2M	Temperatura en 2 metros.	C
6	T2MDEW	Punto de rocío en 2 metros.	C
7	T2MWET	Temperatura de bulbo húmedo en 2 metros.	C
8	T2M_MAX	Temperatura máxima en 2 metros.	C
9	T2M_MIN	Temperatura mínima en 2 metros.	C
10	T2M_RANGE	Rango de temperatura en 2 metros.	C
11	TS	Temperatura de la superficie de la tierra.	C
12	WS10M	Velocidad del viento en 10 metros.	m/s
13	WS10M_MAX	Velocidad máxima del viento en 10 metros.	m/s
14	WS10M_MIN	Velocidad mínima del viento en 10 metros.	m/s
15	WS10_RANGE	Rango de velocidad del viento en 10 metros.	m/s
16	WS50M	Velocidad del viento en 50 metros.	m/s
17	WS50M_MAX	Velocidad máxima del viento en 50 metros.	m/s
18	WS50M_MIN	Velocidad mínima del viento en 50 metros.	m/s
19	WS50_RANGE	Rango de velocidad del viento en 50 metros.	m/s
20	score	Grado de sequía.	-

Tabla 6.2: Resumen de Variables

Fuente: Elaboración Propia

6.2 Python

Según Lutz (2013) Python es un lenguaje de programación de uso general. La popularidad que tiene Python es debido a que es un lenguaje de programación se enfoca en la lectura, coherencia y la calidad del software. Los enfoques que maneja Python son con intención de aumentar la productividad del programador. Este lenguaje de programación cuenta con una gran colección de librerías, conocidas como librerías estándar. Las librerías de terceros son desarrolladas para aplicaciones específicas de Python.

Raschka (2020) nos dice que Python es un lenguaje de programación interprete de alto nivel, que es reconocido por ser fácil de aprender. Es capaz de aprovechar el poder de los sistemas a nivel lenguajes de programación. Python tiene un gran crecimiento de popularidad dentro de la comunidad de computación científica. Las librerías más recientes de *Machine Learning* y *Deep Learning* son basadas en librerías de Python.

6.2.1 Librería *statsmodels*

Según Fulton (2022) la librería *statsmodel* es una librería de Python utilizada para análisis econométricos y estadísticos, es una librería bien establecida. Además, abarca un gran número de clases de modelos como la regresión lineal, ANOVA, modelo lineal generalizado, modelo de efectos mixtos, entre otros.

6.2.2 Librería *linearmodels*

La librería *linearmodels* es una extensión de la librería *statsmodels*, donde se incluyen nuevos aplicativos. La librería incluye modelos como modelo de panel de datos, regresión de alta dimensión, estimadores de sistemas de regresión, entre otros.

6.3 LIMPIEZA DE DATOS

Rahm (2000) dice que la limpieza de datos busca la mejora de la calidad de los datos identificando errores e inconsistencias. La información incompleta

genera incertidumbre durante el análisis de datos. La corrección de inconsistencia y errores permite el aumento de la calidad de los datos. Ridzuan (2019) define limpieza de datos como una operación que se realiza sobre los datos existentes para eliminar anomalías. Esta operación consiste en eliminar errores, inconsistencias y transformar los datos es un formato uniforme.

6.3.1 Resumen Semanal

Los datos cuentan con registros semanales del grado de sequía de cada condado. Los registros meteorológicos son diarios, sin embargo, nuestra variable objetivo es semanal.

En la figura 6.4 tenemos los primeros registros del condado 1001, que representa al condado de Autauga, en el estado de Alabama. Por temas de espacio solo se agregó la precipitación y la presión superficial a la figura 6.4. La variable *score* tiene valores nulos en los registros diarios, pero el grado de sequía se tiene de forma semanal. Por lo tanto, la variable objetivo está en cada condado cada siete días a partir del 4 de enero del año 2000, hasta finales del año 2020.

out[9]:

	fips	date	PRECTOT	PS	score
0	1001	2000-01-01	0.22	100.51	NaN
1	1001	2000-01-02	0.20	100.55	NaN
2	1001	2000-01-03	3.65	100.15	NaN
3	1001	2000-01-04	15.95	100.29	1.0
4	1001	2000-01-05	0.00	101.15	NaN
5	1001	2000-01-06	0.01	101.31	NaN
6	1001	2000-01-07	0.01	101.37	NaN
7	1001	2000-01-08	1.02	100.77	NaN
8	1001	2000-01-09	16.19	99.88	NaN
9	1001	2000-01-10	21.23	99.78	NaN
10	1001	2000-01-11	1.33	100.40	2.0

Figura 6.4: Muestra de datos (5 variables de 20) Fuente: Elaboración propia

Los datos diarios fueron utilizados para obtener un promedio semanal de cada una de las variables, de esta forma no se pierde información. En la figura

6.5 tenemos el código de Python, utilizado para agrupar en el promedio semanal de los distintos registros meteorológicos. La función “*groupby*” en la figura 6.5 nos permite hacer el resumen semanal en un tiempo aceptable (optimizado para Python), ya que si se realiza este proceso por iteraciones tomaría mucho tiempo, dado la cantidad de datos.

```
In [6]: def concat_summary_fips(sample):
        global summary
        summary = pd.concat([summary, sample.groupby([pd.Grouper(key='date', freq='W-TUE')]).mean()])
```

Figura 6.5: Resumen Semanal en Python Fuente: Elaboración propia

6.3.2 Multicolinealidad

Según Shrestha (2020) el término multicolinealidad, en el análisis de regresión lineal múltiple, indica la relación lineal entre las variables independientes. La colinealidad indica dos variables que son combinaciones lineales perfectas o cercanas entre sí. La multicolinealidad ocurre cuando el modelo de regresión incluye variables que están significativamente correlacionadas, no solamente con la variable dependiente sino entre ellas. La multicolinealidad aumenta el error estándar en los coeficientes del modelo, lo cual afecta el análisis. También, vuelve variables significantes a estadísticamente insignificantes. La multicolinealidad aumenta la varianza de los coeficientes de regresión, volviéndolos inestables, trayendo un problema al momento de interpretarlos.

Existen distintos métodos para detectar multicolinealidad, en este estudio utilizaremos el coeficiente de correlación de Pearson y el factor de inflación de varianza (*VIF*).

6.3.2.1 Coeficiente de Correlación de Pearson

El coeficiente de correlación de Pearson está dado por la siguiente ecuación:

$$r = \frac{n(\sum XY) - (\sum X)(\sum Y)}{\sqrt{[n \sum X^2 - (\sum X)^2][n \sum Y^2 - (\sum Y)^2]}}$$

Donde el término r es el coeficiente de correlación, n es el número de observaciones, X es la primera variable utilizada y Y la segunda. Según Barton (2014), hablando de la multicolinealidad, mencionan que “*Idealmente, la correlación entre las covariables debería ser baja, con un valor de r menor a 0.7*”. Por lo tanto, el coeficiente de correlación indica posible colinealidad si el valor absoluto del coeficiente de correlación de Pearson es mayor que 0.7.

La figura 6.6 es un ejemplo de la matriz de coeficientes de correlación de Pearson de nuestras variables. La matriz del coeficiente de correlación de Pearson es una herramienta estadística. Esta nos muestra la fuerza y dirección de las relaciones lineales entre variables, revelando patrones y tendencias en los datos. Los valores varían entre -1 y 1, donde -1 es correlación negativa perfecta, 1 es correlación positiva perfecta y 0 indica ausencia de correlación. El coeficiente debe mantenerse por debajo del valor absoluto de 0.7 para prevenir cualquier potencial problema de colinealidad. Por lo tanto, una de las variables fuertemente correlacionadas debe ser eliminada.

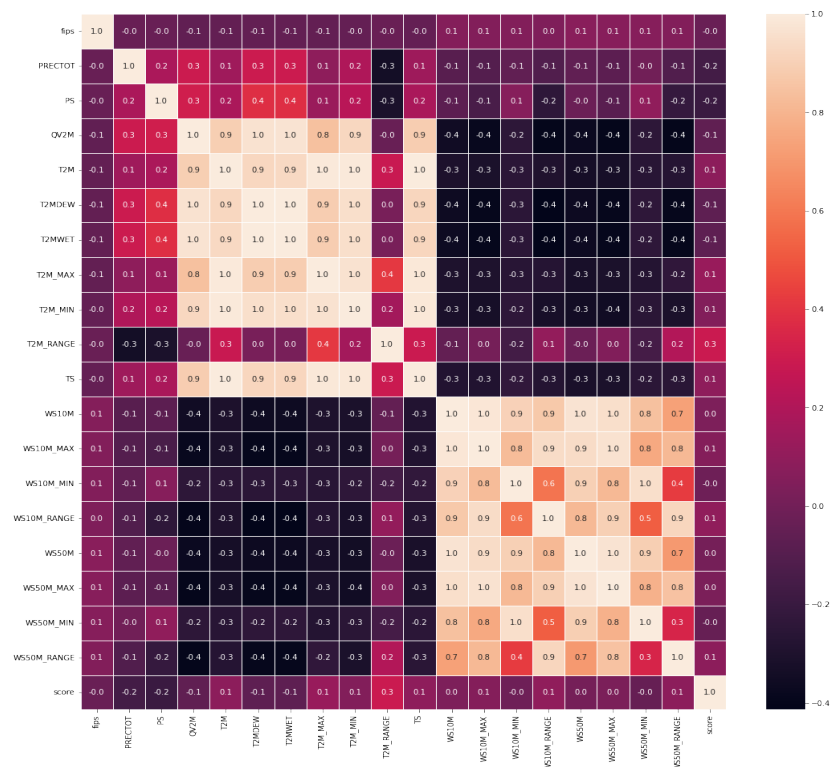


Figura 6.6: Matriz de Correlación de Pearson Fuente: Elaboración Propia

6.3.2.2 Variance Inflation Factor (VIF)

El factor de inflación de varianza o *VIF*, tiene el propósito de cuantificar el incremento estimado, en la variabilidad del coeficiente de regresión, cuando las variables independientes muestran correlación. En esencia, este indicador nos permite comprender cómo la relación entre estas variables puede influir en la precisión de nuestros resultados de regresión. El *VIF* está dado por la siguiente ecuación:

$$VIF = \frac{1}{1 - R^2},$$

donde, $R^2 = \frac{\sum_{t=1}^T (\hat{Y}_t - \bar{Y})^2}{\sum_{t=1}^T (Y_t - \bar{Y})^2}$ representando el ajuste en nuestra regresión.

Las variables independientes exhiben diferentes niveles de colinealidad según sus valores de *VIF*. El $VIF = 1$ sugiere que no existe colinealidad entre estas variables. El rango de $1 < VIF < 5$, representa una colinealidad moderada. El rango de $5 < VIF < 10$ representa una alta colinealidad, lo que indica la presencia de multicolinealidad entre las variables explicativas. Los $VIF > 10$ indican que los coeficientes de regresión se estiman débilmente, con la presencia de multicolinealidad (Shrestha, 2020).

En la figura 6.7 tenemos los factores de inflación de varianza de las variables finales. En la primera columna tenemos el nombre de la variable. Las variables libres de multicolinealidad son:

- El rango de temperatura en 2 metros. (*T2M_RANGE*)
- La temperatura en la superficie de la tierra. (*TS*)
- El rango de velocidad del viento en 50 metros. (*WS50M_RANGE*)
- La presión superficial. (*PS*)
- La velocidad mínima del viento en 50 metros. (*WS50M_MIN*)
- La precipitación. (*PRECTOT*)

Out[116]:

	variable	VIF
3	T2M_RANGE	1.633160
4	TS	1.440139
6	WS50M_RANGE	1.358051
2	PS	1.253535
5	WS50M_MIN	1.246116
1	PRECTOT	1.222214

Figura 6.7: VIFs Finales Fuente: Elaboración Propia

Los VIFs y la matriz de correlación de Pearson fueron utilizados para descartar variables mediante un proceso iterativo. Al descartar una variable se vuelve a generar la matriz de coeficientes de correlación de Pearson y se generan los factores de inflación de varianza, hasta eliminar la multicolinealidad.

6.4 MODELOS

Los modelos utilizados para encontrar factores meteorológicos determinantes de sequía (en condados de Estados Unidos, como se definió en le alcance de la investigación) fueron: un modelo de panel con efectos ajustados y un modelo lineal mixto. En esta sección estudiaremos como implementar ambos modelos en Python. Cada uno de los modelos utiliza las mismas variables explicativas, representadas por el promedio semanal de los registros meteorológicos, después de ser filtradas por la multicolinealidad.

6.4.1 Efectos Ajustados: OLS

El modelo de efectos ajustados fue implementado en Python, utilizando la librería *linearmodels*. La figura 6.8 representa el modelo de efectos ajustados con efectos aleatorios, y efectos en el tiempo. Este modelo es estimado por los mínimos cuadrados ordinarios (OLS). Lo que buscamos es que el modelo identifique variables significativas. También, que nos indique la dirección del coeficiente.

La librería contiene la función *PanelOLS*, la cual nos permite agregar los efectos anteriormente mencionados. Esta función recibe como parámetro la variable objetivo *score*. Además, especificamos nuestras variables explicativas, en una variable creada en Python llamada *exog*. Como atributo tiene los efectos por entidad (efectos aleatorios) y por tiempo, es primordial colocar como *True* estos atributos. En caso de no agregar efectos aleatorios y efectos de tiempo, el modelo se vuelve un *PooledOLS*. El *PooledOLS* está dirigido a los datos transversales agrupados. Estos son similares a los datos transversales, sin embargo, se tiene en cuenta el tiempo en que se generó el registro.

```
In [133]: mod = PanelOLS(summary["score"], exog, entity_effects=True, time_effects=True)
```

Figura 6.8: Efectos Ajustados en Python Fuente: Elaboración Propia

6.4.2 Modelo Lineal Mixto

La figura 6.9 representa el modelo lineal mixto en Python con la librería *statsmodels*. La librería *statsmodels* tiene integrada la librería *Patsy*. Esta librería nos permite describir un modelo estadístico, especialmente los modelos estadísticos lineales. La librería *Patsy* funciona como un intermediario, para desarrollar la ecuación, entre el investigador y el código. La función *mixedlm* recibe como primer parámetro la ecuación lineal de nuestra regresión. Después se especifica el nombre del conjunto de datos utilizado, guardado en la variable *dataset*. Por último, le especificamos el elemento, el cual es cada uno de los condados de Estados Unidos. Este atributo genera el intercepto aleatorio.

```
In [143]: # Random Intercept Model w/out Interaction Term
model = smf.mixedlm("score ~ date + T2M_RANGE + TS + PRECTOT + PS + WS50M_RANGE + WS10M_MIN",
dataset, groups= 'fips').fit()
```

Figura 6.9: Modelo Lineal Mixto en Python Fuente: Elaboración Propia

El proceso de ajuste de los modelos lineales mixtos implica encontrar los valores de los parámetros de máxima verosimilitud de los datos observados. La verosimilitud es una medida de la probabilidad de obtener los datos observados dados los parámetros del modelo. En otras palabras, se buscan los valores de los parámetros que hacen que los datos observados sean más probables bajo el modelo propuesto.

CAPÍTULO 7

RESULTADOS Y CONCLUSIONES

En este capítulo se presentan los resultados del modelo de efectos ajustados y el modelo lineal mixto. La sección de resultados se centra solamente en las variables libres de multicolinealidad. Las tablas presentadas en este capítulo solamente contienen las variables estadísticamente significativas. Además, al igual que en los anteriores, cuando se habla de datos, se hace referencia a los datos pertenecientes a los distintos condados de Estados Unidos. Los resultados de cada modelo constan de dos partes: tabla de resultados y ecuación resultante.

Se utiliza la R^2 como métrica de ajuste y dato complementario, debido a que los objetivos se basan en encontrar la significancia de las variables meteorológicas y no en el ajuste del modelo. La R^2 mide la proporción de la varianza total de la variable dependiente que es explicada por las variables independientes. Además, es una métrica útil para evaluar cuánta variabilidad del fenómeno estudiado es capturada por el conjunto de variables explicativas incluidas en el modelo (Miles, 2005). Una R^2 baja significa que el modelo no está capturando gran parte de la variabilidad de los datos. Por lo tanto, existe la posibilidad de que haya factores, que no se estén considerando en el modelo, que hagan falta para explicar la variable dependiente. Sin embargo, una R^2 baja no invalida la significancia estadística de los coeficientes del modelo (vía los p – valores) de forma individual (Colton, 2002). Según Eisenhauer (2009), los modelos de regresión con un pequeño poder explicativo pueden mostrar significancia. También, existen artículos publicados con menos del 20% de poder explicativo. Además, Lo (2015) presenta un estudio sobre cómo las variables

significativas no necesariamente son fuertes predictores y viceversa. En resumen, a pesar de que se obtenga una R^2 baja, los p – valores siguen siendo útiles para evaluar la significancia de los efectos individuales en el modelo.

Las tablas de resultados contienen 5 columnas, la primera columna es el nombre (significado) de la variable. La segunda columna representa el coeficiente, este indica la magnitud y dirección que tienen las variables respecto a la variable objetivo. El signo representa el impacto, positivo o negativo, que la variable tiene respecto al grado de sequía. La tercera columna muestra el p – valor, resultado de una prueba de hipótesis con un $\alpha = 0.05$ donde:

H_0 : La variable no es significativa

H_1 : La variable es significativa

Si p – valor $< \alpha$ se rechaza la hipótesis nula, siendo esta variable significativa. La cuarta columna representa el error estándar, que refleja la dispersión o variabilidad de los valores observados respecto a la línea de regresión ajustada. La calidad de ajuste de la regresión está representada por el error estándar de cada coeficiente. Por último, la quinta columna es la significancia, donde un asterisco simple (*) representa un p – valor < 0.05 , doble asterisco (**) un p – valor < 0.01 y triple asterisco (***) un p – valor < 0.001 . Los p-valores son el principal criterio utilizado para determinar qué variable contribuye o no al grado de sequía.

Las ecuaciones resultantes de cada modelo están compuestas por los coeficientes de las tablas de resultados y su intercepto estimado. Estas están representadas con el nombre de las variables del conjunto de datos. Se describen todos los coeficientes de las variables en las ecuaciones resultantes, se indica su orientación mediante el signo asociado, y el error estándar.

Los resultados de los modelos se comparan en función de la significancia de las variables y la R^2 . Es importante recalcar que los coeficientes de los modelos se estiman con distintas metodologías. El modelo lineal mixto se estima

mediante máxima verosimilitud y el de efectos ajustados por *OLS*. La diferencia en las estimaciones puede traer consigo diferencias en los coeficientes.

7.1 MODELO DE EFECTOS AJUSTADOS

$R^2 = 0.0468$				
Nombre	Coeficiente	P-value	Error Estandar	Significancia
Intercepto	19.685	0.00	0.2317	***
Rango de temperatura 2m	0.0935	0.00	0.0004	***
Temperatura de la superficie de la tierra	0.0267	0.00	0.0002	***
Rango de Velocidad del viento en 50m	-0.0264	0.00	0.0008	***
Presión de la superficie	-0.2078	0.00	0.0024	***
Velocidad mínima del viento en 50m	-0.0185	0.00	0.0007	***
Precipitación	-0.0207	0.00	0.0002	***

Tabla 7.1: Resultados del Modelo de Efectos Ajustados Fuente: Elaboración Propia

La tabla 7.1 presenta los resultados del modelo de efectos ajustados, con efectos aleatorios y efectos en el tiempo. Los p – valores de cada una de las variables explicativas son 0, se rechaza la hipótesis nula de cada una. Los factores meteorológicos utilizados en el modelo de efectos ajustados resultan ser significativos. Todas las variables explicativas son significativas con un p – valor < 0.001 , representadas con el triple asterisco (***). Los errores estándar de cada variable son muy pequeños, esto representa que las distintas variables se ajustan de manera adecuada a la regresión. La calidad de la regresión aparenta ser buena por los bajos errores estándar y la significancia de las variables explicativas en la severidad de sequía. Sin embargo, el R^2 obtenido es

de 0.0468, un valor muy bajo. Por lo tanto, la proporción de la variabilidad total de la variable *score* es baja explicada por el modelo de regresión.

$$\text{score} = 19.685 + 0.0935 \text{ T2M_RANGE} + 0.0267 \text{ TS} - 0.0264 \text{ WS50M_RANGE} - 0.2078 \text{ PS} - 0.0185 \text{ WS50M_MIN} - 0.02047 \text{ PRECTOT}$$

La ecuación resultante nos indica que el rango de temperatura en 2 metros tiene un efecto positivo en el grado de sequía. Esto significa que al aumentar la temperatura, se produce un aumento en el grado de sequía. El coeficiente de la temperatura es uno de los más altos, representa que la variable explicativa influye en un alto grado a la severidad de sequía. Sin embargo, es mucho menor que el intercepto.

La temperatura de la superficie de la tierra es de signo positivo, al igual que la temperatura en dos metros. Esto implica que un incremento en la temperatura de la superficie de la tierra conlleva a un aumento en la intensidad de la sequía. El modelo de efectos ajustados coincide en los signos de los dos factores de temperatura. El coeficiente de la variable explicativa no es alto, sin embargo, su aumento repercute en el aumento del grado de sequía. Las sequías comienzan con falta de precipitación. Además, el aumento de temperatura en la tierra genera alta evapotranspiración, la cual agota rápidamente la cantidad de agua.

El rango de velocidad del viento en 50 metros tiene un efecto negativo hacia el grado de sequía. Su coeficiente está dentro del promedio de las demás variables explicativas. La humedad puede ser transportada por el viento, también puede influir en la dirección e intensidad de una tormenta. El constante flujo de viento trae consigo el movimiento de nubes las cuales pueden producir precipitación. El choque de vientos puede generar el ascenso o descenso de este. Además, el viento al subir puede generar precipitación.

El coeficiente de la presión superficial es el de mayor magnitud entre las variables explicativas. Por lo tanto, la variación de la presión superficial tiene el

mayor impacto en el grado de sequía. El coeficiente tiene un impacto negativo al grado de sequía. La presión superficial en sí misma no genera precipitación, sin embargo, es un factor importante que puede influir en la formación de sistemas meteorológicos que pueden conducir a la precipitación. La presión superficial baja provoca climas tranquilos y despejados, sin embargo, si la presión es alta produce viento, nubosidad y precipitación. La variable de presión superficial tiene el error estándar mayor entre las variables explicativas.

La velocidad mínima del viento en 50 metros tiene un efecto negativo al grado de sequía, igual que el rango del viento en 50 metros. Estas dos variables van de la mano, su error estándar es casi idéntico, los coeficientes son cercanos y el signo es el mismo.

La precipitación tiene un efecto negativo al grado de sequía. El coeficiente de precipitación está en el promedio de las demás magnitudes de las variables explicativas. El error estándar de la variable de precipitación es de los más pequeños. La literatura nos indica que toda sequía comienza con la falta de precipitación, se espera que el coeficiente de sequía sea negativo. El modelo de efectos ajustados identifica este factor según lo esperado. Sin embargo, el coeficiente que le asigna a la sequía no es destacable en magnitud.

7.2 MODELO LINEAL MIXTO

$R^2 = 0.1739$				
Nombre	Coeficiente	P-value	Error Estandar	Significancia
Intercepto	35.966	0.00	0.210	***
fecha	0.00	0.00	0.00	***
Rango de temperatura 2m	0.088	0.00	0.00	***
Temperatura de la superficie de la tierra	-0.001	0.00	0.00	***
Rango de Velocidad del viento en 50m	-0.010	0.00	0.001	***
Presión de la superficie	-0.023	0.00	0.001	***
Velocidad mínima del viento en 50m	0.003	0.00	0.001	***
Precipitación	-0.020	0.00	0.00	***
Varianza de condados	0.092	0.002		

Tabla 7.2: Resultados del Modelo Lineal Mixto Fuente: Elaboración Propia

La tabla 7.2 presenta los resultados del modelo lineal mixto con un intercepto aleatorio por condado. Las variables explicativas, utilizadas para ajustar el modelo regresión, son significativas. Estas son significativas con un p – *valor* < 0.001 , representado con el triple asterisco (***). El error estándar de cada una de las variables es cercano a cero, incluso variables como el rango de temperatura en dos metros, temperatura de la superficie de la tierra y la precipitación son cero. Los errores estándar son bajos, por lo tanto, las variables se ajustan de manera adecuada a la regresión. Por otro lado, el R^2 obtenido es de 0.1739, un valor bajo. Por lo tanto, la proporción de la variabilidad total del grado de sequía es baja explicada por el modelo lineal mixto.

$$\begin{aligned} score = & 35.966 + 0.092 \textit{fips} + 0.088 \textit{T2M_RANGE} - 0.001 \textit{TS} - 0.010 \textit{WS50M_RANGE} \\ & - 0.023 \textit{PS} + 0.003 \textit{WS50M_MIN} - 0.020 \textit{PRECTOT} \end{aligned}$$

El rango de temperatura en 2 metros tiene la mayor magnitud entre las variables explicativas. La magnitud asignada a esta variable, por el modelo lineal mixto, tiene una diferencia significativa entre las variables explicativas. La temperatura en 2 metros aumenta junto al grado de sequía y viceversa. Por lo tanto, tiene un efecto positivo hacia el grado de sequía. Su error estándar es cero, beneficiando al ajuste del modelo.

El coeficiente negativo de la temperatura de la superficie de la tierra difiere de lo anteriormente mencionado, ya que el aumento de temperatura en un rango de 2 metros aumenta junto a la temperatura de la superficie de la tierra. El modelo identifica las variables de temperatura con distinta afectación hacia el grado de sequía. Sin embargo, la magnitud del coeficiente de la temperatura de la superficie de la tierra es muy pequeña. El modelo lineal mixto, no asigna a la temperatura de la superficie de la tierra como un factor de alto impacto en el grado de sequía.

El rango de la velocidad del viento en 50 metros tiene un efecto negativo en el grado de sequía. El modelo difiere en la dirección del coeficiente de la velocidad mínima del viento en 50 metros. El rango y el mínimo de la velocidad del viento, se espera que tengan la misma dirección en sus coeficientes.

La precipitación tiene un efecto negativo en el grado de sequía. El error estándar de la precipitación es cero, representando su buen ajuste de esta variable en a la regresión.

7.3 COMPARATIVA DE MODELOS

Nombre	EA $R^2 = 0.0468$	MLM $R^2 = 0.1739$
Intercepto	19.685	35.966
Rango de temperatura 2m	0.0935	0.088
Temperatura de la superficie de la tierra	0.0267	-0.001
Rango de Velocidad del viento en 50m	-0.0264	-0.010
Presión de la superficie	-0.2078	-0.023
Velocidad mínima del viento en 50m	-0.0185	0.003
Precipitación	-0.0207	-0.020

Tabla 7.3: Comparativa Entre Modelos Fuente: Elaboración Propia

En modelo de efectos ajustados se obtiene una R^2 de 0.0468, lo que indica una capacidad muy limitada para explicar la variabilidad de los datos. Por otro lado, el modelo lineal mixto alcanza una R^2 de 0.1739, mostrando el modelo una mayor captura en la proporción de la varianza total del grado de sequía. Sin embargo, este valor aún representa una proporción limitada de la variabilidad total. La métrica de ajuste R^2 no invalida la significancia de cada variable. Además, es crucial destacar las complejidades intrínsecas de la modelación estadística y del análisis de sequía, así como la posibilidad de tener en cuenta otras variables para una evaluación integral del modelo y el alcance un mejor ajuste.

Las variables en ambos modelos son significativas, sin embargo, los coeficientes junto a los errores estándar varían. El coeficiente del rango de temperatura en 2 metros en ambos modelos es positivo, también, no hay una diferencia significativa en su magnitud. Los modelos coinciden en que, el aumento de la temperatura tiene un efecto positivo en grado de sequía. La literatura sustenta que esta es uno de los factores que genera sequías (Sundararajan, 2021).

La temperatura de la superficie de la tierra es una de las diferencias destacables entre modelos. El modelo de efectos ajustados le da una magnitud mayor en comparación con el modelo lineal mixto. Incluso, el impacto en relación con la sequía, según el modelo lineal mixto, es pequeño. Mientras que en un modelo le da una mayor magnitud a esta variable, el otro le da menos impacto en la regresión. Además, la dirección de cada coeficiente, según el modelo, es distinta. El modelo lineal mixto le asigna, a esta variable, un signo negativo. En resumen, difiere en la dirección entre la temperatura de la superficie de la tierra y el rango de temperatura en 2 metros.

Los modelos consideran el rango de la velocidad del viento en 50 metros con un impacto negativo hacia el grado de sequía. La magnitud entre coeficientes no tiene una diferencia destacable.

La presión de la superficie es otra diferencia relevante entre modelos. Esta variable explicativa tiene la mayor magnitud dentro del modelo de efectos ajustados. Por otro lado, en el modelo lineal mixto su magnitud no es destacable entre las demás variables explicativas. La dirección de esta variable en ambos modelos es la misma, la presión superficial tiene un efecto negativo en el grado de sequía.

El coeficiente de la velocidad mínima del viento en 50 metros, según el modelo lineal mixto, es muy pequeño. Esta y el rango del viento en 50 metros difieren en dirección. Por otro lado, en el modelo de efectos ajustados, la dirección de estas variables coincide. Además, les asigna una magnitud mayor a estas variables explicativas, se puede decir que hay una diferencia notable.

La precipitación en ambos modelos tiene un impacto negativo y la magnitud es casi idéntica. Los modelos coinciden en la dirección de esta variable, ya que sin duda la precipitación tiene un impacto negativo al grado de sequía.

7.4 DISCUSIÓN

La sequía es uno de los desastres naturales más caros. El entendimiento del comportamiento de este fenómeno trae consigo beneficios extraordinarios. El identificar factores determinantes meteorológicos en la sequía, nos da la oportunidad de mitigarlos. Las sequías no conocen fronteras, es un problema global. Este fenómeno desata problemas en efecto dominó y la falta de precipitación marca el comienzo de afectaciones más severas.

El objetivo general de esta tesina es generar un modelo explicativo capaz de identificar variables meteorológicas determinantes en la generación de sequía. Este modelo debe ser generado basado en los distintos condados de Estados Unidos. Además, se busca cubrir dos objetivos específicos: 1) Elaborar el modelo explicativo en Python. 2) Comparar el comportamiento que tiene un modelo lineal mixto contra un modelo de efectos ajustados.

En el estudio se generan dos modelos en Python: un modelo de efectos ajustados y un modelo lineal mixto. Ambos modelos tienen un efecto aleatorio en los condados de Estados Unidos y efectos en el tiempo. Las variables utilizadas en ambos modelos, después del filtro de multicolinealidad, resultan significativas. Las variables explicativas son significativas con un $p - valor < 0.001$, representado con el triple asterisco (***) . En ambos modelos el error estándar de las variables explicativas es bajo. A pesar de que en algunos coeficientes hay coincidencias tanto en magnitud como en dirección, el modelo de efectos ajustados tiene una interpretación más coherente respecto a las variables.

El modelo lineal mixto difiere, dentro de sí mismo, en ciertas direcciones de variables explicativas que por sentido común deberían coincidir. El rango y el mínimo de velocidad del viento en 50 metros tienen diferente dirección. También difiere en el rango de temperatura y la temperatura de la superficie de la tierra.

Por lo tanto, al ser ambas variables dependientes del viento y temperatura se considera que hay contradicciones en el signo, pero la magnitud de una de ellas es cercana a cero, por lo que se puede considerar volver a hacer el experimento sin ellas. Sin embargo, el modelo de efectos ajustados obtuvo un R^2 de 0.0468, lo cual señala una capacidad muy restringida para explicar la variabilidad de los datos. En contraste, el modelo lineal mixto logró un R^2 de 0.1739, evidenciando una mayor captura en la proporción de la varianza total del grado de sequía. Sin embargo, este valor representa una proporción limitada de la variabilidad total.

Es esencial destacar la complejidad de los modelos de regresión y, de igual manera, la complejidad del fenómeno de la sequía. La literatura científica resalta que la sequía es un fenómeno natural complejo, influido por múltiples factores. En esta investigación no se logra explicar la variación de este fenómeno, pues en ambos modelos se tienen ajustes muy bajos. Este hallazgo subraya la necesidad de considerar otros factores y métricas para mejorar el ajuste. Además, tomar en cuenta la posible relación intrínseca entre variables, para obtener una comprensión integral y precisa del fenómeno de la sequía y su efecto en el *DSCI*. Sin embargo, encontramos variables meteorológicas significativas en un alto grado, con un $p - \text{valor} < 0.001$. Asimismo, sus signos van acorde a la intuición y a la literatura.

7.5 CONCLUSIÓN

Los objetivos iniciales de la investigación se han alcanzado con éxito. Se generó un modelo explicativo, para identificar los factores determinantes de sequía en la región de Estados Unidos. Para la construcción de los modelos estadísticos, se empleó Python como herramienta. Se procedió a comparar el modelo de efectos ajustados y el modelo lineal mixto.

Se encontraron variables con un alto grado de significancia. Según ambos modelos, las altas temperaturas impactan en el aumento del grado de sequía. Por otro lado, la falta de precipitación es un factor clave para el aumento de grado de sequía. Además, la presión en la superficie de la tierra es un factor de impacto

en la formación de sequía. Este factor está relacionado con la precipitación. La precipitación se produce cuando el aire húmedo se eleva y se enfría, lo que hace que el vapor de agua en el aire se condense en nubes y finalmente se forme precipitación.

El entendimiento de la sequía es muy complejo, ya que este fenómeno se desarrolla por una gran cantidad de variables. El estudio desarrollado en este documento abarcó variables meteorológicas, sin embargo, los modelos utilizados no resultan ser explicativos, por lo que se recomienda utilizar más variables y distintos modelos para describir de mejor forma la sequía. Sin embargo, como se menciona a lo largo de los capítulos, existen diferentes niveles de sequía. El estudio de cada uno implica diferentes variables.

REFERENCIAS

- Anshuka, A., Ogtrop, F., & Willem, R. (2019). Drought Forecasting Through Statistical Models Using Standardized Precipitation Index: A Systematic Review and Meta-Regression Analysis. *Natural Hazards*, 97, 955-977. <https://doi.org/10.1007/s11069-019-03665-6>.
- Balestra, P., & Nerlove, M. (1996). Pooling Cross-Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas. *Econometrica*, 34, 585–612. <https://doi.org/10.2307/1909771>.
- Baltagi, B. (2008). *Econometric Analysis of Panel Data* (Vol. 6). Springer. <https://doi.org/10.1007/978-3-030-53953-5>.
- Barton, B., & Peat, J. (2014). *Medical Statistics: A Guide to SPSS, Data Analysis and Critical Appraisal*. (Vol. 2). John Wiley & Sons Ltd.
- BBC. (2022). Spain's Olive Oil Producers Devastated by Worst Ever Drought. <https://www.bbc.com/news/world-europe-62707435>.
- Bittencourt, P., Oliveira, R., da Costa, A., Giles, A., Coughlin, I., Costa, P., & Rowland, L. (2020). Amazonia Trees Have Limited Capacity to Acclimate Plant Hydraulic Properties in Response to Long-Term Drought. *Global Change Biology*, 26(6), 3569-3584. <https://doi.org/10.1111/gcb.15040>.
- Brown, V. (2021). An Introduction to Linear Mixed-Effects Modeling in R. *Advances in Methods and Practices in Psychological Science*, 4(1), 1-19. <https://doi.org/10.1177/2515245920960351>.
- Christensen, S., Ruder, M., Williams, D., Porter, W., & Stallknecht, D. (2020). The Role of Drought as a Determinant of Hemorrhagic Disease in the Eastern

- United States. *Global Change Biology*, 26(7), 3799-3808.
<https://doi.org/10.1111/gcb.15095>.
- Colton, J., & Bower, K. (2002). Some Misconceptions About R². *International Society of Six Sigma Professionals, EXTRAOrdinary Sense*, 3(2), 20-22.
- Cryer, J., & Kung-Sik, C. (2008). *Time Series Analysis with Applications in R* (Vol. 2). Springer. <https://doi.org/10.1007/978-0-387-75959-3>.
- Eisenhauer, J. (2009). Explanatory Power and Statistical Significance. *Teaching Statistics: An International Journal for Teachers*, 31(2), 42-46.
- Fulton, C. (2022). Bayesian Estimation and Forecasting of Time Series in statsmodels. *Proc. of the 21st Python in Science Conf. (SCIPY 2022)*, 83-89.
- Han, J., & Singh, V. (2020). Forecasting of Droughts and Tree Mortality Under Global Warming: A Review of Causative Mechanisms and Modeling Methods. *Journal of Water and Climate Change*, 11(3), 600-632.
<https://doi.org/10.2166/wcc.2020.239>.
- Hao, Z., Singh, V., & Xia, Y. (2018). Seasonal Drought Prediction: Advances, Challenges, and Future Prospects. *Reviews of Geophysics*, 56(1), 108-141. <https://doi.org/10.1002/2016RG000549>.
- Heim J., Richard. R. (2002). A Review of Twentieth-Century Drought Indices Used in the United States. *Bulletin of the American Meteorological Society*, 83(8), 1149-1166. <https://doi.org/10.1175/1520-0477-83.8.1149>.
- Hsiao, C. (2005). Why Panel Data?. *The Singapore Economic Review*, 50(02), 143-154. <https://doi.org/10.1142/S0217590805001937>.
- Hsiao, C. (2022). *Analysis of Panel Data*. (Vol 4). Cambridge University Press.
- Jiménez, I., Villegas, L., Salazar-Urbe, J., & Álvarez, L. (2020). Facial Growth Changes in a Colombian Mestizo Population: An 18-year Follow-up Longitudinal Study Using Linear Mixed Models. *American Journal of Orthodontics and Dentofacial Orthopedics*, 157(3), 365-376.
<https://doi.org/10.1016/j.ajodo.2019.04.032>.

- Jiménez, L. (2020). Impacto de la Investigación Cuantitativa en la Actualidad. *Convergence Tech*, 4(1), 59-68. <https://doi.org/10.53592/convtech.v4i1V.35>.
- Jolly, E. (2018). Pymer4: Connecting R and Python for Linear Mixed Modeling. *Journal of Open Source Software*, 3(31), 862. <https://doi.org/10.21105/joss.00862>.
- Kuwayama, Y., Thompson, A., Bernknopf, R., Zaitchik, B., & Vail, P. (2019). Estimating the Impact of Drought on Agriculture Using the US Drought Monitor. *American Journal of Agricultural Economics*, 101(1), 193-210. <https://doi.org/10.1093/ajae/aay037>.
- Lo, A., Chernoff, H., Zheng, T., & Lo, S. H. (2015). Why Significant Variables Aren't Automatically Good Predictors. *Proceedings of the National Academy of Sciences*, 112(45), 13892-13897.
- Lu, J., Carbone, G., Huang, X., Lackstrom, K., & Gao, P. (2020). Mapping the Sensitivity of Agriculture to Drought and Estimating the Effect of Irrigation in the United States, 1950–2016. *Agricultural and Forest Meteorology*, 292, 108124. <https://doi.org/10.1016/j.agrformet.2020.108124>.
- Lutz, M. (2013). *Learning Python: Powerful Object-Oriented Programming*. (Vol. 5). O'Reilly Media, Inc..
- Maddala, G., & Lahiri, K. (1992). *Introduction to Econometrics*. (Vol. 2). New York: Macmillan. <https://doi.org/10.12660/bre.v9n11989.3082>.
- Malik, A., Tikhamarine, Y., Souag-Gamane, D., Rai, P., Sammen, S., & Kisi, O. (2021). Support Vector Regression Integrated with Novel Meta-Heuristic Algorithms for Meteorological Drought Prediction. *Meteorology and Atmospheric Physics*, 133, 891-909. <https://doi.org/10.1007/s00703-021-00787-0>.
- Miles, J. (2005). R-squared, Adjusted R-squared. *Encyclopedia of Statistics in Behavioral Science*, 4, 1655-1657. <https://doi.org/10.1002/0470013192.bsa526>.
- Mishra, A., & Singh, V. (2011). Drought Modeling—A Review. *Journal of Hydrology*, 403(1-2), 157-175. <https://doi.org/10.1016/j.jhydrol.2011.03.049>.

- National Integrated Drought Information System (NIDIS). (2022). Current Conditions. <https://www.drought.gov/current-conditions>.
- Nerlove, M. (1996). Pooling Cross-Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas. *Econometrica*, 34, 585–612. <https://doi.org/10.2307/1909771>.
- ONU. (2022). Global Issue: Population. <https://www.un.org/en/global-issues/population>.
- Rahm, E., & Do, H. H. (2000). Data Cleaning: Problems and Current Approaches. *IEEE Data Eng. Bull.*, 23(4), 3-13.
- Raschka, S., Patterson, J., & Nolet, C. (2020). Machine Learning in Python: Main Developments and Technology Trends in Data Science, Machine Learning, and Artificial Intelligence. *Information*, 11(4), 193. <https://doi.org/10.3390/info11040193>.
- Richardson, D. (2019). The Predictability of UK Drought Using European Weather Patterns (Doctoral dissertation, Newcastle University).
- Ridzuan, F. (2019). A Review on Data Cleansing Methods for Big Data. *Procedia Computer Science*, 161, 731-738. <https://doi.org/10.1016/j.procs.2019.11.177>.
- Rodríguez, R., Benito, Á., & Portela, A. (2004). Meteorología y Climatología. Fundación Española para la Ciencia y la Tecnología. Villena Artes Gráficas España, 12-16.
- Salvador, C., Nieto, R., Linares, C., Díaz, J., & Gimeno, L. (2020). Effects of Droughts on Health: Diagnosis, Repercussion, and Adaptation in Vulnerable Regions Under Climate Change. Challenges for future research. *Science of the Total Environment*, 703, 134912. <https://doi.org/10.1016/j.scitotenv.2019.134912>.
- Shrestha, N. (2020). Detecting Multicollinearity in Regression Analysis. *American Journal of Applied Mathematics and Statistics*, 8(2), 39-42. DOI:10.12691/ajams-8-2-1.
- Sundararajan, K., Garg, L., Srinivasan, K., Bashir, A. K., Kaliappan, J., Ganapathy, G. P., ... & Meena, T. (2021). A Contemporary Review on Drought Modeling

- Using Machine Learning Approaches. *Computer Modeling in Engineering & Sciences*, 128(2), 447-487. <https://doi.org/10.32604/cmes.2021.015528>.
- The New York Times. (2022). Mexico's Cruel Drought: 'Here You Have to Chase the Water'. <https://www.nytimes.com/2022/08/03/world/americas/mexico-drought-monterrey-water.html>.
- Tufaner, F., & Özbeyaz, A. (2020). Estimation and Easy Calculation of the Palmer Drought Severity Index from the Meteorological Data by Using the Advanced Machine Learning Algorithms. *Environmental Monitoring and Assessment*, 192, 1-14. <https://doi.org/10.1007/s10661-020-08539-0>.
- U.S. Drought Monitor. (2022). Current Map | U.S. Drought Monitor. de <https://droughtmonitor.unl.edu>.
- U.S. Drought Monitor. (2023). Data Download. <https://droughtmonitor.unl.edu/DmData/DataDownload.aspx>.
- UNICEF. (2022). Los Niños y Niñas Afectados por la Sequía en Algunas Zonas de África se Encuentran al Borde de la Catástrofe. <https://www.unicef.org/es/comunicados-prensa/ninos-afectados-por-sequia-zonas-africa-borde-catastrofe>.
- Wooldridge, J. (2009). *Introductory Econometrics: A modern approach*. (Vol. 4) Cengage learning.
- Zargar, A., Sadiq, R., Naser, B., & Khan, F. (2011). A Review of Drought Indices. *Environmental Reviews*, 19, 333-349, <https://doi.org/10.1139/a11-013>.