

# Precipitación Acumulada Mensual en México

## Análisis Exploratorio de Datos

A partir de datos registrados por **CONAGUA** de la precipitación Acumulada Mensual por entidad federativa, se busca hacer un análisis exploratorio con enfoque en Nuevo León para apoyar el desarrollo del "**Modelo de Redes Neuronales del Sistema de Agua en el Área Metropolitana de Monterrey**" como proyecto de tesis.

```
In [1]: 1 import pandas as pd
        2 import os
```

```
In [2]: 1 df = pd.read_excel("data/Precipitacion/1985Precip.xls", header=1)
        2 first_column_name = df.columns[0]
        3 first_column = df.iloc[:, 0]
        4 entidades_federativas = pd.DataFrame({first_column_name: first_column})
```

```
In [3]: 1 df.head(5)
```

Out[3]:

	ENTIDAD	ENE	FEB	MAR	ABR	MAY	JUN	JU
0	AGUASCALIENTES	4.914729	1.588740	1.947601	17.612315	21.458081	137.052753	85.3982
1	BAJA CALIFORNIA	12.150562	23.625055	9.962204	3.723647	0.017576	0.343847	0.7700
2	BAJA CALIFORNIA SUR	30.304723	0.058778	0.334605	0.259919	0.128790	0.331742	19.5520
3	CAMPECHE	20.943153	26.685829	16.509803	4.248756	233.420941	124.723530	142.7139
4	COAHUILA	40.714546	6.152208	4.887759	43.611932	46.949689	46.314182	20.2395

## Limpieza de Datos

Los datos vienen guardados en múltiples archivos .xls por año desde 1985 hasta 2022. Los datos vienen mensuales por cada entidad federativa y nacional. No hay datos faltantes.

El objetivo será guardar los datos en diferentes formatos que facilite su análisis por mes, por año y por estado. Buscamos responder:

1. ¿Qué mes tiene mayor precipitación?
2. ¿Qué estado tiene mayor precipitación?
3. Serie temporal de precipitación por estado.
4. Distribución de probabilidad de precipitación.

El dataframe principal (serie temporal) será ordenado por entidad federativa (fila) en orden cronológico (columnas).

```
In [4]: 1 # Define the directory where your Excel files are located
        2 excel_files_directory = 'data/Precipitacion/'
```

```

In [4]: 1 # Define the directory where your Excel files are located
2 excel_files_directory = 'data/Precipitacion/'
3
4 # Initialize an empty list to store DataFrames
5 dataframes_main = []
6 dataframes_anual = []
7 df_monthly = pd.DataFrame()
8
9 # Loop through the files in the directory
10 for filename in os.listdir(excel_files_directory):
11     if filename.endswith('.xls') or filename.endswith('.xlsx'):
12         # Read each Excel file and append it to the list
13         file_path = os.path.join(excel_files_directory, filename)
14         df = pd.read_excel(file_path, header=1) # Read Excel file, assuming
15         df1 = df.iloc[:, 1:-1]
16         df2 = df.iloc[:, 13]
17
18         dataframes_main.append(df1)
19         dataframes_anual.append(df2)
20
21 # Concatenate all DataFrames into one
22 df = pd.concat(dataframes_main, axis=1)
23 df_anual = pd.concat(dataframes_anual, axis=1)
24
25 for da_frame in dataframes_main:
26     if df_monthly.empty:
27         df_monthly = da_frame # Set the first DataFrame in the list as the
28     else:
29         df_monthly = df_monthly.add(da_frame, fill_value=0) # Sum the DataF
30
31 df_monthly = df_monthly/12

```

```

In [5]: 1 # Create a date range from January 1985 to December 2022\
2 from datetime import datetime
3
4 date_range = pd.date_range("1985-01-01", "2022-12-31", freq="M")
5 df.columns = date_range.strftime("%Y-%m")

```

```

In [6]: 1 df = pd.concat([entidades_federativas, df], axis=1, ignore_index=False)
2 df.to_csv('results\precipitacion_mensual.csv', index=False)

```

```

In [7]: 1 df_anual.columns = [year for year in range(1985, 2023)]
2 df_anual = pd.concat([entidades_federativas, df_anual], axis=1, ignore_index=
3 df_anual.to_csv('results\precipitacion_anual.csv', index=False)

```

```

In [8]: 1 df_monthly = pd.concat([entidades_federativas, df_monthly], axis=1, ignore_i
2 df_monthly.to_csv('results\precipitacion_mensual_promedio.csv', index=False)

```

## Resultado de limpieza de datos

Las siguientes tablas serán exportadas como archivos csv para su uso, se pueden encontrar en la carpeta de resultados.

```

In [9]: 1 print("Precipitación Acumulada (mm) mensual desde 1985 a 2022 por Entidad Fed
2 display(df.head(5))

```

```
In [9]: 1 print("Precipitación Acumulada (mm) mensual desde 1985 a 2022 por Entidad Federativa en México")
2 display(df.head(5))
3 print("Precipitación Acumulada (mm) anual desde 1985 a 2022 por Entidad Federativa en México")
4 display(df_anual.head(5))
5 print("Precipitación Acumulada (mm) promedio mensual por Entidad Federativa en México")
6 display(df_monthly.head(5))
```

Precipitación Acumulada (mm) mensual desde 1985 a 2022 por Entidad Federativa en México

	ENTIDAD	1985-01	1985-02	1985-03	1985-04	1985-05	1985-06	1985-07
0	AGUASCALIENTES	4.914729	1.588740	1.947601	17.612315	21.458081	137.052753	85.3982
1	BAJA CALIFORNIA	12.150562	23.625055	9.962204	3.723647	0.017576	0.343847	0.7700
2	BAJA CALIFORNIA SUR	30.304723	0.058778	0.334605	0.259919	0.128790	0.331742	19.5520
3	CAMPECHE	20.943153	26.685829	16.509803	4.248756	233.420941	124.723530	142.7139
4	COAHUILA	40.714546	6.152208	4.887759	43.611932	46.949689	46.314182	20.2395

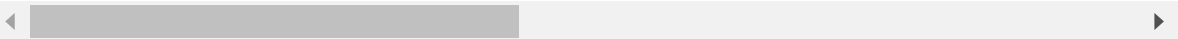
5 rows × 457 columns



Precipitación Acumulada (mm) anual desde 1985 a 2022 por Entidad Federativa en México

	ENTIDAD	1985	1986	1987	1988	1989	1990
0	AGUASCALIENTES	464.358537	434.904801	371.835178	406.821960	404.080402	671.880299
1	BAJA CALIFORNIA	174.391162	174.671611	317.290533	192.406248	89.656432	164.909622
2	BAJA CALIFORNIA SUR	115.634756	140.818859	120.910208	109.012021	214.875187	262.713471
3	CAMPECHE	1015.005528	969.872671	1064.172019	1274.786985	1199.983944	1173.543356
4	COAHUILA	340.317462	412.242148	396.092615	384.099335	322.249047	396.062589

5 rows × 39 columns



Precipitación Acumulada (mm) promedio mensual por Entidad Federativa en México

	ENTIDAD	ENE	FEB	MAR	ABR	MAY	JUN
0	AGUASCALIENTES	42.934488	38.958168	17.017400	17.085431	57.680218	269.452290
1	BAJA CALIFORNIA	93.866521	114.499389	75.166369	30.992015	9.219434	4.098296
2	BAJA CALIFORNIA SUR	30.788542	18.579770	10.560183	3.947551	1.242924	8.651636
3	CAMPECHE	136.512653	93.059961	67.967488	95.252507	286.385839	686.721057
4	COAHUILA	39.878823	35.064211	50.227582	64.223416	137.173028	150.244153



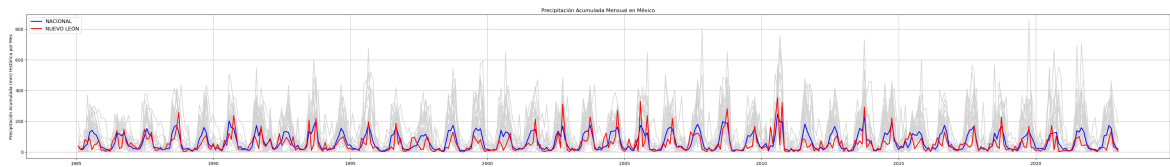
Visualización de datos

## Visualización de datos

In [10]: 1 `import matplotlib.pyplot as plt`

In [11]: 1 `%matplotlib notebook`

```
In [12]: 1 line_labels = (df.iloc[18, 0],df.iloc[32, 0])
2
3 plt.figure(figsize=(48, 6)) # Adjust the width and height as needed
4 # Create a line plot for each line label
5 for i in range(33):
6     if i == 18: # Highlight and label Nuevo Leon
7         continue
8     elif i == 32: # Highlight and label Nacional
9         plt.plot(date_range, df.iloc[i, 1:].values.tolist(), label=line_labels[i],
10            else:
11                plt.plot(date_range, df.iloc[i, 1:].values.tolist(), color='lightgray')
12 plt.plot(date_range, df.iloc[18, 1:].values.tolist(), label=line_labels[0],
13
14 # Customize the plot
15 plt.xlabel('Año')
16 plt.ylabel('Precipitación Acumulada (mm) Histórica por Mes')
17 plt.title('Precipitación Acumulada Mensual en México')
18 plt.legend(prop={'size': 12}, loc = "upper left")
19 plt.grid(True)
20
21 plt.savefig('results\precipitacion_mensual.png')
22 plt.show()
```



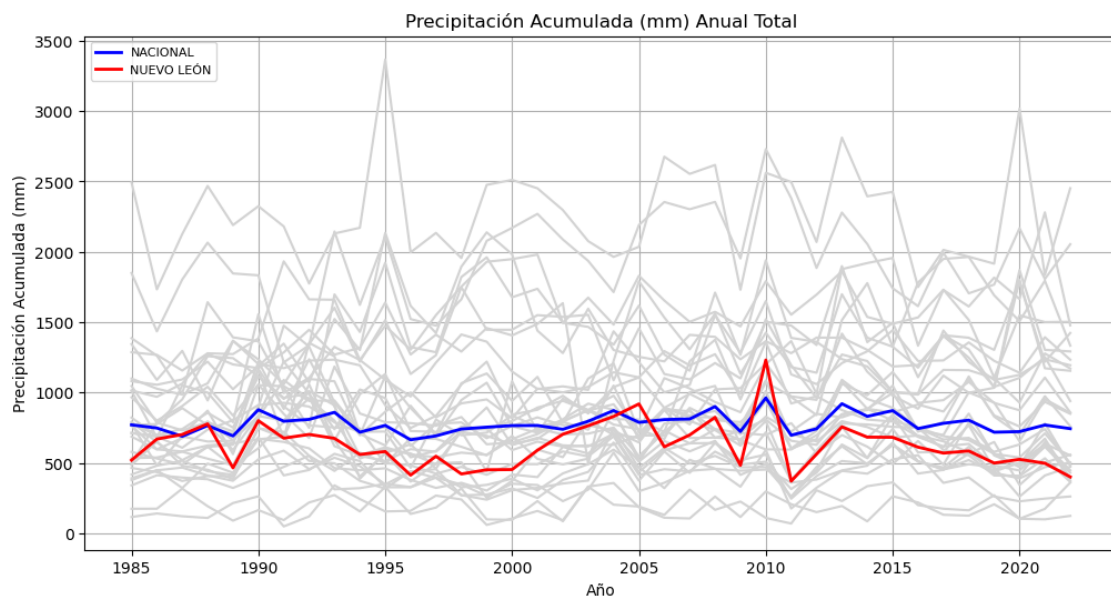
### Observaciones

Se puede observar la periodicidad anual. Hay años con valores atípicos ocasionados por tormentas tropicales, como lo fue para para Nuevo León en el 2010 ocasionado por el huracán Alex.

In [22]: 1 `%matplotlib inline`

```
In [23]: 1 year = df_anual.columns[1:].tolist()
2 plt.figure(figsize=(12, 6)) # Adjust the width and height as needed
```

```
In [23]: 1 year = df_anual.columns[1:].tolist()
2 plt.figure(figsize=(12, 6)) # Adjust the width and height as needed
3 # Create a line plot for each line label
4
5 for i in range(33):
6     if i == 18: # Highlight and Label Nuevo Leon
7         continue
8     elif i == 32: # Highlight and Label Nacional
9         plt.plot(year, df_anual.iloc[i, 1:].values.tolist(), label=line_labels[i],
10                color='lightgray')
11     else:
12         plt.plot(year, df_anual.iloc[i, 1:].values.tolist(), color='lightgray')
13
14 # Customize the plot
15 plt.xlabel('Año')
16 plt.ylabel('Precipitación Acumulada (mm)')
17 plt.title('Precipitación Acumulada (mm) Anual Total')
18 plt.legend(prop={'size': 8}, loc = "upper left")
19 plt.grid(True)
20
21 plt.savefig('results\precipitacion_anual.png')
22 plt.show()
```



### Oservaciones

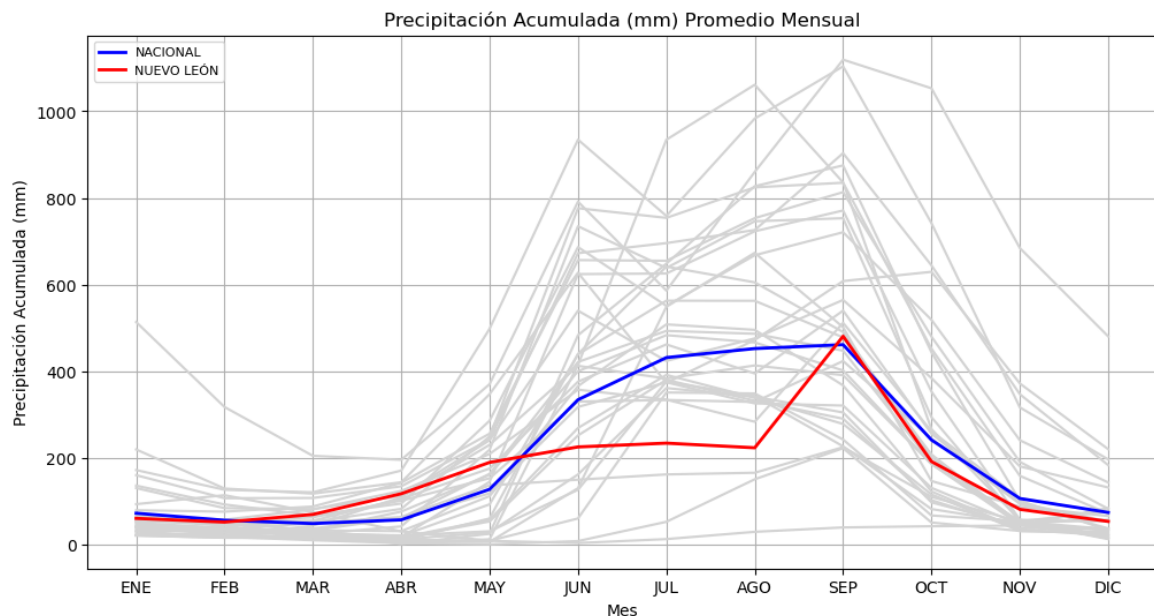
El promedio anual en Nuevo León ha bajado al promedio mientras que el de México se mantiene. Esto indica que el área Noroeste del país ha sido más afectada por sequía que el resto del país (en promedio). Hay mayor varianza en Nuevo León que en el país pero esto es de esperarse debido al menor tamaño de muestra, las características del clima con las estaciones bien delimitadas y condiciones bastante extremas.

```
In [15]: 1 month = df_monthly.columns[1:].tolist()
2 plt.figure(figsize=(12, 6)) # Adjust the width and height as needed
```

```

In [15]: 1 month = df_monthly.columns[1:].tolist()
2 plt.figure(figsize=(12, 6)) # Adjust the width and height as needed
3 # Create a line plot for each line label
4
5 for i in range(33):
6     if i == 18: # Highlight and Label Nuevo Leon
7         continue
8     elif i == 32: # Highlight and Label Nacional
9         plt.plot(month, df_monthly.iloc[i, 1:].values.tolist(), label=line_l
10    else:
11        plt.plot(month, df_monthly.iloc[i, 1:].values.tolist(), color='light
12 plt.plot(month, df_monthly.iloc[18, 1:].values.tolist(), label=line_labels[0
13
14 # Customize the plot
15 plt.xlabel('Mes')
16 plt.ylabel('Precipitación Acumulada (mm)')
17 plt.title('Precipitación Acumulada (mm) Promedio Mensual')
18 plt.legend(prop={'size': 8}, loc = "upper left")
19 plt.grid(True)
20
21 plt.savefig('results\precipitacion_avg_mensual.png')
22 plt.show()

```



### Observaciones

Es notable cómo Septiembre es un mes importante para satisfacer la necesidad hídrica del estado. Además de que para el resto del país en verano se esperan más lluvias de las que hay en Nuevo León.

```

In [16]: 1 #%%matplotlib inline

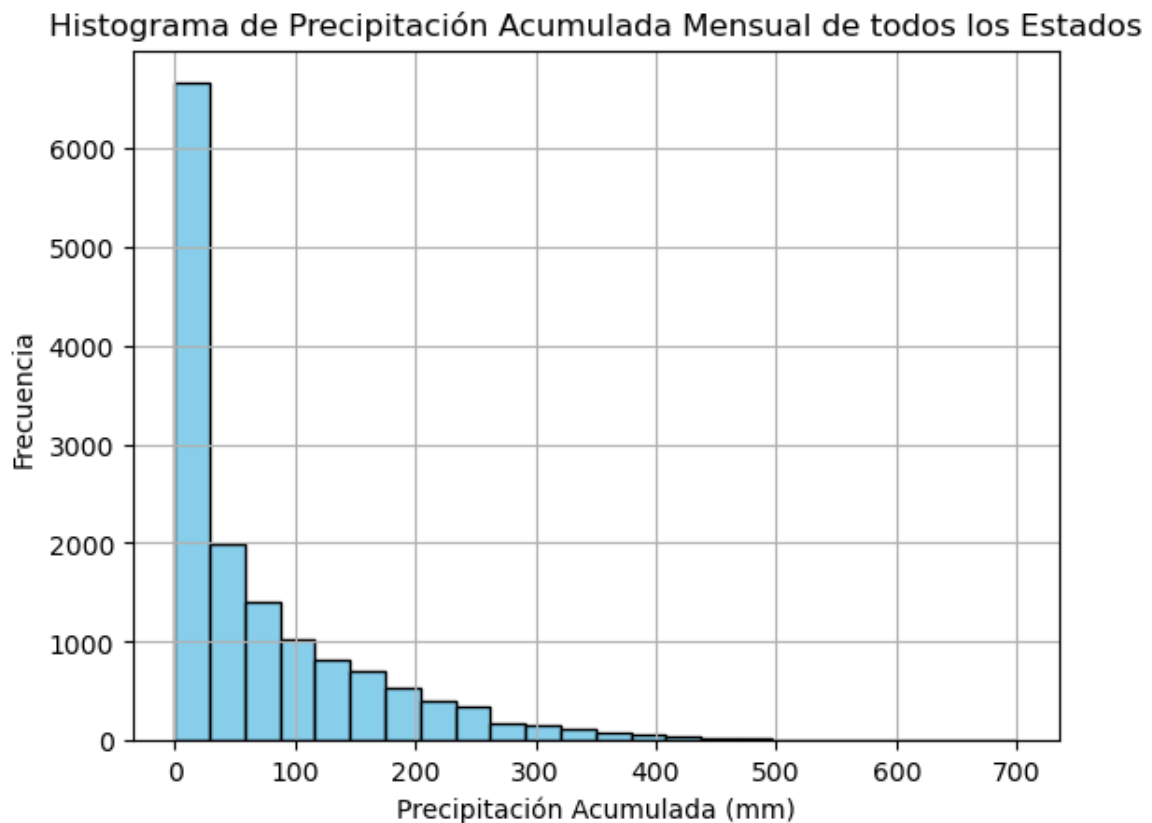
```

```

In [17]: 1 # Extrae todos los valores del DataFrame y conviértelos en una lista
2 data = df.iloc[:-1, 1:].values.flatten()

```

```
In [17]: 1 # Extrae todos los valores del DataFrame y conviértelos en una lista
2 data = df.iloc[: -1, 1:].values.flatten()
3
4 # Crea un histograma de los valores
5 plt.hist(data, bins=24, color='skyblue', edgecolor='black', range=(0, 700))
6 # Personaliza el histograma si es necesario
7 plt.xlabel('Precipitación Acumulada (mm)')
8 plt.ylabel('Frecuencia')
9 plt.title('Histograma de Precipitación Acumulada Mensual de todos los Estado
10 plt.grid(True)
11 # Muestra el histograma
12 plt.savefig('results\histograma.png')
13 plt.show()
```



### Observaciones

Definitivamente no es una distribución normal. Tiene forma de una distribución exponencial. Decae rápidamente indicando que en la mayoría del país hay sequía por gran parte del año.

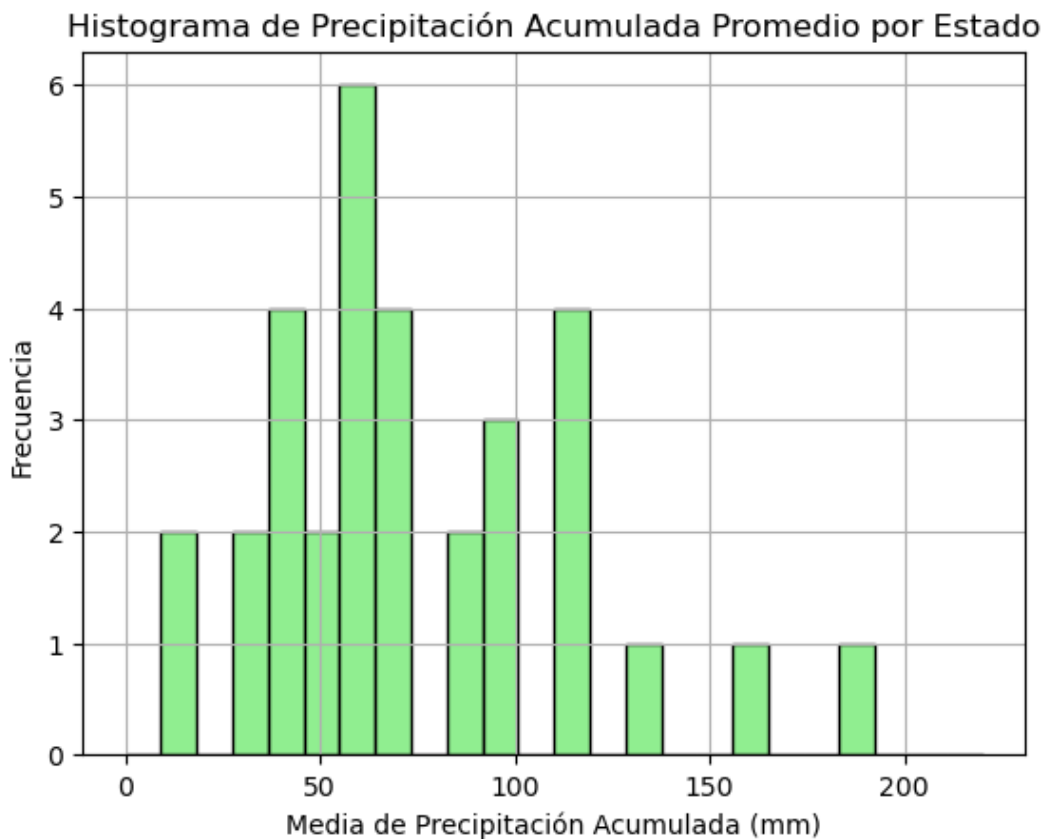
## Medidas de Tendencia Central

```
In [18]: 1 media_filas = df.iloc[: -1, 1:].mean(axis=1)
2 mediana_filas = df.iloc[: -1, 1:].median(axis=1)
```

```
In [19]: 1 media_NL = df.iloc[18, 1:].mean()
2 mediana_NL = df.iloc[18, 1:].median()
```

```
In [19]: 1 media_NL = df.iloc[18, 1:].mean()
2 mediana_NL = df.iloc[18, 1:].median()
3 var_NL = df.iloc[18, 1:].var()
4 media_NAC = df.iloc[32, 1:].mean()
5 mediana_NAC = df.iloc[32, 1:].median()
6 var_NAC = df.iloc[32, 1:].var()
```

```
In [20]: 1 plt.hist(media_filas, bins=24, color='lightgreen', edgecolor='black', range=
2 # Personaliza el histograma si es necesario
3 plt.xlabel('Media de Precipitación Acumulada (mm)')
4 plt.ylabel('Frecuencia')
5 plt.title('Histograma de Precipitación Acumulada Promedio por Estado')
6 plt.grid(True)
7 # Muestra el histograma
8 plt.savefig('results\histograma_media.png')
9 plt.show()
```



```
In [24]: 1 print("\nLa media de precipitación mensual para Nuevo León es de:",media_NL,
2 print("La varianza de precipitación mensual para Nuevo León es de:",var_NL,')
```

La media de precipitación mensual para Nuevo León es de: 52.23474492164372  
Mientras que la Nacional es de: 64.96537280701762

La varianza de precipitación mensual para Nuevo León es de: 3265.6404618333686  
Mientras que la Nacional es de: 2987.4542816189505

Nuevo León al ser un área más pequeña y sumando el factor de las condiciones climatológicas extremas, es más susceptible a variaciones aleatorias, lo que puede afectar la media y la varianza



Nuevo León al ser un área más pequeña y sumando el factor de las condiciones climatológicas extremas, es más susceptible a variaciones aleatorias, lo que puede afectar la media y la varianza de manera más pronunciada.

La mayor varianza indica que los datos en la muestra con la menor media están más dispersos alrededor de la media. Esto comprueba las altas variaciones extremas de un clima semiseco extremo. Aunque cuenta con regiones que registran lluvias anuales mayores de 800 mm. La media general anual del Estado oscila entre 300 y 600 mm.

In [ ]:

1