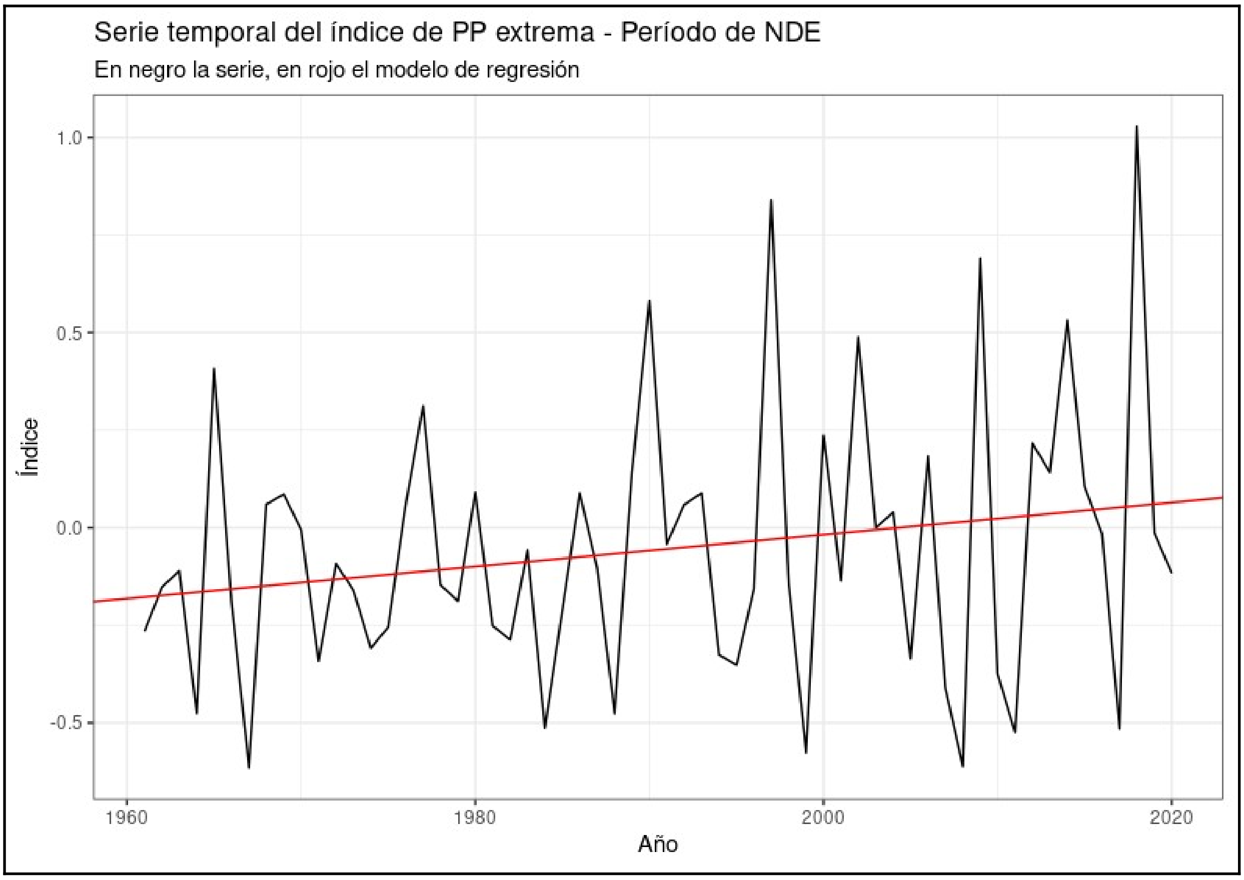
Lorenzo Girotti

1er Parcial Clima II – Segunda Fecha

Ejercicio 2

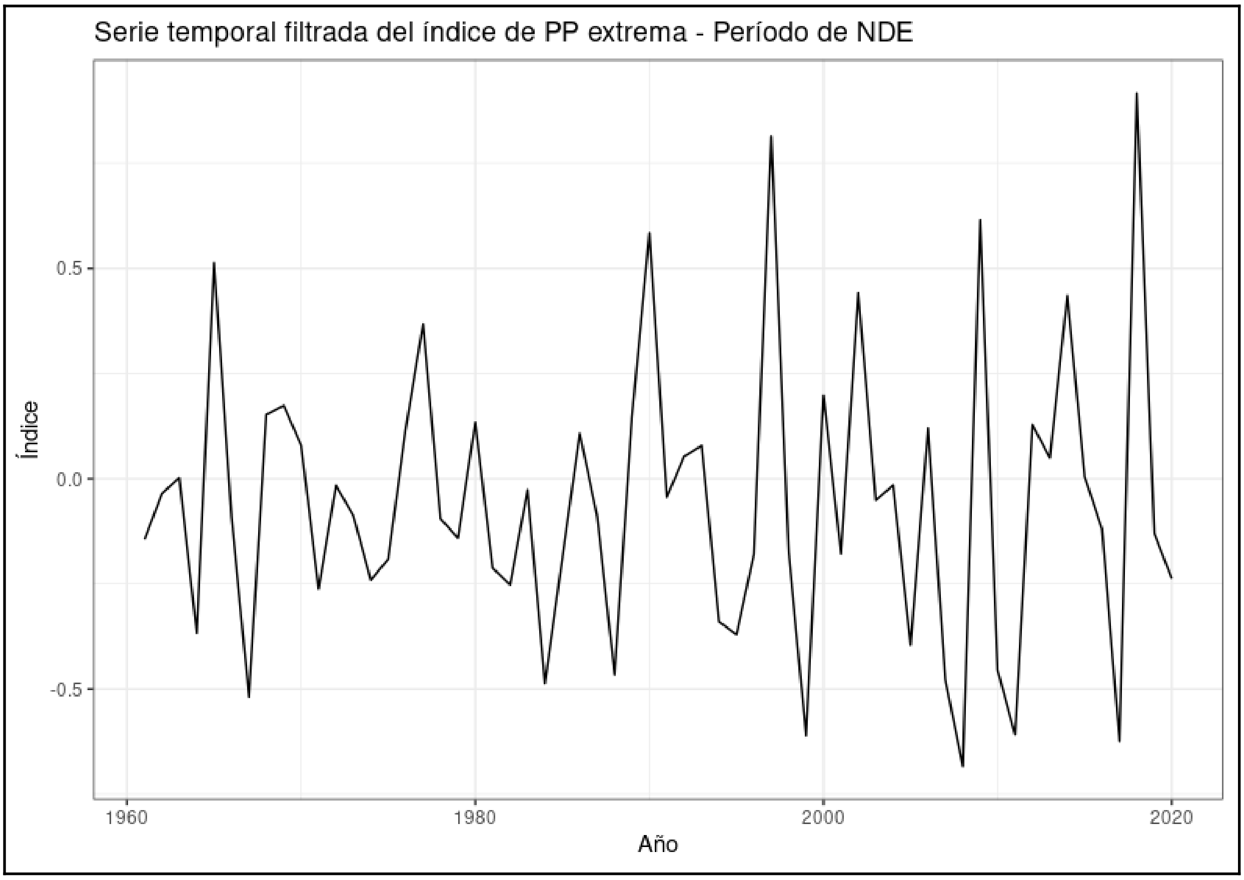
a) Graficamos la serie y superponemos el modelo de regresión lineal para identificar la existencia de tendencia.



Observamos una leve inclinación de la recta, por ende podemos interpretar, de ser estadísticamente significativa, a la pendiente de la misma como la tendencia de la variable. Para poder determinar si la tendencia es o no estadísticamente significativa, necesito establecer un test estadístico. En este caso, utilizo un test de correlación entre los valores del índice y los años, ya que en caso de que exista una correlación distinta de cero, ésto establece una dependencia estadística de las variables (el índice varía con el tiempo).
Para el test necesito establecer una hipótesis nula H_0 y un nivel de significancia α.
Mi hipótesis nula es: "La correlación entre las variables es cero"; mi hipótesis alternativa es que "La correlación entre las variables es positiva"; mi nivel de significancia es $\alpha = 0.1$.
Lo anterior quiere decir que si al realizar el test, obtengo un *p-value* < α entonces puedo rechazar mi hipótesis nula en favor de la alternativa, con una confianza del 90%.

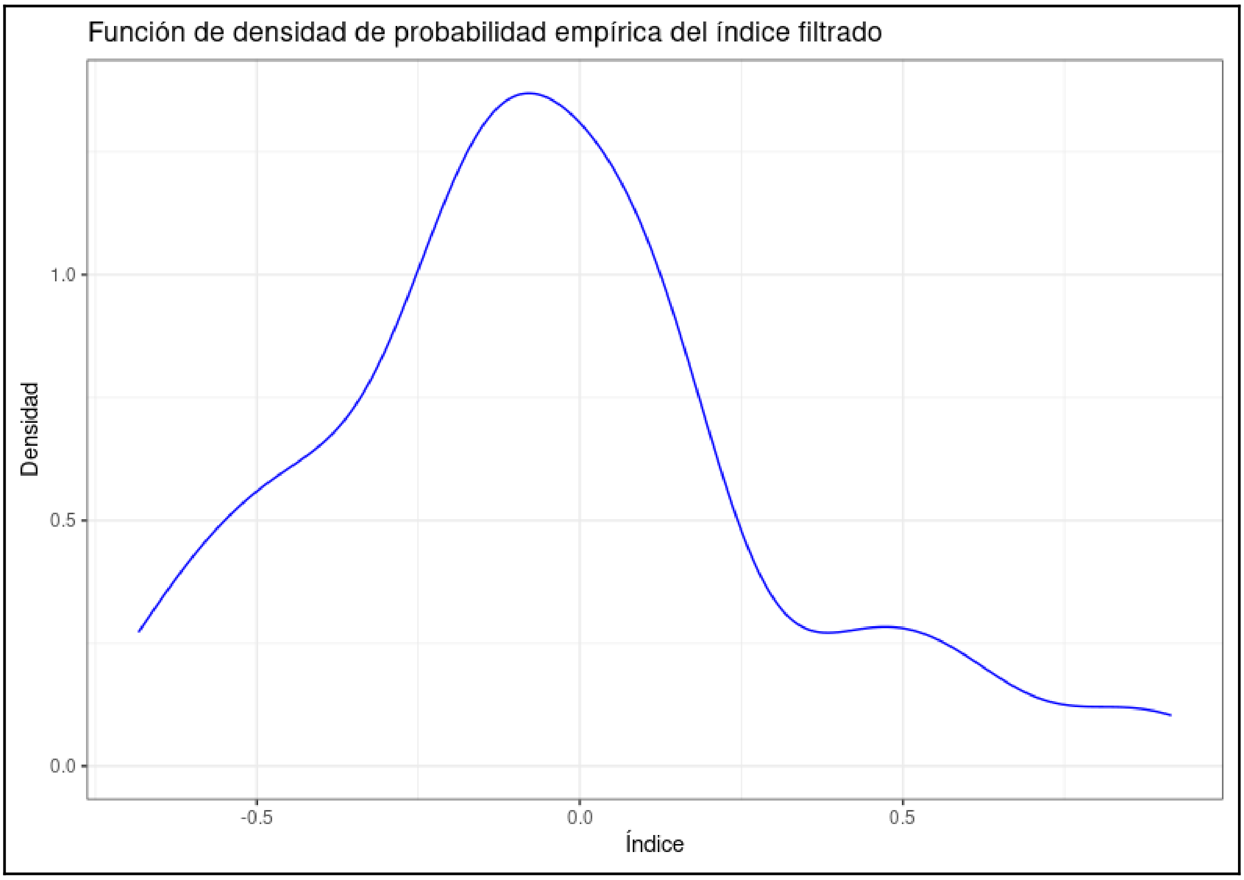
Al realizar el test, obtenemos un *p-value = 0.057* aproximadamente. Entonces puedo rechazar la hipótesis nula de que la correlación es cero, en favor de la alternativa: que la correlación es positiva con una confianza del 90%.

Al existir una tendencia positiva, mi serie deja de ser estacionaria, lo cual implica que a medida que transcurre el tiempo, los datos obtenidos son de poblaciones diferentes; puesto que la población de la que se extrae la muestra no es la misma en el año 1960 que en el año 2000. Por lo tanto, para poder analizar los datos, debo filtrar la tendencia reemplazando a las observaciones por su media más los residuos estimados por el modelo.



De esta manera aseguramos una estacionariedad de la serie y podremos realizar inferencias sobre la población a través de las muestras que disponemos.

b) Graficamos la función de densidad de probabilidad empírica de la variable:



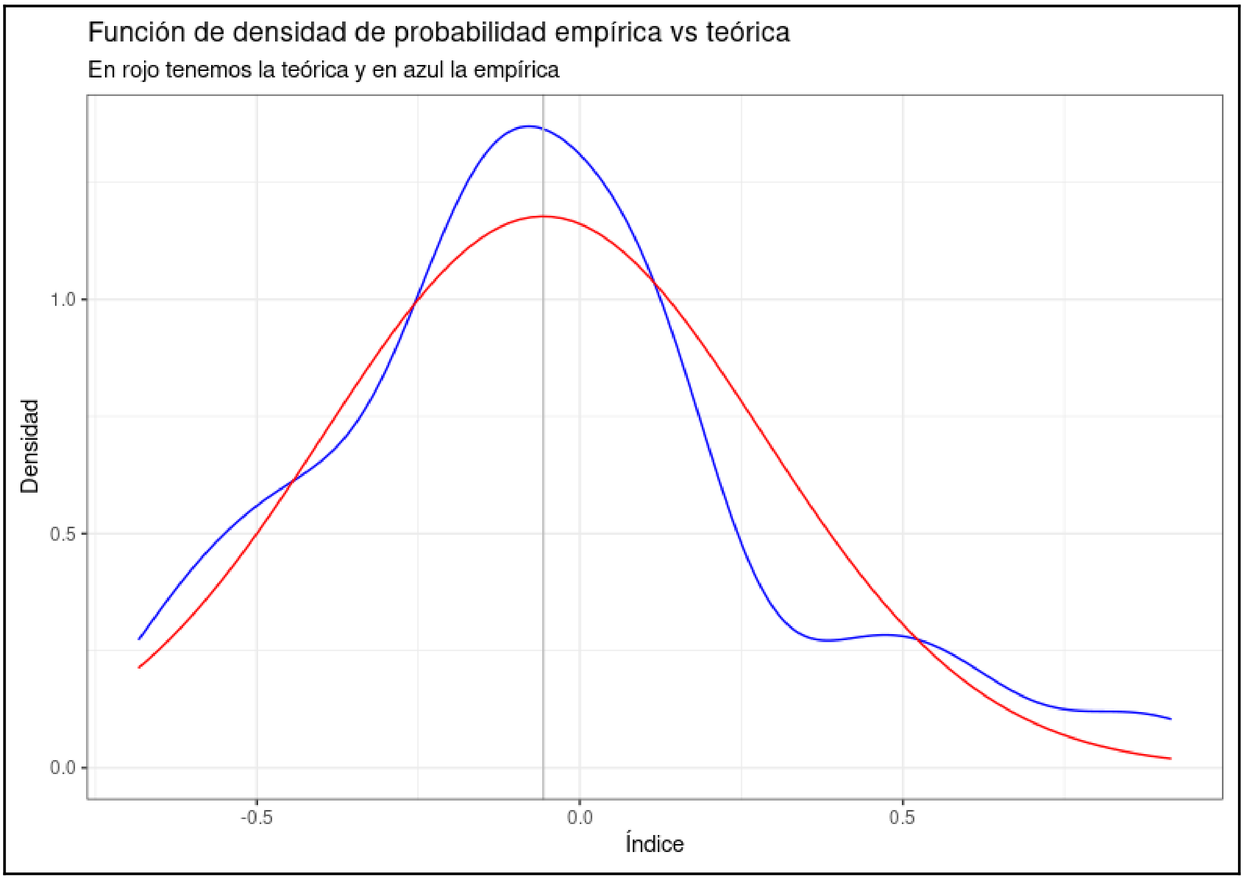
Podemos observar que se asemeja a una normal, con una media cercana al 0, aunque la cola derecha presenta leves perturbaciones y no genera una simetría perfecta.

c) Para ajustar la distribución empírica a una normal teórica, necesitamos los parámetros de *media* y *desvío estándar*. Para obtener la función que mejor ajuste, utilizamos el método de máxima verosimilitud, que se basa en identificar qué distribución teórica es la más probable de que represente mi conjunto de datos: esto lo hace variando los parámetros *media* y *desvío* que caracterizan a la distribución. Una vez que obtiene aquella que cumple con el requisito de ser la que más probabilidad tiene de contener a todos mis datos (la que maximiza la función de verosimilitud), extrae qué parámetros fueron los utilizados para ello.

Utilizo la función *fitdist()* que me permite ajustar un vector de observaciones a una distribución determinada cuyos parámetros son estimdos por el método preferido. En mi caso queda

```
fitdist(data=datos, distr=”norm”, method=”mle”)
```

El ajuste lo vemos cuando graficamos ambas curvas juntas:



Podemos ver que el ajuste, a simple vista, es bueno.

d) Teniendo en cuenta que la serie sigue una distribución normal (luego de ser filtrada) y que ésta depende de parámetros que no son robustos ni resistentes, si quiero evaluar la correlación con los potenciales forzantes debo utilizar un coeficiente de correlación que sea robusto y resistente; como por ejemplo el coefeiciente de correlación de Spearman, que correlaciona los ordinales de ambas series.
Esto me beneficia para poder vincular los valores más altos y bajos de los forzantes con los valores más altos y bajos del índice de PP, sin preocuparme de qué distribución siguen los forzantes.

e) Realizo un test de correlación utilizando el coeficiente de Spearman, bajo la hipótesis nula de que la correlación entre el índice de precipitación y los distintos índices climáticos es cero; con un nivel de significancia de α = 0.1.
Como los indices están desde el año 1991-2020, me quedo sólo con esas fechas en mi serie.
Una vez realizados los tests, obtengo la siguiente tabla:

	Indice	p-value
1	DMI	1.462164e-01
2	Nino12	1.436959e-04
3	Nino34	1.626812e-05
4	Nino4	7.646850e-05
5	SAM	1.196330e-01

Aquí podemos ver que según mi α = 0.1, los únicos índices que pueden rechazar la hipótesis nula son los relacionados al fenómeno de El Niño. Lo cual es esperable, ya que sabemos que la fase positiva y negativa de esta oscilación se relaciona de manera directa con las precipitaciones en el NEA.