


IPC GUATEMALA

PROYECTO MINERIA DE DATOS



Diego Leiva	21752
Maria Ramirez	21342
Gustavo González	21438
Gabriel García	21352
Pablo Orellana	21970





METODO UTILIZADO

METODOLOGÍA CON SPLIT ALEATORIO

PARA LA PREDICCIÓN DEL IPC NACIONAL, SE UTILIZÓ UN SPLIT ALEATORIO, DIVIDIENDO LOS DATOS EN UN 80% PARA ENTRENAMIENTO Y UN 20% PARA PRUEBA. ESTE MÉTODO ASEGURA QUE LOS DATOS SE DISTRIBUYAN DE MANERA ALEATORIA EN AMBOS CONJUNTOS.

METODOLOGÍA CON SPLIT ALEATORIO

PARA CONSIDERAR LA DEPENDENCIA TEMPORAL DE LOS DATOS, SE REALIZÓ UN SPLIT BASADO EN EL TIEMPO, DONDE LOS DATOS SE ORDENARON CRONOLÓGICAMENTE Y SE DIVIDIERON EN UN 80% PARA ENTRENAMIENTO Y UN 20% PARA PRUEBA. ESTE MÉTODO ASEGURA QUE EL MODELO SE ENTRENE CON DATOS PASADOS Y SE PRUEBE CON DATOS FUTUROS.

Selección de la variable de respuesta


LA VARIABLE RESPUESTA SELECCIONADA ES EL ÍNDICE DE PRECIOS AL CONSUMIDOR (IPC) TANTO A NIVEL NACIONAL COMO REGIONAL. LA ELECCIÓN DE ESTA VARIABLE SE DEBE A SU IMPORTANCIA ECONÓMICA, YA QUE EL IPC MIDE LA VARIACIÓN DE LOS PRECIOS DE UNA CANASTA DE BIENES Y SERVICIOS REPRESENTATIVA DEL CONSUMO DE LOS HOGARES.





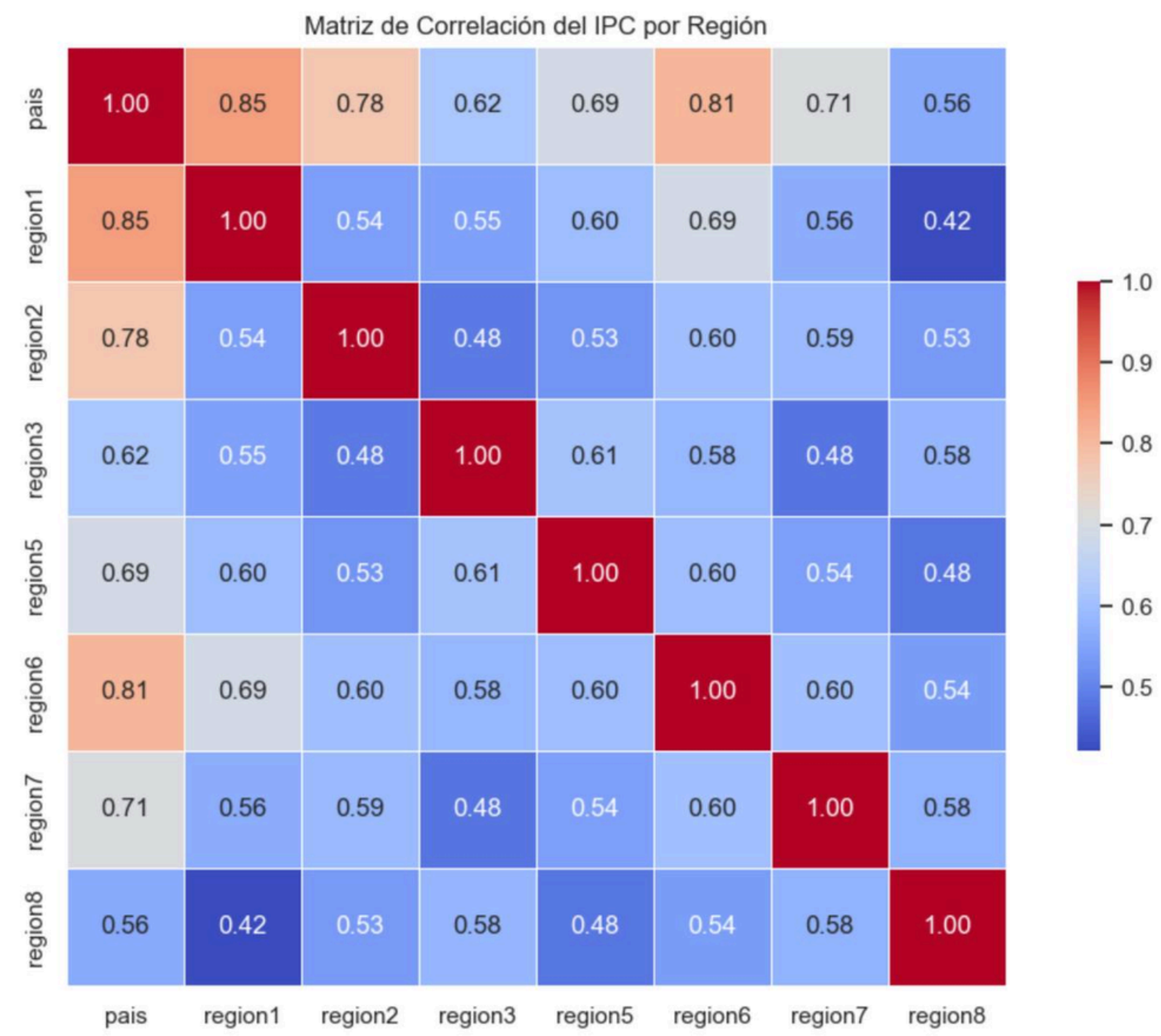
Regresión Lineal

SE APLICARON MODELOS DE REGRESIÓN LINEAL TANTO PARA EL IPC NACIONAL COMO PARA CADA UNA DE LAS REGIONES. LA REGRESIÓN LINEAL FUE ELEGIDA DEBIDO A SU SIMPLICIDAD Y CAPACIDAD DE INTERPRETACIÓN.



Transformación de Datos

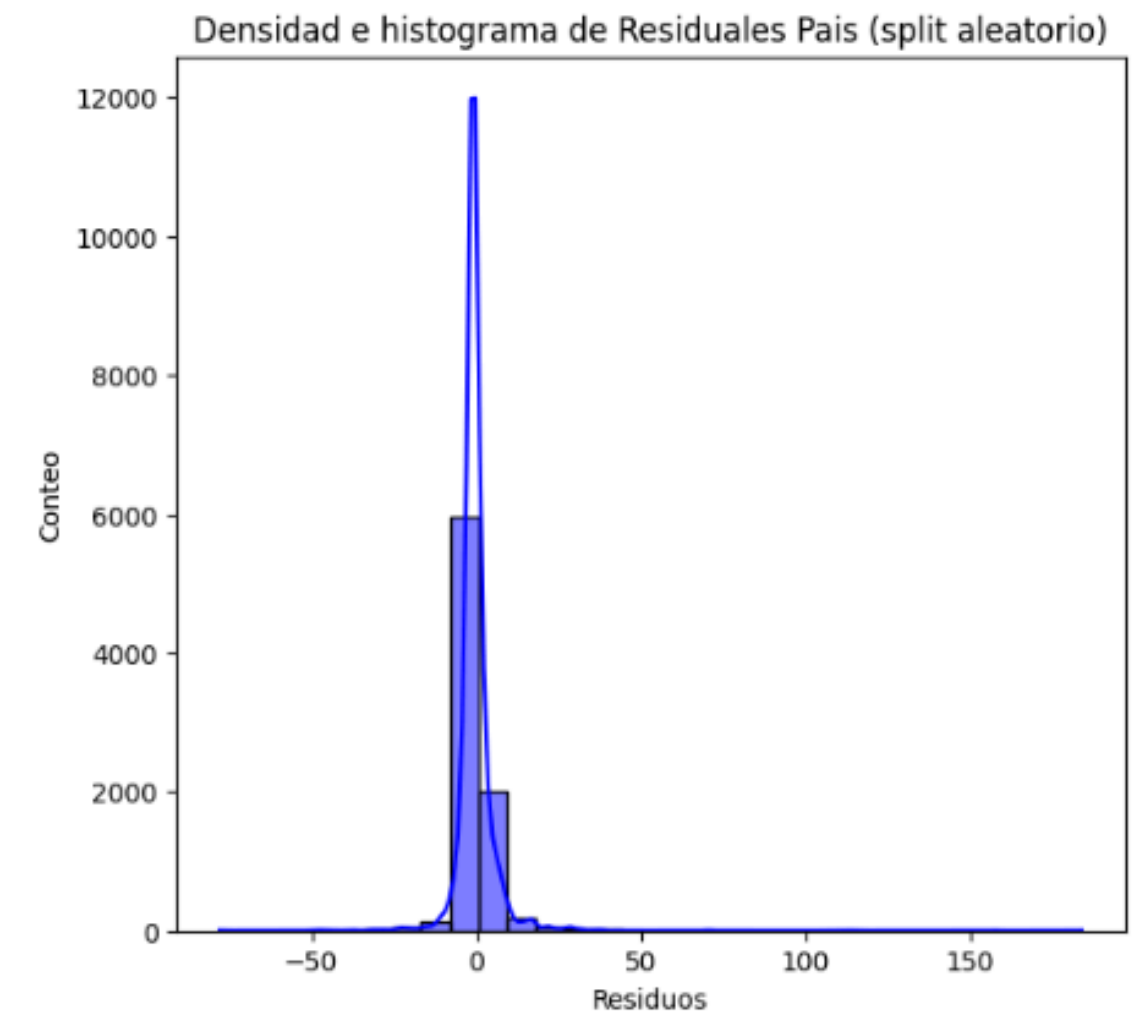
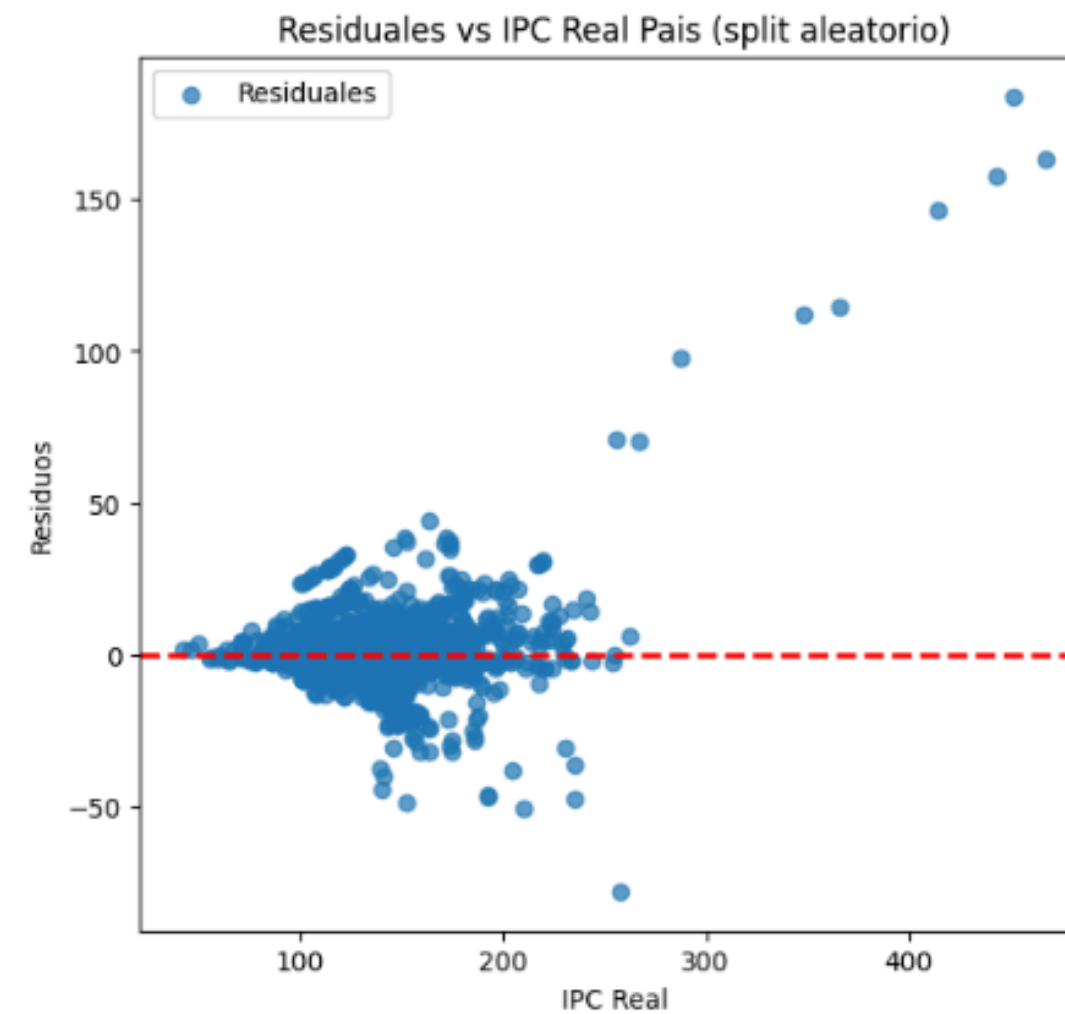
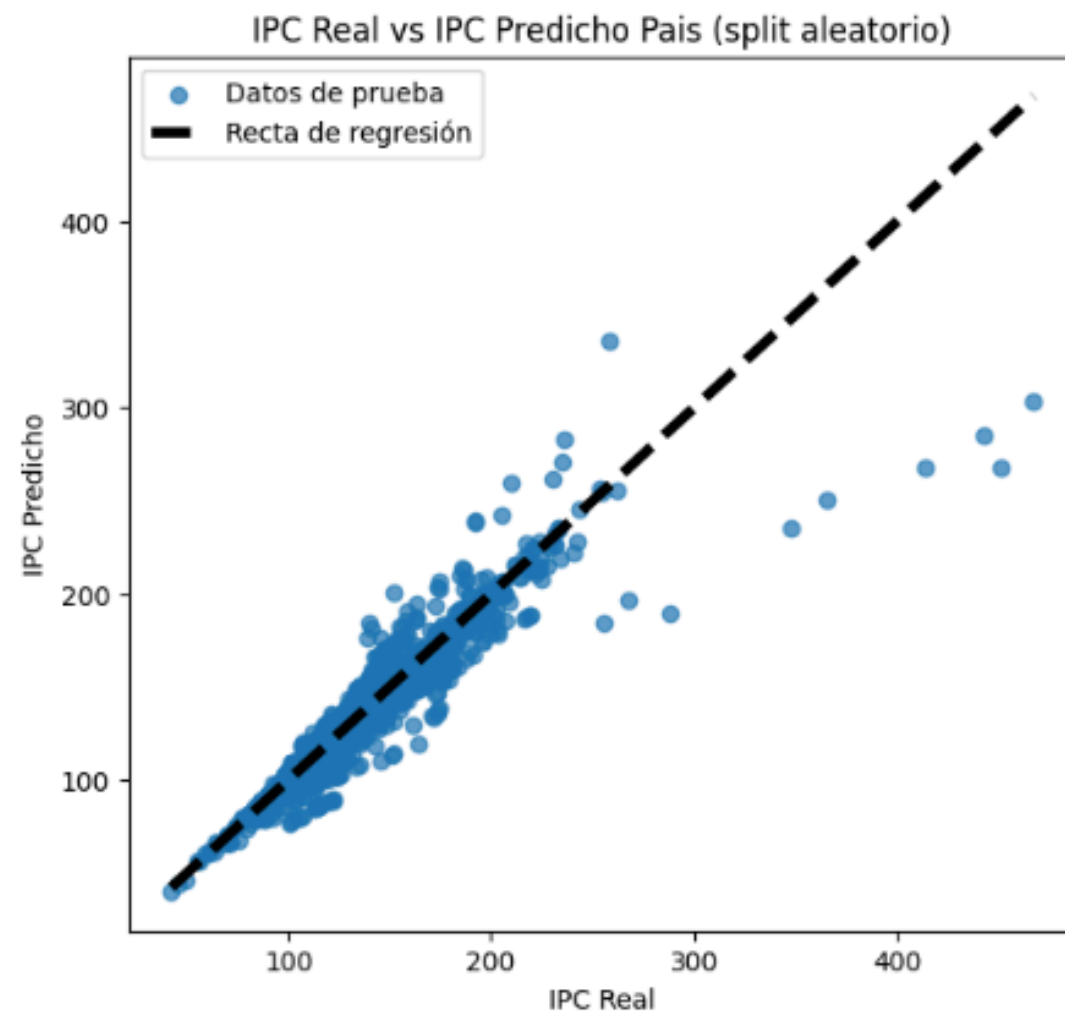
SE ENCONTRÓ QUE LOS VALORES DE IPC POR REGIÓN NO ESTABAN ALTAMENTE CORRELACIONADOS, LO QUE INDICABA QUE LAS REGIONES NO ERAN BUENAS VARIABLES PREDICTORAS DEL IPC GENERAL DE UNA REGIÓN ESPECÍFICA. SE TRANSFORMARON LOS DATOS UTILIZANDO LAS CATEGORÍAS DE PRODUCTOS COMO VARIABLES PREDICTORAS.



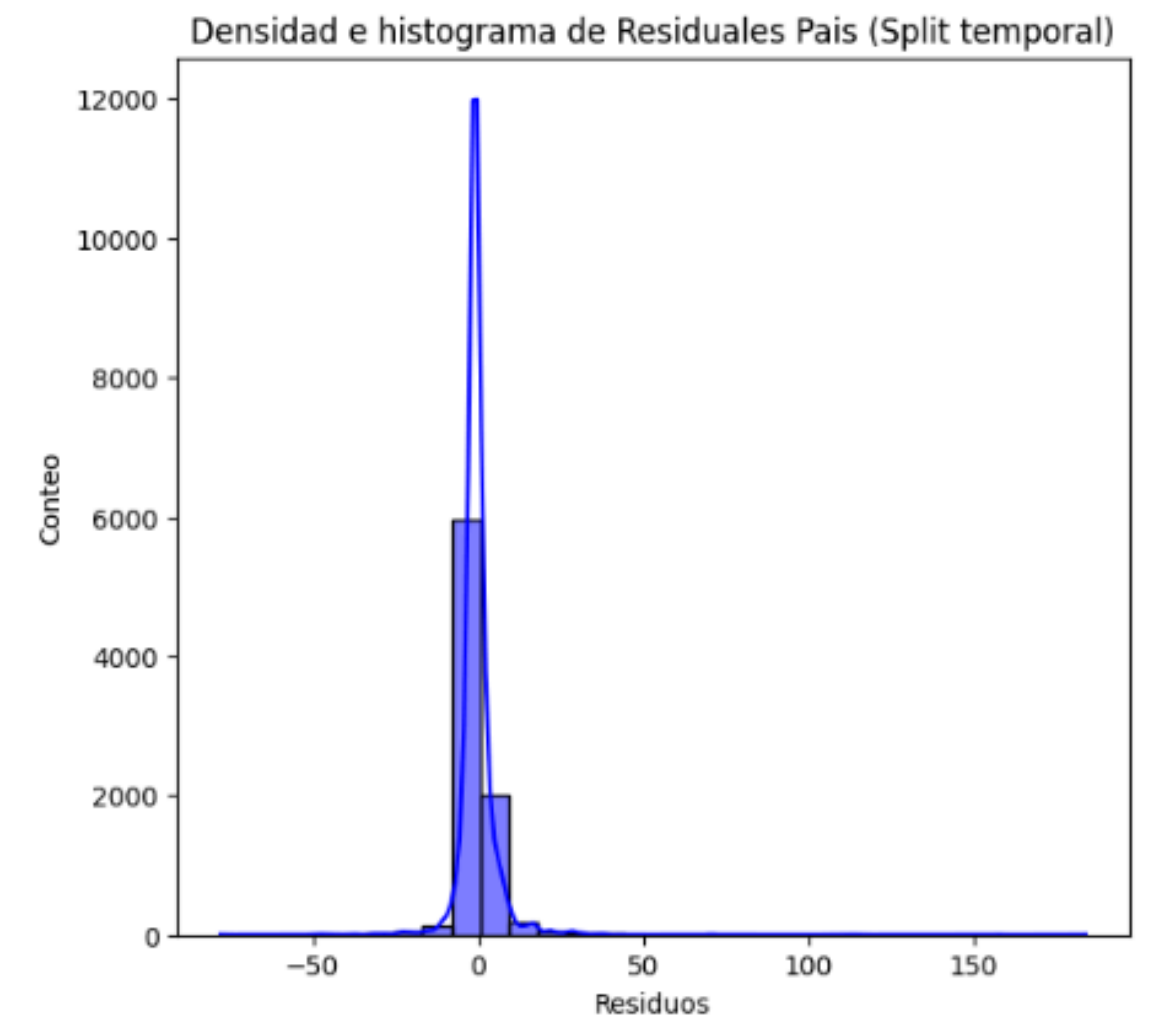
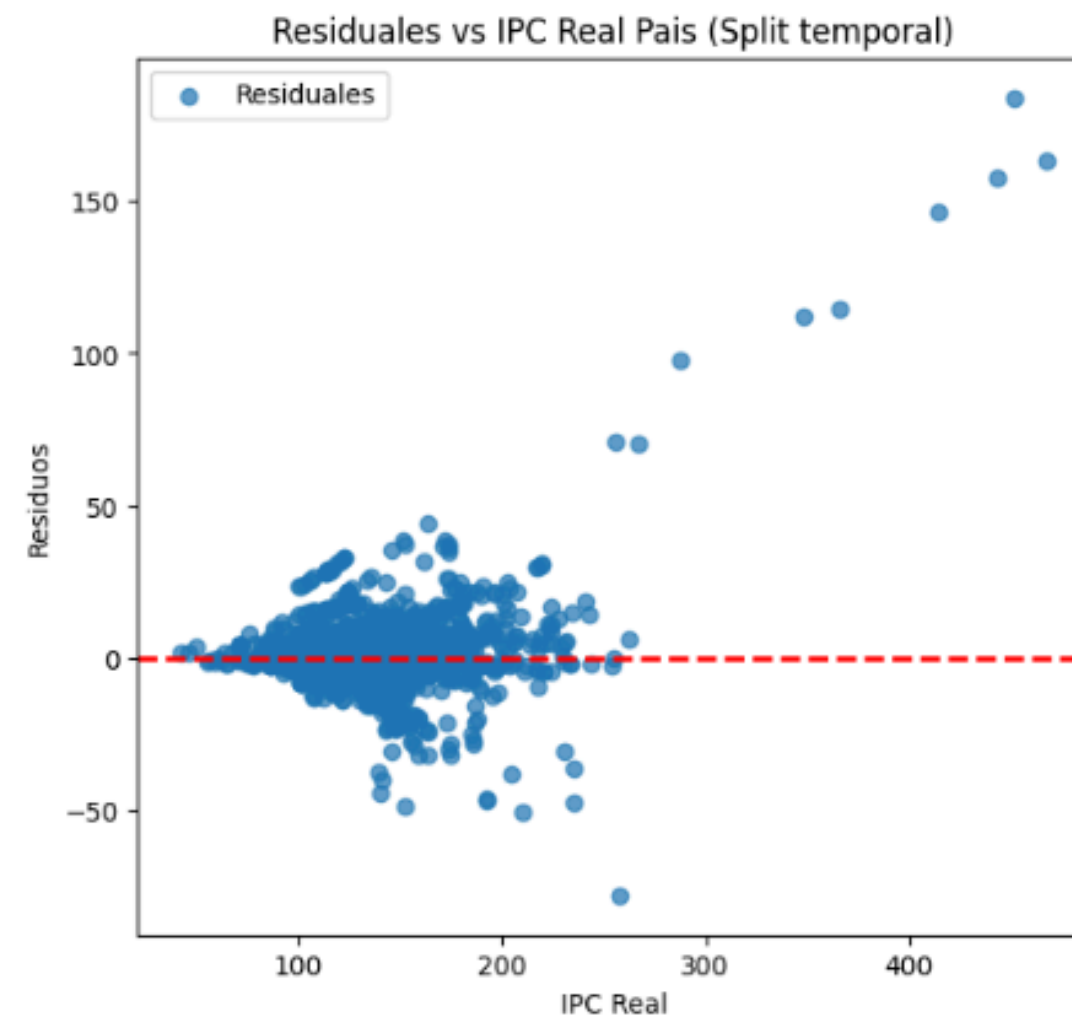
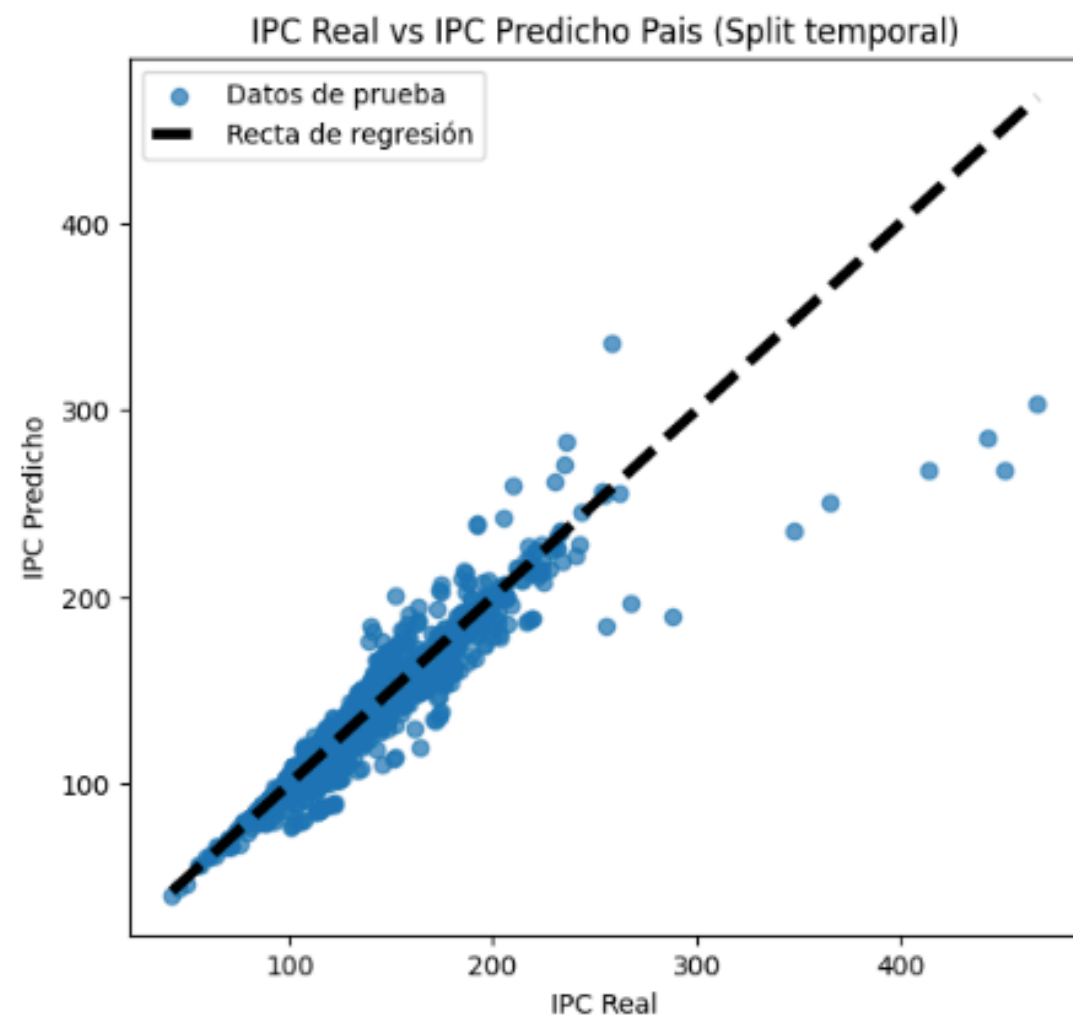
Resultados



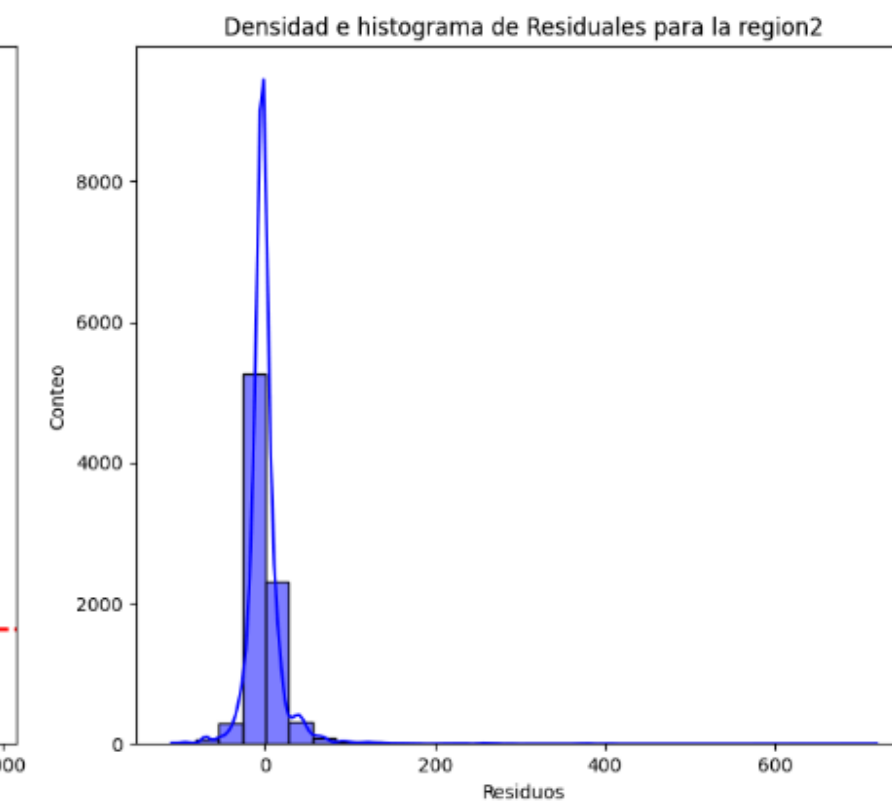
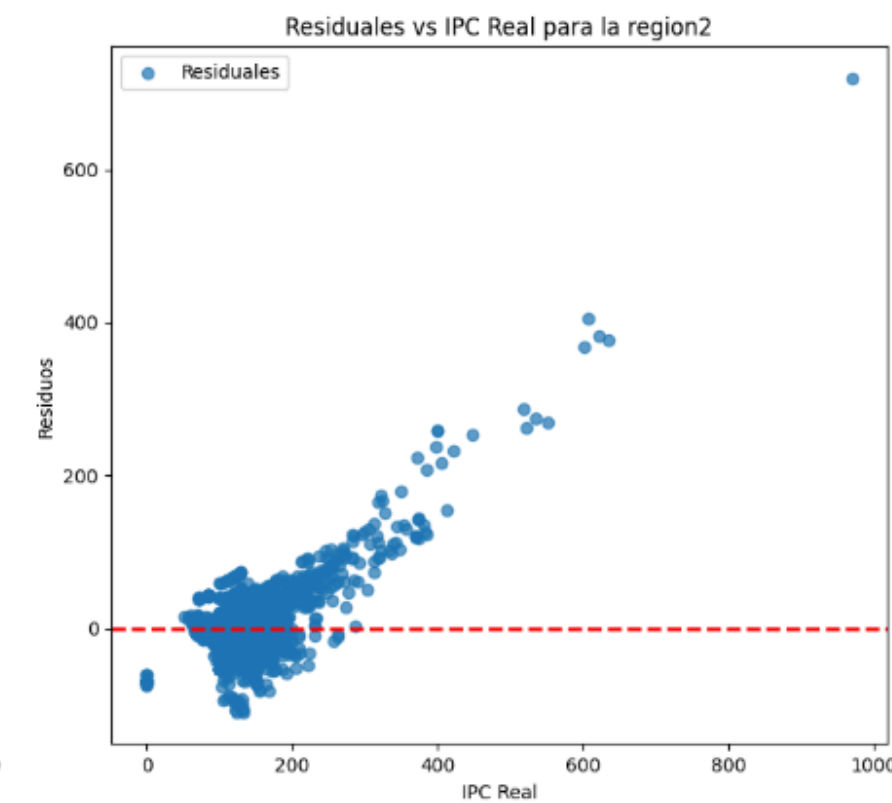
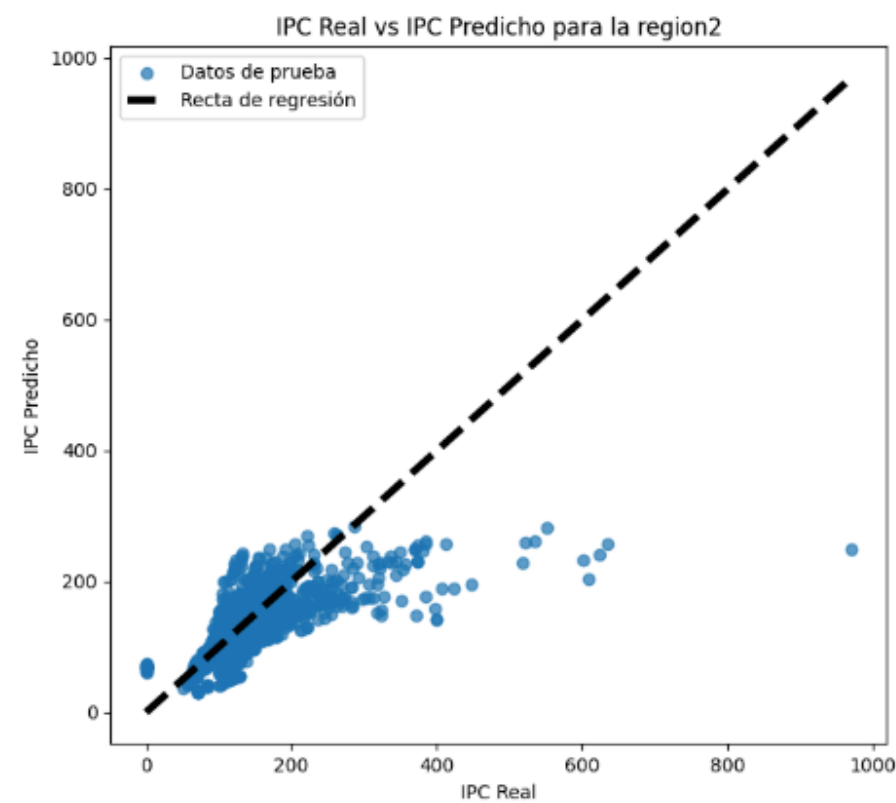
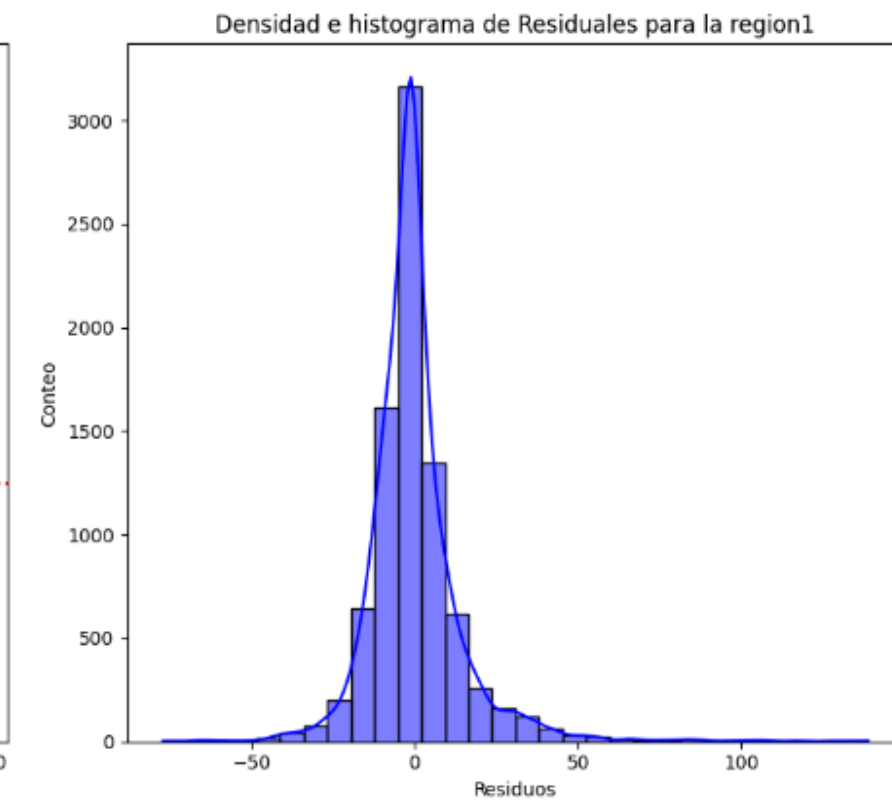
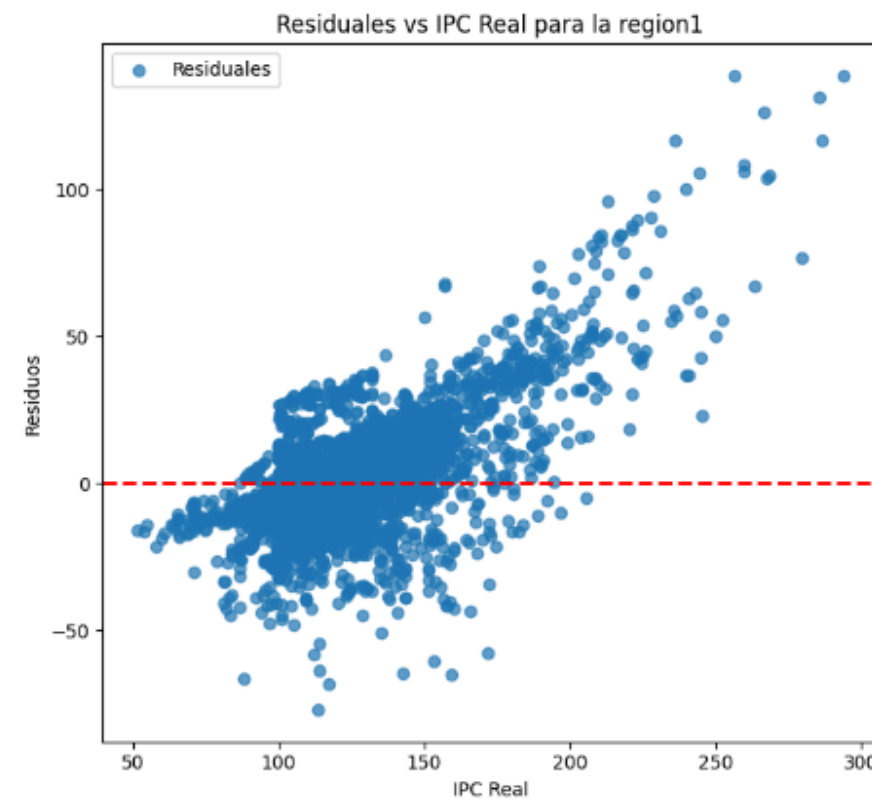
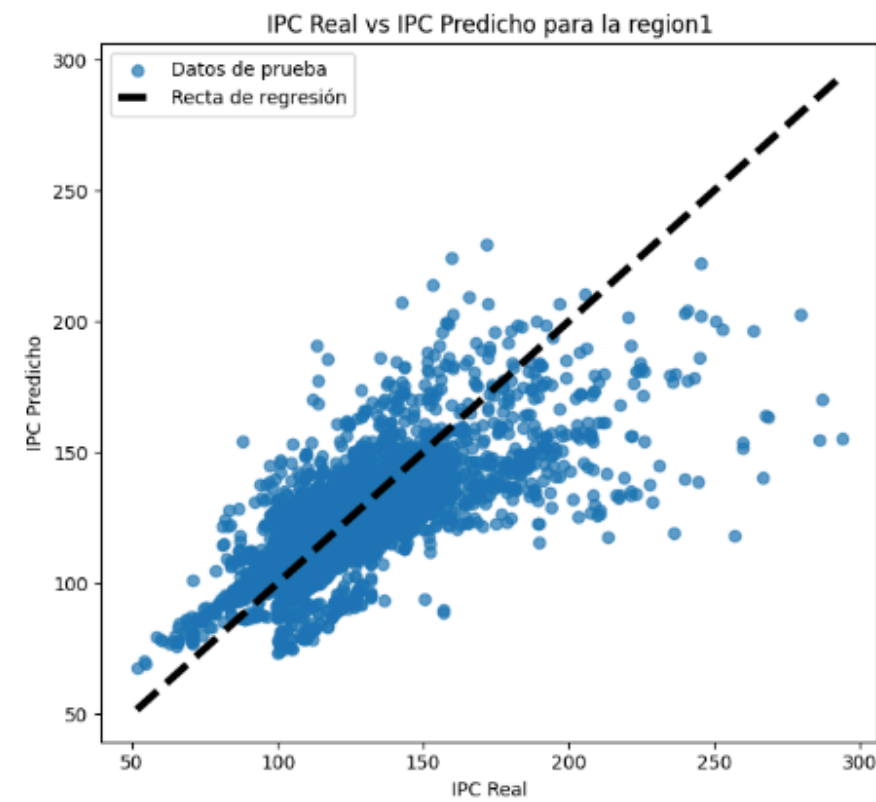
Resultados con Split Aleatorio para el IPC Nacional



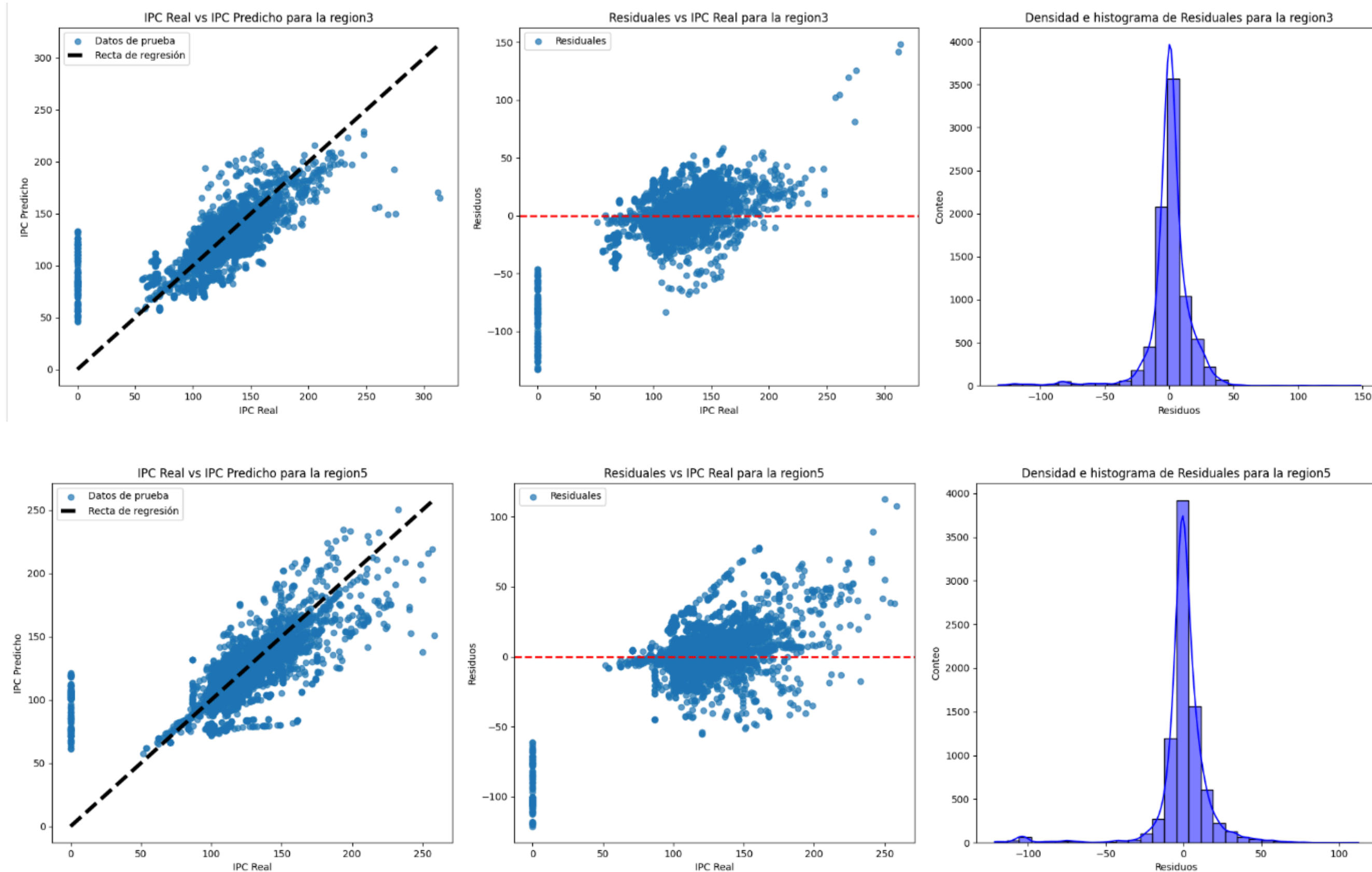
Resultados con Split Basado en Tiempo para el IPC Nacional



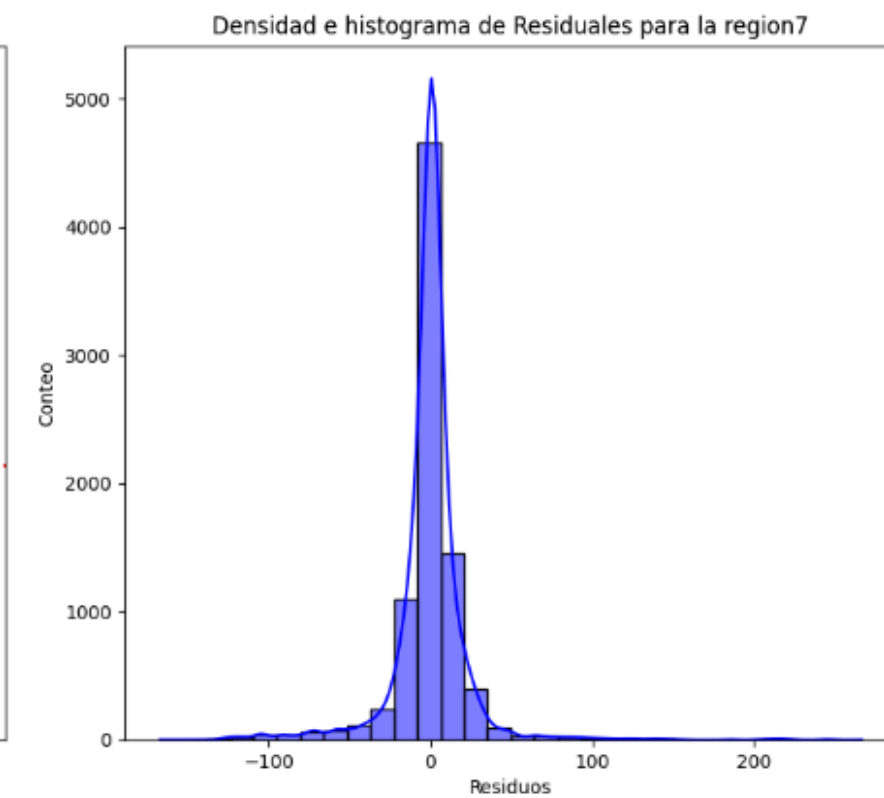
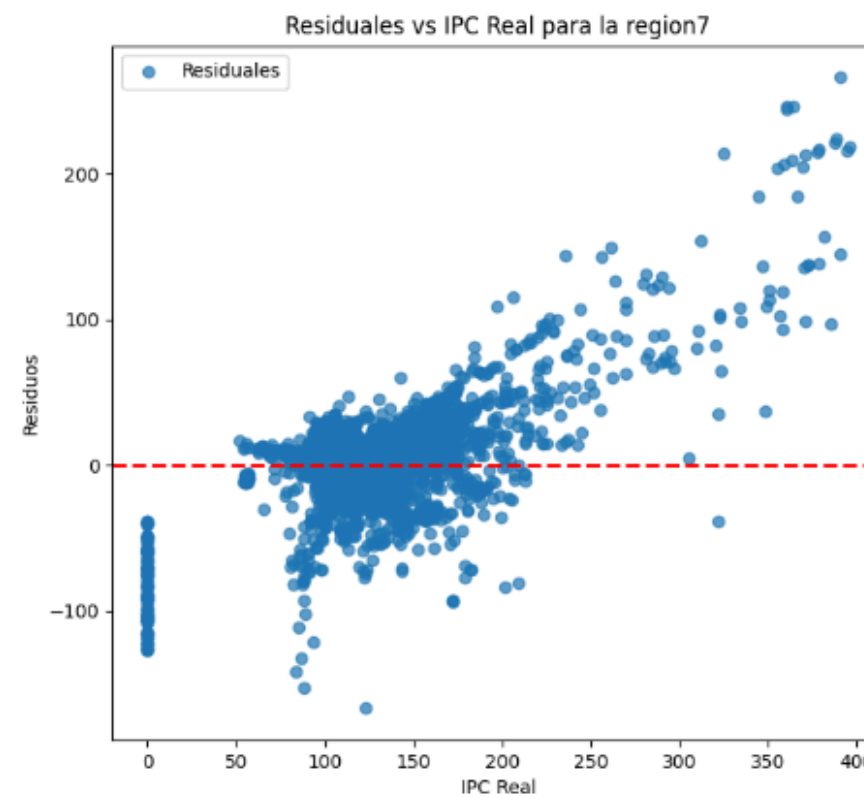
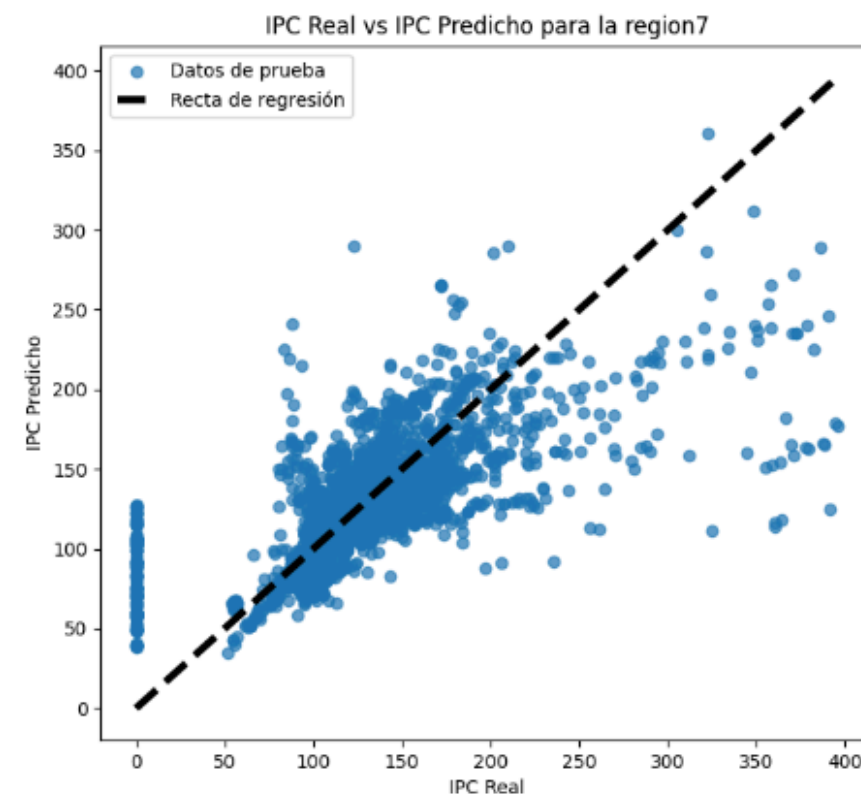
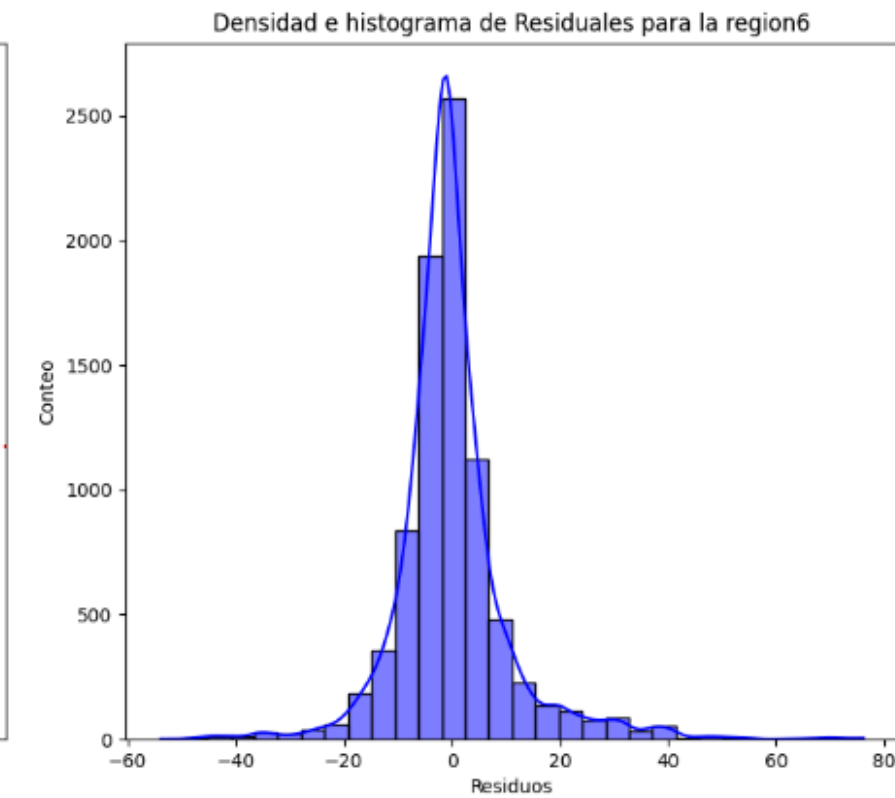
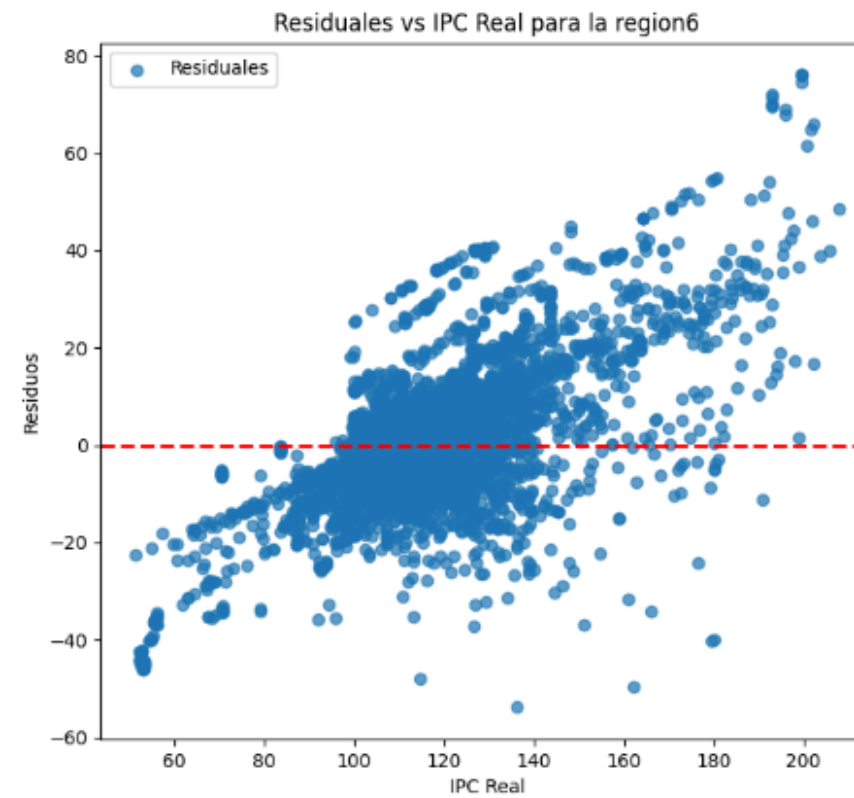
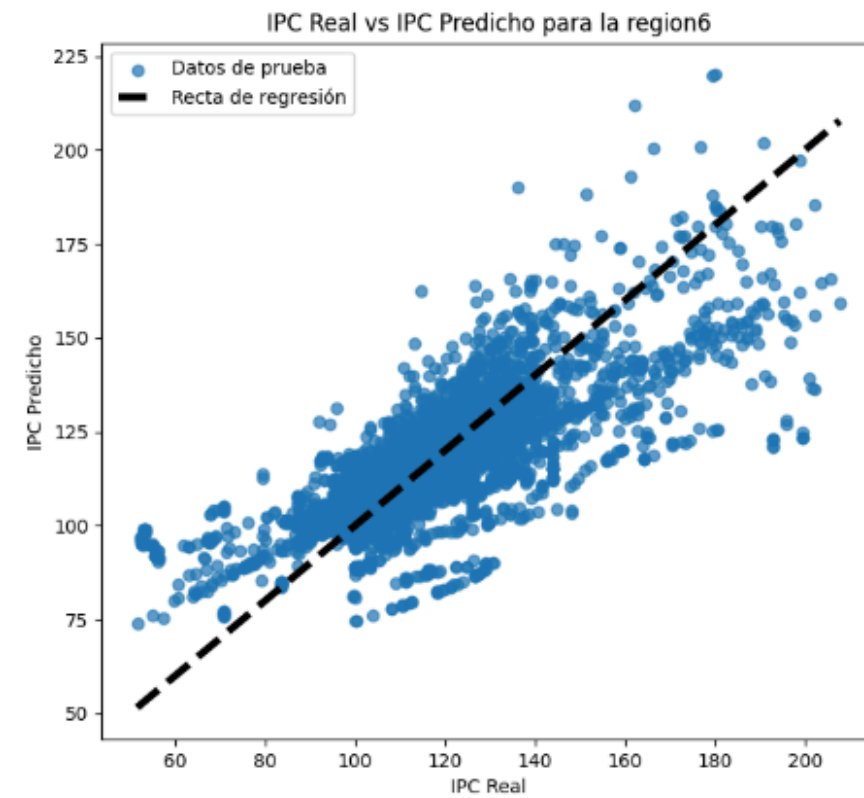
Resultados para el IPC por cada región del país (Región 1 y 2)



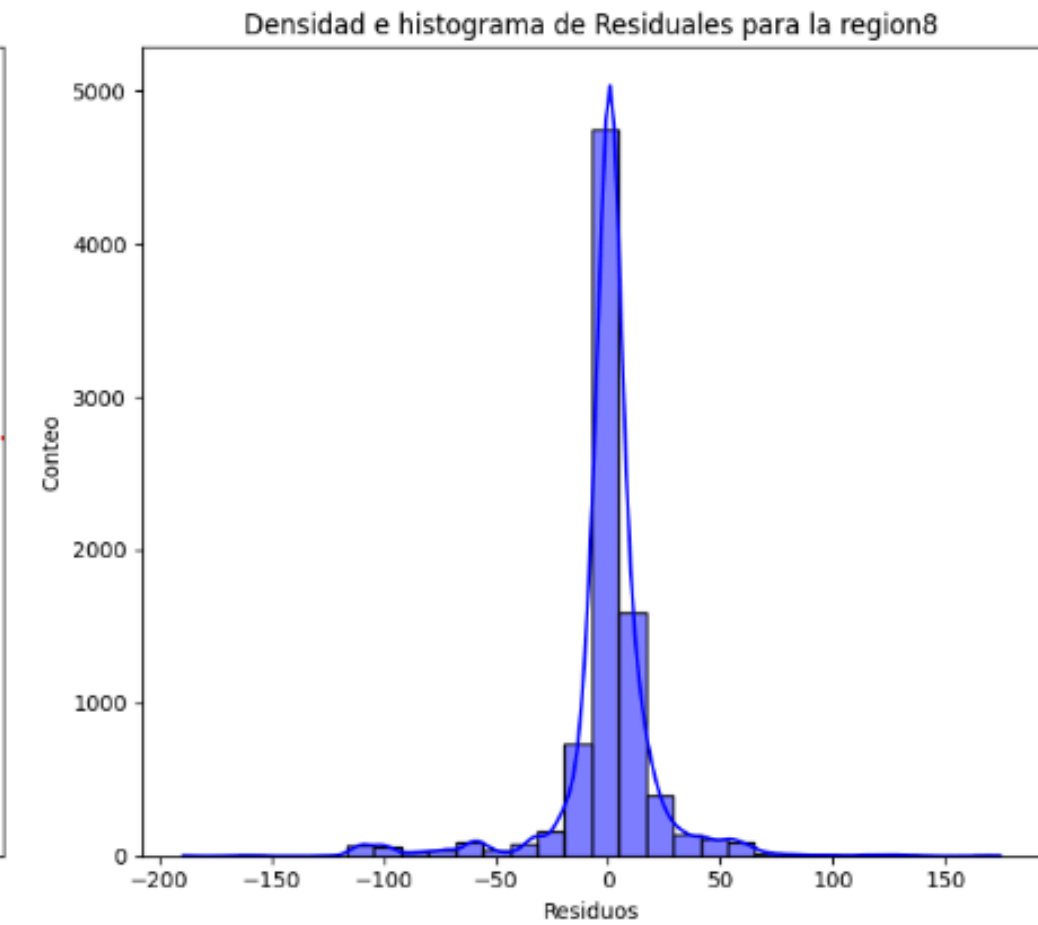
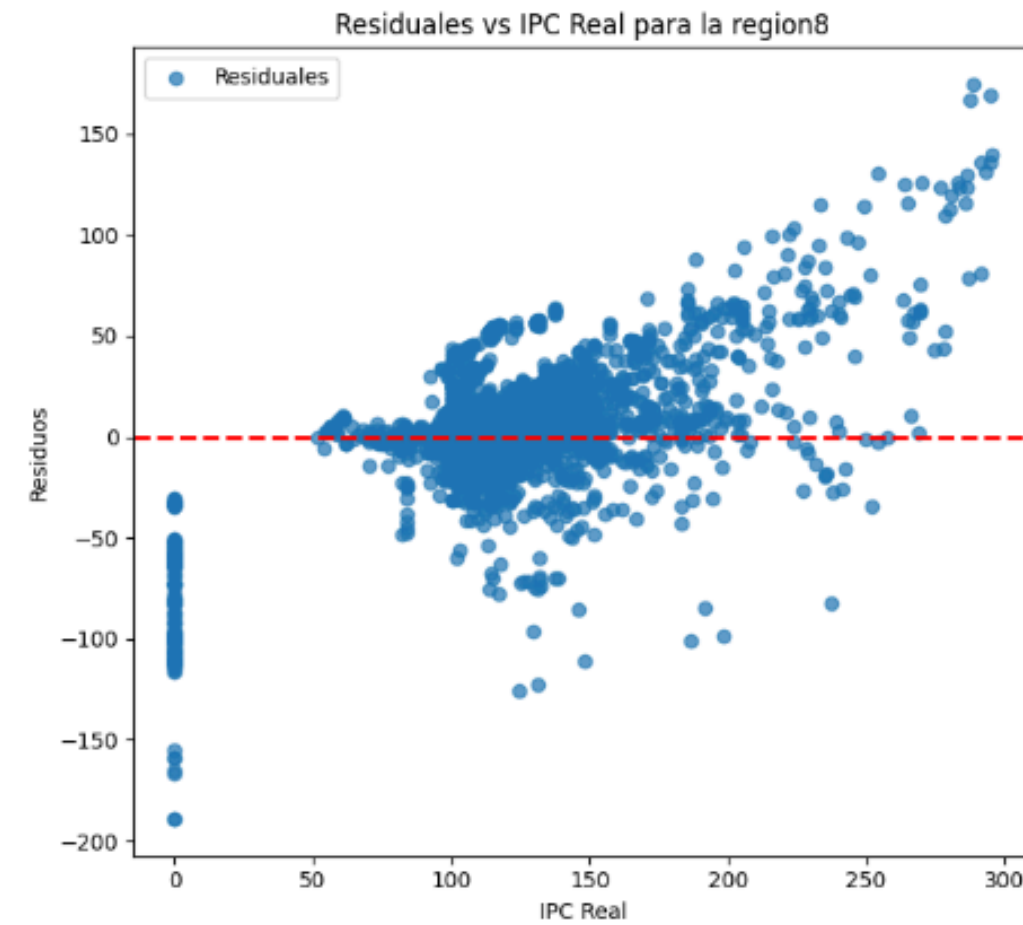
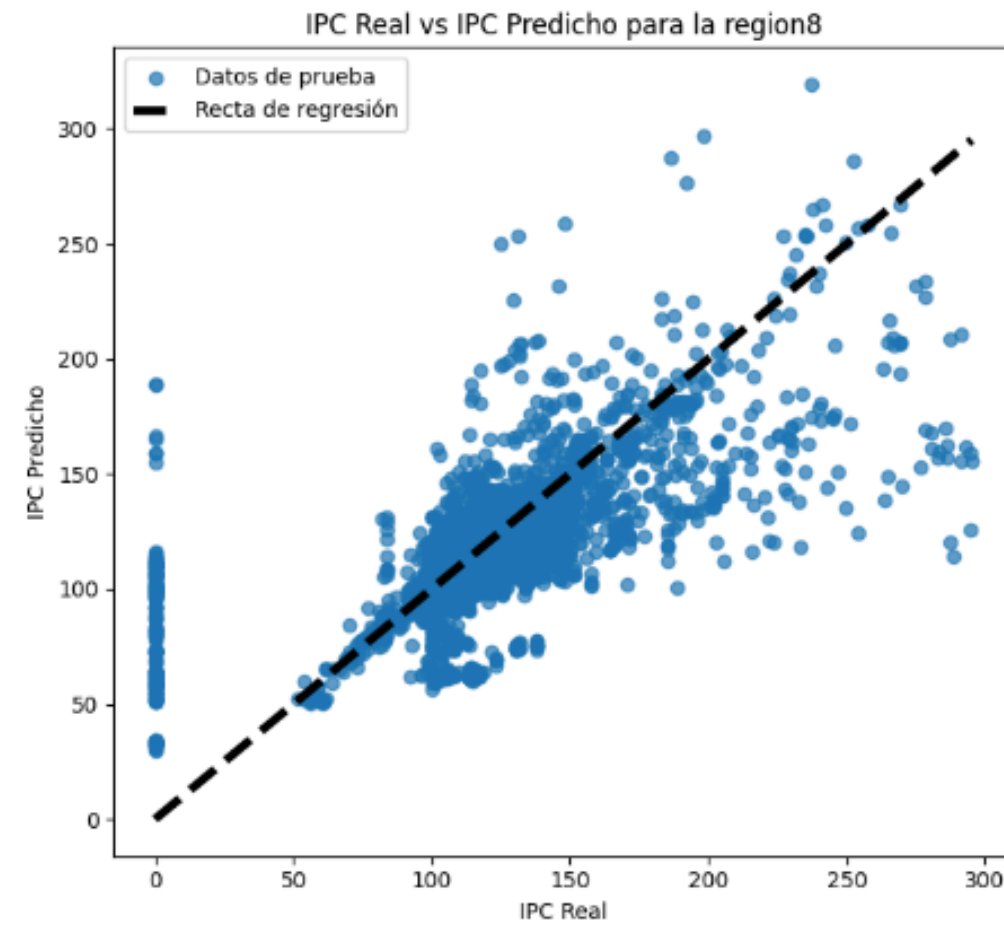
Resultados para el IPC por cada región del país (Región 3 y 5)



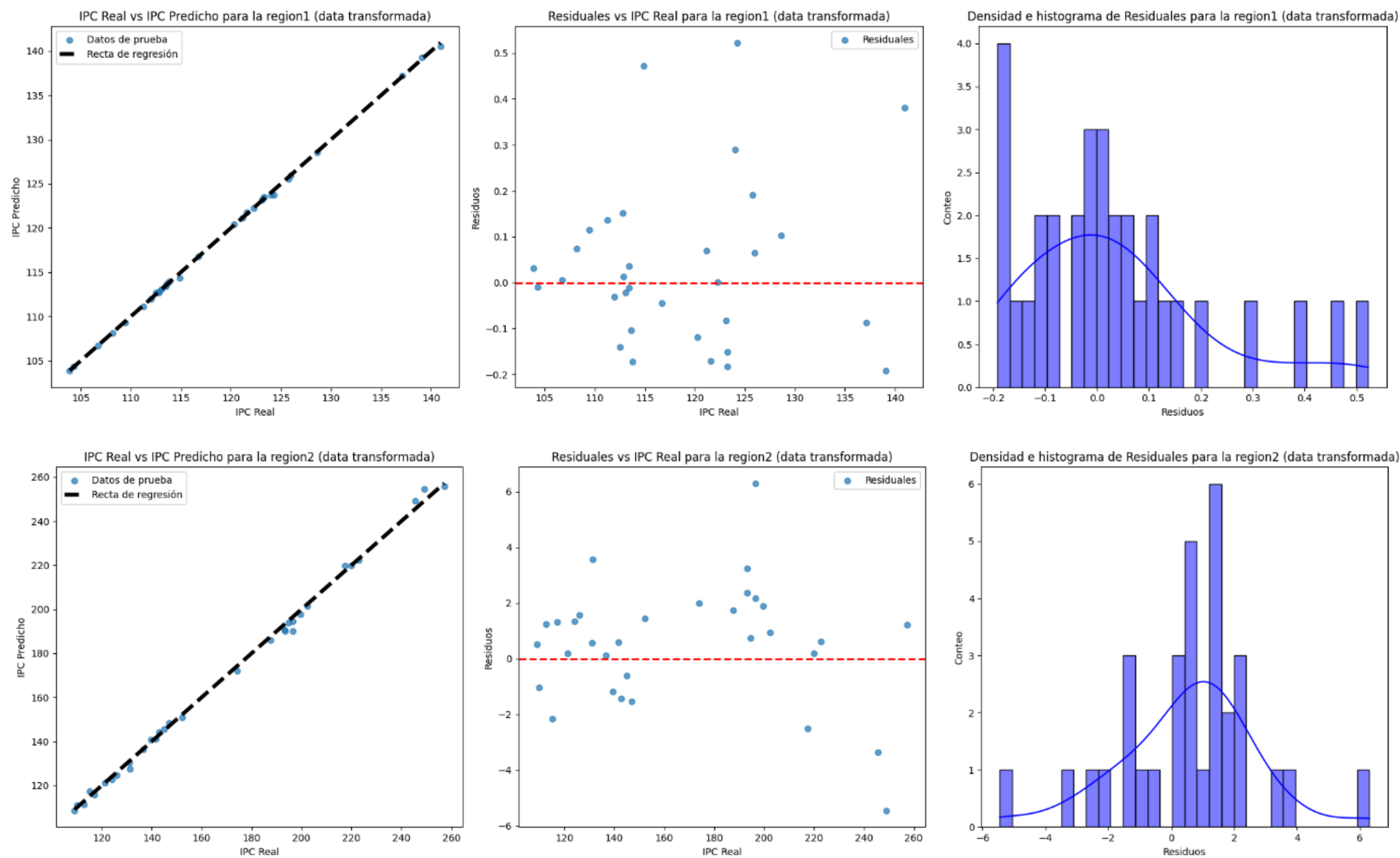
Resultados para el IPC por cada región del país (Región 6 y 7)



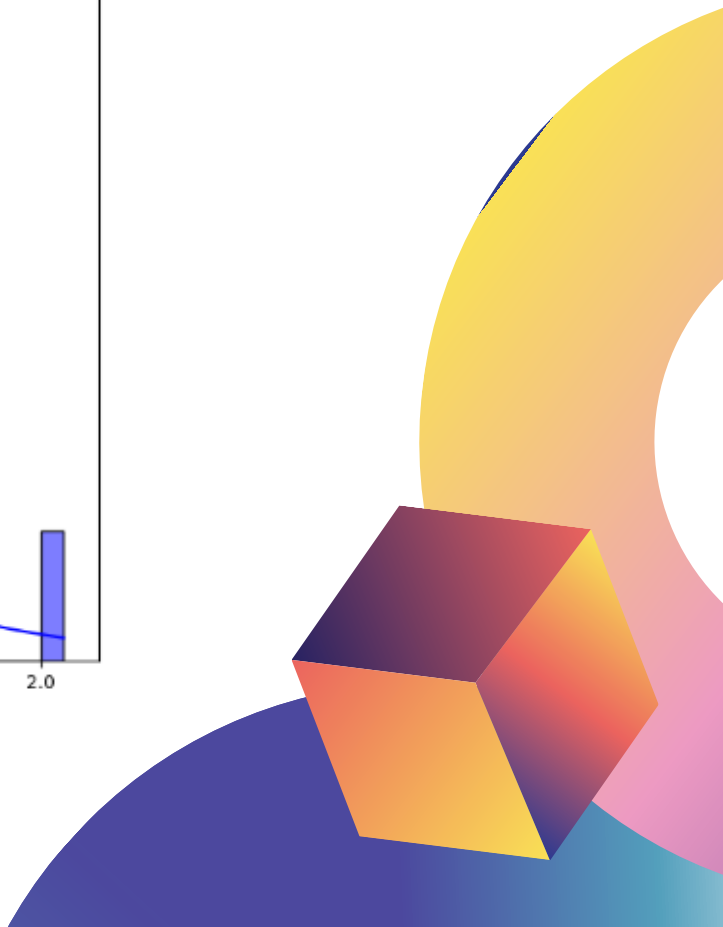
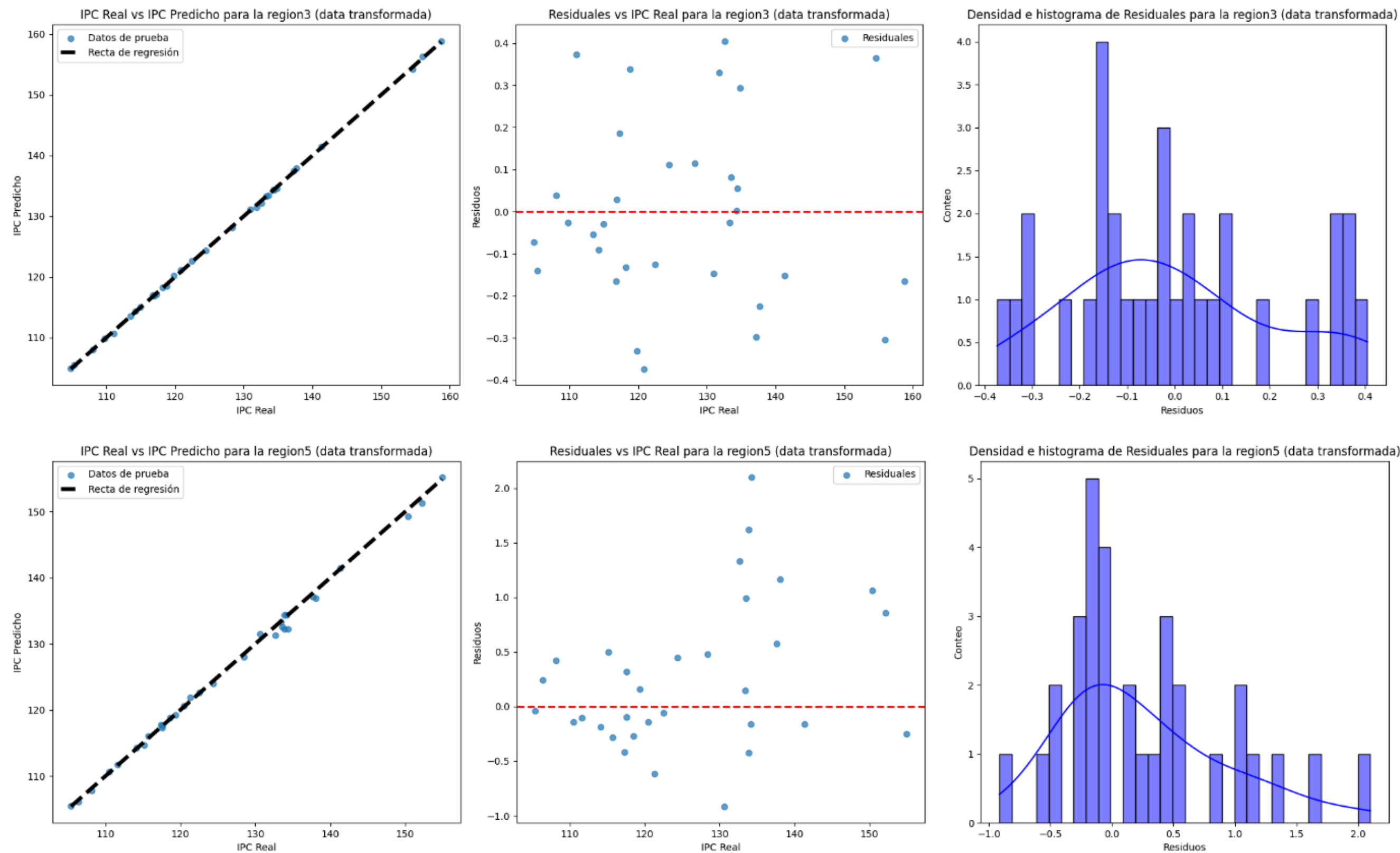
Resultados para el IPC por cada región del país (Región 8)



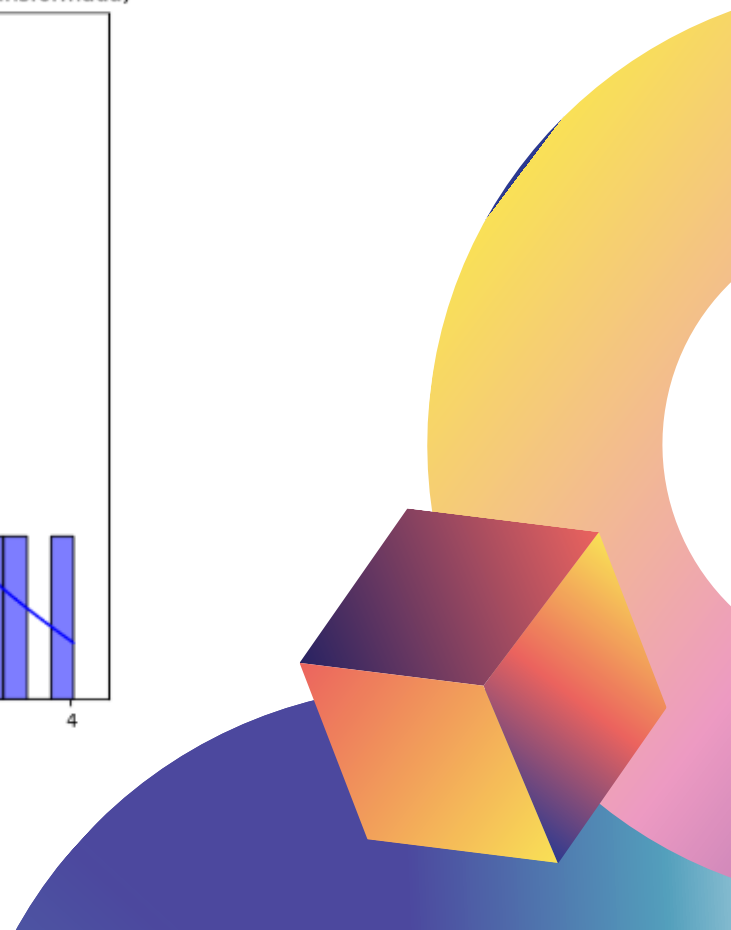
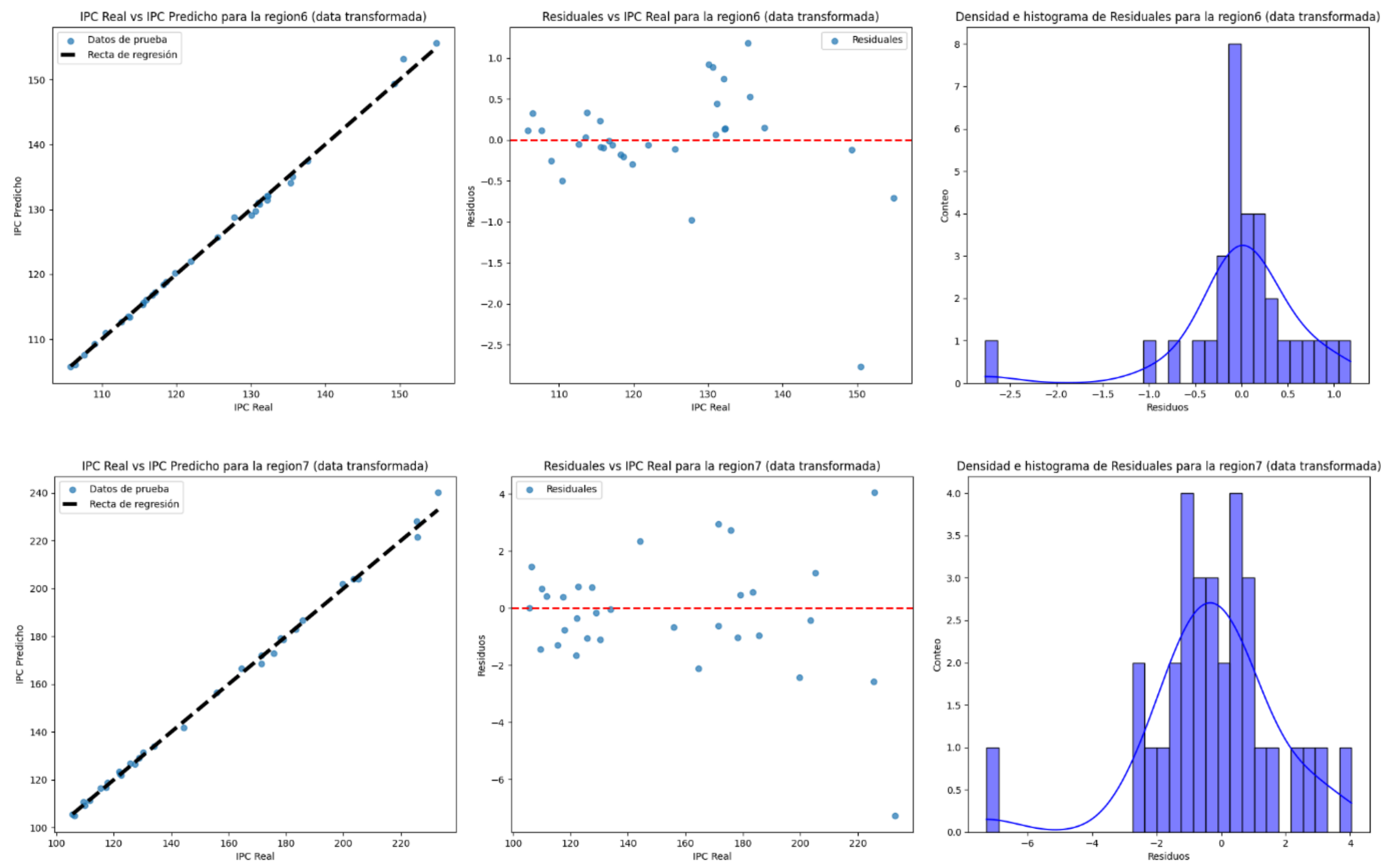
Resultados para el IPC por cada región del país con data transformada (Región 1 y 2)



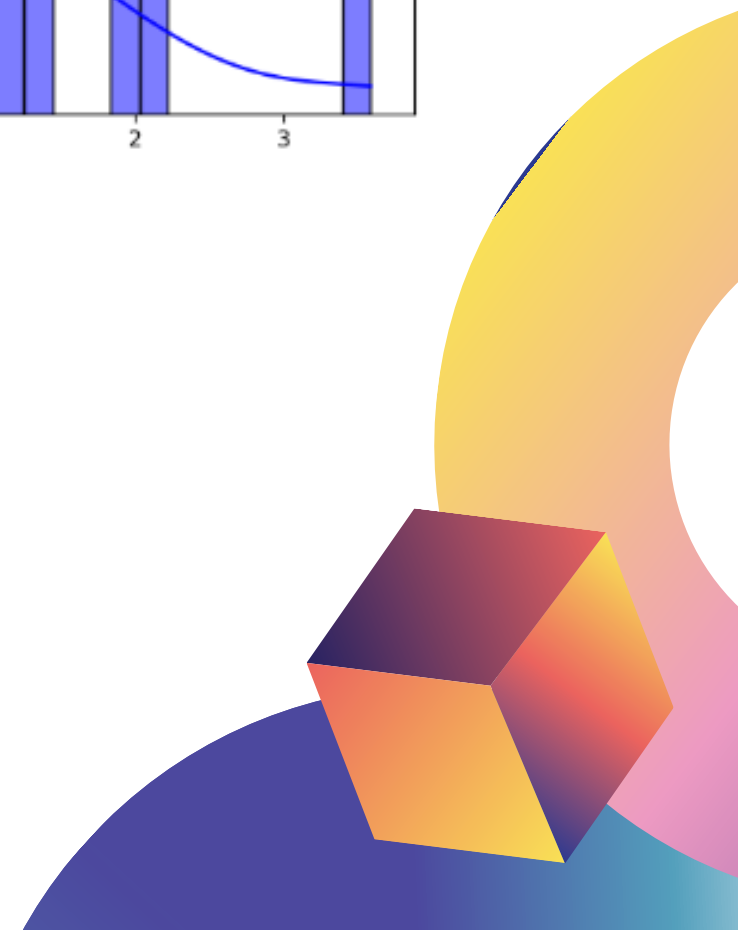
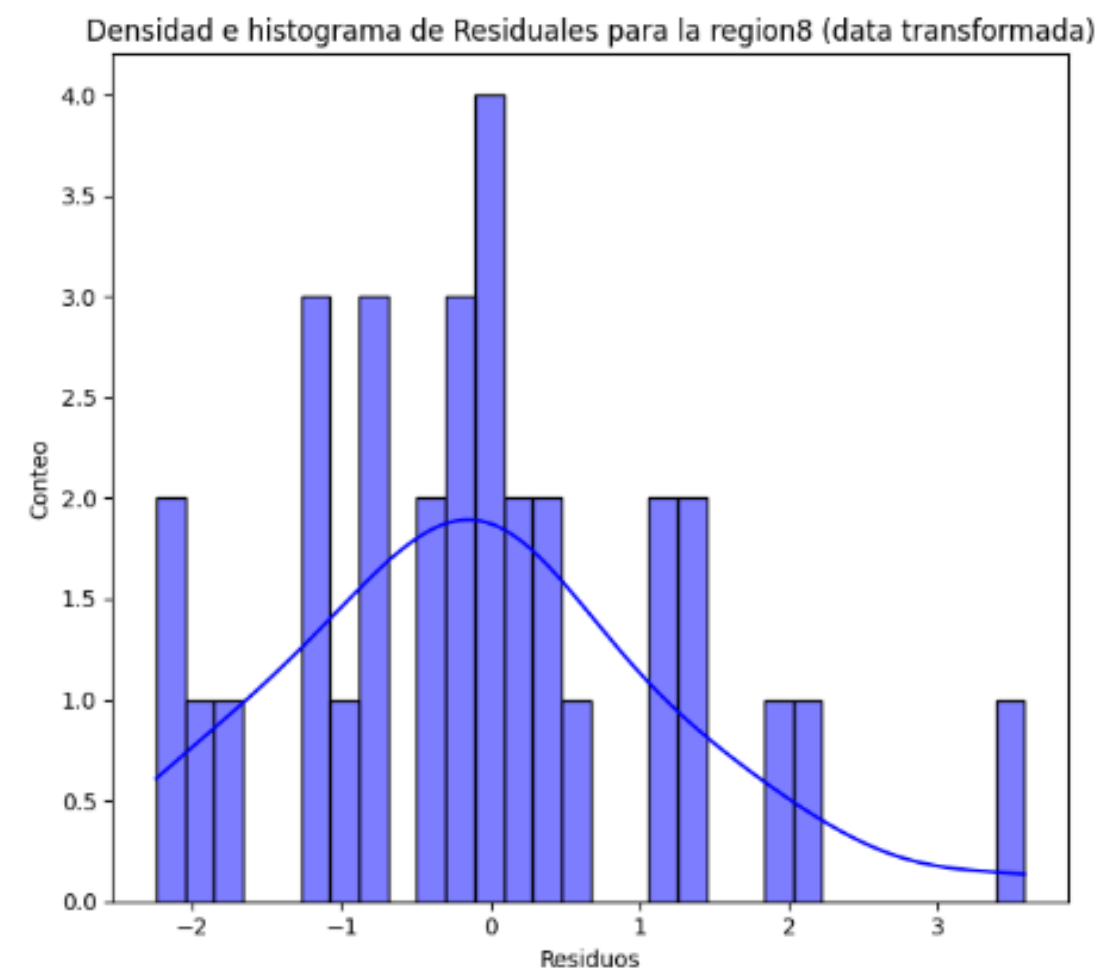
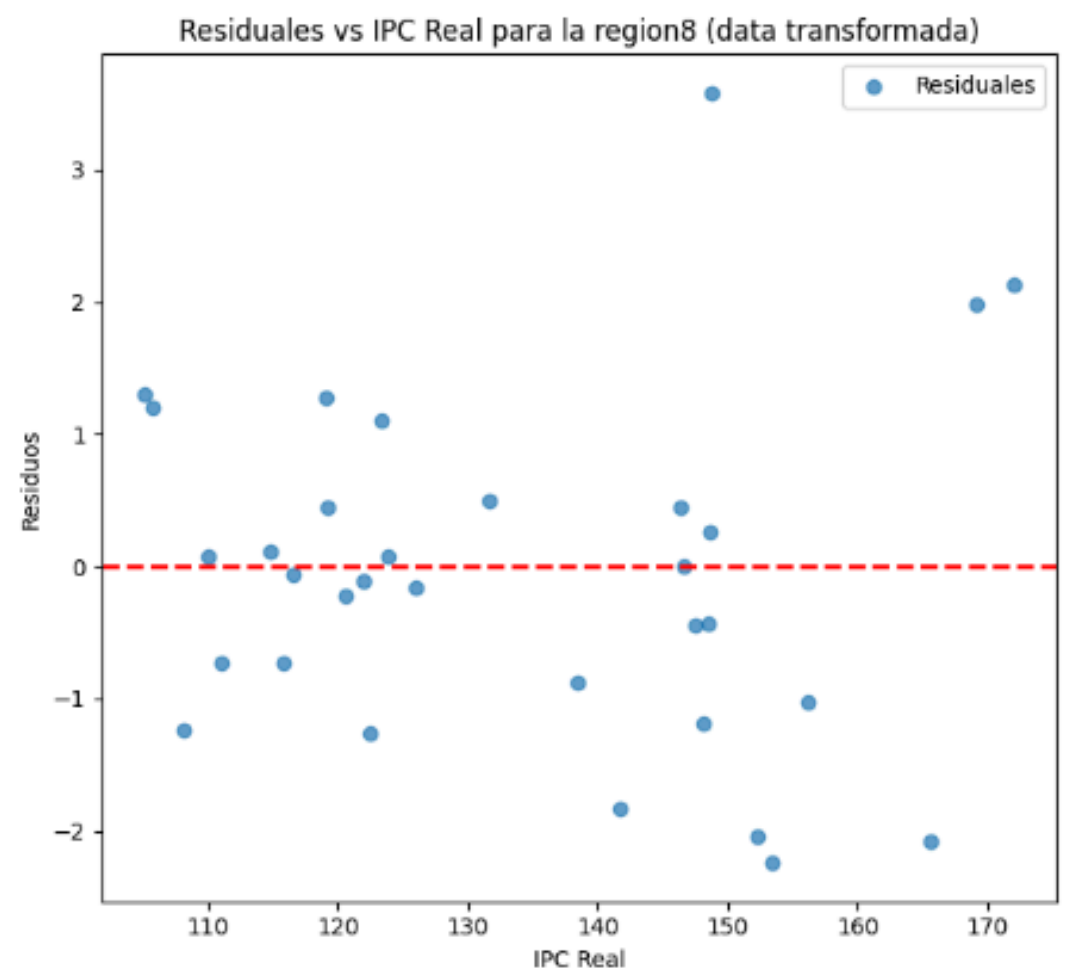
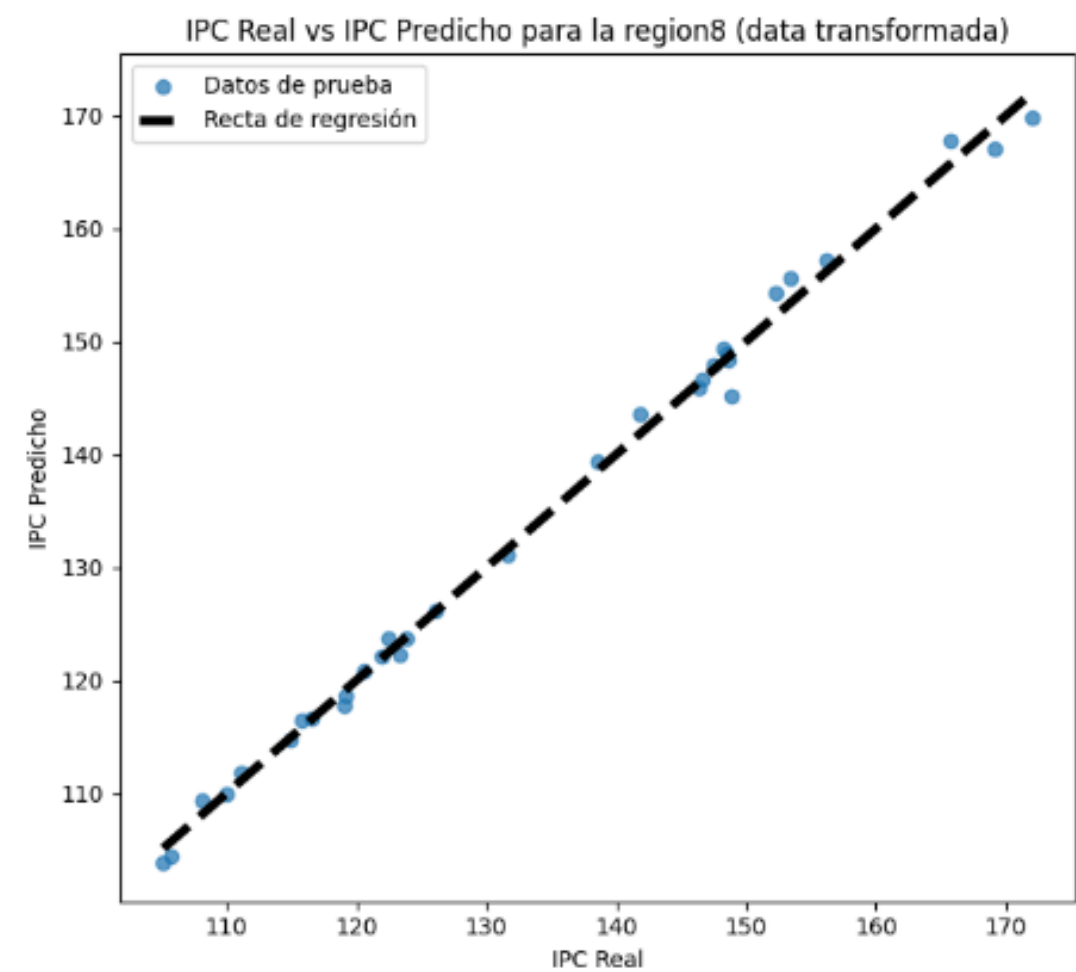
Resultados para el IPC por cada región del país con data transformada (Región 3 y 5)



Resultados para el IPC por cada región del país con data transformada (Región 6 y 7)



Resultados para el IPC por cada región del país con data transformada (Región 8)



Discusión de Resultados

EL ANÁLISIS DEL ÍNDICE DE PRECIOS AL CONSUMIDOR, NOS MUESTRA DISTINTAS CATEGORÍAS LAS CUALES, TIENEN UN IPC PROMEDIO DISTINTO, ES IMPORTANTE MENCIONAR QUE LOS PRECIOS PUEDEN VARIAR POR DISTINTAS RAZONES, POR LA DEMANDA, POR LA COMPETENCIA, POR LA UBICACIÓN GEOGRÁFICA, ENTRE OTROS FACTORES.

SI OBSERVAMOS LAS GRÁFICAS DE IPC REAL VS EL PREDICHO PODEMOS OBSERVAR COMO ALGUNOS MODELOS REGIONALES FUNCIONAN MEJOR QUE OTROS. LAS REGIONES CON MEJOR ALINEACIÓN ENTRE LAS PREDICCIONES Y LOS VALORES REALES INDICAN UN MODELADO EXITOSO DEL COMPORTAMIENTO DEL IPC, POSIBLEMENTE A UN MAYOR HOMOGENEIDAD EN LOS DATOS O A MODELOS MÁS AJUSTADOS A LAS DINÁMICAS REGIONALES.

LOS HISTOGRAMAS Y LAS GRÁFICAS NOS PERMITIERON OBTENER UNA MEJOR PERSPECTIVA CRÍTICA SOBRE LA NORMALIDAD Y LA DISTRIBUCIÓN DE LOS ERRORES DEL MODELO. UN MODELO IDEAL TENDRÍA RESIDUALES NORMALMENTE DISTRIBUIDOS, INDICANDO QUE LOS FACTORES CAPTURADOS POR EL MODELO SON MÍNIMOS Y ALEATORIOS.

LAS GRÁFICAS RESIDUALES PROPORCIONAN DISTINTAS PERSPECTIVAS PARA MEJORAR LOS MODELOS. LAS REGIONES CON RESIDUALES DISPERSOS, SIN PATRÓN REFLEJAN UN BUEN AJUSTE AL MODELO, MIENTRAS QUE PATRONES EVIDENTES EN LOS RESIDUALES SUGIEREN ÁREAS DE MEJORA COMO LA INCORPORACIÓN DE VARIABLES EXPLICATIVAS ADICIONALES O LA REVISIÓN DE LOS VALORES ATÍPICOS QUE PODRÍAN AFECTAR LA PRECISIÓN DE LAS PREDICCIONES.

Conclusiones

- El análisis realizado y mostrado a través de las 48 gráficas revela variaciones en la efectividad del modelo de predicción del IPC entre las diferentes regiones de Guatemala.
- Algunos modelos regionales requieren que la información obtenida sea sumamente precisa, para poder realizar las desviaciones sistemáticas o para mejorar la captura de la volatilidad del IPC.
- Es de suma importancia realizar un buen Análisis Exploratorio de datos (EDA) para tener uniformidad en los datos y realizar un modelo preciso.





Gracias