

INFORME PRACTICA DE ESTADISTICA DESCRIPTIVA

POR JEAN PAUL SANDOVAL

ESTADISTICA INFERENCIAL Y MUESTREO

UNAB

2025

1. INTRODUCCIÓN

El presente informe tiene como objetivo analizar, a partir de una base de datos, las principales herramientas de la estadística descriptiva, empleando instrumentos de programación usando código para la visualización de la información. A lo largo del documento se presentan gráficos ilustrativos, datos relevantes y el análisis correspondiente. Finalmente, se exponen las conclusiones del estudio.

2. DATOS Y METODOLOGÍA

Los datos utilizados para realizar el análisis descriptivo fueron obtenidos de la página gubernamental “*Datos Abiertos para Colombia*”. Se seleccionó una base de datos relacionada con los resultados de las pruebas Saber 11 del municipio de Sabaneta, la cual contiene 6 columnas y un total de 583 registros.

Para la elaboración de los gráficos se empleó el lenguaje de programación Python. La base de datos presenta 6 variables a estudiar: una de tipo numérico y cinco de tipo categórico. Las variables categóricas corresponden a Año, Sector, Colegio, Código DANE e Indicador, mientras que la variable numérica corresponde a Resultados.

Los procedimientos estadísticos aplicados incluyen: mapa de datos faltantes, histogramas, diagramas de cajas y bigotes, diagramas de barras, gráficos de torta y gráficos QQ.

3. RESULTADOS

A continuación, se presentan los gráficos correspondientes a cada variable, iniciando con las categóricas y posteriormente con las numéricas, cada uno acompañado de su respectiva explicación.

Para todas las variables categóricas se analizarán los procedimientos estadísticos diagrama de barras y grafico de torta, en cambio para la variable numérica se analizará los histogramas, diagrama de cajas y bigotes y los gráficos QQ.

3.1. LOS RESULTADOS

En este contexto de las pruebas saber 11, los resultados se refieren al puntaje obtenido por cada colegio en cada indicador de dichas pruebas.

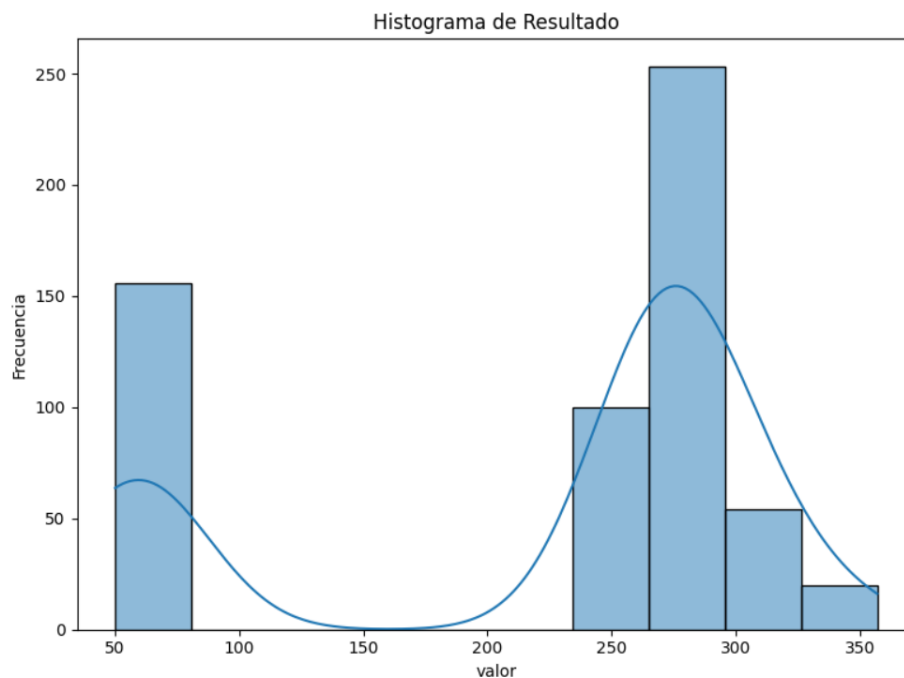


FIGURA 1. Histograma de la variable resultado

El histograma muestra que los valores de la variable *Resultado* oscilan entre aproximadamente 50 y 350 puntos. Sin embargo, se observan rangos particularmente entre 100 y 230 puntos. En el contexto de las pruebas Saber 11, esto sugiere que, en promedio, los colegios incluidos en la muestra no registran puntajes dentro de dicho intervalo, concentrándose los resultados en valores significativamente más bajos o más altos.

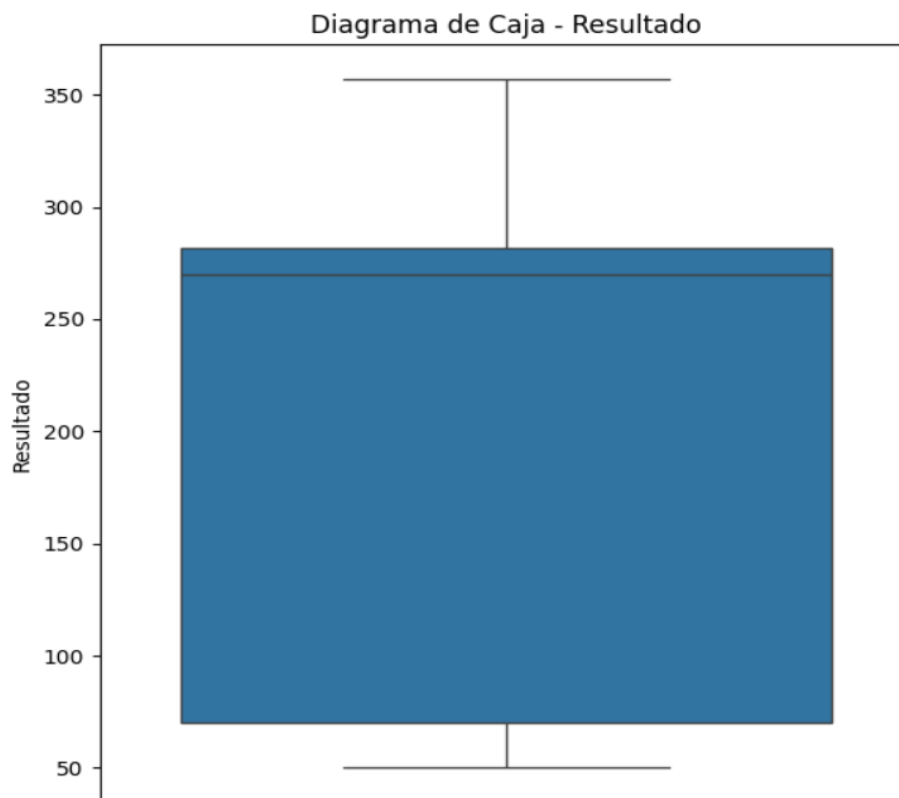


FIGURA 2. Diagrama de cajas y bigotes de la variable resultados

Para comprender mejor el contexto, es importante aclarar que en el diagrama de caja los valores máximos y mínimos corresponden a los extremos de los bigotes, que en este caso se sitúan aproximadamente en 350 y 50 puntos, respectivamente. La mediana, es decir, dicho valor que divide la muestra en 2 partes, se ubica en torno a los 275 puntos. Además, se evidencia la ausencia de datos atípicos, ya que no se observan puntos fuera de los límites establecidos por los bigotes

3.2 AÑO

El año se refiere en este caso a que tanto porcentaje de los registros corresponden a cada año.

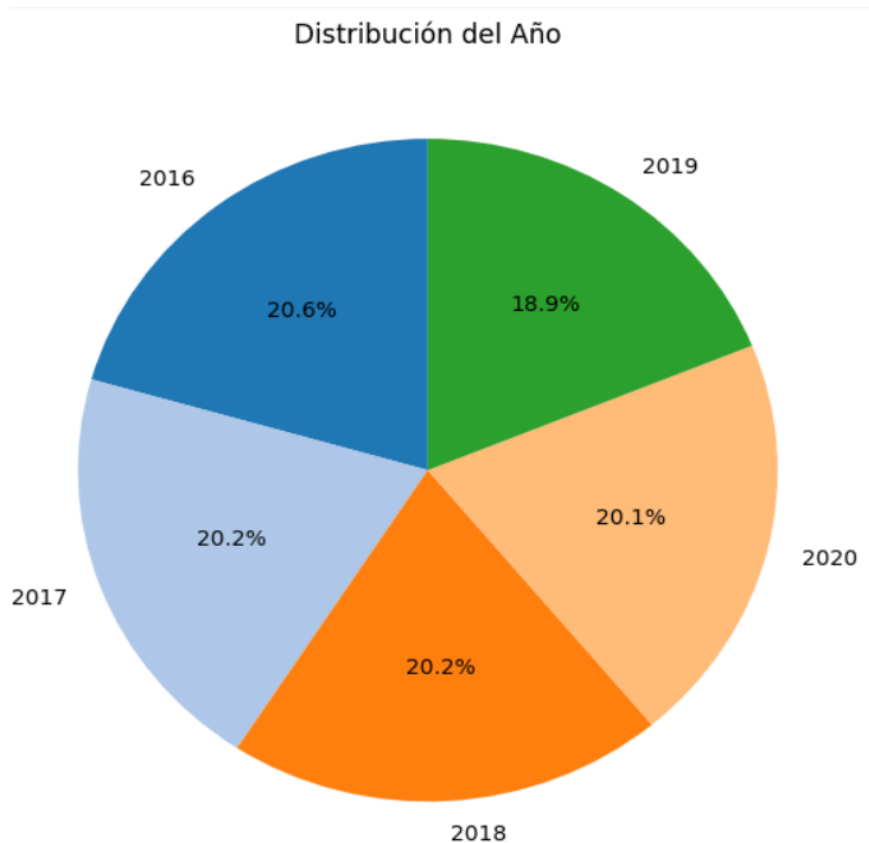


FIGURA 3. Gráfico de torta para la variable año

En este gráfico de torta se puede visualizar que desde el 2016 hasta el 2020, los registros de la base de datos son muy similares a excepción del 2019 que se observa una disminución mucho mayor.

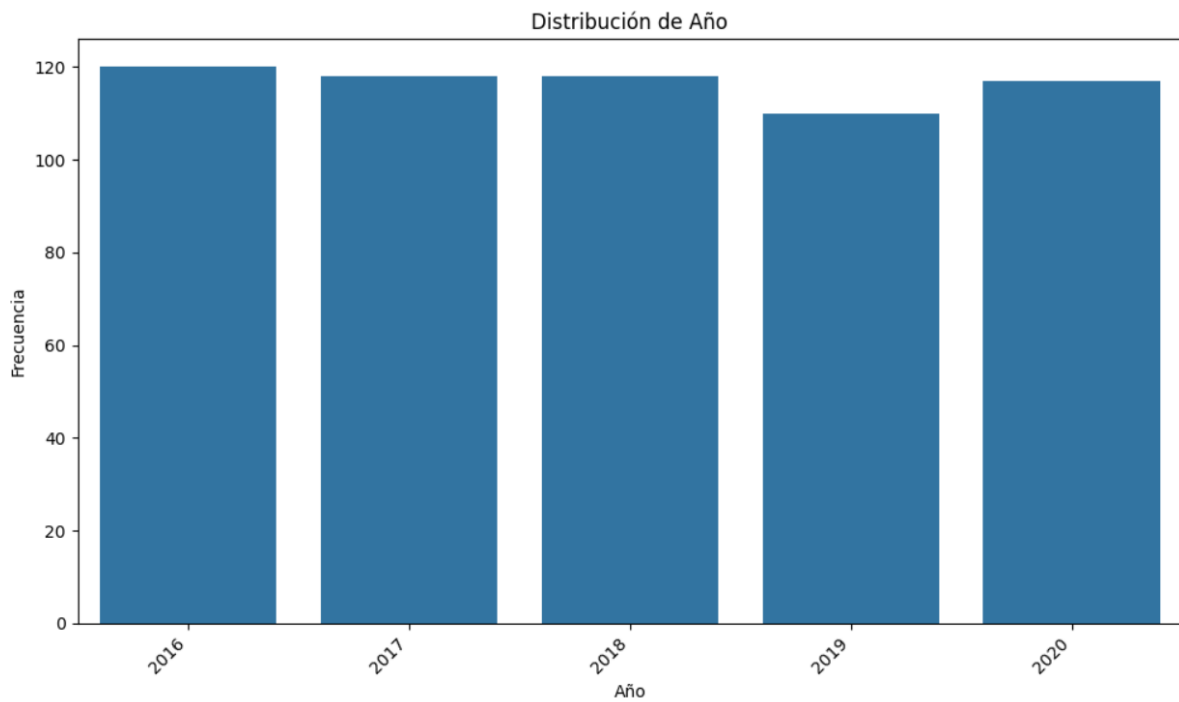


FIGURA 4. Diagrama de barras para la variable año

Como se explicó anteriormente, se observa una frecuencia muy similar entre cada año, sin embargo, el año 2019 enfrenta una frecuencia mucho menor.

3.3 SECTOR

En este estudio se encuentran en los datos únicamente 2 sectores; el privado y el público.

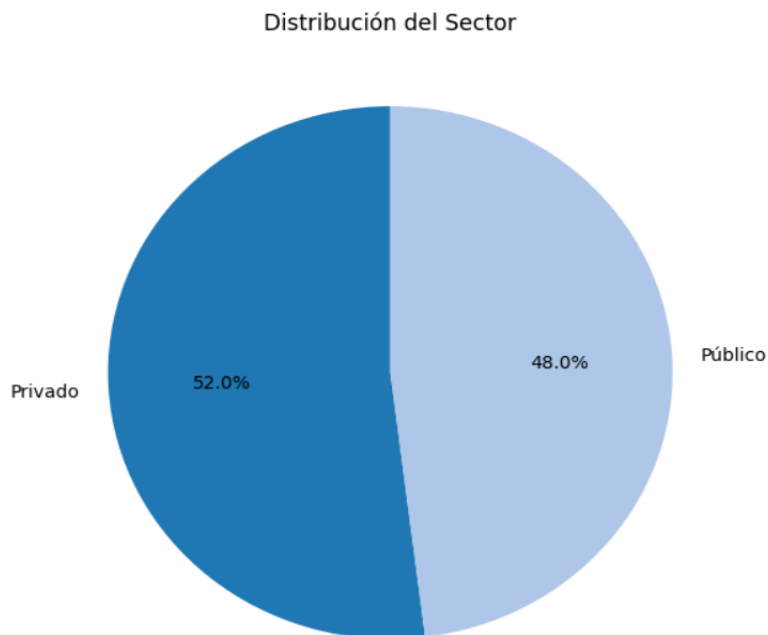


FIGURA 5. Gráfico de torta para la variable sector

Dado que solo existen dos sectores, la gráfica de torta se divide en dos partes. Se observa una ligera predominancia del sector privado, que supera al sector público por aproximadamente 4 puntos porcentuales.

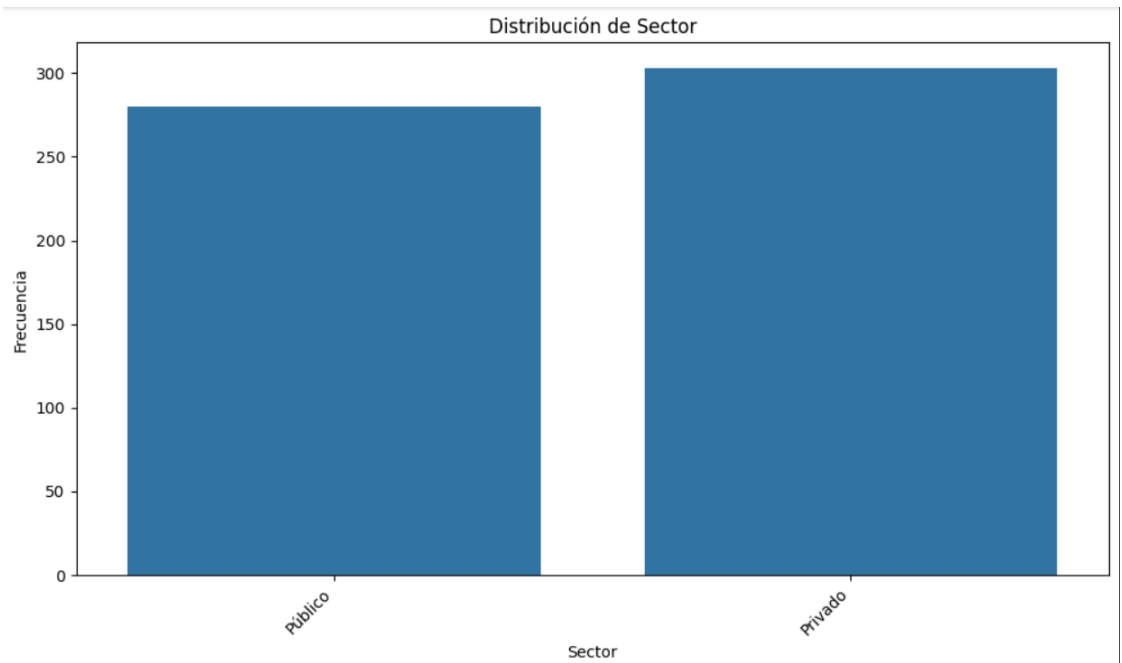


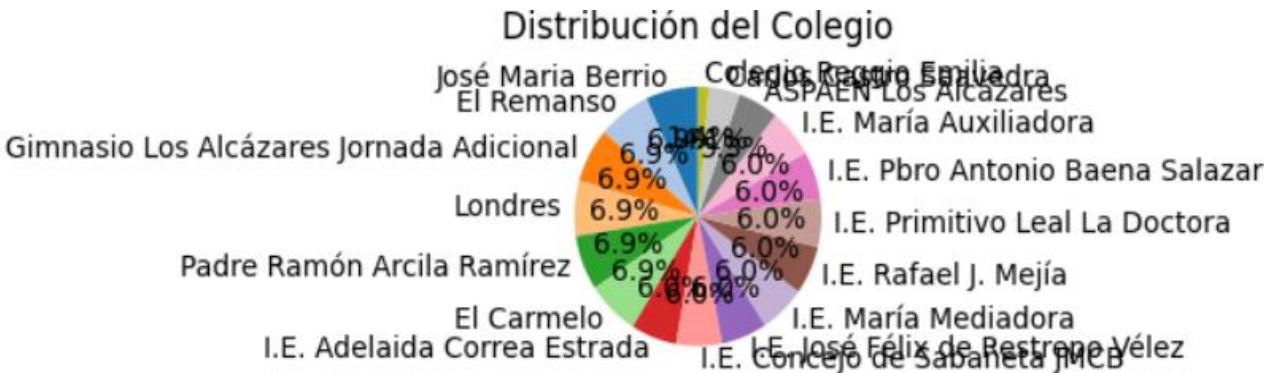
FIGURA 6. Diagrama de barras de la variable sector

Se observa una clara diferencia entre los dos sectores, se puede ver que el sector privado prevalece por encima del publico

3.4 COLEGIOS

A continuación, se presenta la participación de los colegios registrada en la base de datos.

FIGURA 7. Gráfico de torta de la variable colegios



Debido al elevado número de colegios, los datos representados en la gráfica se ven muy agrupados, lo que dificulta distinguir cada valor individualmente. Sin embargo, de manera general, se observa que la participación de los colegios es bastante similar, con la mayoría registrando porcentajes superiores al 6%.

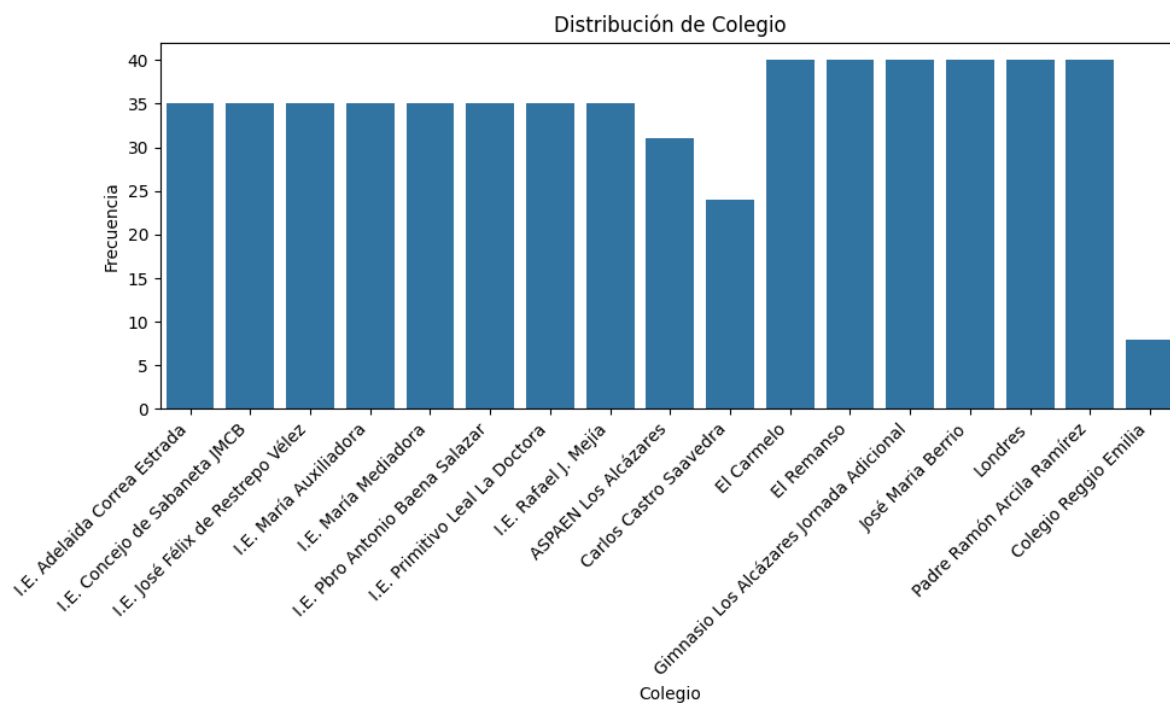


FIGURA 8. Diagrama de barras de la variable colegios

Con este diagrama se observa de manera más clara la distribución porcentual de los colegios en relación con los 583 registros de la base de datos. Se destaca que seis instituciones presentan el mismo nivel de participación, con una frecuencia aproximada de 40.

3.5 CÓDIGO DANE

Este código depende del colegio por ende posee los mismos datos que la variable de los colegios, así mismo los gráficos son exactamente iguales

3.6 INDICADOR

El indicador se refiere precisamente a cada disciplina evaluada en las pruebas saber por cada colegio.

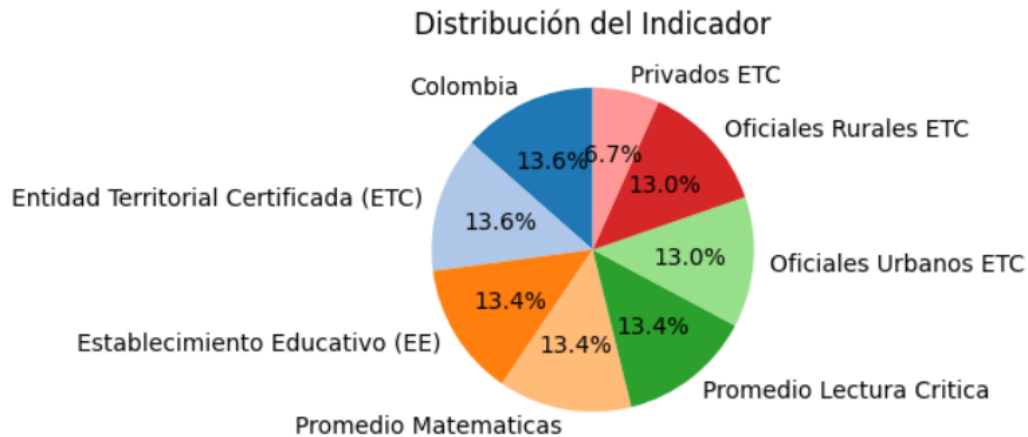
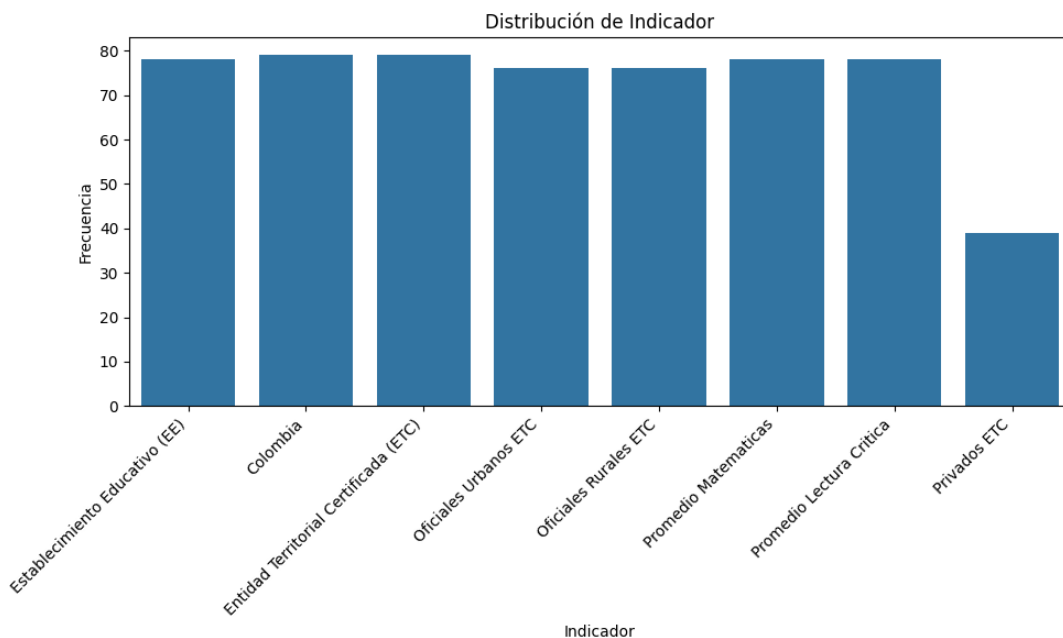


FIGURA 9. Grafico de torta de la variable indicador

Esta variable muestra que los indicadores más predominantes son Colombia y ETC, mostrando el mismo porcentaje, así mismo, el indicador que posee el menor valor es privados ETC, el resto de indicadores posee valores muy similares

FIGURA 10. Diagrama de barras para la variable sector



Como anteriormente se pudo observar todos los indicadores tienen valores significativos a excepción del indicador privados ETC, pues cuenta con la menor frecuencia; así mismo hay 2 indicadores que poseen prevalencia, así como Colombia y ETC.

3.7 GRAFICO QQ

Este procedimiento estadístico en específico es muy importante pues al observar su grafica se puede observar si los datos poseen una distribución normal o sencillamente no tienen dicha distribución.

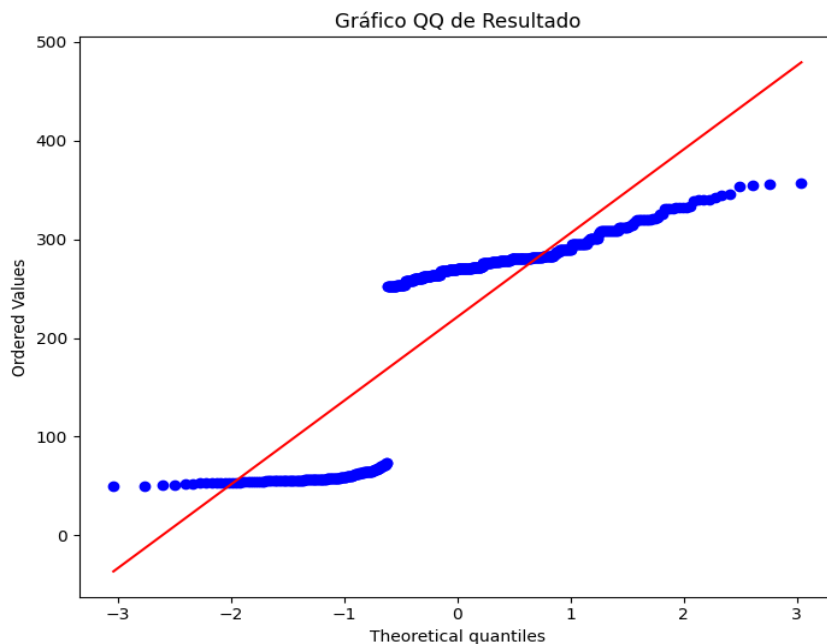


FIGURA 11. Grafico QQ de la variable resultados.

Si en esta grafica los puntos azules tendieran a estar por encima de la línea roja se podría afirmar con seguridad que posee distribución normal, pero como no es el caso, se puede decir con exactitud que no tiene una distribución normal.

4. CONCLUSIÓN

Los datos obtenidos de las pruebas Saber 11 del municipio de Sabaneta, correspondientes al periodo 2016–2020, evidencian que las variables categóricas presentan, en general, una participación bastante igual en cuanto a la distribución de registros, con dos excepciones notables: el año 2019, que muestra una disminución marcada, y el sector privado, que registra una ligera ventaja sobre el público, específicamente de 4 puntos porcentuales. En el análisis de los indicadores, destaca “Privados ETC” por su baja representación, lo que podría indicar una menor relevancia de este indicador en el contexto evaluado. En cuanto a las variables numéricas, se identifican vacíos significativos en el rango de puntajes entre 85 y 230, Asimismo, la mayoría de los colegios concentra sus puntajes entre 275 y 300, sugiriendo que una parte importante de la muestra se ubica en un nivel alto de desempeño, mientras otra queda rezagada, además se puede evidenciar el panorama desigual en cuanto a puntajes en la educación de sabaneta.