



Informe practica #1 estadística

Presentado por:

Ivan Santiago Rodriguez Romero

Docente:

Andres Fabian Leal Archila

Universidad Autónoma de Bucaramanga

Estadística

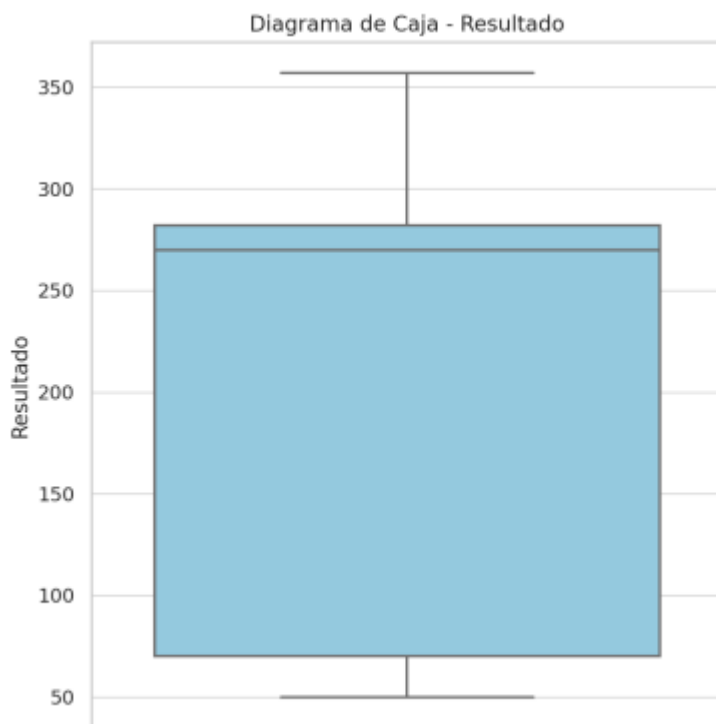
Bucaramanga – Colombia

1. Introducción:

En esta práctica analizamos los resultados de las pruebas Saber 11 de los estudiantes de Sabaneta. El objetivo fue conocer mejor los datos, identificar valores atípicos (outliers), analizar las variables categóricas y numéricas, y revisar si la variable principal seguía una distribución normal. Usamos Python para realizar los gráficos y análisis.

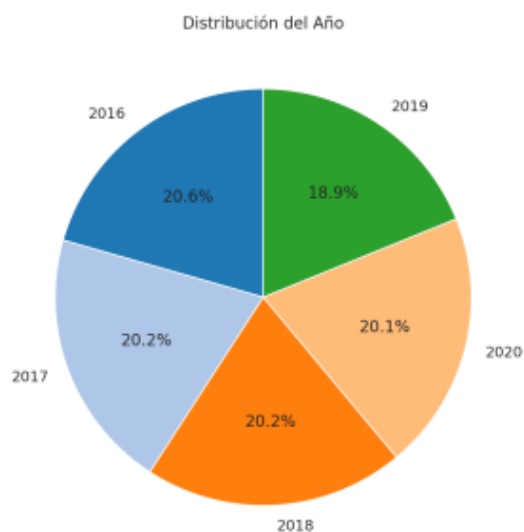
2. Análisis de variables y gráficos:

2.1. Variable numérica - Resultado

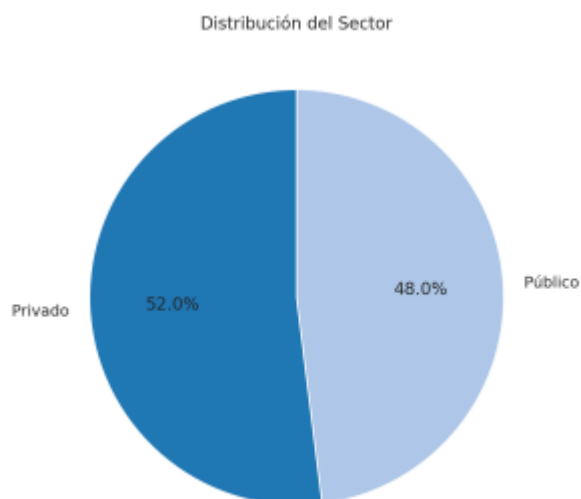


Este diagrama muestra cómo están distribuidos los puntajes. La línea dentro de la caja representa la mediana (270 puntos). Se observa que la mayoría de los valores están concentrados dentro del rango intercuartílico, pero hay varios puntos alejados, que son los outliers.

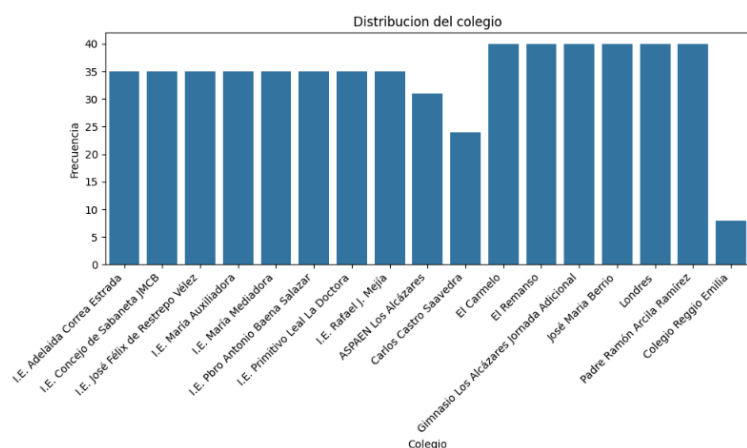
2.2. Variables categóricas



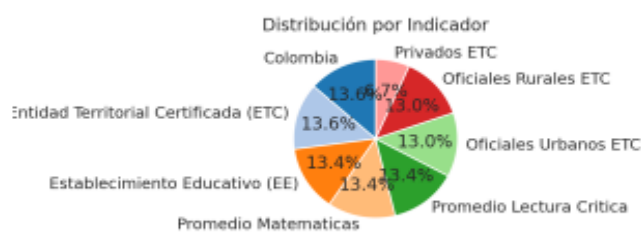
Año: La mayor proporción de datos corresponde al año 2016 (120 registros), seguido por 2017 y 2018 (118 cada uno), 2020 (117) y 2019 (110).



Sector: El sector privado representa el 52% (303 registros) y el público el 48% (280 registros).



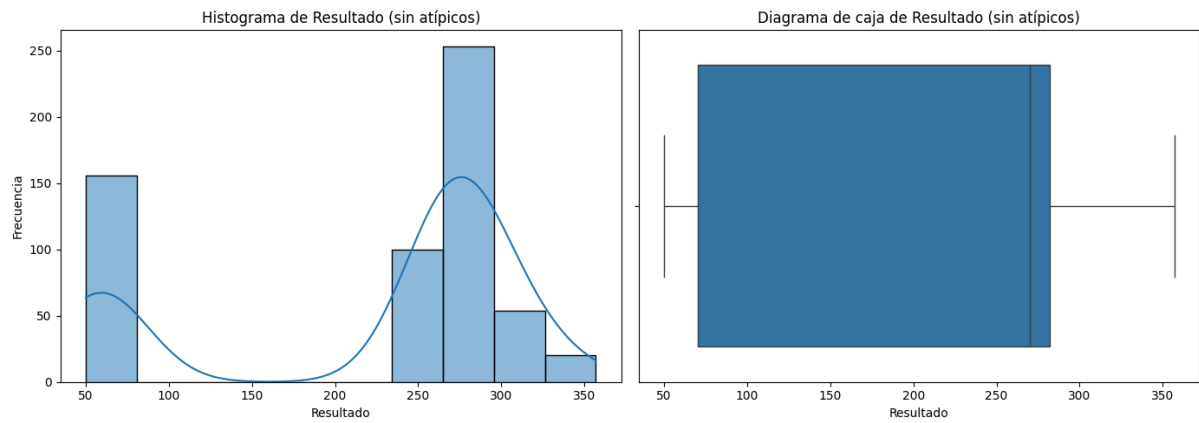
Colegio: Varios colegios tienen la misma representación (40 estudiantes cada uno), como Londres, Padre Ramon Arcila Ramirez, El Carmelo, El Remanso y Gimnasio Los Alcazares Jornada Adicional. Otros tienen menos, como Carlos Castro Saavedra (24) y Colegio Reggio Emilia (8).



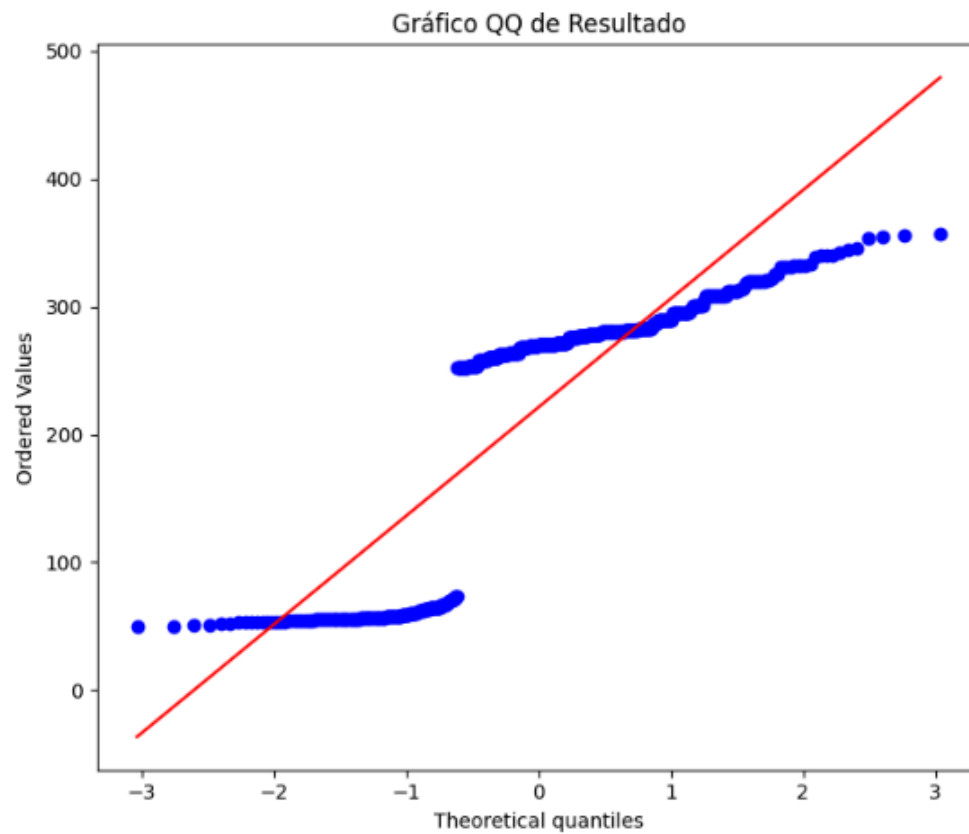
Indicador: Los más frecuentes son Colombia y Entidad Territorial Certificada (ETC) (79 registros cada uno), seguidos de Establecimiento Educativo (EE) y Promedio Matemáticas (78 cada uno).

3. Outliers:

Usamos el método del rango intercuartílico para identificar valores atípicos en la variable Resultado. Después de eliminar los outliers, quedaron 583 registros para el análisis. Esto reduce el sesgo que podrían causar los valores extremos y permite un análisis más representativo.



4. Test de normalidad:



Se aplicaron varias pruebas a la variable Resultado sin outliers: Shapiro-Wilk: $p = 3.56 \times 10^{-30}$ -> No normal Kolmogorov-Smirnov: $p = 1.57 \times 10^{-65}$ -> No normal Anderson-Darling: Estadístico = 80.5745, Crítico(5%) = 0.782 -> No normal Jarque-Bera: $p = 0.0$ -> No normal El gráfico QQ confirma que los datos no siguen una distribución normal.

5. Conclusión:

- En esta práctica aprendimos varios procesos nuevos: Cargar y explorar un conjunto de datos reales, representar variables categóricas y numéricas mediante gráficos. Detectando y eliminando valores atípicos usando el método intercuartílico. Lo que nos lleva a evaluar la normalidad con diferentes pruebas estadísticas y con un gráfico QQ.
- Los resultados mostraron que el sector privado tiene una ligera mayoría, que el puntaje mediano es de 270 puntos, que existen outliers que influyen en la difusión, y que la distribución de la variable Resultado no es normal, incluso sin outliers.