

Estudio del test SAT

Vamos a analizar un estudio que recogió datos de los resultados del test SAT (un test estándar que se utiliza ampliamente para admisiones en USA) por estados, combinados con otros datos.

El test SAT se utiliza para medir el rendimiento de los estudiantes y, compararlo con el gasto en educación. La idea es la de tratar de entender si hay una relación entre el gasto y los resultados. En este caso, el estudio es de un sólo año y se compara entre los diferentes estados de Estados Unidos.

Es importante resaltar que el tema del impacto del gasto en los resultados es un tema recurrente en política educativa, y genera debates sobre la verdadera incidencia del gasto. Hay quien sostiene que no es el gasto lo importante, sino cómo se gasta el dinero, mientras que otras posiciones tienden a intentar igualar el gasto en favor de la igualdad de oportunidades.

Según la documentación del data set (<https://ww2.amstat.org/publications/jse/secure/v7n2/datasets.guber.cfm>) tenemos las siguientes variables:

- **state**: Nombre de los estados
- **expend**: Gasto actual por alumno, en promedio de asistencia diaria en escuelas públicas, primarias y secundarias, 1994-95 (en miles de dólares)
- **ratio**: Promedio del ratio alumno/maestro en escuelas de primaria y secundaria públicas, otoño 1994
- **salary**: Salario anual promedio de maestros en escuelas públicas de primaria y secundaria entre 194-95, en miles de dólares.
- **frac**: Porcentaje de estudiantes elegibles que realizaron el SAT, 1994-95
- **verbal**: Promedio en calificación verbal de la puntuación SAT, 1994-95
- **math**: Promedio en matemáticas de la puntuación SAT, 1994-95
- **sat**: Promedio total de la puntuación SAT, 1994-95

Hemos de tener en cuenta que los atributos **expend**, **ratio** y **salary** se relacionan directa o indirectamente con el gasto, y el gracias al dataset podemos relacionar estos datos con los resultados del SAT totales o sus componentes verbal y math.

Se pide:

- Analizar gráficamente si los scores del test tienen una escala similar y los componentes tienen relación.
- Ordenar los datos y obtener los estados que más (y menos) gastan y los que mejores (y peores) resultados obtienen
- Analizar de manera intuitiva con gráficos si hay relación entre las variables que indican resultados (las del test SAT) y las variables que indican gasto.
- Utilizar correlaciones y un modelo lineal básico como exploración numérica de ese tipo de relación gasto/resultados. ¿Cuál es la conclusión aparente?