

Universidad de Costa Rica

Facultad de Ciencias

Escuela de Matemática

Herramientas de Ciencias de Datos I

Proyecto Individual

Estudiante:

Gabriel Chaves Esquivel

Carné:

C4E273

Profesor:

Joshua Cervantes

II Semestre, Año 2025

Análisis de la Educación Diversificada en Costa Rica (2018–2021)

1. Introducción

El presente análisis evalúa la situación de la educación diversificada (décimo-undécimo) en Costa Rica durante el periodo 2018–2021 utilizando los siguientes datos extraídos de tablas públicas del MEP: matrícula inicial, rendimiento definitivo (posterior a convocatoria), aprobados–aplazados–reprobados (previo a convocatoria) y exclusión intraanual. El objetivo es identificar patrones de rendimiento, diferencias territoriales, evolución temporal e instituciones destacadas, considerando además el impacto de la pandemia en los años 2020 y 2021, pudiendo observar así las afectaciones que pudo o no haber causado la virtualidad.

2. Explicación de código

Se utilizaron cuatro bases de datos oficiales del MEP, como se mencionó anteriormente. Además, se utilizaron las librerías `readxl` y `tidyverse`, para leer los excels y realizar los filtrados y gráficos, respectivamente. En primer lugar, se filtraron y seleccionaron solo las columnas de datos necesarias y deseadas para el análisis, como lo son: Curso lectivo, provincia, nombre de la institución, datos de décimo y undécimo periodo, etc; dentro de este mismo filtrado, también se dejaron solo los años a estudiar. Posteriormente se manejaron los NA, al ser colegios que no poseían ningún dato se optó por eliminarlos del todo, puesto que además solo eran 7 colegios con estas características. Se hizo una revisión de que no hubiera datos sin sentido, como lo podría ser que los aprobados y reprobados sean mayores a la cantidad matriculada de estudiantes, etc; sin embargo, no se mostraron errores. Finalmente, Se crearon indicadores derivados: tasas de aprobación, reprobación, aplazados, exclusión total y matrícula inicial total

Continuando con el análisis de datos, en primer lugar se revisó tabla por tabla y luego un análisis conjunto entre rendimiento previo a convocatoria y posterior a este, además de matrícula inicial y exclusión intraanual. En primer lugar, a la tabla de rendimiento previo a convocatoria se le hizo un summary para ver los datos generales, se hizo un gráfico de líneas para observar la tendencia del promedio anual de la tasa de aprobación, aplazo y reprobación, utilizando `ggplot`; posteriormente se hicieron tres `boxplot` que separaba por año y provincia la tasa de aprobación, aplazamiento y reprobación respectivamente; por último, se vieron los diez colegios con mayor y menos tasa de aprobación entre todos los años. Para la tabla de rendimiento definitivo, matrícula inicial y exclusión se hizo prácticamente lo mismo. Finalmente, para los análisis conjuntos lo que se hizo fue agrupar por año los datos y comparar sus promedios.

En resumen, el análisis incluyó: comparación por provincia, diferencias regionales, evolución temporal y obtención de los 10 colegios con mejor y peor desempeño por año.

3. Resultados

En las tendencias de rendimiento se puede observar que en promedio se mantuvo bastante similar a los años previos a pandemia, teniendo una decaída de tasa de aprobación promedio de 2018 a 2019 y un pico de 2019 a 2020, mientras que se presentó este mismo comportamiento en 2020 y 2021. Un caso similar pasa en la comparación de provincias, mostrando brechas entre provincias fuera y dentro de la GAM sobretodo. En donde realmente se ven más diferencias en el rendimiento académico es en los valores extremos de colegios, pues al menos la mitad de los colegios con las tasas de aprobación más bajas suelen ser de los años 2020 y 2021. Siendo un caso muy contrario a la matrícula inicial, pues está incrementó bastante durante los años de pandemia, siendo Cartago la provincia con mayor matrícula promedio. Finalmente, respecto a la exclusión intraanual, la tendencia incrementó y decrementó en gran medida, pero finalmente terminó en 2021 bastante cercano a como comenzó en 2018; empero, al igual que en el rendimiento, donde se ven mayores diferencias en la exclusión es en sus valores extremos, pues en este caso la mayoría de los datos de colegios con la mayor exclusión y exclusión intraanual se dieron en pandemia, específicamente en 2021; dejando claro que la pandemia fue un fuerte golpe a los estudiantes a la hora de seguir sus estudios.

4. Conclusiones

La educación diversificada muestra una estabilidad promedio en pandemia, sin embargo, sus datos extremos se disparan y muestran que sí hubo graves afectaciones en los colegios que ya presentaban problemas desde antes, pero se vieron aun más perjudicados por los efectos de la pandemia; además, se puede notar la brecha entre educación fuera y dentro de la GAM. Por lo que se concluye que, pese a que la pandemia fue un fuerte impacto al país y al sistema educativo, los que más se vieron afectados fueron los colegios que posiblemente ya presentaban carencias y dificultades previo a pandemia y esta ultima los hizo decaer finalmente.

5. Recomendaciones

1. Implementar medidas de solución post-pandemia, pues el nivel educativo se vió afectado
2. Analizar los colegios que presentamos los valores extremos y tener especial cuidado con ellos.
3. Combatir mediante proyectos, apertura de colegios públicos e inversión la brecha entre provincias.
4. No fijarse solo en el rendimiento numérico de aprobados y desaprobados, sino realizar un estudio del nivel de conocimientos académico.