

Correlación, Causalidad y la Brecha Digital:
La Tensión entre el Acceso a Internet y las Notas en Costa Rica

Dodanim Castillo Aráuz

ORCID: <https://orcid.org/0009-0002-0578-5938>

Correo: dcastillo@unadeca.net

Universidad Adventista de Centro América

Marvin Francisco Moncada Hernández

ORCID: <https://orcid.org/0009-0001-9485-8571>

Correo: marvin.moncada@unadeca.net

Universidad Adventista de Centro América

Resumen

Este estudio analiza el impacto del acceso a Internet en el rendimiento académico de estudiantes costarricenses, utilizando técnicas de regresión lineal y correlación simple. La investigación revela una correlación extremadamente alta ($r \approx 0.99$) en la actividad digital y las notas en matemáticas a nivel provincial.

Sin embargo, el análisis enfatiza que esta fuerte asociación no implica causalidad, sino que señala la presencia de una profunda brecha digital y socioeconómica. El artículo traduce este hallazgo estadístico en recomendaciones de política educativa, subrayando la urgencia de invertir en la equidad digital (tipo de dispositivos) en zonas rurales. Se concluye que la estadística es una herramienta esencial para el diseño de políticas públicas con los pies en la tierra, siempre y cuando el investigador evite la falacia de la causalidad y analice la complejidad del contexto.

Index Terms: Brecha digital, Rendimiento Académico, Regresión Lineal, Políticas educativas, Equidad, Costa Rica.

INTRODUCCIÓN Y PLANTEAMIENTO

El análisis de datos en la educación no es un ejercicio meramente académico; es la brújula que guía la inversión pública [1]. En el contexto de la educación superior, y más aún después de la digitalización forzada por la pandemia, el acceso a Internet se ha consolidado como un factor determinante del éxito o fracaso académico. Sin embargo, persisten grandes diferencias entre el acceso que tienen los estudiantes de zonas urbanas respecto a sus pares en áreas rurales. Esta disparidad, conocida como la Brecha Digital, no solo se refiere a la conexión, sino también a la calidad del servicio, el tipo de dispositivo y la alfabetización digital.

El problema que planteamos es el siguiente: ¿Qué tan fuerte es la asociación entre la conectividad digital y las notas de los estudiantes en Costa Rica?

Para responder a esta pregunta, empleamos el Análisis de Regresión Lineal Simple y Correlación sobre datos oficiales (promedios de notas de matemáticas y porcentajes de hogares con acceso a Internet) desagregados a nivel provincial. La regresión, en esencia, busca dibujar la "línea recta" que mejor explique cómo una variable afecta a la otra: es la simplificación matemática de una realidad compleja. El objetivo inicial era medir la magnitud de esta relación.

El hallazgo fue, a la vez, contundente e incómodo: se observó una correlación estadísticamente significativa con un coeficiente r cercano a 0.99. Un número tan alto sugiere que el acceso a Internet y el rendimiento académico están casi perfectamente ligados. Este resultado, si bien dramático, no puede ser interpretado a la ligera. El principal objetivo de este editorial es discutir por qué una correlación casi perfecta no es sinónimo de causalidad, y cómo este resultado debe ser usado por UNADECA y las autoridades educativas para enfrentar la Brecha Digital como un problema de justicia social.

El artículo se desarrollará en torno a tres argumentos clave que advierten contra la interpretación ingenua de la correlación y proponen caminos claros para la política educativa.

Desarrollo: De la Correlación al Contexto

La investigación reveló una relación ascendente casi perfecta entre la conectividad digital y el rendimiento en matemáticas, con un coeficiente de correlación de Pearson (r) cercano a 0.99[1]. Este valor indica que las provincias con mayor acceso a Internet también tienden a tener los promedios académicos más altos. El modelo de Regresión Lineal Simple (la línea recta) validó este patrón, explicando el 98% de la varianza en las notas ($R^2 = 0.98$).

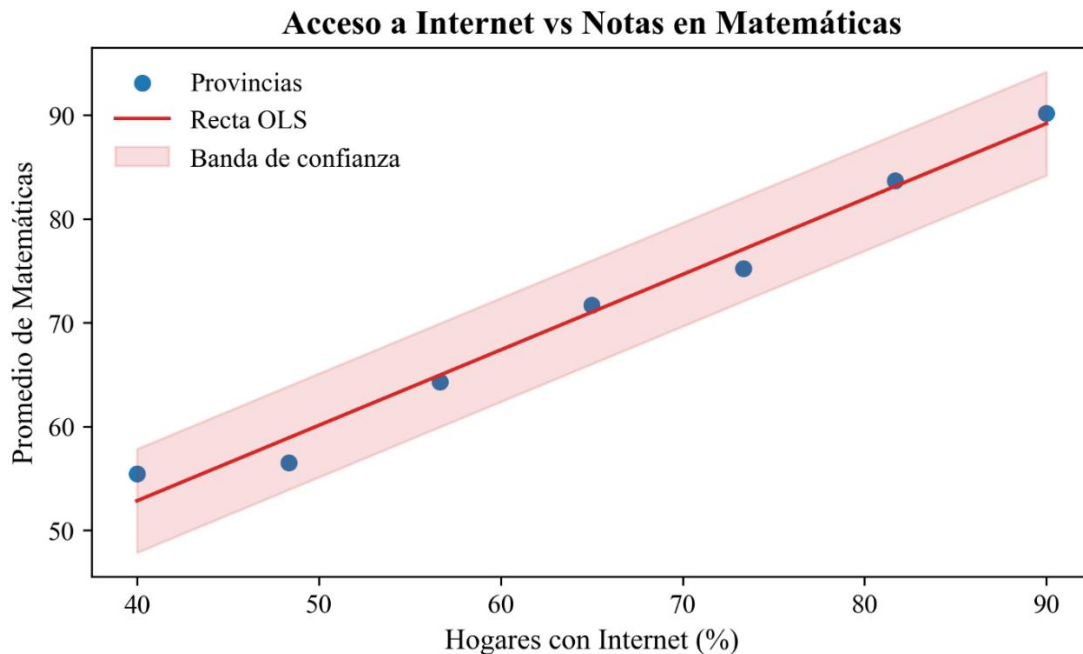


Figura 1. Relación entre acceso a Internet y notas en matemáticas por provincia. La línea recta muestra el ajuste de regresión lineal

La Figura 1 ilustra esta relación. Cada punto representa una provincia costarricense. La línea de regresión muestra el patrón ascendente. La banda de confianza estrecha confirma la fuerza de la asociación estadística.

Un estadístico podría detenerse aquí y celebrar la solidez del modelo. Sin embargo, un educador o un sociólogo debe ir más allá y entender las tres lecciones cruciales que este dato esconde [3].

1. Lección de Estadística: Correlación $\neq$ Causalidad

El error más grave en la interpretación de este análisis es asumir que la instalación de un router (enrutador o punto de acceso) es la causa directa del aumento en las calificaciones [2].

- **El Mito:** Si una provincia sube su conectividad en un 10%, sus notas automáticamente subirán **7.8 puntos** (basado en el coeficiente de la pendiente $\beta_1 \approx 0.78$)
- **La Realidad:** La correlación es simplemente un termómetro que mide si dos variables "bailan al mismo ritmo". En este caso, la alta correlación es un síntoma de variables confusoras no medidas. ¿Y si las familias con mayor acceso a Internet también son aquellas con mayor ingreso familiar, mayor educación parental o acceso a colegios privados?
- **El Desafío:** La Regresión Lineal Simple no puede distinguir el efecto de la conectividad por sí solo [4]. Ignorar estas variables de confusión puede llevar a políticas públicas ineficaces, malgastando recursos en infraestructura sin abordar las desigualdades socioeconómicas subyacentes.

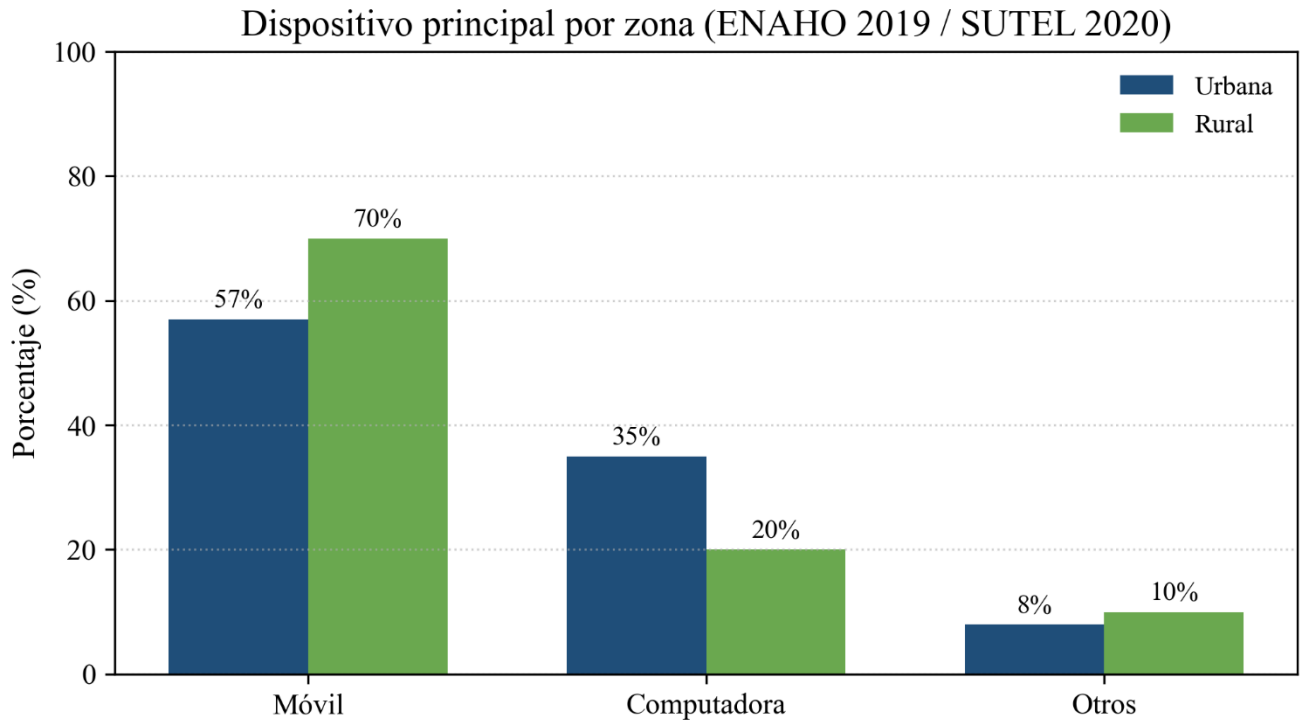
2. Lección de Tecnología: La Brecha Digital Silenciosa

El análisis revela que la desigualdad no solo se mide en términos de acceso, sino en calidad y tipo de dispositivo.

Los datos muestran una marcada disparidad entre las zonas urbanas y rurales. En el ámbito rural, se reporta que hasta un 70% del acceso a Internet es a través de teléfonos móviles.

La Figura 2 y la Tabla I detallan esta brecha. En zonas rurales, el teléfono móvil domina. En zonas urbanas, las computadoras tienen mayor presencia. Esta diferencia afecta directamente las oportunidades de aprendizaje.

Zona	Uso Principal del Dispositivo	Implicación para el Estudio
Urbana	Predominio de computadoras portátiles/escritorio	Permite el desarrollo completo de trabajos de investigación, hojas de cálculo, y recursos en línea.
Rural	Predominio de teléfonos móviles	Limita la calidad del acceso y el uso de herramientas académicas complejas, impactando negativamente el rendimiento académico.



*Figura 2. Dispositivo principal para acceso a Internet por zona geográfica.
Datos de ENAH0 2019 y SUTEL 2020*

Recomendación de Política Educativa (Consejo): Para cerrar la brecha digital, no basta con expandir la señal Wi-Fi. Las instituciones educativas deben priorizar las políticas orientadas a la equidad en el dispositivo. Es urgente invertir en infraestructura tecnológica en instituciones públicas (computadoras, portátiles) para garantizar que el acceso se traduzca en una verdadera herramienta de aprendizaje, y no solo en una herramienta de comunicación.

3. Lección Metodológica: El Diablo Está en los Supuestos

El éxito del modelo ($R^2 = 0.98$) dependió de la rigurosa validación de supuestos; sin este paso, los resultados habrían sido engañosos [4].

Consejo para la Investigación Institucional: Todo investigador debe ser un escéptico de sus propios datos, verificando que se cumplan las condiciones para que la "línea recta" tenga validez.

- **Evitar el Enjambre de Abejas (Linealidad):** Si la relación entre variables fuera curva (no lineal), el modelo de regresión simple sería inadecuado.
- **Asegurar la Distribución Gaussiana (Normalidad):** El análisis de los residuos (los errores no explicados) debe seguir una distribución normal. En este estudio, los gráficos Q-Q confirmaron que los errores se comportaban bien.
- **Cuidado con el Embudo (Homoscedasticidad):** La varianza del error debe ser constante. Si la dispersión de los datos se amplía a medida que aumentan las notas (un "embudo"), la credibilidad de los intervalos de confianza se pierde.

Este proceso de validación evita publicar **conclusiones erróneas**. Si un modelo explica el 98% de la varianza, el 2% restante es un cajón de sastre donde se meten los errores de medición y las variables no observadas.

4. Análisis Complementario: La Desigualdad entre Tipos de Colegio

Los datos revelan otra capa de desigualdad. Los 50 mejores colegios de Costa Rica se concentran en tipos específicos de instituciones [7]. La figura 3 muestra que 34 de los 50 mejores colegios son privados. Solo 8 son públicos o subvencionados. Esto refleja una desigualdad estructural en el sistema educativo costarricense.

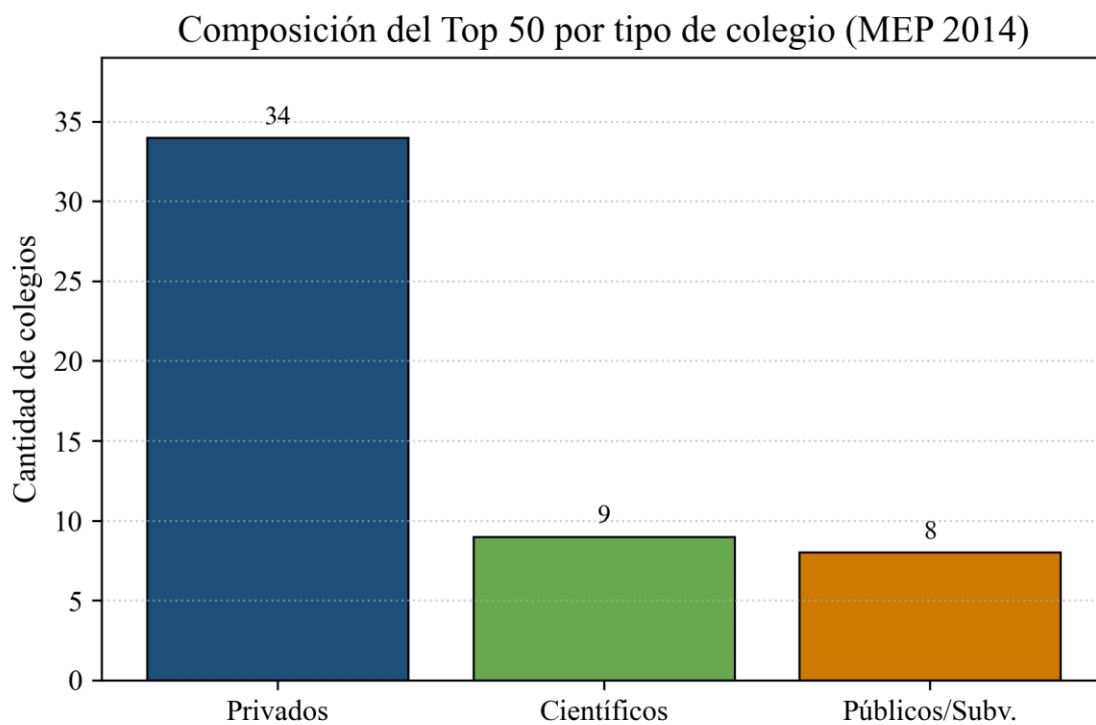


Figura 3. Composición del Top 50 de colegios por tipo de institución. Datos del MEP 2014.

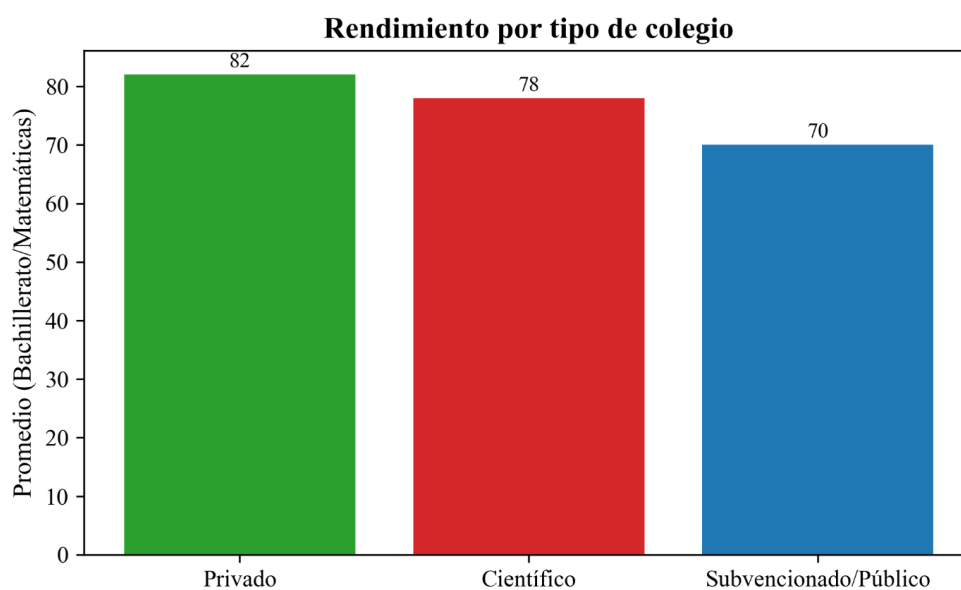


Figura 4. Rendimiento académico promedio por tipo de colegio en bachillerato y matemáticas

La figura 4 confirma esta brecha. Los colegios científicos y privados tienen

promedios más altos. Los públicos y subvencionados muestran rendimiento inferior. Esta diferencia persiste incluso considerando el acceso a Internet. Esto sugiere que otros factores, como recursos económicos, calidad docente y apoyo familiar, juegan roles determinantes

Conclusión y llamado a la acción

Esta investigación confirma que la estadística inferencial es una herramienta poderosa para diagnosticar desigualdades en la educación, revelando la fuerte correlación entre el acceso a Internet y el rendimiento académico [1]. Sin embargo, el valor de este hallazgo radica no en la cifra en sí, sino en la reflexión crítica que exige.

Reflexión final: El modelo de regresión lineal ha actuado como un microscopio, mostrando que la equidad educativa se juega en la calidad del acceso digital, no solo en la disponibilidad. Una política que solo se enfoque en el Internet sin capacitación docente o acceso a dispositivos es, en esencia, dinero malgastado.

Llamado a la acción (Motivación): La tarea para UNADECA y la comunidad educativa es doble: primero, invertir con inteligencia en la infraestructura digital de las zonas más vulnerables para cerrar la brecha; y segundo, exigir que todo análisis estadístico sea conducido con el rigor necesario para distinguir entre correlación y causalidad. Solo así la estadística se convierte en una herramienta ética al servicio de la justicia social.

AGRADECIMIENTOS

Agradecemos al Ministerio de Educación Pública (MEP), al Instituto Nacional de Estadística y Censos (INEC) y a la Superintendencia de Telecomunicaciones (SUTEL) por proporcionar los datos públicos que hicieron posible este análisis.

REFERENCIAS BIBLIOGRÁFICAS

- [1] **P. M. Morales Vallejo, “Correlación y regresión, simple y múltiple,” Universidad Pontificia Comillas, Madrid, 2012.**
- [2] **J. Dagnino, “Regresión lineal,” *Rev. chil. anest*, vol. 43, no. 2, pp. 143–149, 2014.**
- [3] **I. M. Peláez, “Modelos de regresión: lineal simple y regresión logística,” *Revista Seden*, vol. 14, pp. 195–214, 2016.**
- [4] **J. Cortés, N. Bielsa, E. Cobo, P. Muñoz, y A. González, “Regresión lineal simple,” *Universitat Politècnica de Catalunya*, pp. 3–35, 2015.**
- [5] **Eka en Línea, “Estos son los 50 mejores colegios de Costa Rica según las pruebas del MEP,” 2016. [Online]. Disponible: <https://ekaenlinea.com/estos-son-los-50-mejores-colegios-de-costa-rica-segun-las-pruebas-del-mep/>**
- [6] **Superintendencia de Telecomunicaciones (SUTEL), “86 de cada 100 habitantes tiene acceso a Internet,” 2020. [Online]. Disponible: <https://www.sutel.go.cr/noticias/comunicados-de-prensa/86-de-cada-100-habitantes-tiene-acceso-internet>**
- [7] **Instituto Nacional de Estadística y Censos (INEC), “Encuesta Nacional de Hogares (ENAHOG),” 2019.**