

# App para aprender idiomas

## Parcial Computacional - MEFE 2025

Bautista Salvatierra Pérez  
DNI: 44892610 | LU: 529/22  
bautistasalvatierra@gmail.com  
[Link al Colab](#)

Departamento de Física, Facultad de Ciencias Exactas y Naturales - Universidad de Buenos Aires  
31 de julio de 2025

### Resumen

Este trabajo presenta un enfoque estadístico para evaluar el impacto de una aplicación móvil sobre la retención de vocabulario en estudiantes de idiomas. A partir de un diseño experimental simulado, se modeló el desempeño de los usuarios mediante variables binomiales y se construyó un test de hipótesis basado en la diferencia de log-verosimilitudes. La distribución nula del estadístico fue estimada por simulaciones Monte Carlo, lo que permitió definir umbrales críticos y estimar la potencia del test. Los resultados muestran que, con 600 observaciones, el método propuesto permite detectar mejoras pequeñas ( $f \gtrsim 0,015$ ) con alta probabilidad. Estos hallazgos sugieren que la aplicación produce un efecto significativo en la retención y validan el enfoque estadístico adoptado para cuantificar su efecto.

### Introducción

En los últimos años, el uso de aplicaciones móviles para el aprendizaje de idiomas ha experimentado un crecimiento notable. No obstante, la efectividad real de estas herramientas en la mejora del rendimiento cognitivo sigue siendo objeto de estudio. En este trabajo se evalúa cuantitativamente el impacto de una nueva aplicación móvil diseñada para mejorar la retención de vocabulario en estudiantes de idiomas extranjeros mediante un test de hipótesis basado en la comparación de distribuciones binomiales.

Para ello, se llevó a cabo un experimento controlado con dos grupos: un grupo experimental que utilizó la aplicación durante un mes y un grupo control que empleó métodos tradicionales. Cada participante completó 20 pruebas diarias, en las cuales debía recordar 10 palabras nuevas, generando aproximadamente 600 observaciones por grupo. El análisis estadístico se basa en estas observaciones simuladas.

### Metodología

Se modeló la cantidad de palabras recordadas correctamente en cada prueba como una variable aleatoria binomial

$$X \sim \text{Bin}(n = 10, p), \quad (1)$$

donde  $p$  representa la probabilidad de recordar correctamente una palabra. El objetivo del análisis es determinar si el uso de la aplicación genera un aumento significativo en dicha probabilidad. Para ello, se plantea un test de hipótesis:

- $H_0 : p = p_0$  (sin mejora),
- $H_1 : p = p_0 + f$ , con  $f > 0$  (mejora en la retención),

donde  $p_0$  es estimado a partir de los datos simulados del grupo control (se utilizó  $p = 0,6$  para la simulación).

Para llevar a cabo el test se empleó como estadístico la diferencia de log-verosimilitudes entre los modelos bajo cada hipótesis:

$$\Delta \log \mathcal{L} = \log \mathcal{L}(X|p_0 + f) - \log \mathcal{L}(X|p_0), \quad (2)$$

donde la log-verosimilitud para un conjunto de observaciones  $X = \{X_1, \dots, X_N\}$  se define como:

$$\log \mathcal{L} = \sum_{i=1}^N [X_i \log p + (n - X_i) \log(1 - p)]. \quad (3)$$

En nuestro caso,  $n = 10$  es la cantidad de palabras por prueba y  $N = 600$  es el número total de pruebas por grupo.

La distribución nula del estadístico  $\Delta \log \mathcal{L}$  se obtuvo mediante simulaciones Monte Carlo generando conjuntos de datos bajo  $H_0$  (es decir, con  $p = p_0$ ). A partir de esta distribución empírica se determinó el valor crítico correspondiente a un nivel de significancia  $\alpha = 0,05$ .

Posteriormente, se estimó la potencia del test para distintos valores de  $f$  evaluando la proporción de veces que el estadístico supera dicho valor crítico cuando los datos se generan bajo  $H_1$ . También se calcularon los errores tipo I y tipo II asociados a cada configuración simulada [1].

Finalmente, se exploró un aspecto de la sensibilidad del test: la cantidad mínima de pruebas necesarias para detectar distintos tamaños de efecto con ese mismo umbral

de potencia. Asimismo, se realizó una simulación para distintos valores intermedio de  $f$  a fin evaluar si la aplicación genera un efecto estadísticamente significativo en la retención de vocabulario.

## Análisis y resultados

A partir de las simulaciones realizadas, se obtuvo una estimación de la probabilidad de recordar correctamente una palabra bajo la hipótesis nula:  $p_0 = 0,6007 \pm 0,0064$ , a partir de datos generados con  $p = 0,6$ . Esta estimación se utilizó como base para construir todos los tests de hipótesis posteriores. Esta estimación se utilizó en todo el análisis posterior como base para construir los tests de hipótesis.

En la Figura 1 se muestra la log-verosimilitud acumulada  $\log \mathcal{L}(X | p)$  bajo distintos valores de  $p = p_0 + f$ . En la figura se observa cómo la evidencia a favor de  $H_1$  se acumula progresivamente. Para  $f = 0,05$ , la log-verosimilitud acumulada alcanza valores significativamente mayores, mientras que para  $f = 0,01$  la acumulación es más lenta, reflejando la dificultad para detectar efectos pequeños. El recuadro interior muestra el comportamiento en las últimas 50 pruebas, donde se observa mayor separación entre curvas. Esto sugiere que un estadístico sensible al efecto podría ser la diferencia de log-verosimilitudes entre ambos modelos.

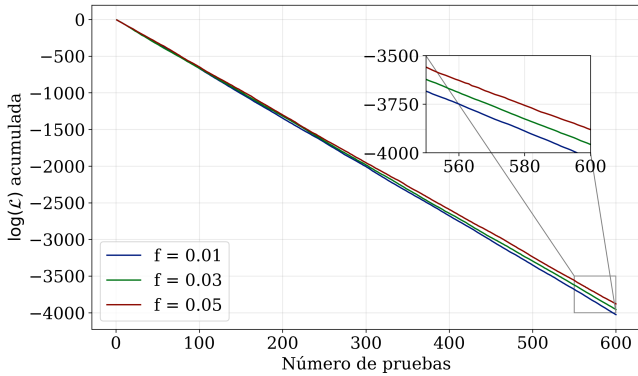


Figura 1: Log-verosimilitud acumulada para distintos valores de  $p = p_0 + f$ .

De este modo, en la Figura 2 se enseña el incremento acumulado en  $\Delta \log(\mathcal{L})$  para diferentes tamaños de efecto. Ahora se puede observar una tendencia sistemática más clara a favor de  $H_1$  a medida que aumentan las pruebas. A partir de este comportamiento se adoptó  $T = \Delta \log \mathcal{L}$  como estadístico de test.

Luego, en la Figura 3 se comparan las distribuciones simuladas del estadístico bajo  $H_0$  y  $H_1$ , para  $f = 0,015$ . La distribución bajo  $H_1$  aparece desplazada hacia valores más altos. A partir del percentil 95 de la distribución nula se definió el valor crítico  $T_c = 1,2$ , correspondiente a un nivel de significancia  $\alpha = 0,05$ . Se establece entonces la regla de decisión: si  $T > T_c$ , se rechaza  $H_0$ ; en caso contrario, no

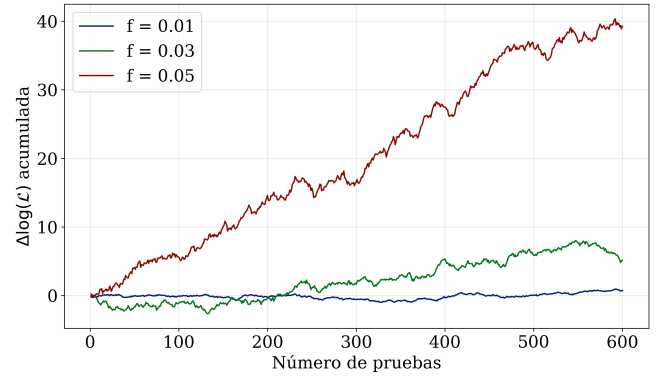


Figura 2: Evolución acumulada del estadístico  $\Delta \log \mathcal{L}$  para distintos  $f$ .

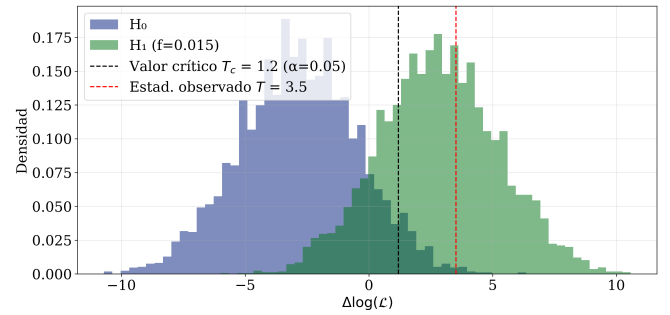


Figura 3: Distribuciones simuladas de  $\Delta \log \mathcal{L}$  bajo  $H_0$  y  $H_1$ , con  $f = 0,015$ .

se rechaza. En este ejemplo, el valor observado supera  $T_c$ , por lo que se rechaza  $H_0$ .

A partir de este test, se estimó la potencia del test (la definimos  $P(\text{Rechazar } H_0 \mid H_1 \text{ verdadera})$ ). La Figura 4 muestra la curva de potencia estimada del test para valores de  $f$  en  $[0, 0,03]$ . La potencia aumenta rápidamente con  $f$ , alcanzando el umbral del 80 % para  $f \approx 0,015$ . También se graficó el error tipo II, definido como  $1 - \text{potencia}$  (se acepta la hipótesis nula cuando esta es falsa). Por otro lado, la probabilidad de cometer errores de tipo I (se rechaza la hipótesis nula cuando esta es verdadera). A través de la simulación se obtuvo que 0.0512. Era esperado que el resultado sea similar al nivel de significancia utilizado para la simulación  $\alpha = 0,05$ .

En la Figura 5, se muestra el tamaño muestral mínimo necesario para distintos valores de  $f$ . El análisis de sensibilidad permite evaluar qué combinación de tamaño de efecto y cantidad de pruebas es necesaria para detectar un efecto con al menos 80 % de potencia. El valor actual de  $n = 600$  permite detectar efectivamente efectos de magnitud  $f \geq 0,015$  aproximadamente.

Finalmente, se presenta en la Figura 6 una visualización detallada de las distribuciones bajo  $H_0$  y  $H_1$  para distintos valores de  $f$ , junto con el valor crítico, el estadístico observado y la decisión del test. Esta visualización permite evaluar gráficamente la separación entre ambas hipótesis y la sensibilidad del estadístico ante pequeñas variaciones

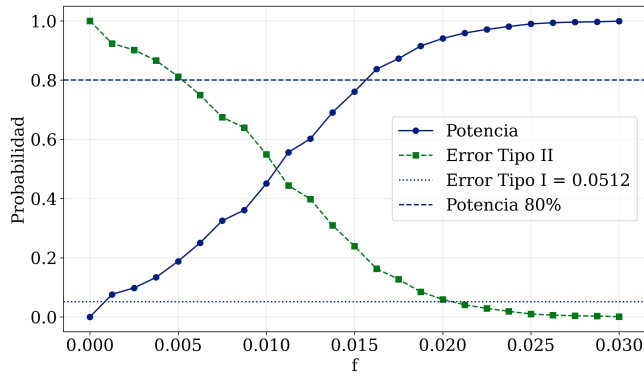


Figura 4: Curvas de potencia y error tipo II del test en función del tamaño del efecto  $f$ .

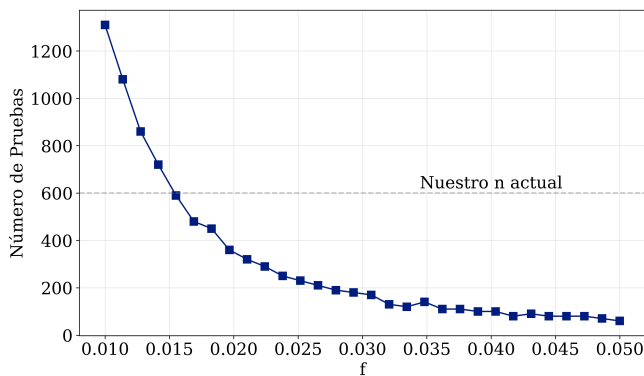


Figura 5: Cantidad mínima de pruebas necesarias para alcanzar 80 % de potencia, en función de  $f$ .

en  $f$ . Se observa que para valores de  $f \geq 0,015$ , el test tiende a rechazar  $H_0$  con alta probabilidad. El único detalle a destacar es que, para visualizar las distribuciones bajo  $H_0$  y  $H_1$ , se utilizó la función `gaussian_kde`, que estima curvas de densidad suaves a partir de los datos simulados. Este método coloca un kernel gaussiano sobre cada punto y los suma, generando una representación continua de las distribuciones. Al evitar el ruido de los histogramas y ajustar automáticamente el ancho de banda, permite comparar con mayor claridad las regiones críticas y la separación entre hipótesis.

## Conclusión

En este trabajo se desarrolló un enfoque estadístico para evaluar el efecto de una aplicación móvil sobre la retención de vocabulario en estudiantes de idiomas. A partir de un diseño experimental simulado, se modeló el desempeño de los participantes mediante variables binomiales y se propuso un test de hipótesis basado en la diferencia de log-verosimilitudes entre los modelos bajo  $H_0$  y  $H_1$ . La distribución nula del estadístico fue estimada mediante simulaciones Monte Carlo, lo que permitió definir un umbral crítico y calcular la potencia del test para distintos

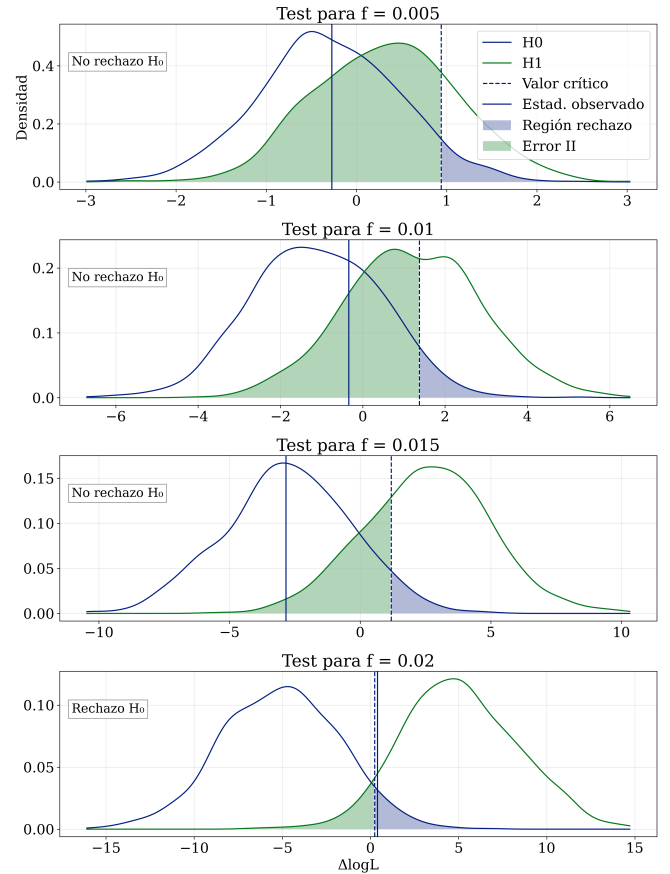


Figura 6: Comparación entre distribuciones bajo  $H_0$  y  $H_1$  para distintos tamaños de efecto.

tamaños de efecto.

El análisis reveló que el estadístico adoptado acumula evidencia de manera consistente a favor de la hipótesis alternativa cuando existe una mejora real en la retención. Se comprobó que, con 600 observaciones, es posible detectar efectos de magnitud  $f \gtrsim 0,015$  con al menos un 80 % de potencia, y que el error tipo I se mantiene alineado con el nivel de significancia nominal. Asimismo, se analizó la sensibilidad del test respecto del tamaño muestral y se validó visualmente su desempeño mediante comparaciones directas entre las distribuciones simuladas. En conjunto, los resultados respaldan tanto la eficacia de la aplicación como la robustez del enfoque metodológico utilizado para cuantificar su impacto.

## Referencias

- [1] Frodesen, A. G., Skjeggstad, O., & Tøfte, H. (1979). *Probability and Statistics in Particle Physics*. Columbia University Press.