

Análisis Univariante: VIH en Malawi

Prof. Miguel Lafuente Blasco

1. Contexto del estudio

El conjunto de datos `Malawi_VIH.csv` procede de un experimento de campo realizado en zonas rurales del sur de Malawi durante 2004–2006. En aquella época, la prevalencia nacional de VIH superaba el 12 %, y el gobierno impulsaba estrategias para mejorar el diagnóstico temprano y el seguimiento clínico.

En este contexto se enmarca el estudio de Rebecca L. Thornton, que combinó pruebas de VIH realizadas a domicilio con dos intervenciones aleatorias diseñadas para aumentar la probabilidad de que los participantes acudieran voluntariamente a conocer sus resultados en un centro de salud:

- (I) **Distancia al centro de resultados.** A cada participante se le asignó aleatoriamente uno de varios centros de asesoramiento y entrega de resultados (VCT), de forma que la distancia a recorrer varió exógenamente entre unos pocos metros y más de una hora a pie.
- (II) **Incentivos económicos.** De manera independiente, se ofrecieron cupones en efectivo de cuantías muy reducidas —entre 0 y 3 USD al tipo de cambio de la época— que solo se entregaban si la persona se presentaba en el VCT para conocer su diagnóstico.

En total se entrevistó y se tomó una muestra de sangre a 2894 adultos. La base de datos incluye información sobre las siguientes variables:

1. **edad:** Edad del participante, en años.
2. **distancia_centro_medico_km:** Distancia (en kilómetros) desde la residencia del participante hasta el centro médico asignado.
3. **recibio_incentivo:** Indica si la persona recibió un incentivo monetario para acudir a conocer sus resultados (sí/no).
4. **incentivo_total:** Monto total del incentivo ofrecido, expresado en moneda local.
5. **obtuvo_resultado:** Indica si la persona acudió al centro médico y obtuvo su resultado de la prueba de VIH (sí/no).
6. **resultado_vih:** Resultado de la prueba de VIH (positivo/negativo).

Fuente: Thornton, Rebecca L. (2008). *The demand for, and impact of, learning HIV status.* *American Economic Review*, 98(5), 1829–1863. <https://doi.org/10.1257/aer.98.5.1829>

Nota: En este documento se realiza únicamente un análisis univariante con el objetivo de describir las principales características individuales de los participantes y las variables clave del estudio. El análisis de relaciones entre variables —incluidas las causas y consecuencias de la intervención— se abordará en un segundo documento, centrado en técnicas de análisis bivariante.

2. Cuestiones univariantes

1. Edad

- ¿Qué edad tiene el participante más joven? ¿Y el de mayor edad?
- ¿Entre qué edades se concentra la mayor parte de la muestra?
- Si agrupamos por décadas (20–29, 30–39, ...), ¿cuál es la franja con más participantes?
- ¿Hay más personas menores de 30 años que personas de 50 años o más?
- ¿Qué edad exacta se repite con mayor frecuencia?
- A partir de los resultados, ¿dirías que la mayoría se encuentra en la adultez joven, madura o avanzada?
- Calcula e interpreta el coeficiente de asimetría de la distribución de edades.
- Describe, en términos reales, a los participantes que están en el cuartil inferior (Q_1) de edad.
- Si tuvieses que describir al participante promedio en términos de edad, ¿cómo lo harías?
- ¿Qué edad supera solo el 15 % de los participantes?

2. Distancia al Centro Médico (km)

- Representa esta variable gráficamente.
- ¿Cuál es la distancia media que los participantes deben recorrer para recibir sus resultados?
- ¿Cuál es la distancia mínima y máxima registrada en el conjunto de datos?
- ¿Existe una distancia que se repita con mayor frecuencia?
- Si agrupas en intervalos de 1 km, ¿qué rango es el más común?
- ¿Qué porcentaje vive a menos de 1 km del centro VCT?
Sugerencia: agrupa en intervalos de 1 km a partir de 0.
- Calcula el coeficiente de variación de esta variable e interpreta su valor.
- Calcula una medida de asimetría e interpreta que quiere decir esta medida en términos reales.
- Resume brevemente cómo es la distancia al centro médico de estas personas.

3. Recibió Incentivo (sí/no)

- ¿Qué porcentaje de los participantes recibió algún tipo de incentivo?
- Realiza un gráfico para mostrar la proporción de personas que recibieron y que no recibieron el incentivo.
- Escribe una única frase, fácil de leer, para explicar la información contenida en esta variable.

4. Incentivo Total

- ¿Cuál es el monto medio ofrecido?

- (b) ¿Cuáles son el mínimo y el máximo?
- (c) ¿Cuál es el valor de incentivo más común?
- (d) Representa los montos de incentivo mediante un histograma y comenta su distribución.
- (e) ¿Cuál es el percentil 25 del incentivo total ofrecido?
- (f) Calcula e interpreta la mediana de los incentivos.
- (g) ¿Qué indica el coeficiente de variación sobre la dispersión de los incentivos?

5. **Obtuvo Resultado (sí/no)**

- (a) ¿Es más frecuente que los participantes hayan optado por obtener sus resultados o que hayan preferido no hacerlo?
- (b) ¿Qué porcentaje de los participantes decidió conocer sus resultados de VIH?
- (c) ¿Cuántas personas recogieron efectivamente su resultado?
- (d) Representa la variable gráficamente y comenta lo observado.

6. **Resultado de VIH (positivo/negativo)**

- (a) ¿Cuántas personas fueron testadas?
- (b) ¿Qué porcentaje de los participantes tuvo un resultado negativo en la prueba de VIH?
- (c) ¿Cuántos participantes tuvieron un resultado positivo en la prueba de VIH?
- (d) Realiza un gráfico de barras que muestre la proporción de resultados positivos y negativos.
- (e) Ídem con un gráfico de sectores. Usa los colores adecuadamente.
- (f) ¿Qué conclusión se puede extraer sobre la prevalencia de resultados positivos en este grupo?