

## *Análisis Bivariante: 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 2 894 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).

De esta manera, se pretende conocer si un incentivo, por pequeño que sea, promueve que las personas acudan a conocer sus resultados médicos, lo que podría mejorar el seguimiento y tratamiento temprano del VIH, ayudando a frenar la propagación y reducir el impacto de la enfermedad en la comunidad.

**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:** Este documento constituye la segunda parte del análisis de este conjunto de datos. Aquí se abordan relaciones entre variables mediante técnicas de análisis bivariante. Para la descripción univariante de las variables y el contexto general del estudio, consúltese el primer documento relativo al análisis univariante.

## 2. Cuestiones

### 2.1. Evaluando el buen diseño del estudio: independencia entre incentivo y distancia al centro médico

Examina si la asignación de incentivos está relacionada con la distancia al centro médico.

1. Representa la variable `distancia_centro_medico_km` con un diagrama de cajas, comparando entre quienes recibieron incentivo y quienes no.
2. Si el experimento está correctamente diseñado, la asignación de incentivos debe ser independiente de la distancia al centro médico. De lo contrario, podrían introducirse sesgos en la interpretación de resultados: por ejemplo, si las personas que viven más lejos tienen más (o menos) probabilidades de recibir incentivos, no se estaría evaluando de forma aislada el efecto real de cada variable sobre el comportamiento de acudir o no al centro.

Calcula y compara medidas de síntesis (media, mediana, dispersión) de la distancia en ambos grupos (con y sin incentivo) y argumenta si parece haber diferencias relevantes. Concluye si, desde este punto de vista, el diseño experimental puede considerarse aleatorio y equilibrado.

### 2.2. Evaluando el buen diseño del estudio: descartando sesgos en el incentivo según resultado de VIH

Examina si existe una relación entre el resultado de la prueba de VIH y el monto del incentivo ofrecido. El objetivo es comprobar si hubo algún sesgo en la asignación de incentivos en función del estado de salud de los participantes, es decir, si quienes dieron positivo o negativo recibieron sistemáticamente incentivos diferentes.

1. Calcula la media del `incentivo_total` por separado para las personas con resultado positivo y negativo en la prueba de VIH.
2. Elabora un gráfico comparativo que represente visualmente ambas medias.
3. En un máximo de tres frases, interpreta si existe alguna diferencia relevante entre los grupos y qué podría implicar en términos de diseño o equidad del experimento.

### 2.3. Efecto del incentivo sobre la recogida de resultados

Analiza si recibir un incentivo influye en la decisión de acudir a recoger el resultado de la prueba de VIH.

1. Elabora una tabla de doble entrada que muestre el número de personas según las combinaciones de las variables `recibio_incentivo` y `obtuvo_resultado`.
2. Presenta una tabla de frecuencias relativas condicionadas que permita evaluar la proporción de personas que obtuvieron su resultado en cada uno de los grupos (con y sin incentivo).
3. Representa gráficamente la proporción de personas que recogieron su resultado según si recibieron incentivo o no (por ejemplo, con un gráfico de barras agrupadas o apiladas).
4. Interpreta la relación observada entre ambas variables en un máximo de tres frases, explicando en términos reales si los incentivos parecen haber tenido efecto sobre la decisión de acudir al centro médico.

#### 2.4. Relación entre el monto del incentivo y la recogida de resultados

Para analizar con más detalle el efecto del incentivo, estudia la relación entre el monto del incentivo ofrecido y la decisión de acudir a recoger el resultado de la prueba de VIH.

1. Calcula la media del `incentivo_total` por separado para quienes sí obtuvieron su resultado y para quienes no lo hicieron.
2. Representa gráficamente estas medias con un gráfico comparativo adecuado.
3. Interpreta en términos reales si la magnitud del incentivo parece tener un impacto sobre la decisión de acudir al centro médico a conocer el resultado.

#### 2.5. Relación entre la distancia y la recogida del resultado

Explora si la distancia al centro médico está relacionada con la probabilidad de acudir a recoger el resultado de la prueba de VIH. El objetivo es evaluar si la accesibilidad física influye en la decisión de los participantes.

1. Calcula la distancia media al centro médico por separado para quienes sí obtuvieron su resultado y para quienes no lo hicieron.
2. Representa gráficamente ambas distribuciones con un gráfico adecuado.
3. Interpreta en términos reales si parece existir una relación entre la distancia y la decisión de acudir al centro médico.
4. Repite el análisis utilizando los cuartiles de distancia y representa los resultados mediante un diagrama de cajas. Comenta si el patrón observado se mantiene.

#### 2.6. Sospechas previas y sociología: Estado serológico y comportamiento frente al diagnóstico

Analiza si existe una relación entre el resultado de la prueba de VIH y la decisión de acudir a recoger dicho resultado. El objetivo es explorar si el estado serológico influye en el comportamiento del participante respecto a la búsqueda de información sobre su salud.

1. Construye una tabla de contingencia entre las variables `resultado_vih` y `obtuvo_resultado`.
2. ¿Qué porcentaje de las personas que obtuvieron sus resultados dieron positivo? ¿Y de las que no los recogieron?
3. Representa visualmente las proporciones del apartado anterior con un gráfico adecuado, que facilite la comparación entre los dos grupos.
4. A partir de los datos estudiados, analiza si las personas con resultado positivo en la prueba de VIH tienden a recoger sus resultados con mayor frecuencia (lo que podría indicar una sospecha previa sobre su estado de salud y una mayor disposición a buscar ayuda), con menor frecuencia (posiblemente por temor al diagnóstico o por estilos de vida menos orientados al cuidado personal), o con una frecuencia similar a quienes obtuvieron un resultado negativo. Expón tu conclusión en términos comprensibles y conectados con el contexto del estudio.

### 3. Conclusiones

En cada uno de los apartados anteriores se ha estudiado un aspecto de un experimento de una intervención médica y sociológica. Resume, de manera sencilla y accesible, las principales conclusiones obtenidas en todos los análisis realizados mediante un breve discurso coherente.