

Desigualdades en salud: Estudio de la asociación entre conductas saludables y nivel socioeconómico a partir de la Encuesta Nacional de Factores de Riesgo

Weill Martina, Fernández Shanahan Tomás, Peralta Martínez Ramón

Introducción

Las enfermedades crónicas no transmisibles (ENT) - e.g., Diabetes, Cáncer, Enfermedades cardiovasculares, etc. - no tienen una única causa y los determinantes que las producen son múltiples. Estos determinantes pertenecen a distintos niveles de organización y van desde lo molecular hasta lo social. El conocimiento de estos determinantes tiene implicancias en la prevención y el control de la enfermedad (Susser y Susser 1996; Susser 2004).

En los últimos años se ha enfatizado la relevancia de los factores sociales como determinantes de enfermedades a través de múltiples factores de riesgo (Phelan et al. 2010; Marmot and Allen 2014). Existe fuerte evidencia de que la mayoría de los factores de riesgo de ENT están estructurados a lo largo del gradiente socioeconómico (Di Cesare et al. 2013). Las ENT afectan en mayor proporción a países de bajos ingresos: las muertes prematuras en las edades de mayor productividad desfavorecen el desarrollo económico.

El objetivo de este trabajo es estudiar la asociación entre conductas saludables (en particular, el Nivel de Actividad Física) y el nivel socioeconómico (ajustado por sexo y edad) en individuos mayores de 18 años. La hipótesis planteada es que el Nivel de Actividad Física está asociado con el Nivel Socioeconómico, y se espera que un menor Nivel Socioeconómico esté vinculado con un menor Nivel de Actividad Física. Debido a que la baja actividad física es un factor de riesgo para las ENT, los sectores más sedentarios serían más propensos a contraer enfermedades no transmisibles.

Nivel Socioeconomico---?---> Nivel de Actividad Física -----> ENT

Figura 1: Hipótesis del trabajo.

-----> : Asociacion.

(?): Asociacion que se busca poner a prueba.

Materiales y Métodos

En la Argentina en el año 2005, el INDEC implementó una Encuesta Nacional de Factores de Riesgo (ENFR) que es realizada cada cuatro años. El objetivo de esta encuesta es estudiar los comportamientos y condiciones de vida que pueden afectar la salud de la población de 18 años y más, así como algunas características de las viviendas y de los hogares donde residen. La encuesta alcanza a individuos residentes de localidades urbanas de 5.000 y más habitantes. En la ENFR del año 2018 se seleccionaron al azar 49.170 viviendas de todas las jurisdicciones de la Argentina, y se obtuvieron 29.224 respuestas. Para este trabajo se seleccionaron de la ENFR del 2018 solo aquellas variables que resultaron relevantes para el objetivo principal de este estudio (Nivel de actividad física, Nivel de estudios, Quintil de ingresos, Tipo de vivienda, etc). Se eliminaron todos los individuos que presentaron ausencia de respuesta (NAs) en la base de datos. Esto se decidió al considerar que representaban menos del 1% de la base de datos.

1. Variables del estudio:

a) Variable respuesta (VR): Nivel de actividad física (Bajo vs Medio y Alto). Es una variable cualitativa que puede tomar tres valores (Baja, Medio o Alto). Pertenecerán al nivel de actividad física Alto aquellos encuestados que cumple alguno de los siguientes criterios:

- 3 o más días de actividad física intensa acumulando al menos 1500 MET-minuto por semana.
- 7 o más días de cualquier combinación (caminata, actividad física moderada, actividad física intensa) que acumulen al menos 3000 MET-minutos por semana.

Pertenecerán al nivel de actividad física Medio aquellos encuestados que cumplan alguno de los siguientes criterios:

- 3 o más días de actividad física intensa, al menos 20 minutos
- 5 o más días de actividad física moderada o caminata que acumulen al menos 30 minutos
- 5 o más días de cualquier combinación (caminata, actividad física moderada, actividad física intensa) que acumulen al menos 600 MET-minutos por semana.

Pertenecerán al nivel de actividad física Bajo aquellos encuestados que cumplan el siguiente criterio:

- Sin actividad reportada o no incluido en niveles Medio o Alto.

Estrategia de Análisis: Dicotomizar la VR en “Bajo” y “Medio-Alto”. Así pues, una posible distribución candidata sería la distribución Bernoulli en donde a cada observación independiente se determina cuál condición de Nivel de Actividad Física cumple (si “Bajo” o “Medio-Alto”).

b) Variables explicatorias (VE):

- VE1: Edad del individuo.
- VE2: Sexo del individuo. Representa el sexo del encuestado. Puede tomar dos posibles valores (1 para varón, 2 para mujer).
- VE3: Tipo de vivienda en la que habita el individuo. Representa el estilo de vivienda del encuestado. Puede tomar siete posibles valores (del 1 al 7), correspondientes a las siguientes categorías: Casa, Casilla, Departamento, Pieza de inquilinato, Pieza de hotel o pensión, Local no construido para habitación, Otros
- VE4: Material del piso de la vivienda. Material predominante de los pisos de la vivienda. Puede tomar cuatro posibles valores (del 1 al 4), correspondientes a las siguientes categorías: Cerámica, baldosa, mosaico, mármol, madera o alfombra; Cemento o ladrillo fijo; Tierra o ladrillo suelto; Otros.
- VE5: Baño. Indicador numérico de presencia (1) o ausencia (2) de baño o letrina.
- VE6: Quintil de ingresos del hogar. Indicador numérico de a qué quintil de hogares pertenece según el ingreso por unidad consumidora (toma valores del 1 al 5).
- VE7: Nivel de estudios del individuo. Representa el nivel de formación del encuestado. Puede tomar ocho posibles valores (del 1 al 8), correspondientes a las siguientes categorías: Sin instrucción, Primario incompleto, Primario completo, Secundario incompleto, Secundario completo, Terciario/universitario incompleto, Terciario/universitario completo, Educación especial
- VE8: Tipo de cobertura médica. Representa el nivel de cobertura de salud del encuestado. Puede tomar dos posibles valores (1 si el encuestado tiene obra social, prepaga o servicio de emergencia médica, 2 si sólo tiene cobertura pública).
- VE9: % de hogares con necesidades básicas insatisfechas de la provincia. Representa el porcentaje de hogares, a nivel provincial, que no tienen las necesidades básicas satisfechas.

Estrategia de Análisis: Todas las VE son de carácter Fijo ya que son informativas. Todas las VE serán analizadas como cualitativas salvo Edad.

Estrategia de modelado:

La VR es del tipo dicotómica donde se podría decir que el nivel bajo de actividad física representa un “fracaso” mientras que el nivel medio-alto representa un “éxito”. Para poder realizar un ajuste matemáticamente robusto en una VR con Distribucion Bernoulli (donde hay una cota inferior y una superior) es necesario modelar una transformación de la VR. La función de enlace utilizada en una regresión logística es del tipo Logit, es decir el logaritmo del cociente entre la probabilidad de éxito y la probabilidad de fracaso (odds). Al transformar la VR a odds esta pasa de estar acotada entre 0 y 1 a estar acotada entre 0 y $+\infty$. Al aplicar \ln , se obtiene una nueva variable que puede adoptar cualquier valor real. Así pues, el modelo consta de tres componentes:

- Componente Aleatorio: La VR, en este caso es Nivel de Actividad Física.
- Predictor Lineal: Combinación lineal entre las distintas VE. Dado que en este caso se cuenta con una variable cuantitativa (Edad), se decide parametrizar como una regresión lineal.
- Función de Enlace: Es una función que vincula el Componente Aleatorio (la VR) con el Predictor Lineal. Esta función es fundamental ya que permite modelar variables respuesta que son acotadas. En el caso de las distribuciones Bernoulli la función de enlace es el Logaritmo Natural (\ln).

La clasificación de las VE implicó un desafío analítico, ya que a priori las VE anteriormente planteadas podrían ser colineales o no tener un efecto predictor en el modelo. Se abordó este problema mediante el planteo de Modelos Univariados; modelos que contienen una VE (sumando además Sexo y Edad, variables control). Se plantearon 5 modelos univariados y a cada uno se le realizó un análisis mediante la función `drop1` de la librería `car`. Esta función realiza una prueba de hipótesis de diferencia de devianzas, comparando la devianza del modelo planteado con la devianza del modelo sin cada una de las VE en cuestión (secuencialmente). Si la prueba para una dada VE es significativa se concluye que la devianza del modelo inicialmente planteado difiere de la del modelo sin esa VE. De lo que se desprende que esa VE aporta para el mejor ajuste del modelo (disminuye la devianza del modelo).

Los modelos univariados permitieron determinar cuáles VE tendrían un efecto sobre la VR. El siguiente paso fue determinar si hay colinealidad (asociación) entre las VE seleccionadas. Para esto se realizó un análisis con la función `drop1` solamente que en este caso en vez de realizarlo con modelos univariados se plantearon modelos con las VE Sexo, Edad, NBI y alguna otra VE. NBI es una variable del bloque provincial, al ver que NBI tenía un efecto sobre la VR fue la primera VE no control incluida en el modelo. Esto fue así ya que NBI es una variable importante que era de interés para el estudio. Por lo tanto, si la prueba de `drop1` para una dada VE es significativa se concluye que la devianza del modelo inicialmente planteado difiere de la del modelo sin esa VE. Por incumplimiento del supuesto de linealidad se convirtió a la variable NBI, originalmente cuantitativa, en una cualitativa de tres niveles, donde cada nivel representa a un tercio de las provincias.

2. Modelos univariados - Estrategia para la selección de VE:

Para determinar cuáles VE (de las mencionadas previamente) incluir en el modelo se debió realizar una selección de VE. Se plantearon modelos Univariados; modelos que presentan una VE (además de Sexo y Edad que son las VE control). El objetivo fue determinar mediante la función `drop1` aquellas VE que resultan significativas. Luego se analizaron modelos sumando variables, el objetivo aquí fue descartar variables que sean colineales. De esta manera se llegó finalmente a un modelo aditivo final. No se plantearon interacciones debido a que este fue un estudio observacional por lo tanto solo se deberían plantear aquellas interacciones que sean de interés, en nuestro caso no se propuso ninguna interacción a priori.

3. Primera aproximación a los datos y edición:

Como primer paso se debió dicotomizar la variable respuesta (nivel de actividad física) pasando de tres niveles (actividad baja, media y alta) a dos (sin actividad / con actividad). Para esto se agrupó actividad

baja por un lado y media y alta por otro, quedando así la variable respuesta con dos niveles - Sin Actividad y Con Actividad respectivamente.

Al analizar nuestros datos podemos observar que hay 254 personas que no reportaron la actividad física (NAs), se puede también ver que no hay valores faltantes en las otras variables explicatorias. Las filas con NAs fueron eliminadas.

```
## nivel_actividad_fisica  provincia          Edad          Sexo
## 0 :13764                Length:29224      Min.   : 18.00    Min.   :1.000
## 1 :15206                Class :character  1st Qu.: 32.00    1st Qu.:1.000
## NA's: 254              Mode  :character  Median : 44.00    Median :2.000
##                               Mean    : 46.49    Mean   :1.569
##                               3rd Qu.: 61.00    3rd Qu.:2.000
##                               Max.    :104.00    Max.   :2.000
##
## tipo_vivienda  material_piso_viv  banio      quintil_hogar
## Min.   :1.000  Min.   :1.000  Min.   :1.000  Min.   :1.000
## 1st Qu.:1.000  1st Qu.:1.000  1st Qu.:1.000  1st Qu.:2.000
## Median :1.000  Median :1.000  Median :1.000  Median :3.000
## Mean   :1.308  Mean   :1.153  Mean   :1.004  Mean   :3.007
## 3rd Qu.:1.000  3rd Qu.:1.000  3rd Qu.:1.000  3rd Qu.:4.000
## Max.   :7.000  Max.   :4.000  Max.   :2.000  Max.   :5.000
##
## nivel_estudios  tipo_cobertura  agua_a_traves_de  nivel_actividad_fisica_num
## Min.   :1.000  Min.   :1.000  Min.   :1.00    Min.   :0.0000
## 1st Qu.:3.000  1st Qu.:1.000  1st Qu.:1.00    1st Qu.:0.0000
## Median :5.000  Median :1.000  Median :1.00    Median :1.0000
## Mean   :4.674  Mean   :1.305  Mean   :1.07    Mean   :0.5249
## 3rd Qu.:6.000  3rd Qu.:2.000  3rd Qu.:1.00    3rd Qu.:1.0000
## Max.   :8.000  Max.   :2.000  Max.   :5.00    Max.   :1.0000
##                               NA's    :254
## porcentaje
## Min.   : 3.788
## 1st Qu.: 7.878
## Median : 8.141
## Mean   :10.523
## 3rd Qu.:14.242
## Max.   :19.731
##
```

La variable “nivel de estudios” tiene 8 categorías. Sin embargo muchas categorías tienen un muy bajo nivel de casos, por ende se decidió reestructurar los datos de manera que haya tres categorías. Estas categorías fueron creadas a partir de los resultados de la ENFR en los que se agruparon las conclusiones de esta manera:

-Categoría 1: sin instrucción, primario incompleto.

-Categoría 2: primario completo, secundario incompleto

-Categoría 3: secundario completo, universitario incompleto, universitario completo

La categoría “Educación especial” se eliminó debido a que no correspondía a ninguna de las categorías generadas (se desconoce hasta que nivel de estudios llegaron) y el número de individuos era muy bajo (45).

Por otro lado, dado que se cuentan con muchas variables que describen el nivel socioeconómico a nivel Hogar/Vivienda, se decidió crear una nueva variable que actuará como un indicador de vulnerabilidad a nivel Hogar/Vivienda. Aquel individuo que cumpla con al menos una de las siguientes condiciones será considerado como Vulnerable (1) en el nivel Hogar/Vivienda.

- Si tipo_vivienda == 6 o 2 (local no construido para habitacion o casilla)
- Si material_piso_viv == 3 (tierra o ladrillo suelto)
- Si agua_a_traves_de == 3 o 4 (obtiene agua a traves de perforacion con bomba manual / aljibe o pozo).
- Si baño == 2 (no tiene baño)

Por incumplimiento del supuesto de linealidad (apéndice) se convirtió a la variable NBI, originalmente cuantitativa, en una cualitativa de tres niveles, donde cada nivel representa a un tercil, siendo el primer tercil el de menor NBI, y el tercer tercil el de mayor.

Resultados:

1. Estadística Descriptiva:

Con el objetivo de desarrollar una mejor comprensión de los datos disponibles se utilizaron distintas herramientas gráficas para poder determinar el comportamiento cualitativo de las distintas variables. Como primer paso se buscó observar la variación en el nivel de actividad física de los individuos dependiendo del sexo (Figura 1). Se observa un mayor porcentaje de individuos varones que realizan actividad física respecto a las mujeres.

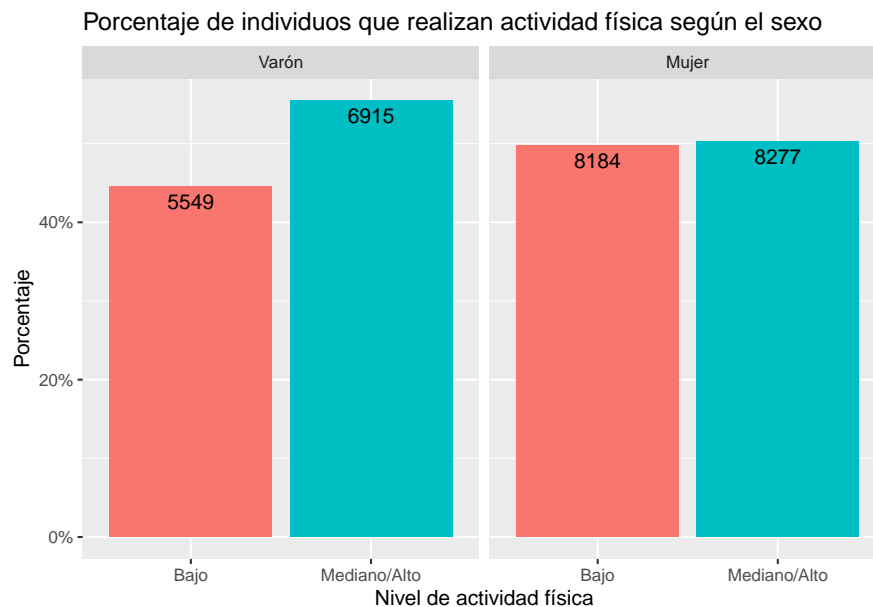


Figura 1: Porcentaje de individuos que realizan actividad física según el sexo. Indicado dentro de cada barra está el número de individuos pertenecientes a esa categoría.

Luego, siguiendo con la hipótesis de este trabajo, se buscó observar la tendencia general del porcentaje de individuos que realizan actividad física al incrementar el poder económico. Para esto se realizaron dos análisis: en el primero se estudió el poder económico. En la Figura 2 se observa el porcentaje de individuos que realizan o no realizan actividad física en función del quintil de ingresos del hogar. La tendencia, más notoria a partir del cuarto quintil, es el incremento del porcentaje de individuos que realizan actividad física a medida que aumenta el poder económico del hogar. El segundo quintil es el único con una mayor proporción de individuos que no realizan actividad física.

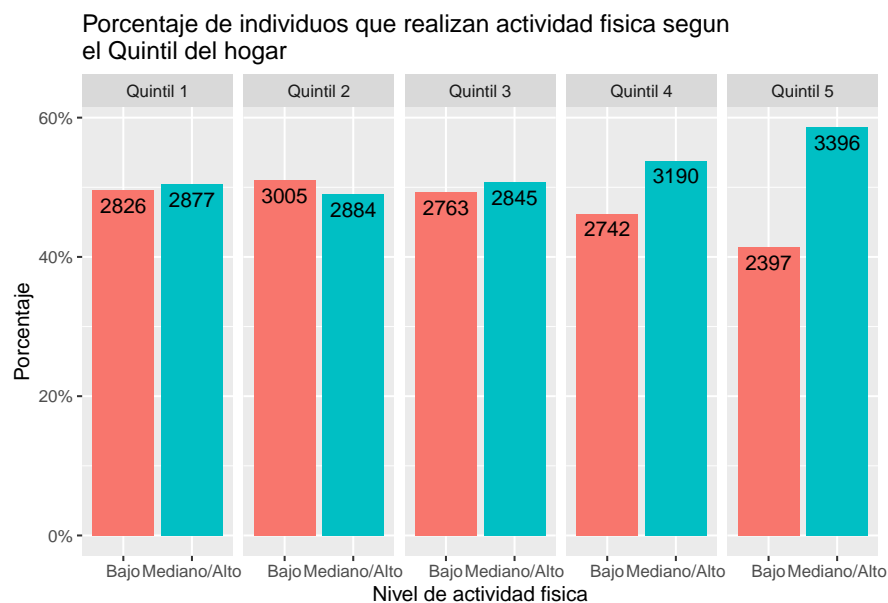


Figura 2: Porcentaje de individuos que realizan actividad física para cada quintil. Indicado dentro de cada barra está el número de individuos pertenecientes a esa categoría.

Como última estrategia se realizó un análisis similar pero utilizando un indicador socio-cultural, el nivel de formación del encuestado (nivel de estudios). El resultado se puede observar en la Figura 3: hay una tendencia de incremento en el porcentaje de individuos que realizan actividad física al incrementar la formación del encuestado. Para los niveles de instrucción más bajos el porcentaje de individuos que no realizan actividad física es mayor al de individuos que si realizan, mientras que para el nivel más alto esta tendencia se revierte.

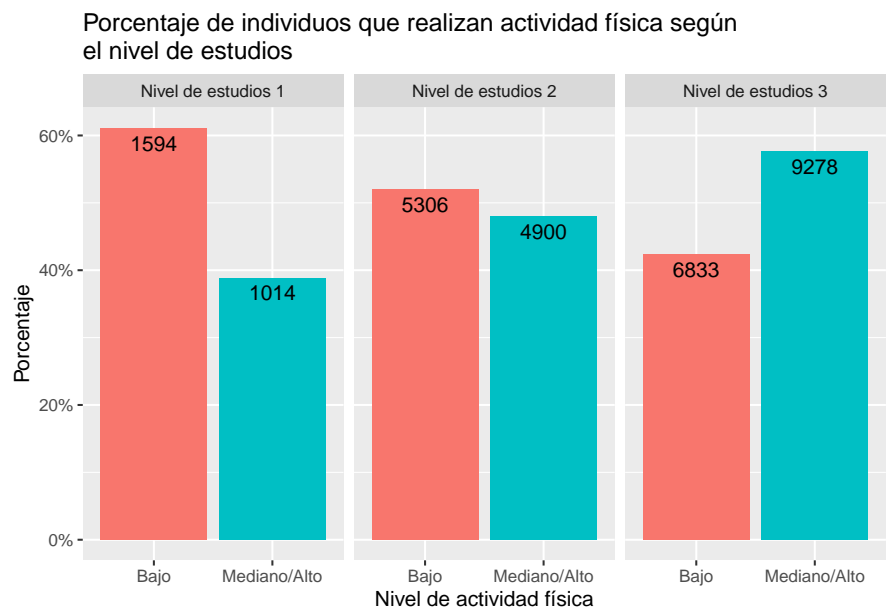


Figura 3: Porcentaje de individuos que realizan actividad física para cada nivel de estudios. Indicado dentro de cada barra está el número de individuos pertenecientes a esa categoría.

En los tres análisis presentados se puede observar una tendencia general de un incremento en la actividad física al aumentar el nivel socioeconómico de los individuos, lo cual apoya la hipótesis planteada. Teniendo esto en cuenta se procedió con el modelado de los datos.

2. Análisis secuencial de modelos:

Habiendo hecho un análisis descriptivo preliminar, y con evidencia cualitativa que apoya la hipótesis planteada, se procedió a realizar un análisis secuencial de los distintos modelos aditivos. Se fijaron las variables control Sexo y Edad. La encuesta presenta variables a distintos niveles (provincial, hogar, individual), se incluyó a la variable NBI como una de pseudo control, con la cual se buscaba incluir en el modelo el nivel económico a nivel provincial, el mayor nivel presente en la base de datos. Los resultados de este análisis se pueden observar en la tabla 1. Se consideraron cumplidos los supuestos de estos modelos (mirar apéndice).

Tabla 1: Resultados obtenidos del análisis secuencial de los modelos aditivos, incluye P-Valores resultado del análisis por la función drop1 y el AIC asociado a ese modelo en particular.

Modelo	VE fijada	p Values (drop1)							
		AIC del modelo	Sexo	Edad	NBI	Vulnerable Hogar	Quintil	Nivel de estudios	Tipo cobertura
m	Sexo y Edad	39090	2.00E-16	2.00E-16					
m1		38934	5.30E-15	2.00E-16	2.20E-16				
m2		39078	3.17E-15	2.00E-16		0.0002256			
m3		38907	1.82E-12	2.00E-16			2.00E-16		
m4		38909	2.00E-16	2.00E-16				2.00E-16	
m5		39018	4.41E-16	2.00E-16					2.00E-16
m6	Sexo, Edad y NBI	38929	4.52E-15	2.00E-16	2.20E-16	0.008372			
m7		38804	1.02E-12	2.00E-16	2.20E-16		2.20E-16		
m8		38779	2.00E-16	2.00E-16	2.20E-16			2.20E-16	
m9		38886	8.47E-16	2.00E-16	2.20E-16				1.16E-12
m10	Sexo, Edad, NBI y vulnerable Hogar	38804	8.70E-13	2.00E-16	2.20E-16	0.1038	2.20E-16		
m11		38780	2.00E-16	2.20E-16	2.00E-16	0.1948		2.00E-16	
m12		38884	7.89E-16	2.20E-16	2.20E-16	0.04975			5.65E-12
m13	Sexo, Edad, NBI y Quintil	38729	1.58E-15	2.20E-16	2.20E-16	-	7.99E-12	2.20E-16	
m14		38795	2.64E-13	2.20E-16	2.20E-16	-	2.20E-16		0.0007505
m15	Sexo, Edad, NBI, Quintil y Nivel de Estudios	38727	2.00E-16	2.20E-16	2.20E-16	-	1.34E-06	2.20E-16	0.042

Se controló la inclusión o exclusión de las distintas variables por el aumento significativo de sus p-valores con el objetivo de descartar variables que sean colineales (que estén asociadas). Debido a esto se concluyó que el modelo más apropiado era el modelo 15 (m15 en la tabla) el cual incluye las variables control, NBI, quintil de ingresos, nivel de estudios y cobertura.

Modelo aditivo Final - Parametrizado como un modelo de Regresión Lineal:

El modelo final utilizado en este trabajo fue:

$$\ln(Odds)_i = \beta_0 + \beta_1 * Sexo2_i + \beta_2 * Edad_i + \beta_3 * nbi_medio_i + \beta_4 * nbi_alto_i + \beta_5 * quintil_4_i + \beta_6 * quintil_3_i + \beta_7 * quintil_2_i + \beta_8 * quintil_1_i + \beta_9 * nivel_estudios2_i + \beta_{10} * nivel_estudios3_i + \beta_{11} * tipo_cobertura2_i$$

$$i = 1 : 28925$$

De esta manera el intercepto se interpreta como el valor de $\ln(Odds)$ para los individuos que estén representados por los niveles: * Hombre (Sexo1) * 18 años de edad. * Tercil de NBI bajo. * Quintil de ingresos n° 5 * Universitario completo o incompleto.

Se eligieron para el intercepto aquellos niveles de las distintas VE que fueran representativas de un nivel socioeconómico alto (salvo para Sexo y Edad que son las variables control), para poder luego ver claramente si pertenecer a un nivel socioeconómico bajo reduce la probabilidad de realizar actividad física. Finalmente se determinó la magnitud del efecto de cada variable mediante el odds ratio.

Resultados y Discusión:

En la Figura 4 se observan los odds ratio ajustados para cada variable, los cuales reflejan su magnitud de efecto.

```
## Scale for 'y' is already present. Adding another scale for 'y', which will
## replace the existing scale.
```

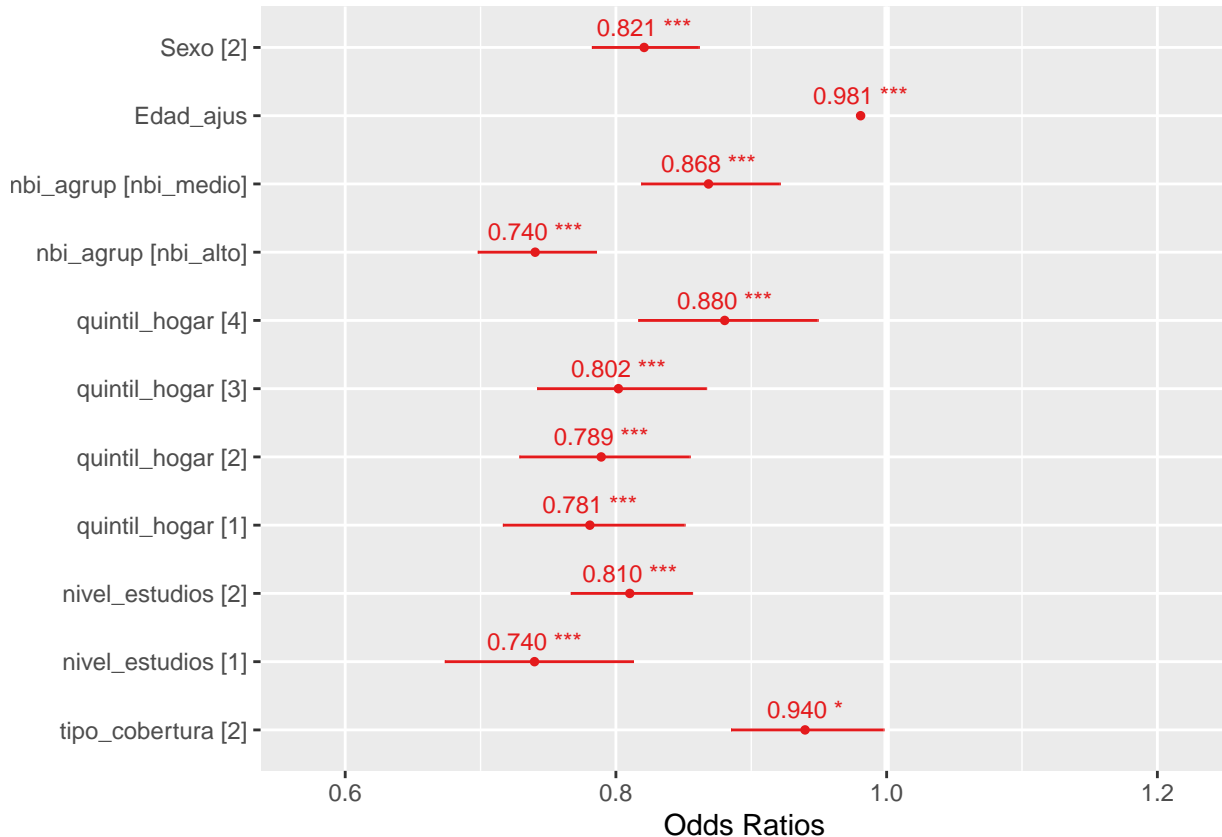


Figura 4: *plot_model* del modelo final.

El nivel socioeconómico es una cualidad difícil de representar en una única variable, así pues este modelo usa como indicadores del nivel económico y social de un individuo las variables Quintil de Ingresos, NBI, Nivel de Estudios y Tipo de Cobertura.

A partir del modelo final se determinó que la probabilidad de realizar actividad física es menor para aquellos individuos que pertenecen a un nivel socioeconómico bajo. Para ilustrar este concepto se calculó la probabilidad de realizar actividad física para individuos de los distintos quintiles ajustando por las demás variables (Figura 5). Allí se puede ver que la probabilidad de realizar actividad física para un individuo del quintil 5 (el nivel de mayor ingresos) es significativamente mayor que para un individuo del quintil 1 (el nivel de menor ingresos).

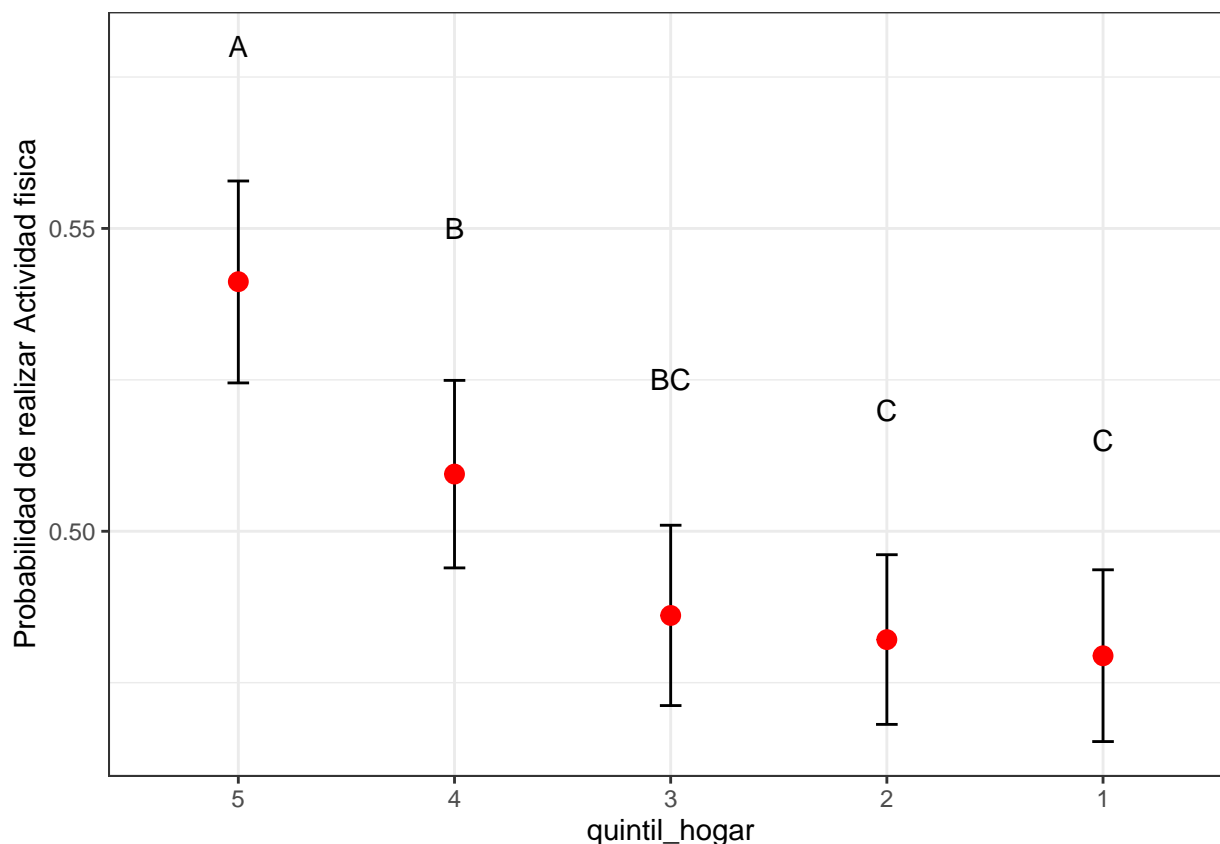


Figura 5: Probabilidad ajustada de realizar actividad física según el quintil.

Observando la Tabla 2 se puede ver que ajustando por las demás variables el modelo predice que el Odds de actividad física para Sexo2 (mujeres) disminuye entre un 14% y 22% respecto al Odds de varones.

A su vez, por cada año de edad que aumenta y ajustando por las demás variables, se predice una disminución en el Odds del 2%.

Por otra parte, los individuos de provincias con NBI alto (alto % de necesidades básicas insatisfechas) presentan un Odds de actividad física que es entre 21% y 30% menor que el Odds de los individuos que viven en una provincia con NBI bajo.

Además se observa que el Odds de actividad física en individuos pertenecientes al quintil de ingresos 1 (el menor nivel de ingresos) disminuye entre 15% y 28% respecto a los individuos del quintil 5 (el mayor nivel de ingresos).

Por otro lado, el Odds de actividad física en individuos con un nivel de estudios 1 (sin instrucción, primario incompleto) disminuye entre 19% y 33% respecto a los individuos con un nivel de estudios 3 (secundario completo, universitario incompleto, universitario completo).

Finalmente, el Odds de actividad física en individuos con un tipo de cobertura 2 (cobertura publica) disminuye significativamente entre 0.22% y 11% respecto a los individuos con cobertura 1 (obra social, prepaga o servicio de emergencia médica).

Tabla 2: Magnitud de efecto para el modelo final (OR de cada nivel referidos al intercepto)

Predictores	OR	(95% IC)	p
Sexo [2]	0.82	(0.78 – 0.86)	<0.001
Edad_ajus	0.98	(0.98 – 0.98)	<0.001
nbi_agrup [nbi_medio]	0.87	(0.82 – 0.92)	<0.001
nbi_agrup [nbi_alto]	0.74	(0.70 – 0.79)	<0.001
quintil_hogar [4]	0.88	(0.82 – 0.95)	0.001
quintil_hogar [3]	0.8	(0.74 – 0.87)	<0.001
quintil_hogar [2]	0.79	(0.73 – 0.85)	<0.001
quintil_hogar [1]	0.78	(0.72 – 0.85)	<0.001
nivel_estudios [2]	0.81	(0.77 – 0.86)	<0.001
nivel_estudios [1]	0.74	(0.67 – 0.81)	<0.001
tipo_cobertura [2]	0.94	(0.89 – 1.00)	0.042

Se procedió a validar el modelo utilizando una tabla de clasificacion cruzada. La tabla de clasificacion cruzada agrupa por un lado a los individuos que realizan o no actividad física (segun lo observado). Y por otro lado dentro de esos grupos, la cantidad de individuos que realizan o no actividad física segun lo predicho por el modelo. Dado que el modelo arroja probabilidades para cada combinacion de niveles de variables explicatorias, es necesario fijar una probabilidad a partir de la cual se considera que el individuo si realiza actividad física. Para esto se realizó una curva ROC que permite determinar el valor optimo de probabilidad de corte tal que se maximice la Tasa de Verdaderos Positivos (TPR) y se minimice la Tasa de Falsos Positivos (FPR).

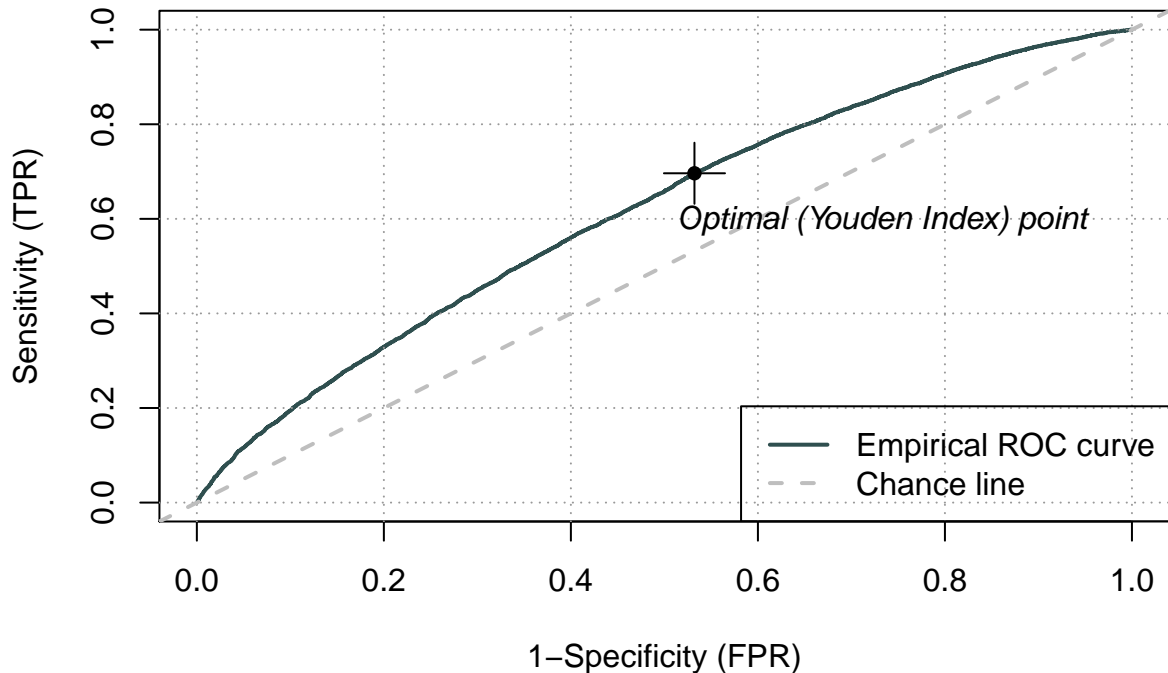


Figura 6: Curva ROC. El eje horizontal representa la Tasa de Falsos Positivos (FPR por sus siglas en ingles). El eje vertical representa la Tasa de Verdaderos Positivos (TPR).

En la Figura 6 se puede observar la curva ROC (linea continua) del modelo que denota como cambian TPR y FPR para distintas probabilidades de corte. Cuanto mayor sea el area debajo de la curva ROC, mas predictivo será el modelo ya que presenta una mayor relacion TPR/FPR. En ese caso se obtuvo un AUC de

0.62 indica que el modelo es poco predictivo, donde el solapamiento de negativos y positivos es grande. El cutoff que minimiza los falsos positivos y maximiza los verdaderos positivos es de 0.497.

Tabla 3: validación para el cutoff óptimo.

Cutoff	TP	FP	TN	FN
0.49751	10579	7308	6425	4613

La sensibilidad del modelo es del 71% es decir, por cada 100 personas que realizan ejercicio físico, el modelo elegido predice correctamente 71. La especificidad es del 54% es decir, por cada 100 personas que no realizan ejercicio, nuestro modelo predice correctamente 54. La exactitud es del 63%, es decir 63 de cada 100 predicciones realizadas por el modelo resultan correctas; por lo tanto la tasa de error es de 0,37, es decir cada 100 predicciones, 37 son incorrectas.

Conclusión

En conclusion, en este trabajo se han observado evidencias de que existe un cierto grado de asociación entre el nivel socioeconómico y el nivel de actividad física; siendo los individuos de menor nivel aquellos que presentan una menor probabilidad de realizar actividad física. Esto se condice con la hipótesis inicialmente planteada. Por lo tanto, dado que la baja actividad física es un factor de riesgo para las ENT, en este trabajo se presentan evidencias para suponer que el bajo nivel socioeconómico es un factor de riesgo para ENF.

Apéndice:

Linealidad para la variable Edad:

```
# Primero tengo que calcular las proporciones de exitos para cada valor de edad:

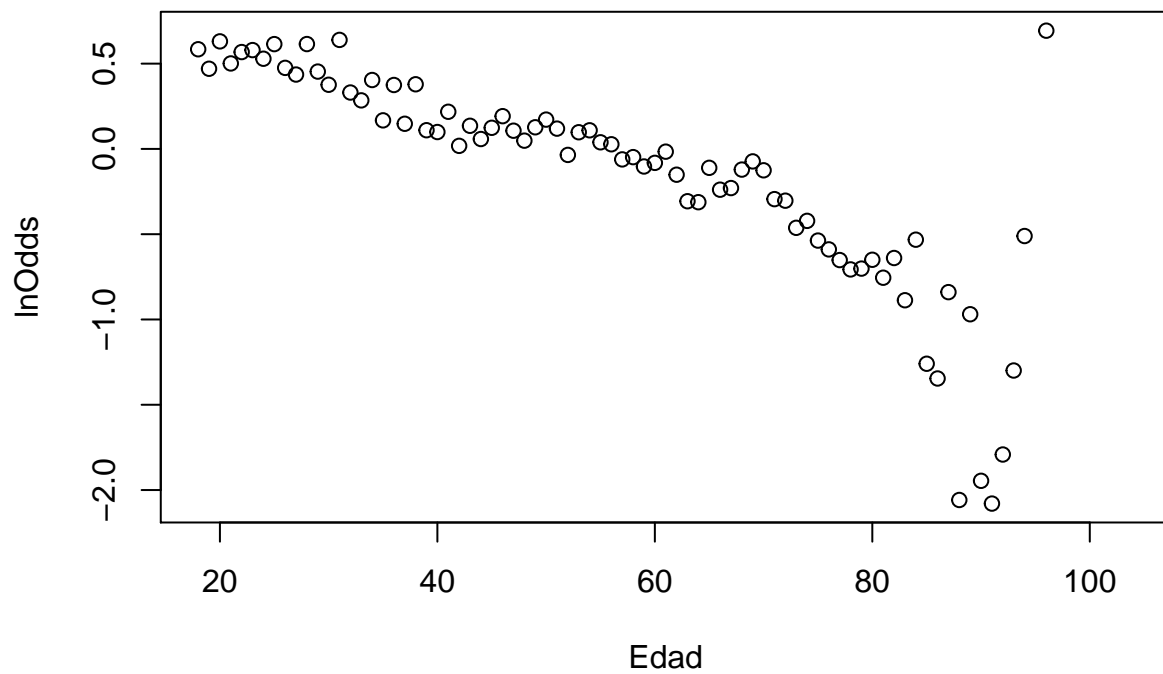
prob_exito_edad<-(tapply(Datos$nivel_actividad_fisica_num, Datos$Edad, mean)) #saco la media para la va

prob_fracaso_edad<-(1-prob_exito_edad)

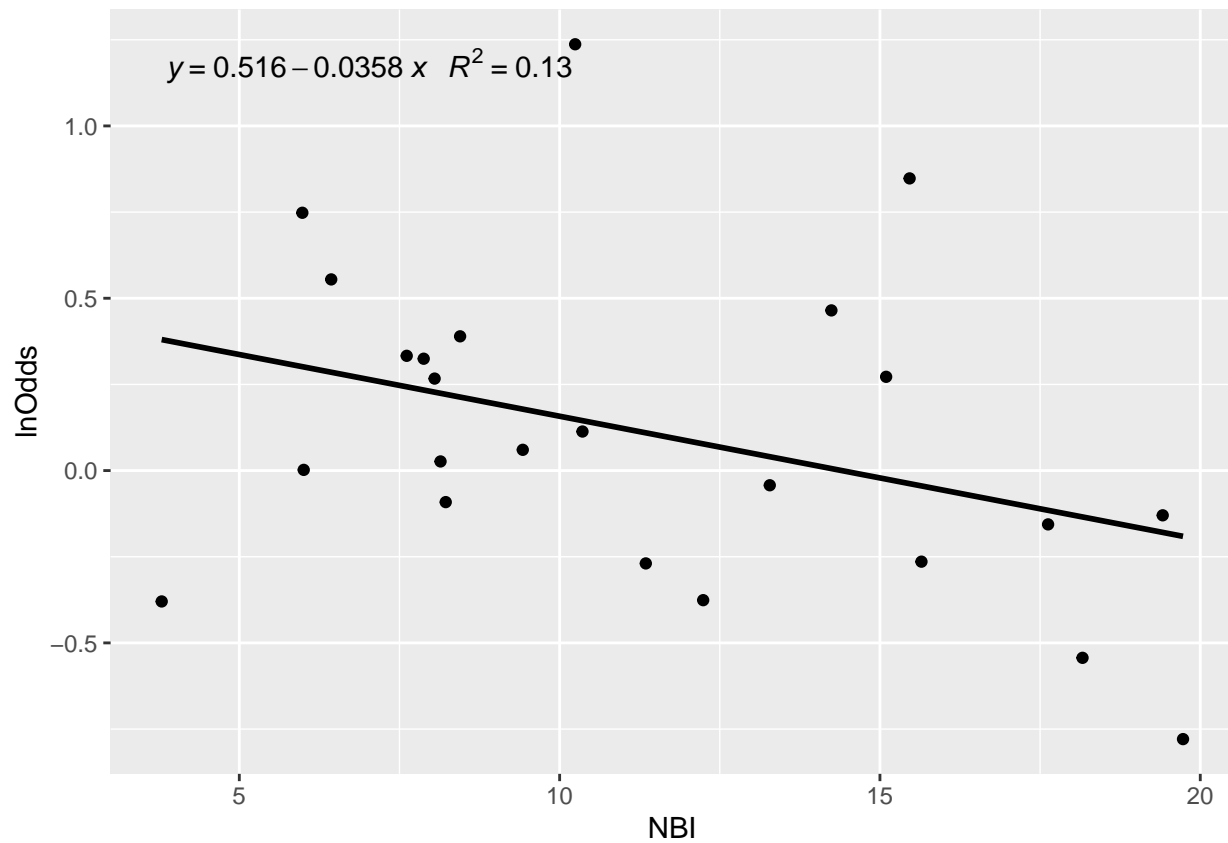
odds_edad<-prob_exito_edad/prob_fracaso_edad #calculo el odds

lnOdds_edad<-log(odds_edad) #calculo logit
Edad_crec <- sort(unique(Datos$Edad),decreasing = FALSE) #eje x es la edad en sentido creciente

linealidad_Edad<-plot(Edad_crec, lnOdds_edad, xlab="Edad",ylab="lnOdds") #grafico predictor lineal en
```



Incumplimiento de la linealidad para la variable NBI:



Modelos:

- Modelo con variables control edad y sexo:

Planteo una funcion para supuestos y Summary:

```
m<-glm(nivel_actividad_fisica_num ~ Sexo + Edad_ajus , data = Datos, family = binomial)
drop1(m, test = "Chisq")
```

```
## Single term deletions
##
## Model:
## nivel_actividad_fisica_num ~ Sexo + Edad_ajus
##           Df Deviance   AIC    LRT  Pr(>Chi)
## <none>          39084 39090
## Sexo         1     39146 39150  61.74 3.917e-15 ***
## Edad_ajus    1     39948 39952 863.97 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(m)
```

```
##
## Call:
## glm(formula = nivel_actividad_fisica_num ~ Sexo + Edad_ajus,
##      family = binomial, data = Datos)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5154  -1.1906   0.9021   1.0991   1.5953
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.7667507  0.0263913  29.053 < 2e-16 ***
## Sexo2       -0.1899477  0.0241905  -7.852 4.09e-15 ***
## Edad_ajus   -0.0194949  0.0006735 -28.947 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 40025  on 28924  degrees of freedom
## Residual deviance: 39084  on 28922  degrees of freedom
## AIC: 39090
##
## Number of Fisher Scoring iterations: 4
```

- Modelo univariado: Nbi

```
m1<-glm(nivel_actividad_fisica_num ~ Sexo + Edad_ajus + nbi_agrup, data = Datos, family = binomial)
drop1(m1,test = "Chisq")
```

```
## Single term deletions
##
## Model:
## nivel_actividad_fisica_num ~ Sexo + Edad_ajus + nbi_agrup
##           Df Deviance   AIC    LRT  Pr(>Chi)
## <none>           38924 38934
## Sexo         1     38985 38993  61.14 5.304e-15 ***
## Edad_ajus    1     39852 39860 927.37 < 2.2e-16 ***
## nbi_agrup    2     39084 39090 159.92 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(m1)
```

```
##
## Call:
## glm(formula = nivel_actividad_fisica_num ~ Sexo + Edad_ajus +
##       nbi_agrup, family = binomial, data = Datos)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5857  -1.1803   0.8812   1.1011   1.7023
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.9225513  0.0299975  30.754 < 2e-16 ***
## Sexo2         -0.1895628  0.0242589  -7.814 5.53e-15 ***
## Edad_ajus     -0.0203509  0.0006796 -29.947 < 2e-16 ***
## nbi_agrupnbi_medio -0.1564944  0.0299253  -5.230 1.70e-07 ***
## nbi_agrupnbi_alto  -0.3675440  0.0292424 -12.569 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 40025  on 28924  degrees of freedom
## Residual deviance: 38924  on 28920  degrees of freedom
## AIC: 38934
##
## Number of Fisher Scoring iterations: 4
```

- Modelo univariado: vulnerable

```
m2<-glm(nivel_actividad_fisica_num ~ Sexo + Edad_ajus + vulnerable_hogar, data = Datos, family = binomial)
drop1(m2,test = "Chisq")
```

```
## Single term deletions
##
## Model:
## nivel_actividad_fisica_num ~ Sexo + Edad_ajus + vulnerable_hogar
##           Df Deviance   AIC    LRT  Pr(>Chi)
```

```
## <none>          39070 39078
## Sexo            1    39133 39139  62.16 3.172e-15 ***
## Edad_ajus       1    39942 39948 871.94 < 2.2e-16 ***
## vulnerable_hogar 1    39084 39090  13.61 0.0002256 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(m2)
```

```
##
## Call:
## glm(formula = nivel_actividad_fisica_num ~ Sexo + Edad_ajus +
##       vulnerable_hogar, family = binomial, data = Datos)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5204  -1.1913   0.8982   1.0981   1.6912
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.7777551  0.0265759  29.265 < 2e-16 ***
## Sexo2          -0.1906386  0.0241975  -7.878 3.31e-15 ***
## Edad_ajus      -0.0196090  0.0006744 -29.075 < 2e-16 ***
## vulnerable_hogar1 -0.2729304  0.0739954  -3.688 0.000226 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 40025  on 28924  degrees of freedom
## Residual deviance: 39070  on 28921  degrees of freedom
## AIC: 39078
##
## Number of Fisher Scoring iterations: 4
```

- Modelo univariado: quintil

```
m3<-glm(nivel_actividad_fisica_num ~ Sexo + Edad_ajus + quintil_hogar, data = Datos, family = binomial)
drop1(m3,test = "Chisq")
```

```
## Single term deletions
##
## Model:
## nivel_actividad_fisica_num ~ Sexo + Edad_ajus + quintil_hogar
##              Df Deviance   AIC    LRT Pr(>Chi)
## <none>          38893 38907
## Sexo            1    38943 38955  49.67 1.822e-12 ***
## Edad_ajus       1    39821 39833 927.40 < 2.2e-16 ***
## quintil_hogar   4    39084 39090 190.84 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(m3)
```

```
##
## Call:
## glm(formula = nivel_actividad_fisica_num ~ Sexo + Edad_ajus +
##       quintil_hogar, family = binomial, data = Datos)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6475  -1.1868   0.8777   1.0972   1.7033
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.0595754  0.0364273  29.087 < 2e-16 ***
## Sexo2         -0.1712215  0.0243064  -7.044 1.86e-12 ***
## Edad_ajus     -0.0205233  0.0006853 -29.948 < 2e-16 ***
## quintil_hogar4 -0.1810715  0.0379163  -4.776 1.79e-06 ***
## quintil_hogar3 -0.3268765  0.0384113  -8.510 < 2e-16 ***
## quintil_hogar2 -0.3942522  0.0379955 -10.376 < 2e-16 ***
## quintil_hogar1 -0.4709237  0.0385855 -12.205 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 40025  on 28924  degrees of freedom
## Residual deviance: 38893  on 28918  degrees of freedom
## AIC: 38907
##
## Number of Fisher Scoring iterations: 4
```

- Modelo univariado: nivel de estudios

```
m4<-glm(nivel_actividad_fisica_num ~ Sexo + Edad_ajus + nivel_estudios, data = Datos, family = binomial)
drop1(m4,test = "Chisq")
```

```
## Single term deletions
##
## Model:
## nivel_actividad_fisica_num ~ Sexo + Edad_ajus + nivel_estudios
##              Df Deviance   AIC    LRT  Pr(>Chi)
## <none>                38899 38909
## Sexo              1    38976 38984  76.64 < 2.2e-16 ***
## Edad_ajus         1    39483 39491 584.16 < 2.2e-16 ***
## nivel_estudios    2    39084 39090 185.15 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(m4)
```



```
##
## Call:
## glm(formula = nivel_actividad_fisica_num ~ Sexo + Edad_ajus +
##      nivel_estudios, family = binomial, data = Datos)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5537  -1.1902   0.8857   1.0895   1.6827
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.8515826  0.0277329  30.707  <2e-16 ***
## Sexo2         -0.2129091  0.0243457  -8.745  <2e-16 ***
## Edad_ajus     -0.0168732  0.0007049 -23.936  <2e-16 ***
## nivel_estudios2 -0.2983768  0.0261479 -11.411  <2e-16 ***
## nivel_estudios1 -0.4603501  0.0454554 -10.128  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 40025  on 28924  degrees of freedom
## Residual deviance: 38899  on 28920  degrees of freedom
## AIC: 38909
##
## Number of Fisher Scoring iterations: 4
```

- Modelo univariado: cobertura

```
m5<-glm(nivel_actividad_fisica_num ~ Sexo + Edad_ajus + tipo_cobertura, data = Datos, family = binomial)
drop1(m5,test = "Chisq")
```

```
## Single term deletions
##
## Model:
## nivel_actividad_fisica_num ~ Sexo + Edad_ajus + tipo_cobertura
##              Df Deviance   AIC    LRT  Pr(>Chi)
## <none>                39010 39018
## Sexo              1    39076 39082   66.04 4.413e-16 ***
## Edad_ajus         1    39945 39951  935.54 < 2.2e-16 ***
## tipo_cobertura    1    39084 39090   74.14 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(m5)
```

```
##
## Call:
## glm(formula = nivel_actividad_fisica_num ~ Sexo + Edad_ajus +
##      tipo_cobertura, family = binomial, data = Datos)
##
```

```
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5762  -1.1840   0.8951   1.0988   1.7160
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.9015740  0.0308563   29.22 < 2e-16 ***
## Sexo2         -0.1968213  0.0242383   -8.12 4.65e-16 ***
## Edad_ajus     -0.0215278  0.0007161  -30.06 < 2e-16 ***
## tipo_cobertura2 -0.2372414  0.0275873   -8.60 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 40025  on 28924  degrees of freedom
## Residual deviance: 39010  on 28921  degrees of freedom
## AIC: 39018
##
## Number of Fisher Scoring iterations: 4
```

- Modelo: Nbi + vulnerable

```
m6<-glm(nivel_actividad_fisica_num ~ Sexo + Edad_ajus + nbi_agrup + vulnerable_hogar, data = Datos, family = "binomial")
drop1(m6,test = "Chisq")
```

```
## Single term deletions
##
## Model:
## nivel_actividad_fisica_num ~ Sexo + Edad_ajus + nbi_agrup + vulnerable_hogar
##              Df Deviance   AIC    LRT  Pr(>Chi)
## <none>                38917 38929
## Sexo                1    38979 38989   61.46 4.519e-15 ***
## Edad_ajus           1    39849 39859  932.08 < 2.2e-16 ***
## nbi_agrup           2    39070 39078  153.27 < 2.2e-16 ***
## vulnerable_hogar    1    38924 38934    6.95 0.008372 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(m6)
```

```
##
## Call:
## glm(formula = nivel_actividad_fisica_num ~ Sexo + Edad_ajus +
##      nbi_agrup + vulnerable_hogar, family = binomial, data = Datos)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5883  -1.1814   0.8791   1.1021   1.6992
##
## Coefficients:
```

```
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)      0.9283515  0.0300876  30.855 < 2e-16 ***
## Sexo2            -0.1900809  0.0242629  -7.834 4.72e-15 ***
## Edad_ajus        -0.0204192  0.0006802 -30.020 < 2e-16 ***
## nbi_agrupnbi_medio -0.1564007  0.0299295  -5.226 1.74e-07 ***
## nbi_agrupnbi_alto  -0.3608550  0.0293520 -12.294 < 2e-16 ***
## vulnerable_hogar1 -0.1964374  0.0745148  -2.636 0.00838 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 40025  on 28924  degrees of freedom
## Residual deviance: 38917  on 28919  degrees of freedom
## AIC: 38929
##
## Number of Fisher Scoring iterations: 4
```

- Modelo: Nbi + quintil

```
m7<-glm(nivel_actividad_fisica_num ~ Sexo + Edad_ajus + nbi_agrup + quintil_hogar, data = Datos, family = binomial)
drop1(m7,test = "Chisq")
```

```
## Single term deletions
##
## Model:
## nivel_actividad_fisica_num ~ Sexo + Edad_ajus + nbi_agrup + quintil_hogar
##               Df Deviance   AIC    LRT Pr(>Chi)
## <none>                38786 38804
## Sexo                1    38837 38853  50.81 1.016e-12 ***
## Edad_ajus           1    39753 39769 966.37 < 2.2e-16 ***
## nbi_agrup           2    38893 38907 107.03 < 2.2e-16 ***
## quintil_hogar       4    38924 38934 137.95 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(m7)
```

```
##
## Call:
## glm(formula = nivel_actividad_fisica_num ~ Sexo + Edad_ajus +
##       nbi_agrup + quintil_hogar, family = binomial, data = Datos)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6936  -1.1770   0.8559   1.0965   1.7323
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)      1.1618300  0.0384608  30.208 < 2e-16 ***
## Sexo2            -0.1735209  0.0243542  -7.125 1.04e-12 ***
```

```
## Edad_ajus      -0.0210488  0.0006892 -30.543 < 2e-16 ***
## nbi_agrupnbi_medio -0.1518183  0.0300018  -5.060 4.19e-07 ***
## nbi_agrupnbi_alto  -0.3041626  0.0299126 -10.168 < 2e-16 ***
## quintil_hogar4    -0.1725835  0.0379948  -4.542 5.56e-06 ***
## quintil_hogar3    -0.3029134  0.0385506  -7.858 3.92e-15 ***
## quintil_hogar2    -0.3535222  0.0383000  -9.230 < 2e-16 ***
## quintil_hogar1    -0.4021519  0.0393559 -10.218 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 40025  on 28924  degrees of freedom
## Residual deviance: 38786  on 28916  degrees of freedom
## AIC: 38804
##
## Number of Fisher Scoring iterations: 4
```

- Modelo: Nbi + cobertura

```
m9<-glm(nivel_actividad_fisica_num ~ Sexo + Edad_ajus + nbi_agrup + tipo_cobertura , data = Datos, fam. = "binomial")
drop1(m9,test = "Chisq")
```

```
## Single term deletions
##
## Model:
## nivel_actividad_fisica_num ~ Sexo + Edad_ajus + nbi_agrup + tipo_cobertura
##              Df Deviance   AIC    LRT Pr(>Chi)
## <none>                38874 38886
## Sexo              1    38938 38948  64.76 8.473e-16 ***
## Edad_ajus         1    39842 39852 968.28 < 2.2e-16 ***
## nbi_agrup          2    39010 39018 136.33 < 2.2e-16 ***
## tipo_cobertura    1    38924 38934  50.55 1.163e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(m9)
```

```
##
## Call:
## glm(formula = nivel_actividad_fisica_num ~ Sexo + Edad_ajus +
##       nbi_agrup + tipo_cobertura, family = binomial, data = Datos)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6325  -1.1753   0.8662   1.1024   1.7026
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.0262408  0.0334995  30.635 < 2e-16 ***
## Sexo2         -0.1953720  0.0242969  -8.041 8.91e-16 ***
```

```
## Edad_ajus          -0.0219938  0.0007197 -30.560 < 2e-16 ***
## nbi_agrupnbi_medio -0.1536394  0.0299577  -5.129 2.92e-07 ***
## nbi_agrupnbi_alto  -0.3412346  0.0294807 -11.575 < 2e-16 ***
## tipo_cobertura2    -0.1980885  0.0278809  -7.105 1.20e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 40025  on 28924  degrees of freedom
## Residual deviance: 38874  on 28919  degrees of freedom
## AIC: 38886
##
## Number of Fisher Scoring iterations: 4
```

- Modelo: Nbi + vulnerable + quintil

```
m10<-glm(nivel_actividad_fisica_num ~ Sexo + Edad_ajus + nbi_agrup + vulnerable_hogar + quintil_hogar,
drop1(m10,test = "Chisq"))
```

```
## Single term deletions
##
## Model:
## nivel_actividad_fisica_num ~ Sexo + Edad_ajus + nbi_agrup + vulnerable_hogar +
## quintil_hogar
##           Df Deviance   AIC    LRT Pr(>Chi)
## <none>                38784 38804
## Sexo           1    38835 38853  51.12 8.704e-13 ***
## Edad_ajus      1    39752 39770 968.00 < 2.2e-16 ***
## nbi_agrup       2    38888 38904 104.76 < 2.2e-16 ***
## vulnerable_hogar 1    38786 38804   2.65  0.1038
## quintil_hogar   4    38917 38929 133.64 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(m10)
```

```
##
## Call:
## glm(formula = nivel_actividad_fisica_num ~ Sexo + Edad_ajus +
## nbi_agrup + vulnerable_hogar + quintil_hogar, family = binomial,
## data = Datos)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6942  -1.1766   0.8554   1.0962   1.7303
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.1632260  0.0384736  30.234 < 2e-16 ***
## Sexo2          -0.1740650  0.0243579  -7.146 8.92e-13 ***
```

```
## Edad_ajus          -0.0210716  0.0006894 -30.567 < 2e-16 ***
## nbi_agrupnbi_medio -0.1517683  0.0300036  -5.058 4.23e-07 ***
## nbi_agrupnbi_alto  -0.3011524  0.0299699 -10.048 < 2e-16 ***
## vulnerable_hogar1  -0.1221271  0.0750889  -1.626  0.104
## quintil_hogar4     -0.1720490  0.0379975  -4.528 5.96e-06 ***
## quintil_hogar3     -0.3020554  0.0385560  -7.834 4.72e-15 ***
## quintil_hogar2     -0.3512931  0.0383262  -9.166 < 2e-16 ***
## quintil_hogar1     -0.3953595  0.0395781  -9.989 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 40025  on 28924  degrees of freedom
## Residual deviance: 38784  on 28915  degrees of freedom
## AIC: 38804
##
## Number of Fisher Scoring iterations: 4
```

- Modelo: Nbi + vulnerable + nivel de estudios

```
m11<-glm(nivel_actividad_fisica_num ~ Sexo + Edad_ajus + nbi_agrup + vulnerable_hogar + nivel_estudios
drop1(m11,test = "Chisq")
```

```
## Single term deletions
##
## Model:
## nivel_actividad_fisica_num ~ Sexo + Edad_ajus + nbi_agrup + vulnerable_hogar +
## nivel_estudios
##
```

	Df	Deviance	AIC	LRT	Pr(>Chi)
<none>		38764	38780		
Sexo	1	38839	38853	75.32	<2e-16 ***
Edad_ajus	1	39410	39424	646.32	<2e-16 ***
nbi_agrup	2	38894	38906	130.74	<2e-16 ***
vulnerable_hogar	1	38765	38779	1.68	0.1948
nivel_estudios	2	38917	38929	153.58	<2e-16 ***

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(m11)
```

```
##
## Call:
## glm(formula = nivel_actividad_fisica_num ~ Sexo + Edad_ajus +
## nbi_agrup + vulnerable_hogar + nivel_estudios, family = binomial,
## data = Datos)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6184  -1.1819   0.8682   1.0914   1.7553
##
```

```
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.995006   0.031195  31.896 < 2e-16 ***
## Sexo2         -0.211584   0.024404  -8.670 < 2e-16 ***
## Edad_ajus     -0.017976   0.000715 -25.142 < 2e-16 ***
## nbi_agrupnbi_medio -0.141555  0.030033  -4.713 2.44e-06 ***
## nbi_agrupnbi_alto -0.336281  0.029565 -11.374 < 2e-16 ***
## vulnerable_hogar1 -0.097398  0.075135  -1.296  0.195
## nivel_estudios2 -0.285908  0.026294 -10.873 < 2e-16 ***
## nivel_estudios1 -0.398458  0.046030  -8.657 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 40025  on 28924  degrees of freedom
## Residual deviance: 38764  on 28917  degrees of freedom
## AIC: 38780
##
## Number of Fisher Scoring iterations: 4
```

- Modelo: Nbi + vulnerable + cobertura

```
m12<-glm(nivel_actividad_fisica_num ~ Sexo + Edad_ajus + nbi_agrup + vulnerable_hogar + tipo_cobertura
drop1(m12,test = "Chisq")
```

```
## Single term deletions
##
## Model:
## nivel_actividad_fisica_num ~ Sexo + Edad_ajus + nbi_agrup + vulnerable_hogar +
## tipo_cobertura
##              Df Deviance   AIC    LRT  Pr(>Chi)
## <none>                38870 38884
## Sexo                1    38935 38947  64.90 7.891e-16 ***
## Edad_ajus           1    39839 39851 968.80 < 2.2e-16 ***
## nbi_agrup            2    39002 39012 132.43 < 2.2e-16 ***
## vulnerable_hogar     1    38874 38886   3.85  0.04975 *
## tipo_cobertura       1    38917 38929  47.45 5.654e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(m12)
```

```
##
## Call:
## glm(formula = nivel_actividad_fisica_num ~ Sexo + Edad_ajus +
## nbi_agrup + vulnerable_hogar + tipo_cobertura, family = binomial,
## data = Datos)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -1.6332 -1.1750 0.8658 1.1031 1.7003
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.0278236  0.0335127  30.670 < 2e-16 ***
## Sexo2          -0.1955994  0.0242990  -8.050 8.30e-16 ***
## Edad_ajus      -0.0220013  0.0007198 -30.567 < 2e-16 ***
## nbi_agrupnbi_medio -0.1536471  0.0299600  -5.128 2.92e-07 ***
## nbi_agrupnbi_alto  -0.3369325  0.0295621 -11.397 < 2e-16 ***
## vulnerable_hogar1 -0.1467884  0.0748309  -1.962 0.0498 *
## tipo_cobertura2  -0.1928309  0.0280117  -6.884 5.82e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 40025  on 28924  degrees of freedom
## Residual deviance: 38870  on 28918  degrees of freedom
## AIC: 38884
##
## Number of Fisher Scoring iterations: 4
```

- Modelo: Nbi + quintil + nivel de estudios

```
m13<-glm(nivel_actividad_fisica_num ~ Sexo + Edad_ajus + nbi_agrup + quintil_hogar + nivel_estudios, data = Datos, family = binomial)
drop1(m13, test = "Chisq")
```

```
## Single term deletions
##
## Model:
## nivel_actividad_fisica_num ~ Sexo + Edad_ajus + nbi_agrup + quintil_hogar +
##      nivel_estudios
##              Df Deviance   AIC    LRT  Pr(>Chi)
## <none>                38707 38729
## Sexo                1    38771 38791  63.53 1.58e-15 ***
## Edad_ajus           1    39386 39406 678.21 < 2.2e-16 ***
## nbi_agrup            2    38812 38830 104.75 < 2.2e-16 ***
## quintil_hogar        4    38765 38779  57.90 7.99e-12 ***
## nivel_estudios       2    38786 38804  78.81 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(m13)
```

```
##
## Call:
## glm(formula = nivel_actividad_fisica_num ~ Sexo + Edad_ajus +
##      nbi_agrup + quintil_hogar + nivel_estudios, family = binomial,
##      data = Datos)
##
## Deviance Residuals:
```



```
##      Min      1Q   Median      3Q      Max
## -1.6867 -1.1790  0.8545   1.0929   1.7702
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.1466524  0.0388435  29.520 < 2e-16 ***
## Sexo2         -0.1953303  0.0245252  -7.964 1.66e-15 ***
## Edad_ajus     -0.0189185  0.0007353 -25.730 < 2e-16 ***
## nbi_agrupnbi_medio -0.1415199  0.0300658  -4.707 2.51e-06 ***
## nbi_agrupnbi_alto -0.3038039  0.0300332 -10.116 < 2e-16 ***
## quintil_hogar4  -0.1310689  0.0383136  -3.421 0.000624 ***
## quintil_hogar3  -0.2283797  0.0394851  -5.784 7.30e-09 ***
## quintil_hogar2  -0.2513057  0.0400155  -6.280 3.38e-10 ***
## quintil_hogar1  -0.2731966  0.0419552  -6.512 7.43e-11 ***
## nivel_estudios2 -0.2191399  0.0278724  -7.862 3.77e-15 ***
## nivel_estudios1 -0.3100848  0.0476748  -6.504 7.81e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 40025  on 28924  degrees of freedom
## Residual deviance: 38707  on 28914  degrees of freedom
## AIC: 38729
##
## Number of Fisher Scoring iterations: 4
```

- Modelo: Nbi + quintil + cobertura

```
m14<-glm(nivel_actividad_fisica_num ~ Sexo + Edad_ajus + nbi_agrup + quintil_hogar + tipo_cobertura, data = datos, family = "binomial")
drop1(m14,test = "Chisq")
```

```
## Single term deletions
##
## Model:
## nivel_actividad_fisica_num ~ Sexo + Edad_ajus + nbi_agrup + quintil_hogar +
##      tipo_cobertura
##              Df Deviance   AIC    LRT Pr(>Chi)
## <none>                38775 38795
## Sexo                1    38828 38846  53.46 2.64e-13 ***
## Edad_ajus           1    39716 39734 940.85 < 2.2e-16 ***
## nbi_agrup           2    38878 38894 103.48 < 2.2e-16 ***
## quintil_hogar       4    38874 38886  98.76 < 2.2e-16 ***
## tipo_cobertura      1    38786 38804  11.36 0.0007505 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(m14)
```

```
##
## Call:
```

```

## glm(formula = nivel_actividad_fisica_num ~ Sexo + Edad_ajus +
##      nbi_agrup + quintil_hogar + tipo_cobertura, family = binomial,
##      data = Datos)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q        Max
## -1.708   -1.175    0.852    1.099    1.729
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.1940356  0.0396819  30.090 < 2e-16 ***
## Sexo2         -0.1783263  0.0244026  -7.308 2.72e-13 ***
## Edad_ajus     -0.0217674  0.0007221 -30.145 < 2e-16 ***
## nbi_agrupnbi_medio -0.1505983  0.0300122  -5.018 5.22e-07 ***
## nbi_agrupnbi_alto -0.2992096  0.0299488  -9.991 < 2e-16 ***
## quintil_hogar4   -0.1641623  0.0381001  -4.309 1.64e-05 ***
## quintil_hogar3   -0.2862414  0.0388829  -7.362 1.82e-13 ***
## quintil_hogar2   -0.3238849  0.0393010  -8.241 < 2e-16 ***
## quintil_hogar1   -0.3526628  0.0420063  -8.395 < 2e-16 ***
## tipo_cobertura2  -0.1016565  0.0301627  -3.370 0.000751 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 40025  on 28924  degrees of freedom
## Residual deviance: 38775  on 28915  degrees of freedom
## AIC: 38795
##
## Number of Fisher Scoring iterations: 4

```