

Tipo de documento: Tesis de Maestría

Maestría en Econometría

Análisis de los determinantes de la obesidad en la Provincia de Buenos Aires

Autoría: Alzola, Agustina

Año de defensa de la tesis: 2023

¿Cómo citar este trabajo?

Alzola, A.(2023) "Análisis de los determinantes de la obesidad en la Provincia de Buenos Aires". [Tesis de Maestría. Universidad Torcuato Di Tella]. Repositorio Digital Universidad Torcuato Di Tella

<https://repositorio.utdt.edu/handle/20.500.13098/12184>

El presente documento se encuentra alojado en el Repositorio Digital de la Universidad Torcuato Di Tella bajo una licencia Creative Commons Atribución-No Comercial-Compartir Igual 4.0 Argentina (CC BY-NC-SA 4.0 AR)

Dirección: <https://repositorio.utdt.edu>



Análisis de los determinantes de la obesidad en la Provincia de Buenos Aires

**Autora: Alzola Agustina
Legajo: 18K2026**

Tutora: Zamparo Sofía

Departamento de Economía

Maestría en Econometría

Junio 2023

Resumen

La obesidad, considerada la enfermedad no transmisible con mayor prevalencia en el mundo, es uno de los factores de riesgo más importantes vinculados al aumento de enfermedades cardiovasculares, la diabetes tipo II, hipertensión, trastornos del aparato locomotor y algunos tipos de cáncer (OMS, 2021). Ésta, puede prevenirse generando entornos y comunidades favorables que permitan influir en las elecciones de las personas. Por este motivo, en el presente trabajo se analizan los factores determinantes de la obesidad en mayores de 18 años de la Provincia de Buenos Aires, de modo que las conclusiones que surjan de este análisis puedan ser de utilidad para quienes gestionan políticas públicas tendientes a promover la salud mediante la prevención.

Para ello, se utilizarán datos provenientes de la cuarta Encuesta Nacional de Factores de Riesgo (ENFR), relevada en Argentina durante el año 2018 por el Ministerio de Salud-Presidencia de la Nación, conjuntamente con el Instituto Nacional de Estadísticas y Censos (INDEC). En primer lugar, se realiza un análisis descriptivo de los datos, y se evalúa asociación entre diversos factores mencionados en la literatura (como región, sexo, situación conyugal, educación, entre otros) y el Índice de Masa Corporal (IMC). Luego, se estima un modelo de regresión lineal múltiple, un modelo *Logit* binario y un modelo *Logit* ordinal.

A partir del análisis de los resultados, se observa que la edad, el nivel educativo, la situación conyugal, la realización de actividad física, el consumo de frutas y verduras y la región en la que vive el encuestado son factores asociados a la prevalencia de obesidad en los mayores de 18 años de la Provincia de Buenos Aires.

Palabras clave

Obesidad – Sobrepeso – IMC – Regresión logística

Tabla de Contenidos

1.Introducción.....	3
2.Fundamentación conceptual y antecedentes de la investigación	4
3.Metodología aplicada.....	9
3.a. Desarrollo de la metodología econométrica.....	9
3.a.i. Modelo de regresión lineal multiple.....	9
3.a.ii. Modelo de regresion logística binaria.....	11
3.a.iii. Modelo de regresión logística ordinal.....	12
3.b. Datos empleados	14
3.b.i. Diseño de la muestra y dominio de estimación.....	14
3.b.ii. Trabajo de campo, cuestionario	16
3.b.iii. Submuestra utilizada	16
3.c. Especificación econométrica	17
4.Resultados.....	19
4.a. Análisis descriptivo	19
4.b. Formulación y análisis de los modelos econométricos	23
4.b.i. Modelo de regresión lineal múltiple	23
4.b.ii. Modelo de regresión logística binaria.....	24
4.b.iii. Modelo de regresión logística ordinal.....	26
5.Consideraciones finales	33
6.Referencias bibliográficas	35
7.Anexos.....	38
7.a. Prueba chi cuadrado de pearson	38
7.b. Cuestionario.....	39
7.c. Análisis descriptivo de los encuestados.....	41
7.d.Prueba de líneas paralelas	42

1. Introducción

Según la Organización Mundial de la Salud (OMS), desde 1975 la obesidad se ha casi triplicado en el mundo. En 2016, el 39% de las personas adultas de 18 o más años tenían sobrepeso, y el 13% eran obesas.

El aumento de la obesidad en Argentina se traduce en un incremento de la incidencia de las patologías asociadas con un impacto significativo sobre los costos sanitarios y un deterioro de la calidad de vida (Temporelli & Viego, 2010). Según Britos (2012), se ha intensificado la elección de alimentos con menor valor nutritivo pero con mayor aporte calórico, consolidándose un cambio epidemiológico nutricional trascendente: el sobrepeso está instalado como el problema nutricional más extendido.

Siguiendo a Monteverde & Tomas (2017), la prevalencia de sobrepeso es similar a nivel Total Argentina y Provincia de Buenos Aires, registrando un incremento del 3% en el período 2005-2013. Con respecto a la obesidad, ésta registró un incremento del 7% en el período mencionado.

Dado que el sobrepeso y la obesidad, así como las enfermedades no transmisibles vinculadas, pueden prevenirse generando entornos y comunidades favorables que permitan influir en las elecciones de las personas, de modo que la opción más sencilla (la más accesible, disponible y asequible) sea la más saludable en materia de alimentos y actividad física periódica, el interés de este tesis se centra en analizar cuáles son los factores determinantes del sobrepeso y la obesidad en los mayores de 18 años de la Provincia de Buenos Aires. De modo que estos resultados puedan ser de utilidad para quienes gestionan políticas públicas tendientes a promover la salud mediante la prevención.

Para ello, se utilizarán datos provenientes de la cuarta Encuesta Nacional de Factores de Riesgo (ENFR), relevada en Argentina durante el año 2018 por el Ministerio de Salud-Presidencia de la Nación, conjuntamente con el Instituto Nacional de Estadísticas y Censos (INDEC), siendo la población objetivo los individuos de 18 años y más. Su propósito fundamental fue brindar información socio-sanitaria a fin de diseñar estrategias referidas a las enfermedades no transmisibles.

La presente tesis se encuentra estructurada de la siguiente manera: en el capítulo 2 se presenta la fundamentación conceptual y los antecedentes de la investigación, en el capítulo 3 se describe la metodología y los datos utilizados. Finalmente, en el capítulo 4 se presentan los resultados obtenidos y en el 5 las consideraciones finales.

2. Fundamentación conceptual y antecedentes de la investigación

Hasta hace algunos años el sobrepeso y la obesidad eran percibidos como un problema de los países desarrollados, pero en la actualidad este escenario se ha ido modificando y la epidemia se ha extendido a los países de bajos y medianos ingresos. Esto adquiere especial relevancia en los países en desarrollo por dos motivos: por un lado, el esfuerzo que debe hacerse para combatir los efectos nocivos que provoca es relativamente mayor, por otro, existe riesgo de que las poblaciones de menores ingresos vean desmejorada su situación de salud cayendo en trampas de pobreza (Temporelli & Viego, *op Cit.*).

Es la primera vez en la historia que la cantidad de individuos con sobrepeso es similar al número de personas que sufren hambre y desnutrición. Ambas problemáticas, de distinto modo, presentan una gran cantidad de patologías asociadas que disminuyen la esperanza y calidad de vida y la productividad de la fuerza laboral, factores que constituyen un impedimento para el crecimiento y desarrollo de los países (Gardner & Halweil, 2000).

Tanto el sobrepeso como la obesidad se definen como una acumulación anormal o excesiva de grasa que puede ser perjudicial para la salud, y son causadas por un desequilibrio entre la energía consumida y la gastada. Éstos pueden identificarse a través del índice de masa corporal (IMC), que es un indicador simple de la relación entre el peso y la talla, y se calcula dividiendo el peso en kg. sobre el cuadrado de la altura en metros (kg/m^2). En el caso de los adultos, la OMS define al sobrepeso como un IMC igual o superior a $25 \text{ kg}/\text{m}^2$ y a la obesidad como un índice de masa corporal igual o superior a $30 \text{ kg}/\text{m}^2$.

Tabla 1: Clasificación del IMC

IMC	CLASIFICACIÓN
< 18.5	Peso insuficiente
18.5 – 24.9	Peso normal o saludable
25 – 29.9	Sobrepeso
>30	Obesidad

Fuente: elaboración propia en base a la ENFR 2018

La obesidad, considerada la enfermedad no transmisible con mayor prevalencia en el mundo (Mokdad et al., 2004; Flegal et al., 2005; Sturm, 2002), es uno de los factores de riesgo más importantes vinculados al aumento de enfermedades cardiovasculares (principalmente cardiopatías y accidentes cerebrovasculares), la diabetes tipo II, hipertensión, trastornos del aparato locomotor (en especial la osteoartritis) y algunos tipos de cáncer (endometrio, mama, ovarios, próstata, hígado, vesícula biliar, riñones y colon) (OMS, 2021). Además, está relacionada a problemas psicosociales como menor productividad laboral en la vida adulta y menor probabilidad de conseguir empleo, sumado a que incrementa la posibilidad de sufrir discriminación de pares y estigmatización, pudiendo causar trastornos psicológicos relacionados con la baja autoestima (UNICEF, 2019).

La Organización Panamericana de la Salud (2017) alerta sobre los niveles alarmantes de sobrepeso y obesidad alcanzados en América Latina, donde más del 50% de la población de sus países (con excepción de Haití, Paraguay y Nicaragua) lo padecen. En el caso de Argentina, Colombia, México, Paraguay, Perú y

Uruguay, los datos muestran una población donde más de la mitad tiene exceso de peso y más del 15% es obesa (Isunza Vera, 2004).

La obesidad, para la OMS, es una condición compleja, con dimensiones psicológicas y sociales que afectan a las personas de todas las edades, en todos los niveles socioeconómicos de naciones desarrolladas y en vías de desarrollo, por lo que es importante trabajar en su prevención como un problema de salud pública de alta prioridad (Isunza Vera, *op. cit.*).

Actualmente, la obesidad es una enfermedad prevalente en todos los niveles de ingreso constituyendo un serio problema de salud pública en los países en vías de desarrollo, debido al alto costo que representan para los servicios de salud el tratamiento y la rehabilitación de pacientes tanto con obesidad, como de otras ENT asociadas a ella (Peña & Bacallao, 2001).

El incremento de peso puede atribuirse al consumo de alimentos de alto contenido calórico que son ricos en grasa, sumado a un descenso en la actividad física debido a la naturaleza cada vez más sedentaria de muchas formas de trabajo, nuevos modos de transporte y creciente urbanización (OMS, *op. cit.*). En definitiva, se consume más energía y se gasta menos (Temporelli & Viego, *op. cit.*).

Dichos cambios en los hábitos alimentarios y de actividad física se presentan como consecuencia de cambios ambientales y sociales asociados al desarrollo, y a la falta de políticas de apoyo en ciertos sectores como la salud, la agricultura, el transporte, la planificación urbana, el medio ambiente, el procesamiento, distribución y comercialización de alimentos, y la educación (OMS, *op. cit.*). Estos factores adquieren en cada región características propias íntimamente relacionadas a su cultura y modos de vida (Temporelli & Viego, *op. cit.*).

Entre los causantes históricos del incremento del peso a nivel global suelen señalarse:

- Disminución de los precios de los alimentos con mayor aporte calórico y encarecimiento relativo de frutas y verduras.
- Incremento del costo de oportunidad del tiempo de las mujeres por inserción en el mercado laboral, incentivando así la ingesta de alimentos de rápida preparación y/o su adquisición fuera del hogar.
- Incremento de las actividades sedentarias como mirar TV, realizar tareas en computadora o videojuegos.
- Cambios tecnológicos que facilitan las tareas domésticas y laborales, requiriendo menor consumo energético.
- Incremento del número de establecimientos de comidas rápidas.
- Campañas publicitarias masivas tendientes a incrementar el consumo de alimentos ricos en azúcares y grasas, especialmente dirigidas al segmento infantil.

La genética puede considerarse un factor importante y predisponente a esta enfermedad, pero no puede explicar *per se* el notable aumento de la obesidad en las últimas décadas (Temporelli & Viego, *op. cit.*).

Cada uno de los factores mencionados anteriormente, adquiere diferente significado según el lugar donde se presenten. Sin embargo, hay acuerdo entre los estudios de salud pública que las diferencias en el perfil epidemiológico de las enfermedades que afectan a las poblaciones de países desarrollados y en vía de desarrollo podrían deberse a las diversas características geográficas, medioambientales, demográficas, socioeconómicas y étnicas (Serrano, 2005). Así es como varios estudios llegan a la conclusión que existe una estrecha relación entre las condiciones socioeconómicas y la prevalencia de obesidad. Si bien esta relación parece existir no hay acuerdo en el sentido de la misma (Temporelli & Viego, *op. cit.*). Un buen entendimiento de dicha asociación tiene importantes implicancias en las políticas de salud pública, particularmente en la prevención y manejo de la obesidad (Wang & Zhang, 2006).

Uno de los primeros estudios sobre este tema fue realizado en Manhattan en 1965. En él se relaciona la obesidad con algunos factores socioeconómicos, en especial el ingreso. Dentro de sus resultados, se observa que las mujeres que provenían de sectores de ingresos más bajos tenían 6 veces mayores chances de ser obesas que las que provenían de hogares de mayores ingresos (Goldblatt *et al.*, 1965).

Una revisión de 144 estudios publicados sobre la relación entre obesidad y nivel socioeconómico realizada en 1989 revela una relación inversa robusta entre ingreso y salud en el caso de las mujeres en los países desarrollados mientras que esta relación varía en significación y signo en hombres y niños (Temporelli & Viego, *op. cit.*). Sin embargo, en ámbitos en desarrollo, los trabajos evidenciaron una relación directa y fuerte entre las condiciones socioeconómicas y la obesidad tanto en hombres, mujeres y niños (Sobal y Stunkard, 1989).

La pobreza ha sido identificado también como un factor determinante de obesidad aunque su incidencia puede variar entre regiones. La dieta de los hogares pobres está conformada por alimentos que pueden conseguir o comprar y no necesariamente coincide con sus deseos o recomendaciones nutricionales. Bajos niveles de ingreso favorecen el consumo de alimentos ricos en grasas y carbohidratos, cuyos precios son reducidos y poseen mayor poder de saciedad.

En resumen, la obesidad está influenciada por las condiciones socioeconómicas a través de la educación y el ingreso, que inciden sobre los patrones de consumo y gasto de energía de los sujetos y, además, la obesidad puede afectar asimismo las condiciones socioeconómicas futuras al incidir negativamente sobre la salud de los individuos disminuyendo su calidad de vida y productividad laboral (Temporelli & Viego, *op. cit.*).

Wang & Zhang (*op. cit.*) analizaron la relación entre el nivel socioeconómico y el sobrepeso en niños y adolescentes de Estados Unidos, y encontraron diferencias por raza, sexo y edad en la asociación

entre sobrepeso y nivel socioeconómico. Además, encontraron que dicha asociación se ha ido debilitando en los últimos años, por lo tanto, solamente reduciendo la diferencia de ingresos no reducirá la disparidad en obesidad.

Fortich Mesa & Gutierrez (2011) exploraron las asociaciones entre el IMC y las características socioeconómicas de los hogares en Colombia, con datos de la Encuesta Nacional de Demografía y Salud de 2010. Llegaron a la conclusión de que una mayor educación está asociada significativamente con un menor IMC y menor prevalencia de sobrepeso y obesidad, tanto en hombres como mujeres. Además, el IMC también está relacionado positivamente con el índice de riqueza, la región, el clima, la edad. También analizan la situación conyugal, encontrando que una persona casada tiene una mayor probabilidad de estar excedida de peso que una persona no casada, mientras que una persona divorciada tienen mayor probabilidad de tener sobrepeso u obesidad.

Ortiz-Moncada *et al.* (2011), analizaron los determinantes sociales del sobrepeso y la obesidad en España, a partir de la Encuesta Nacional de Salud (ENS) de 2006. Se calculó el IMC como variable a analizar, y se incluyeron las siguientes variables explicativas sociodemográficas: edad, sexo, estado civil, nivel de estudios, clase social e ingresos mensuales. Se encontró asociación significativa entre obesidad y estar casado o viudo, sin estudios, e ingresos menores a 600 euros. También se encontró que el género y la clase social son determinantes sociales relevantes de la obesidad en España.

Manios *et al.* (2005) hallaron un aumento en la prevalencia de la obesidad a medida que aumenta la edad de los adultos en Grecia. Los autores evaluaron hábitos nutricionales y concluyeron que el indicador de dietas está positivamente relacionado con el nivel socioeconómico, de manera que los grupos con menor nivel socioeconómico presentan mayor probabilidad de consumir dietas poco saludables en relación a los grupos con más alto nivel.

García Milian & Creus García (2016) mencionan entre los factores que aumentan los riesgos de padecer obesidad a la falta de actividad, tener una dieta no saludable, genética, hábitos familiares, y también incluyen a las condiciones sociales y económicas.

En Argentina, Temporelli & Viego (*op. cit.*) analizaron los resultados de la Encuesta Nacional de Factores de Riesgo del 2005, llegando a las siguientes conclusiones:

- La realización de actividad física tiene una relación negativa con el exceso de peso, especialmente en presencia de sobrepeso. En presencia de obesidad, esta variable es menos relevante.
- El nivel de ingreso *per cápita* se relaciona positivamente con el exceso de peso, especialmente de obesidad. Niveles de ingreso bajo inciden con más notoriedad en la aparición de eventos de obesidad.
- La educación permite menguar la aparición de obesidad en la población.

- La posesión de seguros de salud alienta la aparición de problemáticas de exceso de peso, espacialmente de obesidad.

Lara & Serio (2015) analizaron los resultados de la misma encuesta tanto a nivel nacional como a nivel provincial en Mendoza, encontrando que los principales determinantes socioeconómicos que aumentan la probabilidad de ser obeso son: la edad (a mayor edad, mayor probabilidad), el sexo (hombre) y el nivel educativo (tener un menor nivel educativo aumenta la probabilidad). Entre los factores que disminuyen la probabilidad de tener obesidad se encuentran la situación conyugal (una persona soltera tendrá menor probabilidad de padecerla), tener una dieta balanceada y realizar actividad física.

3. Metodología aplicada

3.a. Desarrollo de la metodología econométrica

Para analizar los datos de la encuesta, en primer lugar, se realiza un análisis descriptivo conforme la naturaleza de las variables. Para las variables numéricas o cuantitativas, se calculan medidas de resumen y para las variables cualitativas o categóricas, frecuencias absolutas y relativas. El objetivo es conocer las principales características de los encuestados y evaluar la asociación entre el IMC (o sus categorías) y diversas variables mencionadas en la literatura como posibles determinantes del mismo. Para ello, se implementa un análisis bivariado, la Prueba Chi Cuadrado de Pearson (ver Anexo a).

Luego, se aplica metodología econométrica, con el objetivo de identificar factores demográficos y socioeconómicos que inciden sobre los indicadores de sobrepeso y la obesidad en la Provincia de Buenos Aires. Por este motivo, se estiman tres modelos:

- Modelo de regresión lineal múltiple: considerando al IMC como variable dependiente numérica.
- Modelo de regresión logística binaria: considerando como variable dependiente a una variable dummy, que surge de clasificar al IMC en dos categorías (peso saludable o exceso de peso).
- Modelo de regresión logística ordinal: considerando como variable dependiente a una variable categórica y de naturaleza ordinal, que surge de clasificar al IMC en tres categorías de acuerdo a la OMS (peso saludable, sobrepeso y obesidad).

A continuación se desarrolla la metodología correspondiente a estos modelos.

3.a.i. Modelo de regresión lineal múltiple

Siguiendo a Wooldridge (2010), el análisis de regresión múltiple permite inferir causalidad entre más de una variable explicativa (o regresores) sobre la variable dependiente. La forma funcional del modelo es la siguiente:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i + \mu \quad (1)$$

Dónde:

- β_0 es el intercepto, es decir, el valor predicho de Y cuando todas las variables (X) son igual a 0.
- β_1 mide el cambio en Y cuando X_1 se incrementa en una unidad, manteniendo constantes todos los demás factores (efecto marginal).
- β_2 mide el cambio en Y cuando X_2 se incrementa en una unidad, manteniendo constantes todos los demás factores, y así sucesivamente.
- μ es el término de error, y contiene otros factores no observados que afectan a Y.

Este modelo plantea una serie de supuestos que, si se cumplen, los estimadores de Mínimos Cuadrados Ordinarios (MCO) de los parámetros poblacionales obtenidos serán los mejores estimadores lineales insesgados:

- Linealidad en los parámetros (Dada por la forma funcional del modelo especificada en la ecuación 1). Este supuesto implica que el efecto marginal no depende del nivel de los regresores.
- Muestreo aleatorio: se tiene una muestra aleatoria de n observaciones, que sigue el modelo poblacional de la ecuación 1.
- No hay colinealidad perfecta: en la muestra, y por lo tanto en la población, ninguna de las variables independientes es constante y no hay ninguna relación lineal exacta entre ellas.
- Media condicional cero: para cualquier valor de las variables independientes en la población, el promedio del efecto de los factores no observables es igual a cero. Este supuesto implica que los estimadores MCO son insesgados.

$$E(\mu|x_1, x_2, \dots, x_i) = 0 \quad (2)$$

- Homocedasticidad: dado cualquier valor de las variables explicativas, el error μ tienen la misma varianza. Es decir, $Var(\mu|X_1, \dots, X_j) = \sigma^2$. Este supuesto implica que los estimadores MCO son eficientes.

Este modelo es estimado mediante el método de MCO, que elige las estimaciones que minimizan la suma de residuos al cuadrado:

$$\begin{aligned} SRC(\hat{\beta}) &= (y - x\hat{\beta})'(y - x\hat{\beta}) \\ &= y'y - \hat{\beta}'x'y - y'x\hat{\beta} + \hat{\beta}'x'x\hat{\beta} \\ &= y'y - 2\hat{\beta}'x'y + \hat{\beta}'x'x\hat{\beta} \quad (3) \end{aligned}$$

De la condición de primer orden surge el estimador de MCO:

$$\frac{\partial SRC(\hat{\beta})}{\partial \hat{\beta}} = -2x'y + 2x'x\hat{\beta} = 0 \quad (4)$$

$$\hat{\beta} = (x'x)^{-1}x'y \quad (5)$$

Para analizar la bondad de ajuste del modelo, se definen la suma total de cuadrados (STC), la suma explicada de cuadrados (SEC) y la suma de residuales cuadrados (SRC) como:

$$STC = \sum_{i=1}^n (Y_i - \bar{Y})^2 \quad (6)$$

$$SEC = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 \quad (7)$$

$$SRC = \sum_{i=1}^n (\hat{\mu}_i)^2 \quad (8)$$

De modo que $STC = SEC + SRC$.

Entonces:

$$R^2 = \frac{SEC}{STC} = 1 - \frac{SRC}{STC} \quad (9)$$

Y se interpreta al R^2 (coeficiente de determinación) como la proporción de la variación muestral en Y_i que es explicada por la línea de regresión de MCO. Por definición, éste es un número entre 0 y 1.

3.a.ii. Modelo de regresión logística binaria

El modelo de regresión logística binaria se utiliza en el caso en que la variable dependiente sólo toma dos valores: 0 -ausencia de la característica que se analiza -o 1- presencia de la característica que se analiza- y, por este motivo, al no cumplirse los supuestos en los que se basa la estimación por MCO, no resulta adecuada dicha estimación. Además, los valores predichos de la variable dependiente se ven afectados, ya que no puede asegurarse que estén comprendidos entre 0 y 1 (Wooldridge, *op. cit.*).

Como describe Wooldridge (*op. cit.*), la probabilidad de que la variable dependiente tome valor 1 -probabilidad de éxito- es una función no lineal de un conjunto de variables X_i , de manera que $\Pr(Y=1) = F(X\beta)$. En el Modelo *Logit*, G es la función logística:

$$G(z) = \frac{\exp(z)}{1+\exp(z)} \quad (10)$$

Esta función está entre 0 y 1 para todos los números reales z y es la función de distribución acumulada para una variable aleatoria logística estándar. Dicha función es creciente y aumenta con más rapidez cuando $z=0$. $G(z) \rightarrow 0$ a medida que $z \rightarrow -\infty$, y $G(z) \rightarrow 1$ a medida que $z \rightarrow \infty$.

El método de estimación utilizado en este caso es el de Máxima Verosimilitud (MV). Como la estimación de MV está basada en la distribución de Y dada X , la heteroscedasticidad en $\text{Var}(Y/X)$ automáticamente se toma en cuenta. El estimador de máxima verosimilitud (EMV) es consistente y está distribuido normalmente en muestras grandes por lo que puede construirse de forma habitual a los estadísticos t y los intervalos de confianza (IC) de los coeficientes.

Por su parte, el EMV de los coeficientes desconocidos está compuesto por los valores de los coeficientes que maximizan la función de verosimilitud y elige los valores de los parámetros que maximizan la probabilidad de haber obtenido los datos que realmente se observan (Stock & Watson, 2012).

Para obtener el EMV, condicional sobre las variables explicativas, se necesita la densidad de y_i dado x_i . Esto se puede escribir como:

$$f(y|x_i; \beta) = [G(x_i\beta)]^y [1 - G(x_i\beta)]^{1-y} \quad \text{con } y = 0 \text{ o } y = 1 \quad (11)$$

dónde, por simplicidad, se absorbe el intercepto en el vector x_i . La función de log-verosimilitud para la observación i es una función de los parámetros y los datos (x_i, y_i) , y se obtiene al aplicar el log a la ecuación anterior:

$$\mathcal{L}_i(\beta) = y_i \log[G(x_i\beta)] + (1 - y_i) \log[1 - G(x_i\beta)] \quad (12)$$

debido a que $G(\cdot)$ está estrictamente entre cero y uno, $\mathcal{L}_1(\beta)$ está bien definida para todos los valores de β .

La log-verosimilitud para un tamaño de muestra n , se obtiene al sumar la ecuación anterior a través de todas las observaciones: $\mathcal{L}(\beta) = \sum_{i=1}^n \mathcal{L}_i(\beta)$. La EMV de β maximiza esta log-verosimilitud.

Este modelo, permite la interpretación de los $\exp(\beta)$ de la salida de regresión, en términos de cocientes de chances, de manera que representa el efecto multiplicativo en las chances de que $Y=1$, cuando la variable explicativa varía en una unidad manteniendo las demás variables constantes (Agresti, 2002).

Cabe destacar que, los valores de los Coeficientes de Determinación, llamados Pseudo R^2 en formulaciones con variable a explicar no numérica, no son elevados. En este sentido, Wooldridge (*op. cit.*) señala que, en series de corte transversal, el R^2 es bajo. Agrega que un R^2 pequeño refleja lo que se sospecha en las ciencias sociales, y es que en general, es muy difícil predecir el comportamiento de los individuos. A su vez, tanto Wooldridge (*op. cit.*) como Sosa Escudero (2015) relativizan su importancia como medida de bondad del ajuste aunque sugieren analizarlo centrando el interés en el uso que se le quiera dar al modelo, recomendando mayor nivel de exigencia si se desea predecir que si sólo se desea estimar efectos.

Por su parte, Norusis (2005) también afirma que los Pseudo R^2 no son utilizados como el R^2 de las regresiones estimadas mediante el Método de Mínimo Cuadrados Ordinarios (MCO). Siguiendo a Menard (2000), en este último caso, hay consenso sobre el empleo del R^2 pero, cuando se trata de un Modelo *Logit* hay controversia. Como indica Efron (1978), para variables a explicar cuantitativas, sólo hay un único criterio razonable para calcular la variación de los residuos: la suma de los errores al cuadrado. Lo anterior, no ocurre si la variable a explicar es cualitativa.

3.a.iii. Modelo de regresión logística ordinal

Si se considera como variable dependiente a una variable cuyos valores pueden ser 0-Peso normal, 1-Sobrepeso, y 2-Obesidad, se trata de una variable dependiente categórica y de naturaleza ordinal¹, y es posible aplicar una Regresión *Logit* Ordinal o una Regresión *Probit* Ordinal. Ambas se pueden estimar mediante el Método de Máxima Verosimilitud. Sin embargo, la principal diferencia entre ellas radica en que la primera asume una distribución logística mientras que la segunda surge de una distribución normal. En este caso, no es posible aplicar un modelo de regresión multinomial debido a que este tipo de modelos ignora el orden de las categorías (Norusis, *op. cit.*).

Según Amemiya (1985), el Modelo *Logit* Ordinal es mejor que el Modelo *Probit* Ordinal para captar la magnitud de los efectos de las variables explicativas en formulaciones econométricas polinomiales

¹Si bien sus valores pueden ser ordenados, la verdadera distancia entre las categorías es desconocida.

cuando dichas variables tienen naturaleza ordinal. Por su parte, Theil (1971) sostiene que la distribución normal que subyace a un Modelo *Probit* es poco frecuente en aplicaciones empíricas económicas.

Finalmente, Underhill & Figueroa (1996) afirman que una regresión *logit* es adecuada cuando la variable dependiente tiene más de dos categorías y, en aplicaciones econométricas, debido a que el supuesto de normalidad no es muy fuerte en las mismas.

La Regresión *Logit* Ordinal es una extensión de la Regresión *Logit* Binaria que contempla la comparación simultánea de más de un contraste y, a menudo, involucra modelos logísticos acumulativos. Este modelo no considera la probabilidad de un evento individual, sino que considera la probabilidad de ese evento y todos los eventos ordenados antes de éste (Norusis, *op. cit.*). Siguiendo a Agresti (*op. cit.*), los modelos logísticos acumulativos asumen la siguiente forma general:

$$\text{logit}[P(Y \leq j|x)] = \log \left[\frac{P(Y \leq j|x)}{1-P(Y \leq j|x)} \right] = \log \left[\frac{\pi_1(x)+\dots+\pi_j(x)}{\pi_{j+1}(x)+\dots+\pi_J(x)} \right] \quad (13) \quad \text{con } j = 1, \dots, (J - 1)$$

Siendo j la categoría ordinal de la variable dependiente y $\pi_j(x) = P(Y \leq j|x)$.

Por definición, las probabilidades acumuladas deben sumar 1 cuando $j = J$ –siendo J el total de categorías-. Por lo tanto sólo $(J - 1)$ probabilidades acumuladas son identificadas.

Un modelo que, simultáneamente, emplea todos los *logits* acumulativos es este:

$$\text{logit}[P(Y \leq j|x)] = \alpha_j + \beta'x \quad (14) \quad \text{con } j = 1, \dots, (J - 1)$$

En donde α es el intercepto desconocido, y β el vector de los coeficientes desconocidos que corresponde a x .

Cuando $\beta > 0$, el *logit* acumulativo se incrementa cuando x aumenta. Esto implica que Y tiende a ser menor para valores altos de X . Por el contrario, cuando $\beta < 0$ el incremento en x está asociado con niveles más altos de Y (Powers & Xie, 1999).

Cada *logit* acumulativo tiene su propia ordenada al origen (α). Los α_j se incrementan en j , dado que $P(Y \leq j | x)$ se incrementa en j para valores fijos de x y el *logit* es una función creciente de dicha probabilidad. Estos valores reciben el nombre de “valores umbral o frontera”. A excepción de que se quieran computar probabilidades de respuesta, dichos valores no son de interés.

Cada modelo tiene el mismo efecto β para cada *logit*. Los efectos de x no varían con respecto a las categorías de respuesta, esto es, que β no está indexado por j . Dichos coeficientes, reciben el nombre de “coeficientes efectos” ya que reflejan los efectos de las variables explicativas.

El Modelo *Logit* Acumulativo (14) satisface:

$$\text{logit}[P(Y \leq j|x_1)] - \text{logit}[P(Y \leq j|x_2)] = \log \frac{[P(Y \leq j|x_1)]/[P(Y > j|x_1)]}{[P(Y \leq j|x_2)]/[P(Y > j|x_2)]} = \beta'(x_1 - x_2) \quad (15)$$

Un cociente de chances (*odds ratio*) de probabilidades acumuladas se denomina “cociente de chances acumulativo”. Las chances (*odds*) de respuestas $\leq j$ en $x = x_1$ son $e^{[\beta'(x_1 - x_2)]}$ veces las chances en x

= x_2 . El logaritmo de los cocientes de chances acumulativos es proporcional a la distancia entre x_1 y x_2 . La misma proporcionalidad se mantiene en cada *logit*.

Por su parte, la probabilidad predicha acumulada se puede calcular de acuerdo a la siguiente fórmula:

$$P(Frec_j) = \frac{1}{[1+e^{-(\alpha_j-\beta X)}]} \quad (16)$$

3.b. Datos empleados

El Instituto Nacional de Estadística y Censos (INDEC) conjuntamente con el Ministerio de Salud y Desarrollo Social (MSyDS) realizaron la Cuarta Encuesta Nacional de Factores de Riesgo (ENFR) en 2018 con los siguientes objetivos:

1. Proporcionar información válida, confiable y oportuna sobre los factores de riesgo, procesos de atención en el Sistema de salud y prevalencias de las principales ENT (hipertensión, diabetes, obesidad y otras) en la población de 18 años y más, residente en viviendas particulares de áreas urbanas de 5.000 y más habitantes de la República Argentina.
2. Monitorear la evolución de los principales factores de riesgo de las ENT.
3. Evaluar el impacto de políticas de prevención y control realizadas a nivel nacional y provincial.
4. Proporcionar parámetros objetivos de los principales factores de riesgo y las ENT a partir de las mediciones físicas y bioquímicas.

3.b.i. Diseño de la muestra y dominio de estimación

Con el objetivo de maximizar los recursos, integrar y coordinar las operaciones estadísticas, el INDEC emplea un diseño de muestra bajo el esquema de muestra maestra. Es decir, una única gran muestra probabilística que mantiene fijas las unidades de área que la conforman y su estructura probabilística asociada, y que permite subseleccionar las muestras de viviendas para todas las encuestas a hogares del Instituto durante aproximadamente un decenio, o período intercensal. A dicha muestra se la conoce como Muestra Maestra Urbana de Viviendas de la República Argentina (MMUVRA).

La MMUVRA es de alcance nacional y urbano, permite subseleccionar muestras para las encuestas que tienen como principales dominios de estimación a las provincias y a los aglomerados que participan de la Encuesta Permanente de Hogares (EPH) que lleva a cabo el instituto.

Su diseño emplea dos etapas de selección probabilística. Cada unidad de primera etapa de muestreo (UPM) del diseño está definida por un aglomerado o localidad de al menos 2.000 habitantes según en Censo Nacional de Población, Hogares y Viviendas 2010 (CNPHV 2010); el conjunto de todas las UPM

constituyen el marco de maestro para la selección probabilística de primera etapa. Estas son estratificadas de acuerdo al total de población según el CNPHV 2010 y aquellas UPM formadas por aglomerados o localidades de 50.000 habitantes o más son incluidas en la MMUVRA con probabilidad 1 por diseño, y se las denomina UPM autorrepresentadas.

Del resto de las UPM, un conjunto fue seleccionado por provincial mediante un maestro sistemático con probabilidad proporcional a la cantidad total de habitantes. Tanto las UPM autorrepresentadas como las seleccionadas conforman la muestra de aglomerados o localidades de la MMUVRA.

Para la segunda etapa, en las UPM seleccionadas, y solo para ellas, se definieron las unidades de segunda etapa de maestro (USM) o Áreas MMUVRA con base en radios censales, y a la cartografía del CNPHV 2010. En cada UPM, todas sus USM en conjunto cubren territorialmente y determinan la envolvente o área de cobertura asociada a dicha unidad, conformando el arco de muestreo para la selección de segunda etapa. Esta se completa con la selección de una muestra probabilística de USM, que emplea un diseño estratificado definido a partir de variables sociodemográficas y mediante un muestreo sistemático proporcional a la cantidad total de viviendas particulares ocupadas, según el CNPHV 2010.

Por último, en cada una de las USM seleccionadas, se confeccionó inicialmente un listado exhaustivo de viviendas particulares, lo que da origen al marco de selección de viviendas de la MMUVRA y sobre el cual se realizan las subselecciones para las muestras de todas las encuestas a hogares del Instituto.

El diseño muestral de la ENFR se apoya en el diseño de la MMUVRA ajustado a los requerimientos de la encuesta, que como población objetivo incluye a las localidades y aglomerados de 5.000 o más habitantes. Para la muestra definitiva de viviendas de la encuesta se realiza una nueva etapa de selección probabilística de un tercer tipo de unidades de muestreo, denominados “segmentos”. Estos están constituidos por cinco viviendas particulares contiguas o próximas entre ellas dentro del listado de la MMUVRA, y cuyo principal objetivo es concentrar los desplazamientos en terreno de los encuestadores y del personal de salud, para reducir el costo del operativo. Una selección sistemática con igual probabilidad de estos segmentos permitió conformar la muestra definitiva de viviendas de la encuesta.

La encuesta por diseño tiene distintos dominios de estimación para cada uno de los pasos. El 100% de las viviendas seleccionadas en la muestra forman parte del Paso 1, lo que permite, además de resultados al nivel del total país, desagregarlos a nivel de 6 regiones:

- ◆ Gran Buenos Aires (incluye la Ciudad Autónoma de Buenos Aires y los partidos del Gran Buenos Aires)
- ◆ Noroeste
- ◆ Noreste
- ◆ Cuyo

- ◆ Pampeana (incluye Córdoba, Santa Fe, Entre Ríos, La Pampa y el resto de los partidos de la provincial de Buenos Aires)

- ◆ Patagonia

Que además, pueden ser desagregadas en 24 jurisdicciones (23 provincias y CABA), y en ocho aglomerados urbanos de más de 500.000 habitantes:

- ◆ Gran Buenos Aires

- ◆ Gran Córdoba

- ◆ Gran La Plata

- ◆ Gran Mendoza

- ◆ Gran Rosario

- ◆ Gran Salta

- ◆ Gran Tucumán

- ◆ Mar del Plata

3.b.ii. Trabajo de campo, cuestionario

La encuesta se relevó mediante la aplicación de un cuestionario digital en un dispositivo electrónico, durante una entrevista presencial en la Vivienda, con intervención de un encuestador para la entrevista y un representante de salud para realizar las mediciones físicas y bioquímicas. Los equipos de relevamiento contaban además con una versión del cuestionario en papel, para su uso excepcional en caso de que no pudiera utilizarse el dispositivo digital. El cuestionario está conformado por dos bloques: un bloque Hogar y un bloque Individual. En el primero se relevaron datos sobre las condiciones habitacionales, así como características demográficas y educativas de todas las personas que conforman el hogar y datos laborales del jefe de hogar, mientras que el bloque individual estaba destinado a registrar datos laborales, así como información sobre los factores de riesgo de las ENT (Anexo b).

3.b.iii. Submuestra utilizada

Dado que el objetivo de esta tesis será identificar qué factores demográficos y socioeconómicos influyen en la prevalencia de obesidad en mayores de 18 años en la Provincia de Buenos Aires, no se utilizarán todas las observaciones de la muestra. Sólo se analizarán a los encuestados de la Provincia de Buenos Aires, que hayan respondido las preguntas que permiten el cálculo del IMC.

3.c. Especificación econométrica

Los modelos estimados toman la forma:

$$Y = f(\text{actividad física, edad, situación conyugal, alimentación, nivel educativo, región, sexo, } \mu)$$

Donde Y representa a las distintas variables dependientes incluídas en cada modelo, cuya definición puede observarse en la siguiente tabla:

Tabla 2: Variables dependientes analizadas

MODELO	VARIABLE	DEFINICIÓN	CATEGORIA
REGRESIÓN LINEAL MÚLTIPLE	IMC	Variable numérica continua que indica el IMC del encuestado.	
REGRESIÓN LOGÍSTICA BINARIA	EXCPESO	Variable categórica que indica la presencia de exceso de peso	0 = Peso saludable 1 = Sobrepeso u obesidad
REGRESIÓN LOGÍSTICA ORDINAL	CAT_IMC	Variable ordinal que indica la clasificación del IMC en rangos (definida por la OMS)	0 = Peso saludable 1 = Sobrepeso 2 = Obesidad

A partir de la bibliografía y de los modelos analizados, las variables explicativas son las siguientes:

Tabla 3: Regresores analizados

VARIABLE	DEFINICIÓN	CATEGORIA
ACT_FIS	Variable categórica nominal que indica si el encuestado realiza actividad física	0 = No 1 = Si
ADULTO	Variable categórica nominal que indica si el encuestado tiene entre 30 y 59 años	0 = No 1 = Si
CASADO	Variable categórica nominal que indica si la situación conyugal del encuestado es Casado.	0 = No 1 = Si
CONS_FV	Variable categórica nominal que indica si el encuestado consume las 5 porciones diarias recomendadas de frutas o verduras.	0 = No 1 = Si
EDUC1	Variable categórica nominal que indica si el nivel de educación de encuestado es hasta primaria incompleta.	0 = No 1 = Si
EDUC2	Variable categórica nominal que indica si el nivel de educación de encuestado es hasta primaria completa.	0 = No 1 = Si
EDUC3	Variable categórica nominal que indica si el nivel de educación de encuestado es hasta secundario completo.	0 = No 1 = Si
EDUC4	Variable categórica nominal que indica si el nivel de educación de encuestado es terciario/universitario completo.	0 = No 1 = Si
JOVEN	Variable categórica nominal que indica si el encuestado tiene entre 18 y 29 años	0 = No 1 = Si
REGION	Variable categórica nominal que indica la región dentro de la Provincia de Buenos Aires	0 = Pampeana 1 = Metropolitana
SEPARADO	Variable categórica nominal que indica si la situación conyugal del encuestado es Separado.	0 = No 1 = Si

SEXO	Variable categórica nominal que indica el sexo del encuestado	0 = Mujer 1 = Varón
SOLTERO	Variable categórica nominal que indica si la situación conyugal del encuestado es soltero.	0 = No 1 = Si

La edad fue preguntada como variable numérica, pero para este análisis fue clasificada en 3 grupos: Joven (18 a 29 años), Adulto (30 a 59 años) y Adulto Mayor (60 años o más).

La variable CONS_FV fue creada en base a la recomendación de diversos organismos, entre los que se puede mencionar a la OMS, la Asociación Americana del Corazón (AHA por sus siglas en inglés), y al Ministerio de Salud que a través de las Guías Alimentarias para la Población Argentina recomienda el consumo de al menos 5 porciones diarias de frutas o verduras.

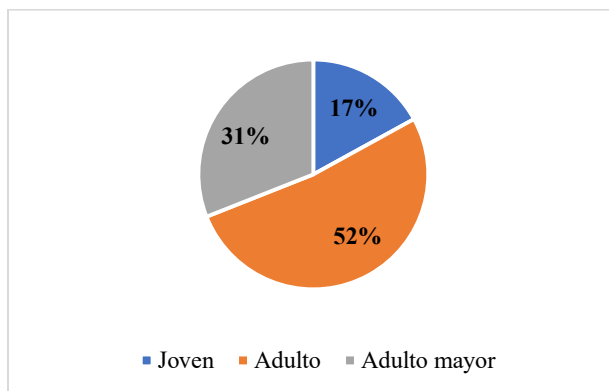
4. Resultados

4.a. Análisis descriptivo

4.a.i. Descripción de la submuestra:

Entre las principales características de los encuestados se observa que el 55% de ellos son mujeres, mientras que el 45% son hombres. La edad promedio de los mismos es de 49 años, habiendo un 17% de jóvenes, 31% de adultos mayores y un 52% de adultos.

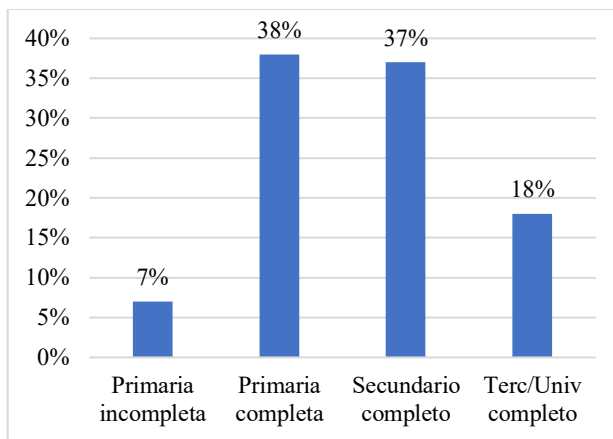
Figura 1: Rango de edad de los encuestados



Fuente: elaboración propia en base a la ENFR 2018

Con respecto a la educación, el 38% de los encuestados tiene hasta nivel primario completo, y un 37% completó hasta el nivel secundario. Sólo un 18% de ellos completó el nivel universitario o terciario.

Figura 2: Educación de los encuestados



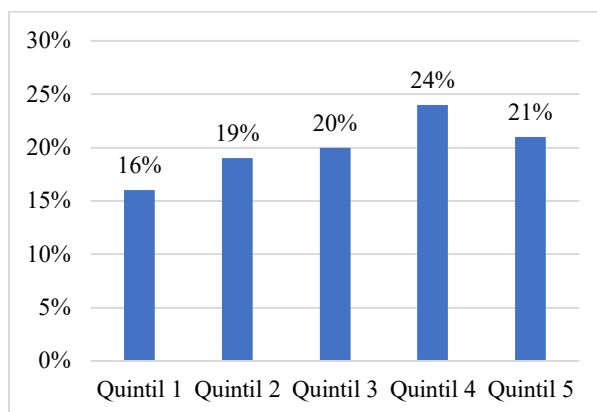
Fuente: elaboración propia en base a la ENFR 2018

Entre otras características también se puede mencionar que el 60% de los encuestados tiene un empleo, mientras que un 5% es desocupado y un 35% inactivo, y que con respecto a la cobertura de salud, el 75% de los encuestados posee una cobertura privada mientras que el 25% restante tiene acceso únicamente a la cobertura pública.

Dentro de las características de los hogares analizadas, se observa que un 61% de ellos se encuentra en la región pampeana de la Provincia de Buenos Aires, es decir, fuera del Área Metropolitana de Buenos Aires (AMBA). Además, el 76% de los hogares de los encuestados corresponden a hogares multipersonales, mientras que el 24% restante son hogares unipersonales.

Con respecto al ingreso, el mayor porcentaje de encuestados se ubica en el quintil 4 (24%), seguido por el quintil 5 (21%) y quintil 3 (20%). El menor porcentaje de encuestados se encuentra en el quintil 1 (16%). Además, sólo un 15% de los hogares encuestados recibe algún tipo de subsidio o programa social.

Figura 3: Quintil de ingreso de los encuestados



Fuente: elaboración propia en base a la ENFR 2018

4.a.ii. Análisis del IMC y sus determinantes:

Al analizar el **IMC**, se observa que el promedio en mayores de 18 años de la Provincia de Buenos Aires es de 27,55; siendo 14,7 el mínimo y 62,4 el máximo observado. No se observan grandes diferencias al analizar su promedio por región o aglomerado.

A continuación se analizará la asociación entre el IMC, agrupada en intervalos definidos por la OMS, y los posibles factores determinantes mencionados en la bibliografía sobre el tema.

Se observa que la mayor proporción de encuestados tiene sobrepeso, tanto a nivel provincial como por región. También se observa una menor proporción de encuestados con sobrepeso en el AMBA que en el resto de la provincia, mientras que hay una mayor proporción de encuestados con obesidad en el AMBA. La proporción de encuestados con peso saludable es muy similar a nivel región o provincial.

Tabla 4: IMC por Región

IMC / REGIÓN	METROPOLITANA	PAMPEANA	TOTAL PROVINCIA BS. AS.
PESO INSUFICIENTE (< 18,5)	1%	1%	1%
PESO SALUDABLE (18,5 – 24,9)	33%	34%	34%
SOBREPESO (25 – 29,9)	36%	38%	37%
OBESIDAD (>30)	30%	27%	28%

Fuente: elaboración propia en base a la ENFR 2018

Al analizar la relación entre el rango de IMC y el **quintil de ingreso**, es posible indicar que, según la Prueba Chi Cuadrado de Pearson, ambas variables se encuentran estadísticamente asociadas (Valor “p” = 0,005). En la siguiente tabla se observa que dentro de los encuestados del quintil 1, hay una mayor proporción que tiene peso saludable, pero ese porcentaje es apenas superior al de los encuestados con sobrepeso. En los quintiles 2 a 5, el porcentaje de encuestados con sobrepeso es muy similar (entre 36,8% y 37,9%), sin embargo se observa que dentro de cada quintil, el porcentaje de encuestados con obesidad disminuye a medida que el quintil de ingreso es más alto.

Tabla 5: IMC por Quintil de Ingreso

IMC / QUINTIL	QUINTIL	QUINTIL	QUINTIL	QUINTIL	QUINTIL
	1	2	3	4	5
PESO INSUFICIENTE (< 18,5)	1,9%	1,4%	0,9%	0,9%	1,2%
PESO SALUDABLE (18,5 – 24,9)	33,8%	31,3%	31,6%	34%	37,2%
SOBREPESO (25 – 29,9)	33,6%	36,8%	37,5%	37,9%	37,5%
OBESIDAD (>30)	30,7%	30,5%	30%	27,2%	24,2%

Fuente: elaboración propia en base a la ENFR 2018

Con respecto al **sexo del encuestado**, también se encuentra estadísticamente asociado al nivel de IMC según la Prueba Chi Cuadrado de Pearson (Valor “p” = 0,000). Un 42,4% de los hombres encuestados tienen sobrepeso y un 29,1% obesidad, mientras que dentro de las mujeres encuestadas, el mayor porcentaje de ellas posee un peso saludable (38,4%), seguido por un 32,4% que presenta obesidad.

Tabla 6: IMC por Sexo del encuestado

IMC / SEXO	VARON	MUJER
PESO INSUFICIENTE (< 18,5)	0,6%	1,7%
PESO SALUDABLE (18,5 – 24,9)	27,8%	38,4%
SOBREPESO (25 – 29,9)	42,4%	32,4%
OBESIDAD (>30)	29,1%	27,6%

Fuente: elaboración propia en base a la ENFR 2018

También se observa asociación entre las variables IMC y **rango de edad** (Valor “p” = 0,000 de la prueba Chi Cuadrado de Pearson). En la tabla 7 se observa que dentro de los jóvenes encuestados, más de la mitad de ellos poseen un peso saludable, mientras que entre los adultos y adultos mayores, la mayor proporción de ellos poseen sobrepeso, siendo mayor el porcentaje entre los adultos mayores (42,2% vs. 36,2%). Además, puede observarse que el porcentaje de encuestados con peso saludable va disminuyendo al aumentar el rango de edad, y lo opuesto ocurre con el sobrepeso y la obesidad.

Tabla 7: IMC por Rango de edad

IMC / RANGO DE EDAD	JOVEN	ADULTO	ADULTO MAYOR
PESO INSUFICIENTE (< 18,5)	3,1%	1,1%	0,3%
PESO SALUDABLE (18,5 – 24,9)	52,2%	33,0%	24,4%
SOBREPESO (25 – 29,9)	29,4%	36,2%	42,2%
OBESIDAD (>30)	15,4%	29,7%	33,1%

Fuente: elaboración propia en base a la ENFR 2018

La **situación conyugal** es otra de las variables que se encuentra asociada al IMC (Valor “p” = 0,000). Entre los solteros encuestados, el 46,4% de ellos posee un peso saludable, seguido por un 32,7% con sobrepeso y un 18,7% obesidad. Este grupo posee los menores porcentajes de sobrepeso y obesidad

comparado al resto de las categorías. La mayor proporción de encuestados con sobrepeso se encuentra en el grupo de los divorciados, mientras que la mayor proporción de encuestados con obesidad se encuentra en el grupo de los casados.

Tabla 8: IMC por Situación conyugal

IMC / SITUACIÓN CONYUGAL	UNIDO	CASADO	SEPARADO	DIVORCIADO	VIUDO	SOLTERO
PESO INSUFICIENTE (< 18,5)	1,0%	0,8%	2,0%	0,4%	0,5%	2,3%
PESO SALUDABLE (18,5 – 24,9)	34,7%	26,1%	33,9%	30,2%	29,1%	46,4%
SOBREPESO (25 – 29,9)	36,0%	39,2%	34,9%	43,9%	38,2%	32,7%
OBESIDAD (>30)	28,3%	33,9%	29,1%	25,5%	32,2%	18,7%

Fuente: elaboración propia en base a la ENFR 2018

Con respecto a la **educación**, se observa asociación estadística entre ella y el IMC y (Valor “p” = 0,000). Dentro las diferentes categorías, el mayor porcentaje de encuestados con obesidad o sobrepeso se encuentra en el grupo de quienes tienen hasta primario incompleto, y esta proporción disminuye a medida que aumenta el nivel educativo. Sin embargo, con respecto al peso saludable, se observa una mayor proporción de encuestados entre quienes tienen nivel terciario o universitario completo, y ese porcentaje disminuye a medida que se tiene un nivel educativo inferior.

Tabla 9: IMC por Educación

IMC / EDUCACIÓN	HASTA PRIMARIO INCOMPLETO	PRIMARIA COMPLETA	SECUNDARIO COMPLETO	TERCIARIO /UNIVERSITARIO COMPLETO
PESO INSUFICIENTE (< 18,5)	0,3%	0,9%	1,6%	1,3%
PESO SALUDABLE (18,5 – 24,9)	21,1%	28,3%	37,7%	41,3%
SOBREPESO (25 – 29,9)	39,3%	38,2%	34,8%	37,4%
OBESIDAD (>30)	39,3%	32,6%	25,9%	20%

Fuente: elaboración propia en base a la ENFR 2018

También se observa asociación entre las variables IMC y **cobertura de salud** (Valor “p” = 0,010), encontrando que dentro de quienes tienen una cobertura privada, casi el 38% posee sobrepeso. Ese porcentaje desciende a 33,7% entre quienes tienen cobertura pública. Además, dentro de este grupo, la mayor proporción de encuestados tiene un peso saludable.

Tabla 10: IMC por Cobertura de Salud

IMC / COBERTURA DE SALUD	OBRA SOCIAL/ PREPAGA / SERVICIO DE EMERGENCIA MED.	COBERTURA PÚBLICA
PESO INSUFICIENTE (< 18,5)	1%	1,7%
PESO SALUDABLE (18,5 – 24,9)	33%	35,7%
SOBREPESO (25 – 29,9)	37,9%	33,7%
OBESIDAD (>30)	28,1%	28,8%

Fuente: elaboración propia en base a la ENFR 2018

Puede observarse que la **condición de actividad** es otra de las variables asociadas al IMC (Valor “p” = 0,000). La proporción de encuestados que presentan obesidad en las distintas categorías es similar (entre 36,6% y 37,2%), pero se observa una mayor proporción de encuestados con obesidad entre los inactivos (30,8%), seguidos por los ocupados (27%) que entre los desocupados (25,9%).

Tabla 11: IMC por Condición de Actividad

IMC / CONDICIÓN DE ACTIVIDAD	OCUPADO	DESOCUPADO	INACTIVO
PESO INSUFICIENTE (< 18,5)	1%	3,2%	1,2%
PESO SALUDABLE (18,5 – 24,9)	35,4%	34%	30,7%
SOBREPESO (25 – 29,9)	36,6%	36,9%	37,2%
OBESIDAD (>30)	27%	25,9%	30,8%

Fuente: elaboración propia en base a la ENFR 2018

La **actividad física** es otra de las variables asociadas significativamente al IMC (Valor “p” = 0,000). Tanto entre los encuestados que realizan actividad, como entre quienes no realizan, la mayor proporción posee sobrepeso (37,2% vs. 35,4%). Sin embargo se observa que la proporción de encuestados con obesidad es mayor entre quienes no realizan actividad física.

Tabla 12: IMC por Actividad Física

IMC / ACTIVIDAD FÍSICA	SI	NO
PESO INSUFICIENTE (< 18,5)	1,3%	0,7%
PESO SALUDABLE (18,5 – 24,9)	35,4%	26,9%
SOBREPESO (25 – 29,9)	37,2%	35,4%
OBESIDAD (>30)	26,1%	37%

Fuente: elaboración propia en base a la ENFR 2018

Por último, se analiza la asociación entre las variables IMC y **consumo de 5 porciones de frutas o verduras diarias**, tal como lo recomiendan las Guías Alimentarias para la Población Argentina (GAPA). Existe asociación significativa entre ambas variables (Valor “p” = 0,024). La proporción de encuestados con peso saludable o sobrepeso es mayor entre quienes consumen las 5 porciones diarias de frutas o verduras, mientras que dicha proporción es menor entre quienes tienen obesidad.

Tabla 13: IMC por Consumo de frutas o verduras

IMC / QUINTIL	SI	NO
PESO INSUFICIENTE (< 18,5)	0,3%	1,3%
PESO SALUDABLE (18,5 – 24,9)	36,2%	33,2%
SOBREPESO (25 – 29,9)	40%	36,7%
OBESIDAD (>30)	23,6%	28,8%

Fuente: elaboración propia en base a la ENFR 2018

4.b. Formulación y análisis de los modelos econométricos

4.b.i. Modelo de regresión lineal múltiple

En primer lugar, se realizó la estimación de una regresión lineal múltiple, con el objetivo de evaluar qué variables explican el IMC (variable continua). En la tabla 14 se pueden observar las estimaciones del modelo. En ella puede observarse que todas las variables son estadísticamente significativas, y que presentan los signos esperados de acuerdo a la bibliografía, y que los regresores que tienen un mayor impacto sobre el IMC son la edad, el nivel educativo, la actividad física y la situación conyugal.

- Si una persona tiene entre 18 y 29 años (JOVEN), su IMC será 2,017 puntos más bajo que el de una persona de 30 años o más, manteniendo constantes el resto de las variables.

- Una persona con nivel universitario o terciario completo (EDUC4) tendrá un IMC 1,724 puntos más bajo que una persona con un nivel educativo inferior, manteniendo constantes el resto de las variables. Tener como máximo nivel educativo alcanzado al secundario completo (EDUC3) también tiene un efecto negativo sobre el IMC aunque en menor medida.
- Alguien que realiza actividad física (ACT_FIS) también tendrá un IMC 1,228 más bajo que alguien que no realiza actividad, manteniendo el resto de las variables constantes.
- Si una persona se encuentra soltera se espera que su IMC sea 1,059 más bajo que alguien que no lo es, manteniendo el resto de las variables constantes. Estar separado tiene un impacto similar sobre el IMC (-0,999).

Tabla 14: Estimación modelo MCO

VARIABLES	COEFICIENTES NO ESTANDARIZADOS	COEFICIENTES TIPIFICADOS
CONSTANTE	29,321	
REGION	0,443***	0,040
SOLTERO	-1,059***	-0,082
SEPARADO	-0,999***	-0,049
EDUC3	-0,770***	-0,069
EDUC4	-1,724***	-0,123
SEXO	0,651***	0,060
ACT_FIS	-1,228***	-0,092
JOVEN	-2,017***	-0,141
R ²	0,071	
R ² AJUSTADO	0,070	

*** valor “p” < 1%, ** valor “p” < 5%, * valor “p” < 10%.

Fuente: elaboración propia en base a la ENFR 2018

Al realizar el contraste de regresión ANOVA, para verificar que las variables explicativas, de forma conjunta, aportan información en la explicación de la variable dependiente, se rechaza la siguiente hipótesis con un nivel de significación menor a 1%:

H0) Las variables explicativas no están relacionadas linealmente con Y

Sin embargo, al analizar la bondad del ajuste (a través del R² y R² ajustado) se observa que este modelo explica solamente un 7% de los cambios en el IMC, lo que puede dar idea de que haya otras cuestiones no analizadas en la encuesta que influyan sobre el IMC.

4.b.ii. Modelo de regresión logística binaria

Dado que el objetivo no es analizar los determinantes del bajo peso o desnutrición, para este análisis se eliminarán todas las observaciones correspondientes a los individuos con IMC menor a 18,5; de modo que los modelos estimados permitan comparar a los individuos con sobrepeso y obesidad contra individuos de peso saludable.

En primer lugar SPSS estima el modelo reducido, que incluye solamente la constante. Se observa que este modelo clasifica correctamente al 65,9% de las observaciones. Para este modelo se realiza el test Chi Cuadrado con la siguiente hipótesis nula, que es rechazada a un nivel de significación del 1%.

H0) El modelo sin la inclusion de regresores es adecuado

En la siguiente tabla se presentan los resultados del modelo estimado incluyendo las variables explicativas.

Tabla 15: Resultados de la regresión logística

VARIABLES	B	EXP(B)	IC 95% PARA EXP(B)	
			INFERIOR	SUPERIOR
CONSTANTE	0,683***	1,979		
SEXO	0,537***	1,710	1,513	1,933
JOVEN	-1,001***	0,368	0,300	0,450
ADULTO	-0,280***	0,756	0,653	0,876
SOLTERO	-0,392***	0,676	0,579	0,789
EDUC1	0,776***	2,174	1,610	2,934
EDUC2	0,528***	1,696	1,432	2,010
EDUC3	0,294***	1,342	1,135	1,586
ACTFIS	-0,223***	0,800	0,684	0,936

Fuente: elaboración propia en base a la ENFR 2018

Puede observarse que los intervalos de confianza no incluyen al 1, lo que indica que los *odds ratio* son estadísticamente significativos. Del análisis del modelo se desprende que las variables con mayor incidencia sobre el exceso de peso son el nivel educativo, y el sexo del encuestado.

- Quienes tienen hasta primaria incompleta (EDUC1) tienen 2,174 chances más de tener exceso de peso que quienes tienen un nivel educativo superior. Estas chances disminuyen a 1,696 si se tiene primaria completa (EDUC2) como máximo nivel educativo alcanzado, y a 1,342 si se tiene hasta secundario completo (EDUC3).
- Un hombre (SEXO) tiene 1,71 más chances de tener exceso de peso que una mujer.

Además, se observa que quienes tienen menos chances de tener exceso de peso son jóvenes, adultos (en menor medida), solteros y personas que realizan actividad física.

- Una persona con más de 29 años tiene 2,717 (1/0,368) más chances de tener exceso de peso que alguien con 29 años o menos (JOVEN).
- Alguien que no es soltero, tiene 1,479 (1/0,676) más chances de tener exceso de peso que alguien que sí lo es.
- Quienes no realizan actividad física tienen 1,25 más chances de tener exceso de peso que quienes si realizan actividad.

Este modelo es estadísticamente significativo al realizar la Prueba Chi Cuadrado y rechazar la siguiente hipótesis nula:

H0) Las variables explicativas no tienen efecto sobre la variable dependiente

Como se mencionó anteriormente, en el caso de las regresiones logísticas se calculan los pseudo R^2 , sin embargo, éstos no tienen el mismo significado que los estimados mediante MCO, y dado que en este caso sólo se desea estimar efectos, los valores obtenidos no son alarmantes y por lo tanto no se recomienda interpretarlos.

Sin embargo, sí se puede analizar la capacidad de predicción del modelo. De este análisis se observa que clasifica correctamente al 67,8% de las observaciones (un 2% superior al modelo que incluye solamente constante), y que tiene una mejor predicción del exceso de peso que del peso saludable.

Tabla 16: Matriz de confusión

OBSERVADO/PRONOSTICADO	PESO SALUDABLE	EXCESO DE PESO	PORCENTAJE CORRECTO
PESO SALUDABLE	395	1370	22,4%
EXCESO DE PESO	297	3118	91,3%

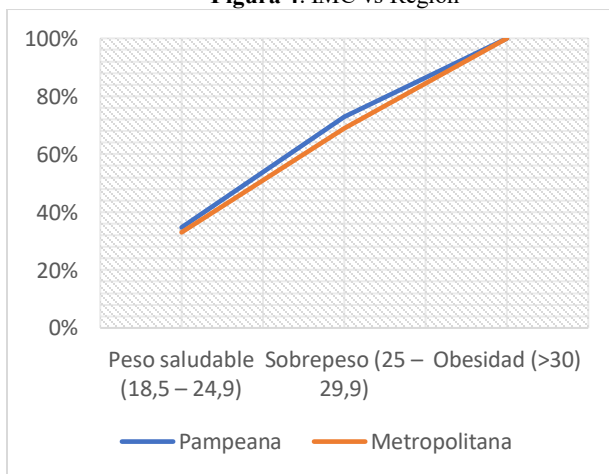
Fuente: elaboración propia en base a la ENFR 2018

4.b.iii. Modelo de regresión logística ordinal

La idea de aplicar este modelo surge a través de la recodificación de la variable dependiente, en una variable de naturaleza ordinal, que puede tomar los valores 0 (peso saludable), 1 (sobrepeso) y 2 (obesidad), de modo que se puedan identificar factores que inciden sobre la obesidad.

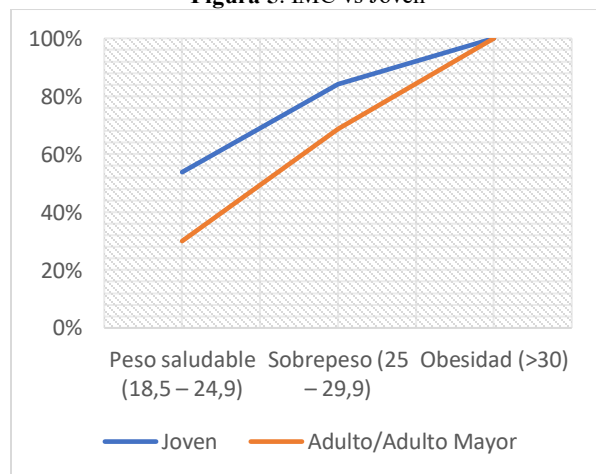
En las figuras que se exponen seguidamente, se muestran los porcentajes acumulados observados para cada uno de los regresores incluidos en el modelo, en relación a la variable que se desea explicar. Se observan dos curvas en cada figura, correspondiente a cada una de las categorías de la variable explicativa. Estas figuras ayudan a visualizar la especificación ordinal, que modela una función de dos curvas (Norusis, *op. cit.*).

Figura 4: IMC vs Región



Fuente: elaboración propia en base a la ENFR 2018

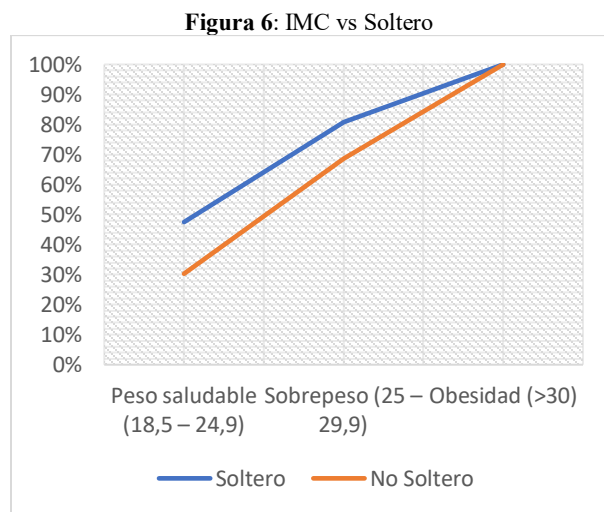
Figura 5: IMC vs Joven



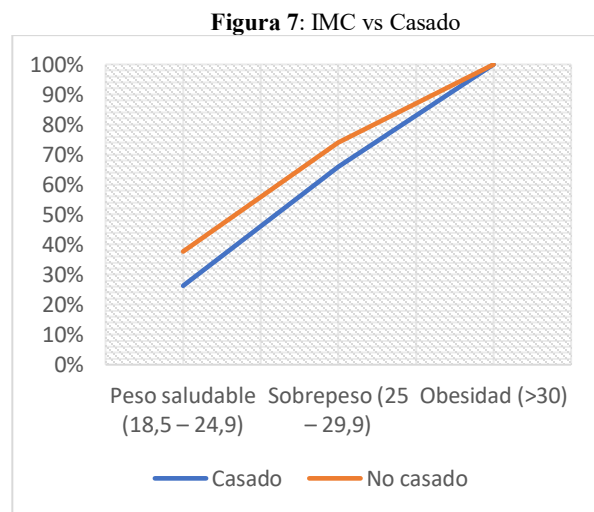
Fuente: elaboración propia en base a la ENFR 2018

En la Figura 4, la línea azul representa los porcentajes acumulados de los residentes en la región pampeana de la provincia de Buenos Aires, y la línea naranja los de quienes residen en la región metropolitana. Se observa que la región pampeana, tiene un mayor porcentaje observado de residentes con peso saludable. A medida que se acumulan los porcentajes, el porcentaje acumulado de quienes viven en la región pampeana se mantiene por encima de quienes viven en el área metropolitana, y al final ambos alcanzan el 100%. Por lo tanto, dado que quienes viven en la región pampeana tienen mayor proporción de habitantes con peso saludable o sobrepeso, es de esperar un signo **negativo** del coeficiente estimado para la variable REGIÓN.

En la Figura 5, la línea azul representa los porcentajes acumulados de los jóvenes y la línea naranja a los encuestados de 30 años o más (adultos y adultos mayores). Se observa una menor proporción de adultos y adultos mayores con peso saludable con respecto al grupo de los jóvenes. Al acumularse los porcentajes, el porcentaje acumulado de los adultos y adultos mayores se mantiene por debajo de los jóvenes, y al final ambos alcanzan el 100%. Así es que, dado que el grupo de adultos y adultos mayores tienen menor proporción de encuestados con peso saludable o sobrepeso, es de esperar un signo **positivo** del coeficiente estimado para la variable JOVEN.



Fuente: elaboración propia en base a la ENFR 2018

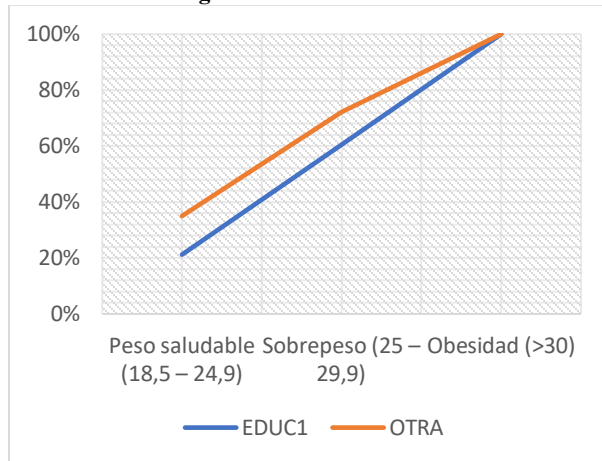


Fuente: elaboración propia en base a la ENFR 2018

La Figura 6 analiza los grupos de solteros y no solteros con respecto a la categoría del IMC. La línea naranja representa los porcentajes acumulados de quienes no están solteros y la azul de quienes si lo están. Se observa que hay una menor proporción de encuestados no solteros con peso saludable en relación a los solteros. A medida que se acumulan los porcentajes, la línea naranja se mantiene por debajo de la línea que representa a los solteros y al final ambos alcanzan el 100%. Por lo tanto, dado que el grupo de no solteros tienen menor proporción de encuestados con peso saludable o sobrepeso, es de esperar un signo **positivo** del coeficiente estimado para la variable SOLTERO.

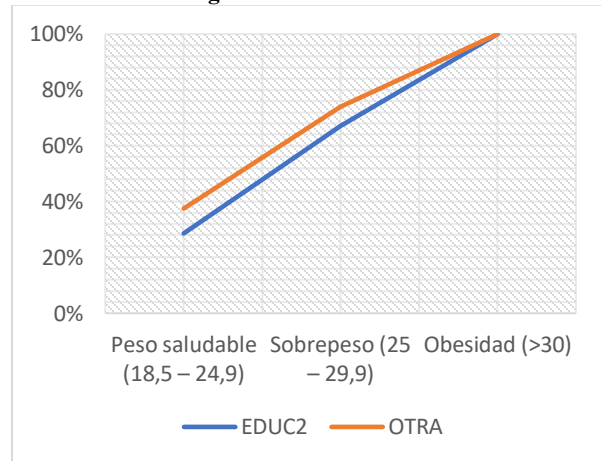
En la Figura 7, la línea naranja representa los porcentajes acumulados de quienes no están casados. Se observa que éstos tienen una mayor proporción de encuestados con peso saludable. A medida que se acumulan los porcentajes, el porcentaje acumulado de este grupo se mantiene por encima de la línea que representa a quienes están casados y al final ambos alcanzan el 100%. Por lo tanto, dado que el grupo de no solteros tienen mayor proporción de encuestados con peso saludable o sobrepeso, es de esperar un signo **negativo** del coeficiente estimado para la variable CASADO.

Figura 8: IMC vs EDUC1



Fuente: elaboración propia en base a la ENFR 2018

Figura 9: IMC vs EDUC2

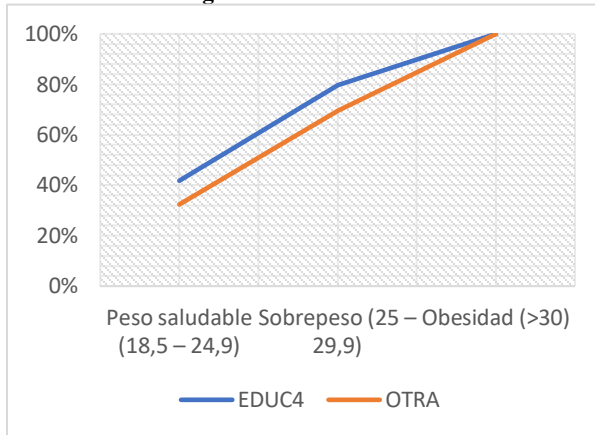


Fuente: elaboración propia en base a la ENFR 2018

En la Figura 8, la línea naranja representa los porcentajes acumulados de quienes tienen un nivel educativo superior a primaria incompleta. Se observa que éstos tienen una mayor proporción de encuestados con peso saludable. A medida que se acumulan los porcentajes, el porcentaje acumulado de este grupo se mantiene por encima de la línea de quienes tienen primaria incompleta y al final ambos acumulan el 100%. Por lo tanto, dado que el grupo de quienes tienen un nivel educativo superior a primaria incompleta tienen mayor proporción de encuestados con peso saludable o sobrepeso, es de esperar un signo **negativo** del coeficiente estimado para la variable EDUC1.

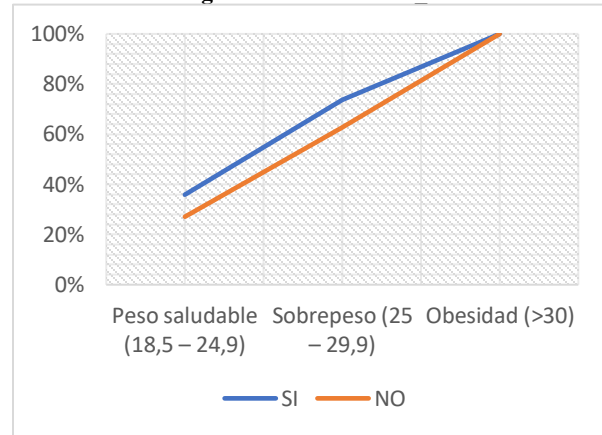
En la Figura 9, la línea naranja representa los porcentajes acumulados de quienes tienen un máximo nivel educativo alcanzado distinto a primaria completa. Se observa que éstos tienen una mayor proporción de encuestados con peso saludable. A medida que se acumulan los porcentajes, el porcentaje acumulado de este grupo se mantiene por encima de la línea de quienes tienen hasta primaria completa y al final ambos alcanzan el 100%. Por lo tanto, dado que el grupo de quienes tienen un nivel educativo distinto a primaria completa tienen mayor proporción de encuestados con peso saludable o sobrepeso, es de esperar un signo **negativo** del coeficiente estimado para la variable EDUC2.

Figura 10: IMC vs EDUC4



Fuente: elaboración propia en base a la ENFR 2018

Figura 11: IMC vs ACT_FIS

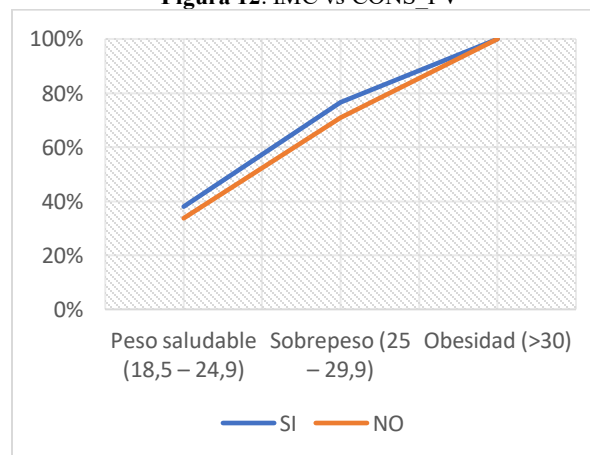


Fuente: elaboración propia en base a la ENFR 2018

En la Figura 10, la línea naranja representa los porcentajes acumulados de quienes tienen un nivel educativo inferior a terciario o universitario completo. Se observa que éstos tienen una menor proporción de encuestados con peso saludable. A medida que se acumulan los porcentajes, el porcentaje acumulado de este grupo se mantiene por debajo de la línea de quienes si tienen terciario o univarsitario completo y al final ambos acumulan el 100%. Por lo tanto, dado que el grupo de quienes tienen un un nivel educativo inferior a terciario o universitario completo tienen menor proporción de encuestados con peso saludable o sobrepeso, es de esperar un signo **positivo** del coeficiente estimado para la variable EDUC4.

En la Figura 11, la línea naranja representa los porcentajes acumulados de quienes no realizan actividad física. Se observa que éstos tienen una menor proporción de encuestados con peso saludable. A medida que se acumulan los porcentajes, el porcentaje acumulado de este grupo se mantiene por debajo de la línea de quienes si realizan actividad y al final ambos alcanzan el 100%. Por lo tanto, dado que el grupo de quienes no realizan actividad física tienen menor proporción de encuestados con peso saludable o sobrepeso, es de esperar un signo **positivo** del coeficiente estimado para la variable ACT_FIS.

Figura 12: IMC vs CONS_FV



Fuente: elaboración propia en base a la ENFR 2018

Por ultimo, en la Figura 12, la línea naranja representa los porcentajes acumulados de quienes no consumen las 5 porciones diarias recomendadas de frutas y verduras. Se observa que éstos tienen una menor proporción de encuestados con peso saludable. A medida que se acumulan los porcentajes, el porcentaje acumulado de este grupo se mantiene por debajo de la línea de quienes si consumen la porción recomendada. Por lo tanto, dado que el grupo de quienes no siguen la recomendación tienen menor proporción de encuestados con peso saludable o sobrepeso, es de esperar un signo **positivo** del coeficiente estimado para la variable CONS_FV.

En la siguiente tabla se presentan los resultados de la estimación del modelo y se observan que los signos de los coeficientes son los esperados:

Tabla 17: Resultados de la estimación del Modelo *Logit* Ordinal

ORDENADA AL ORIGEN -VALORES FRONTERA- PENDIENTES -COEFICIENTES EFECTOS-	ESTIMADORES	INTERVALO DE CONFIANZA 95%	
		INFERIOR	SUPERIOR
		[CAT_IMC = 0]	-0,108
[CAT_IMC = 1]	1,558***	1,141	1,974
[REGION = 0]	-0,112**	-0,217	-0,007
[REGION = 1]	0 ^a		
[JOVEN = 0]	0,732***	0,574	0,890
[JOVEN = 1]	0 ^a		
[SOLTERO = 0]	0,283***	0,139	0,427
[SOLTERO = 1]	0 ^a		
[CASADO = 0]	-0,258***	-0,376	-0,141
[CASADO = 1]	0 ^a		
[EDUC1 = 0]	-0,461***	-0,679	-0,244
[EDUC1 = 1]	0 ^a		
[EDUC2 = 0]	-0,248***	-0,367	-0,129
[EDUC2 = 1]	0 ^a		
[EDUC4 = 0]	0,382***	0,232	0,531
[EDUC4 = 1]	0 ^a		
[ACT_FIS = 0]	0,316***	0,188	0,444
[ACT_FIS = 1]	0 ^a		
[CONS_FV = 0]	0,228**	0,053	0,403
[CONS_FV = 1]	0 ^a		

*** valor "p" < 1%, ** valor "p" < 5%, * valor "p" < 10%.

Fuente: elaboración propia en base a la ENFR 2018

Se observa que todas las variables son estadísticamente significativas a un nivel de significación menor a 1%. Y además, los IC del coeficiente estimado para las variables significativas no contienen al número 0.

De la estimación anterior surge que:

- Quienes viven en la región pampeana de la provincia de Buenos Aires tienen menos probabilidad de tener obesidad que quienes viven en el área metropolitana.

- Quienes tienen 30 años o más, tienen más probabilidad de tener obesidad que quienes pertenecen al grupo etáreo de entre 18 y 29 años.
- Quienes no se encuentran solteros, tienen una probabilidad mayor de tener obesidad que quienes sí lo están.
- Quienes no se encuentran casados, tienen una menor probabilidad de tener obesidad en comparación a quienes sí están casados.
- Quienes no tienen como máximo nivel educativo alcanzado la primaria incompleta, tienen menor probabilidad de tener obesidad. Lo mismo ocurre con quienes no tienen como máximo nivel educativo alcanzado la primaria completa.
- Quienes no tienen el nivel terciario o Universitario completo, tienen mayor probabilidad de tener obesidad que quienes si completaron dicho nivel educativo.
- Quienes no realizan actividad física tienen mayor probabilidad de tener obesidad comparado a quienes sí realizan actividad.
- Quienes no consumen las 5 porciones diarias recomendadas de frutas o verduras tienen mayor probabilidad de tener obesidad que quienes siguen la recomendación.

Con respecto a la evaluaciones del modelo, en primer lugar se realiza la prueba de ajuste Global del modelo, que evalúa la siguiente hipótesis nula, y es rechazada con un nivel de significación menor al 1%:

H0) El modelo sin la inclusion de variables independientes es adecuado

Luego, se analiza la bondad de ajuste del modelo, a través del estadístico Chi Cuadrado de Pearson y de Deviance y no se rechaza la siguiente hipótesis para un nivel de significación del 10%:

H0) El modelo ajusta adecuadamente

Como se mencionó anteriormente, los pseudo R^2 que se calculan en el caso de las regresiones logísticas no tienen el mismo significado que los estimados mediante MCO, y por lo tanto no se recomienda interpretarlos. Otros trabajos empíricos que presentan Pseudo R^2 bajos, al especificar Modelos *Logit* Ordinal, son Grebitus *et al.* (2007), Lupín *et al.* (2007), Rodríguez, *et al.* (2009), Navarro Manotas *et al.* (2014) y Rodríguez Donate *et al.* (2007).

Finalmente, la Prueba de Líneas Paralelas (Anexo d) permite verificar el supuesto de que los coeficientes estimados son los mismos para todos los *logits* –coeficientes proporcionales–; es decir, que el resultado es un conjunto de líneas paralelas –una por categoría–. Si las líneas son paralelas el valor observado del nivel de significación debe ser alto de modo que no se rechace la siguiente hipótesis nula:

H0) Los coeficientes de regresión β –coeficientes de pendientes– son los mismos entre las categorías de respuesta

El no rechazo de la H_0), a un nivel de significación del 10%, proporciona evidencia de que la función enlace empleada *-logit-* es apropiada (Norusis, *op. cit.*).

A partir de los resultados estimados, las probabilidades predichas acumuladas indican que hay un 47% de probabilidad de que un residente de la Provincia de Buenos Aires de 18 años o más tenga un peso saludable.

$$P(CAT_IMC_0) = \frac{1}{1+e^{-(-0,108)}} = 0,4730$$

Además, hay un 83% de probabilidad de que un residente de la Provincia de Buenos Aires de 18 años o más tenga peso saludable o sobrepeso.

$$P(CAT_IMC_{0 \text{ o } 1}) = \frac{1}{1+e^{-(-1,558)}} = 0,8260$$

Dado que $P(CAT_IMC = 0, 1 \text{ o } 2) = 1$, y a partir de las probabilidades calculadas anteriormente, por diferencia, se puede calcular la probabilidad para cada categoría de la variable a explicar:

$$\text{Prob}(CAT_IMC = 0) = 0,4730$$

$$\text{Prob}(CAT_IMC = 1) = \text{Prob}(CAT_IMC = 0 \text{ o } 1) - \text{Prob}(CAT_IMC = 0) = 0,353$$

$$\text{Prob}(CAT_IMC = 2) = \text{Prob}(CAT_IMC = 0, 1 \text{ o } 2) - \text{Prob}(CAT_IMC = 0 \text{ o } 1) = 0,174$$

5. Consideraciones finales

En este trabajo se analizan los factores demográficos y socioeconómicos determinantes de la obesidad en mayores de 18 años de la Provincia de Buenos Aires. Para ello, se utilizaron datos provenientes de la cuarta Encuesta Nacional de Factores de Riesgo (ENFR), relevada en Argentina durante el año 2018. En particular, se analizaron las respuestas de los encuestados de la Provincia de Buenos Aires, que habían respondido las preguntas que permiten el cálculo del IMC.

Con esos datos, se realizó un análisis descriptivo, observándose que un 28% de los encuestados en la Provincia de Buenos Aires presenta obesidad y un 37% sobrepeso. Luego, se evaluó la asociación entre el peso saludable, sobrepeso y obesidad con distintas variables mencionadas en la bibliografía sobre el tema, y se evidenció asociación entre las mencionadas categorías del IMC y región, quintil de ingreso, sexo, rango de edad, situación conyugal, nivel educativo, cobertura de salud, condición de actividad, realización de actividad física y consumo de las 5 porciones diarias recomendadas de frutas o verduras.

Luego se estimaron tres modelos econométricos, analizando que factores inciden en el IMC y sus categorías. Del modelo inicial se desprende que estas variables no explican en gran medida la variabilidad en el IMC, pero sí que quienes viven en el área metropolitana o son hombres tienden a tener un IMC más elevado que quienes viven en la región pampeana o son mujeres respectivamente. Además muestra evidencia de que ser joven, realizar actividad física o tener un nivel educativo mayor son algunos factores que ayudan a reducir el IMC.

En el segundo modelo se analizaron los factores que inciden en el sobrepeso y la obesidad (analizados en conjunto), y se observa que las chances de tener sobrepeso u obesidad aumentan a medida que aumenta la edad, y para niveles educativos más bajos. Además quienes son solteros o realizan actividad física tienen menos chances de tener sobrepeso u obesidad que quienes tienen otra situación conyugal o no realizan actividad, respectivamente, y nuevamente se observa que los hombres tienen más chances de tener un IMC mayor al saludable.

Por último, al estimar la regresión logística ordinal, se pueden analizar los factores que inciden en la obesidad puntualmente. Como en los modelos anteriores, se observa que no ser joven o soltero y no realizar actividad física implica una mayor probabilidad de tener obesidad, mientras que a mayor nivel educativo esa probabilidad disminuye. Como en el primer modelo, ser residente de la región pampeana presenta evidencia en contra de tener obesidad. Y, además, surge un nuevo factor que es el consumo de las 5 porciones diarias recomendadas de frutas o verduras. Quienes no cumplen esta recomendación, tendrán una probabilidad mayor de tener obesidad.

En ninguno de los modelos se encontró evidencia de que el ingreso sea un factor determinante de la obesidad (tampoco del sobrepeso o del IMC), esto coincide con lo mencionado por Peña & Bacallao (*op Cit.*) a cerca de que actualmente la obesidad es una enfermedad prevalente en todos los niveles de ingreso.

Si bien estos resultados están alineados con los antecedentes empíricos citados oportunamente, es importante aclarar que tienen un carácter orientativo, y los resultados podrían ajustarse actualizando la información, principalmente teniendo en cuenta que estos datos fueron relevados previamente a la pandemia de Covid-19, que generó cambios de hábitos en muchas personas. También podría agregarse información más específica, tal como precios de los alimentos, consumo de comidas rápidas, cantidad de calorías consumidas y/o gastadas.

De todos modos, estos resultados pueden guiar a quienes gestionan políticas públicas tendientes a promover la salud mediante la prevención, para lograr reducir la obesidad y las enfermedades no transmisibles asociadas a ella. Por ejemplo, podría mencionarse promover la realización de actividad física, reforzar la difusión de las Guías Alimentarias para la Población Argentina, de modo que se fomente el consumo de las 5 porciones diarias de frutas o verduras, generar canales de educación alimentaria, y, además, generar políticas públicas que apunten a mejorar la calidad de vida de los adultos mayores.

Otra cuestión importante es que resulta interesante para estudios futuros analizar la obesidad en niños, ya que los cambios nutricionales durante la niñez pueden impactar sobre la salud en etapas posteriores de la vida, lo cual tiene implicaciones de política social.

6. Referencias bibliográficas

1. Agresti, A. (2002). *Categorical Data Analysis*. USA: John Wiley&Sons.
2. Amemiya, T. (1985). *Advanced Econometrics*. Cambridge, MA: Harvard University Press.
3. Britos, S. (2012). Reflexiones sobre la seguridad de una alimentación saludable en la Argentina de fines de 2012. Ciudad Autónoma de Buenos Aires (CABA)-Argentina: Centro de Estudios sobre Políticas y Economía de la Alimentación (CEPEA).
4. Efron, B. (1978). Regression and ANOVA with zero-one data: measures of residual variation. *Journal of the American Statistical Association*, 73(361): 113-121.
5. Flegal, K.M.; Graubard, B.I.; Williamson, D.F.; Gail, M.H. (2005). Excess deaths associated with underweight, overweight and obesity. *Journal of the American Medical Association* 293.
6. Fortich Mesa, R.; Gutierrez, J.D. (2011). Los determinantes de la obesidad en Colombia.
7. García Milian, A.J.; Creus García, E.D. (2016). La obesidad como factor de riesgo, sus determinantes y tratamiento. Recuperado de: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21252016000300011
8. Gardner, G.; Halweil, B. (2000). Underfed and overfed: the global epidemic of malnutrition. *Worldwatch*, Paper 150.
9. Grebitus, C.; Yue, C.; Bruhn, M.; Jensen, H. H. (2007). What affects consumption patterns of organic and conventional products?. AAEA Annual Meeting, Portland-USA. Recuperado de: <http://ageconsearch.umn.edu/bitstream/9819/1/sp07gr03.pdf>
10. Goldblatt, P.; Moore, M.; Stunkard, A. (1965). Social factors in obesity. *JAMA* 192.
11. INDEC: www.indec.gov.ar/
12. Isunza Vera, A. (2004). La obesidad en la pobreza. Violación al derecho a la alimentación. Recuperado de: http://academia.edu/32840308/La_obesidad_en_la_pobreza_violación_al_derecho_a_la_alimenta_ción
13. Lara, M. I.; Serio, M. (2015). La obesidad en adultos en Argentina y Mendoza: un análisis de condicionantes socioeconómicos. Universidad Nacional de Cuyo. *Revista de la Facultad de Ciencias Económicas*. ISSN: 0041-8668. Recuperado de: <https://ri.conicet.gov.ar/handle/11336/61040>
14. Lupín, B.; Lacaze, V.; Rodríguez, E.M. (2007). Las percepciones de riesgo de los consumidores en alimentos lácteos: aplicación de una Regresión Logística Ordinal. XII Reunión Científica del Grupo Argentino de Biometría (GAB)-I Encuentro Argentino-Chileno de Biometría, San Martín de los Andes-Argentina. Recuperado de: http://nulan.mdp.edu.ar/758/1/JDifInvEcon_2008_7_14-18.pdf

15. Manios, Y.; Panagiotakos, D. B.; Pitsavos, C.; Polychronopoulos, E.; Stefanadis, C. (2005). Implication of socio-economic status on the prevalence of overweight and obesity in Greek adults: the ATTICA study. *Health Policy*, Vol. 74, No.2, págs 224-232.
16. Menard, S. (2000). Coefficients of Determination for Multiple Logistic Regression Analysis. *The American Statistician*, 54(1): 17-24.
17. Mokdad, A.H.; Marks, J.S.; Stroup, D.F.; Gerberding, J.L. (2004). Actual causes of death in the United States (2000). *Journal of the American Medical Association* 291.
18. Monteverde, M.; Tomás, M.S. (2017). Principales resultados de la encuesta nacional de factores de riesgo en la Provincia de Buenos Aires. *Estudios de población de la Provincia de Buenos Aires*. Año 3, Número 4, Junio 2017. ISSN 2451-6511.
19. Navarro Manotas, E.; Verbel Castellar, A.; Robles García, D.; Hurtado Ibarra, K.(2014). Regresión logística ordinal aplicada a la identificación de factores de riesgo para cancer de cuello uterino. *Ingeniare*, ISSN-e 2390-0504, N°. 17, 2014, págs. 87-105. Recuperado de: <https://dialnet.unirioja.es/servlet/articulo?codigo=5980486>
20. Norusis, M. J. (2005). Ordinal Regression (Chapter 4), SPSS 13.0 Advanced Statistical Procedures Companion. Prentice-Hall. Recuperado de: http://www.norusis.com/pdf/ASPC_v13.pdf
21. Organización Mundial de la Salud (OMS): <http://www.who.int/es/news-room/fact-sheets/detail/obesity-and-overweight>
22. Ortiz-Moncada, R.; Álvarez-Dardet, C.; Miralles-Bueno, J.; Ruíz-Cantero, M.; Re-Saavedra, M.; Villar-Villalba, C.; Perez-Farinos, N.; Serra-Majem, L. (2011). Determinantes sociales de sobrepeso y obesidad en España 2006. Recuperado de: <https://www.sciencedirect.com/science/article/abs/pii/S0025775311006567>
23. Peña, M.; Bacallao, J. (2001). La obesidad y sus tendencias en la región. En *Revista Panamericana. Salud Pública, OPS*, 10,2,2001. Pp 75-78. Recuperado de: <http://scielosp.org/article/rpsp/2001.v10n2/45-78/es/>
24. Powers, D. A.; Xie, Y. (1999). *Statistical Method for Categorical Data Analysis*. San Diego-USA: Academic Press Inc.
25. Rodríguez Donate, M. C.; Cáceres Hernández, J. J. (2007). Modelos de Elección Discreta y especificaciones ordenadas: una reflexión metodológica. *Estadística Española*, 49(166): 451-471. Recuperado de: file:///C:/Users/usuario/Downloads/552_818_166_2.pdf
26. Rodríguez, E. M.; Lupín, B.; Lacaze, V. (2009). Consumers preferences for potatoes with quality attributes in Argentina. *Triennial Symposium of the International Society for Tropical Root Crops*, Lima-Peru.

27. Serrano, R. M. (2005). Metabolic síndrome: a modern variant of stress-related disease? *Rev Esp Cardiol.* 58
28. Scott Long, J. (1997). *Regression models for categorical and limited and dependent variables.* California-USA: Sage Publications Inc.
29. Sobal, J.; Stunkard, A. J. (1989). Socioeconomic status and obesity: a review of the literature. *Psychol Bull* :105(2).
30. Sosa Escudero, W. (2015). *El lado oscuro de la Econometría.* CABA-Argentina: Temas Grupo Editorial.
31. Stock, J. H.; Watson, M. M. (2012). *Introducción a la Econometría.* Madrid-España: Pearson Educación SA.
32. Sturm, R. (2002). The effects of obesity, smoking, and drinking on medical problems and costs. *Health Affairs* 21.
33. Temporelli, K.; Viego, V. (2010). Socioeconomic factors as determinants of overweight and obesity. The Argentinian case. Recuperado de: <https://mpra.ub.uni-muenchen.de/26880/>
34. Theil, H. (1971). *Principles of Econometrics.* New York-USA: Wiley.
35. UCLA: <http://www.stats.oarc.ucla.edu/other/dae/>
36. Underhill, S.; Figueroa, E. (July 1996). Consumer preferences for non-conventionally grown produce. *Journal of Food Distribution Research*, 27(2): 56-66. Recuperado de: <http://ageconsearch.umn.edu/record/27067/files/27020056.pdf>
37. UNICEF: <https://www.unicef.org/argentina/publicaciones-y-datos>
38. Wang, Y.; and Zhang, Q. (2006). Are American children and adolescents of low socioeconomic status at increased risk of obesity? Changes in the association between overweight and family income between 1971 and 2002. *Am J Clin Nutr* 2006; 84:707-16. American Society for Nutrition.
39. Wooldridge, J. M. (2010). *Introducción a la Econometría. Un enfoque moderno.* México D F-México: Cengage Learning Editores, SA de CV.

7. Anexos

7.a. Prueba chi cuadrado de pearson

Seguendo a Agresti (*op. cit.*), las tablas de contingencia permiten el análisis simultáneo de dos o más variables categóricas. Éstas tablas contienen “I” filas y “J” columnas y las celdas resultantes contienen el recuento observado de las posibles combinaciones IxJ (n_{ij}). Con el objetivo de explorar estadísticamente la relación entre diversas variables de interés, a partir de las tablas anteriores, es posible aplicar la Prueba Chí Cuadrado de Pearson, que es una prueba no paramétrica².

Aplicando esta Prueba, se evalúa la siguiente hipótesis:

H₀) Las variables son independientes.

Es decir, que no existe asociación entre las variables en cuestión.

Para realizar la evaluación de dicha H₀, se utiliza el estadístico Chi-Cuadrado de Pearson, que se distribuye como una Chi Cuadrado con [(i-1) (j-1)] grados de libertad (gl).

$$X^2 = \sum \frac{(n_{ij} - \mu_{ij})^2}{\mu_{ij}}$$

Siendo n_{ij} el recuento muestral de la celda ij de la tabla de contingencia, y μ_{ij} las frecuencias esperadas en la celda ij de la tabla de contingencia.

Dicho estadístico toma su mínimo valor de cero, cuando $n_{ij} = \mu_{ij}$. Para un tamaño de muestra fijo, a mayor diferencia $\{n_{ij} - \mu_{ij}\}$ produce un mayor valor de X^2 y una mayor evidencia contra la H₀. El P-value es la probabilidad nula de que X^2 es al menos, tan grande como el valor observado.

La H₀) se rechaza si:

- El valor observado (X^2_o) es superior que el valor teórico -o de tabla- (X^2_t). El valor X^2_o cae en la zona de rechazo.
- El valor “p” es menor que el nivel de significación establecido ($\alpha = 0,01; 0,05$ o $0,10$, por lo general).

El tamaño de muestra requerido establece que todos los valores esperados bajo la H₀) deberán ser mayores o iguales que 5.

²Los procedimientos no paramétricos se aplican cuando la prueba no depende de la distribución que generó la muestra y no necesariamente se relaciona con parámetros poblacionales. Estos procedimientos resultan de gran utilidad en las Ciencias Sociales pues, en general, no se cumple la totalidad de los supuestos que sustentan la Estadística Clásica. Una de las ventajas de estos procedimientos es que son fáciles de aplicar sobre todo si el tamaño de la muestra es pequeño (Fernández Loureiro de Pérez, 2000)

7.b. Cuestionario

A continuación se describen en detalle ambos bloques del cuestionario.

- ◆ **Bloque Hogar (BH):** Destinado a relevar datos sobre las condiciones habitacionales, así como características demográficas y educativas de todas las personas que conforman el hogar y datos laborales del jefe de hogar, con la siguiente estructura de capítulos:

- Detección de viviendas y hogares (DV)
- Componentes del hogar (CH)
- Características de la vivienda (CV)
- Características del hogar (CH)
- Ingresos del hogar (IH)
- Situación laboral del jefe o jefa del hogar (SL)

- ◆ **Bloque Individual (BI):** Destinado a registrar datos laborales, así como información sobre los factores de riesgo de las ENT.

Este bloque se realice a través de una entrevista personal a un único componente seleccionado entre los integrantes del hogar de 18 años y más. La selección del encuestado se realizó mediante la aplicación de la tabla de selección de Kish.

El bloque individual se subdividió en tres pasos:

- Paso 1: Capítulos correspondientes a la información por autorreporte:
 - Situación laboral
 - Salud general
 - Actividad física
 - Tabaco
 - Hipertensión arterial
 - Peso corporal
 - Alimentación
 - Colesterol
 - Consumo de alcohol
 - Diabetes
 - Lesiones
 - Prácticas preventivas
 - Prevención de cancer colorectal
 - Embarazo

- Paso 2: Capítulo correspondiente a información obtenida a través de mediciones antropométricas. Se utilizó para registrar las mediciones de la presión arterial, peso, talla y perímetro de la cintura, realizadas con el instrumental correspondiente de acuerdo a las recomendaciones internacionales.

Por razones presupuestarias, disponibilidad de instrumentos para la toma de las mediciones, y la complejidad de la coordinación del trabajo entre los encuestadores y el personal de salud, este paso por diseño se aplica a una submuestra probabilística del 75% de las viviendas de la encuesta.

La submuestra es una selección de USM que componen el Paso 1, bajo un muestreo simple al azar (MSA) y con una fracción de muestreo de 75% en cada estrato de las USM definida para la MMUVRA, que determina la muestra para el Paso 2 de aproximadamente el 75% de los segmentos y viviendas de la original afectada al Paso 1.

Dicha selección se realice al momento de la extracción de la muestra de viviendas de la encuesta, lo que permitió, en forma anticipada, la distribución y asignación de las cargas de trabajo a los encuestadores.

- Paso 3: Capítulo correspondiente a información obtenida a través de mediciones bioquímicas, en las que se registraron los valores de glucemia capilar y colesterol total obtenidos mediante la utilización de una lanceta para extraer una gota de sangre aplicada en tiras reactivas que se introdujeron en un analizador de química seca. Dichas mediciones se realizaron al miembro del hogar seleccionado bajo su consentimiento, y bajo la condición necesaria de que se encuentre en ayunas, incorporando una complejidad adicional al equipo de campo.

No todos los miembros afectados al Paso 2 continúan al Paso 3, solo aquellos que habitan en un aglomerado o localidad de 150.000 habitantes y más lo completan por decisión operativa.

7.c. Análisis descriptivo de los encuestados

Tabla 18: Caracterización de los encuestados de la Provincia de Buenos Aires analizados

VARIABLES	PARTICIPACIÓN RELATIVA EN LA SUBMUESTRA
SEXO DEL ENCUESTADO	
VARONES	45%
MUJERES	55%
EDAD DEL ENCUESTADO	
18-29 AÑOS	17%
30-59 AÑOS	52%
60 AÑOS Y MÁS	31%
EDAD PROMEDIO: 49 AÑOS	
NIVEL DE EDUCACIÓN DEL ENCUESTADO	
SIN INSTRUCCIÓN/PRIMARIA INCOMPLETA	7%
PRIMARIA COMPLETA	38%
SECUNDARIO COMPLETO	37%
TERCIARIO/UNIVERSITARIO COMPLETO	18%
OCUPACIÓN DEL ENCUESTADO	
OCUPADO	60%
DESOCUPADO	5%
INACTIVO	35%
COBERTURA DE SALUD DEL ENCUESTADO	
OBRA SOCIAL/PREPAGA/MUTUAL	75%
COBERTURA PÚBLICA	25%

Fuente: elaboración propia en base a la ENFR 2018

Tabla 19: Características del hogar del encuestado

VARIABLES	PARTICIPACIÓN RELATIVA EN LA SUBMUESTRA
REGION	
METROPOLITANA	39%
PAMPEANA	61%
TIPO DE HOGAR	
UNIPERSONAL	24%
MULTIPERSONAL	76%
QUINTIL DE INGRESO	
QUINTIL 1	16%
QUINTIL 2	19%
QUINTIL 3	20%
QUINTIL 4	24%
QUINTIL 5	21%
SUBSIDIOS / PROGRAMAS SOCIALES	
SI	15%
NO	83%
NS/NR	2%

Fuente: elaboración propia en base a la ENFR 2018

7.d. Prueba de líneas paralelas

Esta prueba evalúa la posibilidad de que los coeficientes de regresión sean los mismos para todas las categorías. Si se rechaza el supuesto de paralelismo, se debería considerar emplear un modelo de regresión multinomial, que estima distintos coeficientes para cada categoría (Norusis, *op. cit.*).

Como se mencionó anteriormente, la probabilidad acumulada es la función de distribución acumulada F evaluada en $\alpha_j - x\beta$. Dado que β es el mismo para todas las j categorías, la siguiente ecuación define un conjunto de modelos de respuesta con diferentes interceptos:

$$P(Y \leq j|x) = F(\alpha_j - x\beta)$$

Por lo tanto:

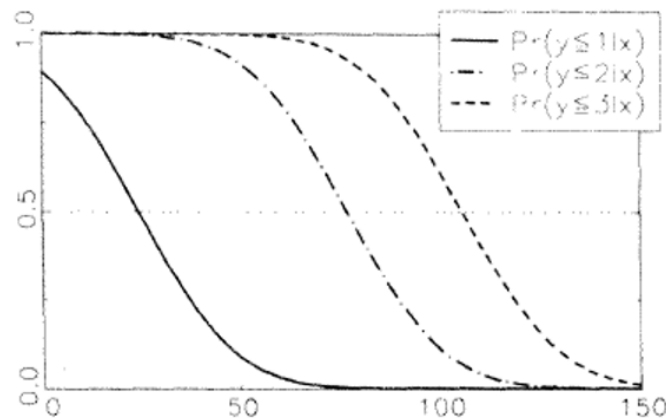
$$\alpha_j - x\beta = (\alpha_j - \beta_0) - \sum \beta_k x_k$$

Entonces, el modelo para $Y \leq 1$ es: $P(Y \leq 1|x) = F((\alpha_1 - \beta_0) - \sum \beta_k x_k)$ con intercepto $\alpha_1 - \beta_0$

Del mismo modo, el modelo para $Y \leq 2$ es: $P(Y \leq 2|x) = F((\alpha_2 - \beta_0) - \sum \beta_k x_k)$

En este modelo, el intercepto cambió a $\alpha_2 - \beta_0$, pero los coeficientes para las variables x_k no cambiaron. El cambio en el intercepto mueve la curva de probabilidad a la derecha o a la izquierda, pero no cambia la pendiente.

Figura 13. Ilustración del supuesto de regresión paralela.



Fuente: Regression models for categorical and limited dependent variables. Long (1997)

En la figura anterior se observan las curvas de probabilidad acumulada en el caso de cuatro categorías ordenadas, resultando tres curvas con interceptos $\alpha_1 - \beta_0$, $\alpha_2 - \beta_0$, y $\alpha_3 - \beta_0$. Al examinar la pendiente de las tres curvas de probabilidad en el punto en que la probabilidad 0.5 -gráficamente indicada en la figura anterior por una línea horizontal punteada- se observa que:

$$\frac{\delta P(Y \leq 1|x)}{\delta x} = \frac{\delta P(Y \leq 2|x)}{\delta x} = \frac{\delta P(Y \leq 3|x)}{\delta x}$$

Por lo tanto se comprueba que las curvas de regresión son paralelas.