



EFFECTO DE LA OBESIDAD EN EL RENDIMIENTO ACADÉMICO

**TESIS PARA OPTAR AL GRADO DE
MAGÍSTER EN ECONOMÍA**

Alumno: Hernán Herrera Velásquez

Profesor Guía: Fabián Duarte

Santiago, Septiembre 2016.

Agradecimientos:

Primero que todo, quiero agradecer el apoyo de Conicyt, Junaeb y COES. Sin la ayuda de estas instituciones y personas dentro de ellas la tesis no habría podido ser terminada. Muchas gracias en especial a Pía por toda la paciencia y los consejos entregados para la realización de esta tesis!

Agradezco a toda mi familia por apoyarme a lo largo de toda mi vida escolar y universitaria. Gracias por las palabras de aliento, las comidas ricas!, el espacio y la oportunidad de estudiar, los consejos, por todo!! Los quiero mucho.

También quiero agradecer a mis amigos, tanto de la U como de tenis, el colegio, Alpha, la Trini y los Serruchos!. Gracias por toda su ayuda y por los momentos divertidos y de relaxo. Mención honrosa a Herl que me apaño dándole más de alguna vuelta a la tesis para ver que resultaba. Gracias a todos! Los Five, Zuñiga, Camilo, Duende, Nati, Pipe, AQMF y otros tantos! Son bknes todos 😊.

Muchas gracias a mi profesor guía Fabián Duarte. Sin duda fue un agrado realizar todo este trabajo junto a él. Muy positivo, perseverante, sencillo, honesto, sin ningún problema para hacerse el tiempo para trabajar codo a codo para lograr los objetivos finales y por sobretodo una gran persona, la cual está ahí si es que uno tiene problemas más allá del índole académico.

Además, agradezco a Christian Ferrada por su ayuda brindada en el proyecto de verano, por todo el trabajo realizado en conjunto, por la cercanía y por la excelente disposición a siempre ayudar cada vez que uno lo necesitaba! Muchas gracias Profe!

En especial quiero agradecer a Rafi por haberme acompañado durante todo este proceso. Sin duda fuiste, eres y serás muy importante en mi vida. Gracias por los consejos, los abrazos, las oraciones, el cariño y por más. Te amo.

Finalmente, gracias a Dios por existir, por las capacidades que me regalo, por todo lo ocurrido en mi vida, sea bueno o malo, de una forma u otra me llevo a ser el hombre que soy actualmente.

"Yo soy el camino, la verdad y la vida - contesto Jesús- Nadie llega al padre si no por mí"

Juan 14:6

Resumen

La obesidad es un fenómeno transversal a toda edad, país y condición socioeconómica. Chile no es la excepción y un total de 64,5% de la población presenta problemas de exceso de peso, donde un 39,3% tiene sobrepeso y un 25,1% es obesa (ENS 2009/2010). Este trabajo estima el efecto del IMC sobre el rendimiento escolar para alumnos de media, medido a través de sus puntajes Simce de matemática y lectura. Sin embargo, existe un problema de endogeneidad explicado por una doble causalidad entre las variables mencionadas y por variables relevantes omitidas, en particular, factores psicológicos como autoestima o ansiedad, para las cuales no es posible obtener una medida confiable y que se pueda relacionar a datos de obesidad para Chile. Este problema se resuelve utilizando una variable instrumental inédita en la literatura, ya que utiliza información del registro de patentes a nivel comunal de la Región Metropolitana, en particular de 29 comunas, y las coordenadas geográficas de establecimientos educacionales subvencionados y municipales, datos no explotados en esta área de investigación. La variable instrumental corresponde al número de locales de comida chatarra entorno a un kilómetro a la redonda de establecimientos educacionales y rezagado un periodo. Los resultados indican que al abrirse un nuevo local de comida chatarra entorno a un colegio, el IMC promedio de los alumnos aumenta en promedio un 10%. Además, se encontró una relación negativa y significativa entre IMC y los puntajes Simce de matemáticas que va entre 18,2 y 31,9 puntos por cada una unidad de IMC que aumente el individuo. A diferencia de matemáticas, la relación entre IMC y los puntajes Simce de lectura no es tan clara, pese a que se registra una relación negativa y significativa en algunos casos. Sin embargo, estos resultados deben ser tomados con cautela debido a limitaciones de la base de datos. Finalmente, se comentarán el resto de las limitaciones y desafíos futuros entorno a esta rema de estudio y se realizan recomendaciones de política para combatir la obesidad.

Índice

1.- Introducción.....	5
2.- Revisión de la Literatura.....	9
3.- Datos.....	13
4.- Metodología.....	15
4.1.- Lógica de la variable instrumental.....	16
4.2.- Viabilidad variable instrumental.....	16
4.3.- Área a la redonda de los establecimientos educativos.....	18
4.4.- Modelos.....	20
4.5.- Construcción de las bases.....	23
4.5.1.- Información Geográfica.....	23
4.5.2.- Combinaciones entre Simce y Mapa Nutricional.....	24
4.5.3.- Proceso de Pegado de las Bases.....	25
5.- Resultados.....	27
5.1 Modelo Principal.....	27
6.- Conclusiones y Discusión.....	32
Anexos.....	35
Anexo A: Combinaciones factibles entre Simce y Mapa Nutricional.....	35
Anexo B: Definiciones de Comida Chatarra.....	35
Anexo C: Temas relacionados a patentes municipales.....	36
Anexo D: Comunas de la Región Metropolitana utilizadas.....	39
Anexo E: Información de variables utilizadas según base de datos.....	40
Anexo F: Estadística Descriptiva.....	44
G.I.- Proceso de Geocodificación.....	46
G.II.- Resultados Geocodificación:.....	49
Anexo H: Identificar locales de comida entorno a establecimientos educativos.....	51
Anexo I: Proceso de match de las bases.....	52
Bibliografía.....	53

Gráficos

Gráfico 1: Evolución de la Obesidad para una muestra de países OCDE	7
Gráfico 2: Distribución espacial de locales de comida chatarra y colegios para RM	17
Gráfico 3: Representación gráfica variable instrumental. Oferta de comida chatarra entorno a un kilómetro del Instituto Nacional	20
Gráfico 4: Proceso para combinar las bases de información.	25
Gráfico 5: Distribución de patentes municipales según comuna	37
Gráfico 6: Distribución de patentes, Grupo I.....	38
Gráfico 7: Distribución de patentes, Grupo II.....	38
Gráfico 9: Representación Gráfica Variable Instrumental, Ñuñoa.....	51
Gráfico 8: Zona de Cobertura de Localizador de Dirección.....	47

Tablas

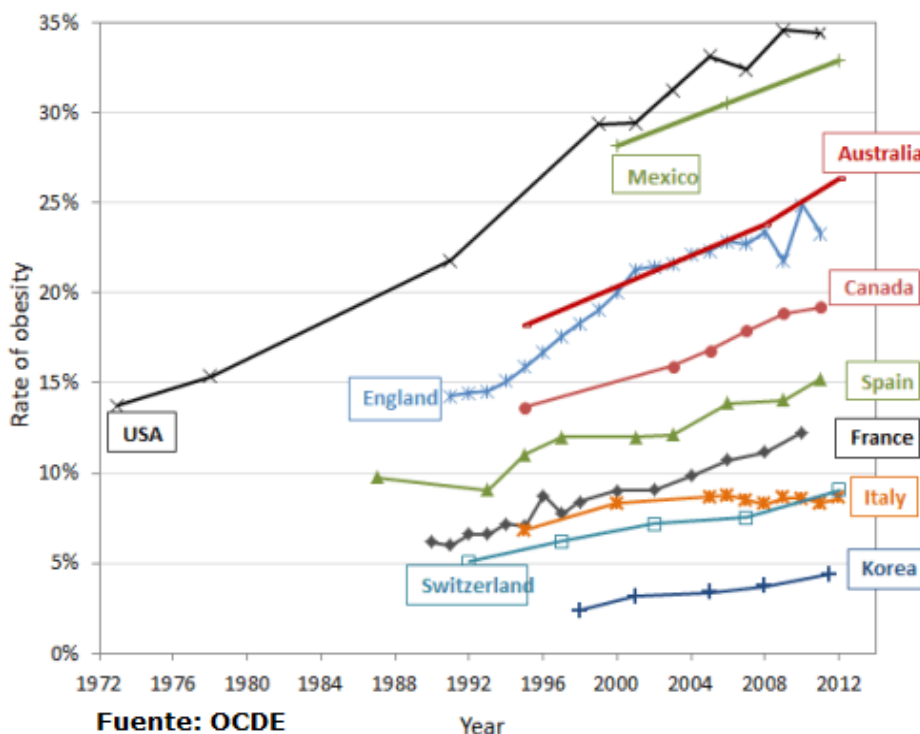
Tabla 1: Promedio de locales por año y distancia	18
Tabla 2: Porcentaje de colegios con al menos un local de comida chatarra, por año y distancia	19
Tabla 3: Modelos a Estimar	22
Tabla 4: Distribución de comunas según Provincia.....	23
Tabla 5: Combinaciones de Mapa Nutricional y Simce a utilizar en el estudio	24
Tabla 6: Número de datos a través del proceso de combinación de bases	26
Tabla 7: Número de datos a través del proceso de combinación de bases	26
Tabla 8: Número de datos a través del proceso de combinación de bases	26
Tabla 9: Número de datos a través del proceso de combinación de bases	27
Tabla 10: Resultados estimaciones MCO puntajes Simce matemáticas	27
Tabla 11: Resultados estimaciones IV puntajes Simce matemáticas	28
Tabla 12: Resultados estimaciones MCO puntajes Simce lectura	29
Tabla 13: Resultados estimaciones IV puntajes Simce lectura	30
Tabla 14: Resultados estimaciones no lineales	31
Tabla 15: Combinaciones factibles entre Simce y Mapa Nutricional	35
Tabla 16: Distribución de patentes comunales según tipo	36
Tabla 17: Uso de comunas Región Metropolitana.....	39

Tabla 18: Bases SIMCE-RBD.....	40
Tabla 19: Bases SIMCE- Alumnos.....	40
Tabla 20: Bases SIMCE - Cuestionario de Padres.....	41
Tabla 21: Conversión del nivel educacional en años de escolaridad	41
Tabla 22: Base Patentes Municipales	42
Tabla 23: Base Mapa Nutricional	42
Tabla 24: Base Registro de Estudiantes Chilenos (RECH).....	43
Tabla 25: Estadística Descriptiva 2013-2012.....	44
Tabla 26: Estadística Descriptiva 2012-2011.....	45
Tabla 27: Resultados Geocodificación por Comuna.....	49
Tabla 28: Geocodificación comida chatarra por comuna	50

1.- Introducción

Desde 1980 a la fecha, los índices de obesidad se han más que duplicado en el mundo, donde actualmente un 39% de la población mundial adulta (18 años o más) presenta sobrepeso¹ y un 13% es obesa². Esta situación se ve reflejada en el gráfico 1, en el que se observa por ejemplo a Estados Unidos, quien pasó de tener tasas de obesidad de un 15% en 1980 a tasas alrededor del 35% de obesidad en 2012.

Gráfico 1: Evolución de la Obesidad para una muestra de países OCDE



Obesidad y sobrepeso son definidas por la Organización Mundial de la Salud (OMS) como la acumulación anormal o excesiva de grasa que puede ser perjudicial para la salud y que es provocada por un desequilibrio entre calorías consumidas y gastadas, donde las primeras son mayores que las segundas. Usualmente se utiliza el índice de masa corporal (IMC) para evaluar el estado nutricional de las personas, donde:

$$IMC = \frac{\text{Peso}}{\text{Altura}^2} \quad \text{Estado Nutricional: } \begin{cases} \text{Sobrepeso } 25 \leq IMC < 30 \\ \text{Obesidad } IMC \geq 30 \end{cases}$$

La rápida evolución de estos índices deja en evidencia que es un fenómeno transversal a toda edad, país y condición socioeconómica, como indica Christopher Murray, director del IHME³. En efecto, este fenómeno no es exclusivo de la población adulta, sino que también afecta a la población infantil, alcanzado cifras cercanas al 30% a nivel mundial⁴.

¹ Una persona con sobrepeso tiene un IMC mayor a 25 y menor que 30, mientras que una obesa un IMC mayor a 30, por lo que son condiciones excluyentes entre sí.

² Organización Mundial de la Salud. Nota descriptiva N° 311.

³ Institute for Health Metrics and Education, centro de investigación de la Universidad de Washington que provee estadísticas médicas de los problemas más importantes de salud a nivel mundial.

⁴ Organización Mundial de la Salud. Datos y cifras sobre obesidad infantil.

Con respecto a Chile, el 39,3% de la población adulta tiene problemas de sobrepeso y un 25,1% es obesa, teniendo así en total un 64,5% de la población con problemas de exceso de peso, según cifras de la *Encuesta Nacional de Salud del año 2009/2010* (ENS). Por otro lado, un 24% de la población infantil tiene sobrepeso y un 10% es obeso, según cifras del Ministerio de Salud (MINSAL)⁵. De hecho, a nivel OCDE Chile es el quinto país con mayor tasa de obesidad⁶, mientras que a nivel infantil el segundo país de Latinoamérica⁷.

Una de las razones de que los índices de obesidad crecieran rápidamente es el aumento del consumo de alimentos hipercalóricos o ricos en grasas y/o azúcares. Cutler (2003) indica que esto se debió principalmente a un cambio tecnológico que existió en la industria alimentaria, el que permitió; disminuir los tiempos de preparación de alimentos; aumentar la oferta de alimentos procesados o no saludables debido a su producción en masa; conservar los alimentos por más tiempo, disminuyendo así el número de veces que se cocina y la apertura de locales o cadenas de comida de manera mucho más sencilla, por ejemplo comida rápida.

Además, este aumento en consumo se vio potenciado por el menor precio relativo de este tipo de alimentos con respecto a los alimentos *sanos*⁸. Por otro lado, la actividad física ha ido disminuyendo, lo cual se explica por formas más sedentarias de trabajo, jornadas laborales más extensas y medios de transporte más sedentarios, lo que resta aún más tiempo a la cocina. En resumen, las calorías consumidas han aumentado, mientras que las gastadas han disminuido.

Otro punto a considerar son los efectos tanto directos como indirectos de la obesidad en la salud y el ámbito laboral. Por el lado de los efectos directos, existe un sinnúmero de artículos médicos que indican que la obesidad es un *desencadenante de enfermedades no transmisibles*: como cardiopatías, asma, diabetes (Must y Strauss, 1999), problemas de sueño (Snell, Adam y Duncan, 2007), resistencia a la insulina, trastornos del sistema musculoesquelético, problemas escolares (Falkner et al, 2001), problemas psicológicos como ansiedad y depresión (Puhl y Latner, 2007), entre otros. De esta forma, un aumento en los índices de obesidad aumenta la probabilidad de contraer una de estas enfermedades, que de no ser bien tratadas pueden llevar a la muerte de una persona⁹.

También, se asocia con gastos en salud un 25% más altos, salarios un 18% más bajos (OMS) y un menor rendimiento académico, tópico en el cual centraremos nuestra atención. Por el lado de los efectos indirectos, encontramos la pérdida de oportunidades laborales, un mayor ausentismo laboral y una menor tasa de asistencia a clases, entre otros.

Este trabajo estudiará la relación entre rendimiento académico y obesidad. La literatura no ha alcanzado un consenso en la materia, debido a que la mayoría de los datos son autorreportados y al problema de endogeneidad presenta en la estimación, que puede ser explicado por dos razones: La primera es la doble causalidad presente en la estimación, es decir, no es obvio que sea la obesidad la que explique un menor rendimiento académico o que el menor rendimiento académico lleve a los individuos a ser más obesos. En segundo lugar, existe un problema de variables relevantes omitidas, es decir, hay aspectos relevantes para la estimación que no están siendo medidos o capturados, en particular los factores psicológicos como ansiedad o autoestima. A grandes rasgos se ha buscado resolver estos problemas de tres formas. Controlando por

⁵ La Tercera (2014) "Minsal: 34% de los niños menores de 6 años sufre obesidad o sobrepeso"

⁶ La Tercera (2014) "Chile es el quinto país con más obesidad en OCDE"

⁷ La Tercera (2015) "Chile, el país con menores desnutridos pero el segundo con más obesos de la región"

⁸ Overseas Development Institute (2015) "The rising cost of a healthy diet"

⁹ Según IHME, 3,4 millones de personas al año mueren debido a un problema de sobrepeso, siendo el sexto factor de riesgo con mayor tasa de mortalidad en el mundo.

factores psicológicos lo que permite solventar el problema de variables omitidas; Ocupando modelos de panel con efecto fijo, donde todo lo que no es medible se asume constante para el individuo y por lo tanto capaz de ser controlado; Empleando variables instrumentales las cuales deben estar correlacionadas con la obesidad, pero no con el rendimiento académico. Donde los instrumentos utilizados hasta el momento por la literatura son medidas relacionadas al IMC, como IMC del individuo un año atrás o IMC de los padres el IMC. O por otro lado, factores genéticos del individuo o sus familiares relacionados a la obesidad.

El presente trabajo es un aporte a la literatura por las siguientes razones: Empleo de datos no autorreportados de peso y altura, a diferencia de la mayoría de las encuestas internacionales en esta materia; Utilización de una variable instrumental inédita en la literatura, la cual utiliza información georreferenciada tanto de colegios como de locales de comida chatarra, en particular el instrumento corresponde a la densidad de locales de comida chatarra entorno a un colegio.

Se utilizarán datos del Mapa Nutricional de la *Junta Nacional de Auxilio Escolar y Becas* (JUNAEB)¹⁰ y los puntajes de las pruebas SIMC¹¹E (Sistema de medición de la calidad de la educación) para obtener medidas de obesidad y rendimiento académico, respectivamente. Además, para construir la variable instrumental se utilizó información geográfica tanto del registro de patentes municipales de 29 comunas¹² de la Región Metropolitana, como de los establecimientos educacionales¹³.

El trabajo se divide en seis secciones. La primera es una introducción a temas de obesidad y sobrepeso. Una segunda sección se revisa cómo se ha abordado la relación entre obesidad y rendimiento académico en la literatura internacional. Una tercera sección que se enfoca en revisar los datos disponibles para este trabajo y una breve estadística descriptiva pertinente a estos. Una cuarta sección que presenta la metodología a emplear en este trabajo y las principales precauciones a la hora de construir tanto la base de datos como el instrumento a utilizar en la estimación. La quinta sección indica los principales resultados de las estimaciones realizadas. Por último, la sección seis describe las principales conclusiones y se discute entorno a futuros desafíos y cuestiones de política pública.

2.- Revisión de la Literatura

El aumento sostenido y explosivo en las cifras de obesidad impulsó una serie de nuevas investigaciones que buscaron analizar los impactos de este fenómeno, es en este contexto que surge la literatura de obesidad y rendimiento académico.

Taras y Datema (2005) realizaron un resumen de los estudios más importantes hasta esa fecha, concluyendo que la evidencia soporta el hecho de una relación negativa entre obesidad y rendimiento académico. Sin embargo, la mayoría de la literatura estaba enfocada en realizar análisis de correlaciones o regresiones de mínimos cuadrados lineales, dejando de lado el problema de endogeneidad inherente a la estimación, o dicho de otra forma, la violación del supuesto de exogeneidad de los regresores con respecto al error. Esto genera problemas de

¹⁰ Encuesta anual aplicada a diferentes niveles educacionales para recolectar información antropométrica de alumnos de colegios subvencionados y municipales. Para más detalles revisar la sección de Datos.

¹¹ Prueba estandarizada con cobertura nacional, aplicada anualmente a diferentes niveles educacionales. Para más detalles revisar la sección de Datos.

¹² Debido a problemas de geocodificación y de datos incompletos por parte de algunas comunas, se decidió realizar el estudio para este número de comunas. Para más detalles revisar sección 4.5.

¹³ Se utilizó el programa ArcGis para procesar y analizar la información geográfica.

identificación que posteriormente sesgarían los resultados y podría llevar a concluir que existen efectos que en verdad no existen. Consciente de este problema, Sabia (2007)¹⁴ indica las tres causas principales de la endogeneidad:

1. Causalidad inversa:

Al momento de estudiar la relación entre obesidad y rendimiento académico, no es clara la relación causal entre estas, es decir, cuál de las dos explica a la otra. De hecho, cualquiera de los siguientes cuatro escenarios puede ocurrir:

- i. Un menor desempeño académico explica mayores niveles de obesidad. Ej: Se come en exceso para compensar este mal desempeño.
- ii. Un mayor nivel de obesidad explica un menor rendimiento académico. Ej: Discriminación, efectos psicológicos o físicos adversos, etc. Cabe destacar que este será esta relación causal la que se busca analizar a lo largo de este trabajo.
- iii. Existe una relación positiva. Ej: Un menor rendimiento académico reduce el apetito, por lo que la obesidad disminuye.
- iv. No existe ninguna relación entre ambas variables.

2. Factores no observables:

Al momento de estimar la relación entre obesidad y rendimiento académico existe una serie de variables que no necesariamente pueden ser capturadas o medidas en datos. Entre ellas se encuentran: Habilidad, variables psicológicas correlacionadas con ambas variables (autoestima, autocontrol), factores genéticos, entre otros. Esto provoca que estas variables estén incorporadas en el término de error del modelo y que por lo tanto este se correlacione con la variable que mide obesidad sesgando así la estimación.

3. Datos autorreportados:

Este tipo de datos son ocupados frecuentemente en este tipo de estudios. El problema de estos es que están sujetos a errores de medición, es decir, al momento de pedirle a un individuo que indique su altura y peso puede ocurrir una de las siguientes situaciones: No conocer sus datos, entregar datos aproximados, entregar datos erróneos pese a conocer los originales o simplemente entregar bien los resultados. Debido a que no se puede tener un 100% de certeza que los datos entregados son los correctos, los resultados también estarán en un cierto grado sesgados.

Los autores conscientes de esta endogeneidad, adoptaron diferentes metodologías para solucionar esta problemática. Una fracción de los autores se enfocó en los factores no observables, en particular los controles psicológicos. En esta línea destacan Neumark-Sztainer (2002) y Krukowski et al (2009). El primero indica que el sobrepeso en adolescentes se asocia a mayores niveles de bullying. Controlando por este factor y resolviendo el problema de variable omitida encuentra que es este factor el que explica un menor rendimiento académico y no así la obesidad o el sobrepeso. Krukowski va un poco más allá e introduce el concepto de *weight based teasing* o bullying específico al sobrepeso. Controlando por esta nueva medida concluye que es este el factor que explica el menor rendimiento académico y no el sobrepeso, lo que indicaría que se debería a un efecto social.

¹⁴ Basándose en Cawley (2004), quien estudió la relación entre obesidad y salarios, revisa da más adelante.

Por otro lado, hubo autores que se centraron en el uso de datos de panel y efectos fijos para analizar las dinámicas entre obesidad y rendimiento académico. Destacan entre ellos Datar y Sturm (2006), Gable et al (2008) y Gable, Chang y Krull (2012). Los autores presentan ciertos puntos en común, como la utilización de la misma base de datos y la identificación de los mismos grupos de individuos para estudiar la dinámica de la obesidad. Los grupos se dividen en aquellos que siempre presentaron sobrepeso, aquellos que nunca tuvieron sobrepeso y aquellos que pasaron a tener sobrepeso.

Datar y Sturm ocupan modelos multinivel de regresiones logísticas para alumnos entre kínder y tercer grado, concluyendo que sólo las mujeres del grupo que pasan a tener sobrepeso registran una penalización en término de puntajes y sólo los hombres con sobrepeso registran un mayor ausentismo escolar. Gable (2008) también estimó modelos multinivel, pero con coeficientes aleatorios de tres niveles. A diferencia del estudio anterior concluye que independiente del sexo, es el grupo con sobrepeso el que progresa menos que sus pares. Finalmente Gable, Chang y Krull utilizan un set de nuevas variables psicológicas de control, habilidades interpersonales y comportamientos internalizados, concluyendo que son estas últimas las que permiten explicar el rendimiento académico de los niños y no así la obesidad¹⁵. Alguna de las deficiencias de estos estudios es que las variables psicológicas a utilizar son reportadas por los profesores, los cuales no son especialistas en estos temas, razón por lo cual podrían estar sesgadas y afectar de esta forma la estimación final.

Por otra parte, la literatura también hizo uso de variables instrumentales (VI) para tratar el problema de endogeneidad. Existen dos aristas en esta materia, aquella que ocupa información relacionada al IMC para instrumentalizar y otra que utiliza información genética.

Sabia (2007) corresponde al primer grupo de autores y ocupa datos de la National Longitudinal Study of Adolescent Health (NLSAH) para estudiantes entre 7° y 12° grado. El autor utiliza la variable instrumental auto-reportada "obesidad de los padres, medido en IMC", concluyendo que existe un efecto negativo y significativo entre obesidad y rendimiento académico. Sin embargo el instrumento es criticable debido a que el IMC de los padres esta correlacionado directamente con los ingresos familiares y por lo tanto con los niveles educacionales, violando el supuesto de exclusión, lo que lo haría un mal instrumento.

En este mismo grupo se encuentran Kaestner y Grossman (2009), quienes ocupando datos de la National Longitudinal Survey of Youth (NLSY) para estudiantes de 5 a 12 años y el IMC del alumno dos períodos atrás como variable instrumental estiman el cambio de peso del estudiante sobre su rendimiento académico. Concluyen que no existe una diferencia significativa entre ambas. Sin embargo, la misma crítica realizada al instrumento de Sabia (2007) puede ser realizada, el IMC del individuo está correlacionado con el nivel de ingresos de su hogar y debido a que este no se mueve drásticamente entre años, al menos en promedio, también afectará los niveles educacionales.

Con respecto al segundo grupo de autores, los que utilizan información genética como variable instrumental, destacamos dos estudios, Fletcher y Lehrer (2008) y Ding (2009). Fletcher y Lehrer ocupan la base NLSAH entre 2001 y 2002 que tiene la particularidad de recolectar información genética para estudiantes entre 7° y 12° grado. En materia genética, en general la información es poligénica, es decir, se compone por la interacción de varios marcadores genéticos. Donde obesidad y sobrepeso se componen por aproximadamente 160 genes, como indica Perusse

¹⁵ Esto sólo fue posible porque conforme pasaron los años la encuesta agregó nuevas preguntas a su repertorio.

et al (2005). De esta forma, los autores estiman un modelo de efectos fijo por familia y diferentes variables instrumentales, las que corresponden a genes asociados a la obesidad, estos son: Transporte de dopamina (DAT), DRD4, receptor de dopamina D2, CYP2D6 y MAOA. Sus resultados indican que existen efectos positivos, pero no significativos de obesidad sobre rendimiento académico, sin embargo, la muestra sólo está compuesto por gemelos y mellizos lo que le resta representatividad a sus resultados.

Ding (2009) ocupa datos de la Georgetown Adolescent Tobacco Research, estudio longitudinal que va desde 1999 hasta 2003, para alumnos de 9° y 12° grado y variables instrumentales similares, las cuales son: D2, DAT, triptófano hidroxilasa y CYP2B6. A diferencia de los autores anteriores, sus resultados indican que existen efectos negativos y significativos del IMC sobre el GPA (grade point average) sólo de las mujeres. Sin embargo, la dificultad y el costo de tomar muestras genéticas de los alumnos recae en que no existan bases de datos de este tipo para estudiar esta relación en Chile.

Si bien la literatura internacional es abundante al momento de analizar la relación entre IMC y rendimiento académico RA, en Chile sólo Ivanovic et al (2001, 2002, 2004, 2008, 2009, 2011, 2013, 2014) han estudiado este tema. Metodológicamente ellos realizan un muestreo aleatorio por colegio de diferentes alumnos y miden factores como: tamaño de la circunferencia de la cabeza, actividad física, hábitos alimenticios, peso, altura, sexo, edad, ingresos familiares, educación de los padres, variables del colegio, etc. A grandes rasgos, utilizan mayoritariamente modelos multi-causales logísticos y concluyen que el factor antropométrico más importante es el tamaño de la circunferencia de la cabeza y que es este el que permite explicar un rendimiento negativo en las pruebas Simce e incluso un menor desempeño en la Prueba de Aptitud Académica y no así la obesidad. Además destacan que la actividad física se relaciona positivamente con el rendimiento académico.

Por otro lado, al momento de estimar la relación entre obesidad y rendimiento académico por lo general se toma como medida de obesidad el IMC del individuo y se realiza una estimación lineal entre estas dos variables. Sin embargo, más de algún autor ha destacado la importancia de estudiar la no linealidad de esta relación, es decir, analizar cómo se comporta la relación entre IMC y rendimiento académico en función de los diferentes estados nutricionales: sobrepeso, obeso, desnutrición y normal. Debido a problemas que serán analizados en la sección metodológica hubo que revisar también la literatura de obesidad y salarios, a partir de la cual se pueden extrapolar ciertas ideas para el estudio de la no linealidad. A grandes rasgos esta literatura sugiere tres opciones para estudiar este tema: Regresiones cuantílicas, grados superiores de IMC y modelos semiparamétricos.

Los modelos de regresiones cuantílicas en vez de buscar la media de la variable a estudiar, buscan la mediana o cualquier otro cuantíl o percentil para estimar una relación entre variables. De esta forma y en particular para los modelos de obesidad y salarios se analiza si es que el IMC de los individuos afecta más o menos dependiendo si los salarios son altos o bajos. Atella et all (2008) ocupó esta metodología para estudiar el impacto de la obesidad sobre los salarios de nueve países en Europa, encontrando un efecto negativo y heterogéneo entre países y a lo largo de la distribución de salarios. Lo importante de este resultado es que no necesariamente la relación negativa y las magnitudes encontradas entre obesidad y rendimiento académico se mantendrán para todos los países, apoyando el hecho de obtener más evidencia para países como Chile.

Otra opción es la utilización de grados mayores de IMC para modelar la no linealidad de esta relación. Wada y Tekin (2010) son los pioneros en esta materia y utilizando IMC^2 , encuentran una relación positiva para IMC y negativa para IMC^2 , es decir, ganar más peso ayuda a aumentar el puntaje, pero sólo hasta cierto punto, a partir del cual este comienza a disminuir. Si bien se podrían utilizar mayores grados de IMC para obtener mayor información, la interpretación se complejizaría por sobre la información que se estaría ganando, motivo por el cual no se avanza más allá de este punto. Además, debido a que IMC ya es una variable endógena, utilizar grados mayores de este incorporará mayor endogeneidad a la estimación y por lo tanto la necesidad de utilizar nuevas variables instrumentales, las cuales no necesariamente se tienen.

Con respecto a los modelos semi-paramétricos, estos tienen la ventaja de que no suponen formas funcionales a priori, como sí lo requiere la utilización de grados superiores de IMC. Sin embargo, son intensivos en el uso de modelos computacionales, por lo que es costoso estimarlos. Shimokawa (2008) fue quien primero utilizó estos modelos en la literatura y encontró que en las colas de las distribuciones de salarios se registran menores puntajes, tanto para hombres como para mujeres. Gregory y Ruhm (2011) también utilizaron esta metodología y concluyeron que mayores niveles de IMC se relacionan negativa e inequívocamente con menores salarios para las mujeres, a diferencia de los hombres donde la relación es menos clara, ya que depende de los supuestos que se realicen en los modelos. Además destacaron que las mujeres comienzan a ver penalizaciones en sus salarios a niveles más bajos que el IMC de corte de sobrepeso médico (23), lo que a juicio de los autores significaría que IMC es un proxy de belleza y por lo tanto mujeres más obesas ganarían menos debido a esta razón.

Finalmente, la literatura de obesidad y salarios es concluyente a la hora de indicar que son sólo las mujeres con sobrepeso las que sufren una penalización en sus salarios, como indican Cawley (2000) y Gotmaker (1993). Diversas teorías explican este hecho: Pagan y Davila (2007) indican que esto se debe a que los trabajos a los que acceden las mujeres obesas, son trabajos que de por sí tienen una menor remuneración; Cawley (2004) indica que esto se debe a una discriminación de género; finalmente Averet y Korenman (1996), explican que es la menor productividad la que explica este suceso.

Si consideramos que un tercio de los niños preescolares y la mitad de los escolares obesos terminan siendo adultos obesos (Serdula et al 1993), lo dicho por la literatura no sólo refleja mayores costos o perjuicios a futuros para este grupo de la población, sino que también podría explicar la menor productividad de este grupo y aportar información a la teoría del menor rendimiento académico de mujeres obesas.

3.- Datos

Al momento de revisar las fuentes de información disponibles en Chile y teniendo en mente la literatura, se buscaron datos de obesidad, rendimiento académico, variables psicológicas y variables genéticas. Sin embargo, no se encontraron bases de datos que permitieran relacionar variables psicológicas y/o genéticas con rendimiento académico de una manera confiable. Tampoco fue posible obtener datos de panel a nivel de educación básica o media. Debido a todo esto es que se optó por seguir una metodología de variable instrumental, para así resolver el problema de endogeneidad inherente a la estimación. Esta variable ocupará información georreferenciada y su construcción será explicada en mayor detalle en la siguiente sección.

Para obtener una medida de obesidad, se ocupará el *Mapa Nutricional*¹⁶ de JUNAEB. Este es una encuesta anual aplicada a establecimientos municipales y particulares subvencionados de pre-kínder, kínder, primero básico y primero medio. Se tendrá información a nivel de individuo y se contará con el establecimiento, comuna, provincia y región al cual asiste. Además, se cuenta con el IMC (variable de nuestro interés), sexo y fecha de nacimiento de cada estudiante. No obstante, no se dispone de un identificador por alumno que permita relacionar esta base a otras.

Se utilizarán los datos de 2009 a 2012, para primero medio. Además, el total de datos recolectados representa aproximadamente el 90% del total de matrícula de estos niveles del sistema público y un 93% del total del sistema educacional, incluyendo el sector privado, según JUNAEB.

Debido a que no existe un identificador en las bases JUNAEB que permita unir esa base con otras, fue necesario utilizar dos fuentes de información en materia de rendimiento académico, cuestión que se aclarará en la siguiente sección. La primera corresponde a SIMCE, una prueba estandarizada con cobertura nacional, desarrollada por el Ministerio de Educación y aplicada anualmente a diferentes niveles educacionales, los cuales no necesariamente son encuestados todos los años¹⁷. Su propósito, es generar y recopilar información relevante para mejorar la calidad y la equidad de la educación. Además, sitúa los logros de los alumnos en un contexto nacional permitiendo evaluar los resultados de aprendizaje de los establecimientos. Por último, también recoge información sobre docentes, estudiantes y padres y apoderados a través de diferentes cuestionarios¹⁸.

La segunda fuente de información sobre rendimiento académico es elaborada por el Centro de Estudios del Ministerio de Educación y corresponde al Registro de Estudiantes Chilenos (RECH). Esta es información estadística sobre distintos niveles y actores del sistema pre-escolar y escolar chileno, la cual abarca las siguientes temáticas: Matrícula única, matrícula consolidada de educación parvularia, rendimiento escolar, asistencia declarada de alumnos, SIMCE TIC, alumnos prioritarios y beneficiarios SE. Nos centraremos en los datos de rendimiento escolar¹⁹, donde se tendrá la siguiente información de los alumnos: sexo, fecha de nacimiento, tipo de enseñanza, RBD²⁰, promedio final, asistencia, mrun²¹ y situación de promoción al finalizar cada año escolar.

Por último, se utilizará información geográfica tanto de colegios²², como de patentes comunales o municipales de la Región Metropolitana (RM). Se busca así capturar puntos de oferta de comida alrededor de colegios a través del tiempo, cuestión que será explicada con mayor detalle en la sección metodológica. Por otro lado, las patentes son un permiso necesario para emprender cualquier actividad comercial que necesita un local fijo y que es otorgado por la municipalidad respectiva. Existen cuatro tipos de patentes: Comercial, industrial, profesional y de

¹⁶ Informe elaborado por la Junta Nacional de Auxilio Escolar y Becas (JUNAEB) desde el año 2001 hasta 2013

¹⁷ Para un mayor detalle de los niveles y años que han rendido el Simce desde 1998 a la fecha, revisar: http://www.agenciaeducacion.cl/wp-content/uploads/2013/01/calendario_pruebas_simce_simbologia30-07-15-v2.jpg

¹⁸ <http://www.agenciaeducacion.cl/simce/que-es-el-simce/>

¹⁹ Disponibles desde 2002 hasta 2014.

²⁰ RBD es el Rol Base de Datos del Colegio, es decir, es un código exclusivo para cada colegio el cual es entregado una vez que este ha sido creado.

²¹ MRUN, en Chile existe el Rol Único Nacional (RUN) que es el número identificador único e irrepetible que posee todo chileno. Debido a que conocer el RUN del individuo violaría la confidencialidad de las personas es que se crea el MRUN que corresponde a un RUN enmascarado o dicho de otra forma, se asigna un número a cada RUN que permite identificar al alumno a través de las bases educacionales, pero sin saber cuál es su RUN, ya que se trabaja con el MRUN.

²² Coordenadas geográficas obtenidas de: <http://recursos.datos.gob.cl/datastreams/69978/colegios-y-liceos-de-chile-en-region-metropolitana/>. Inicialmente había 2229 colegios en la base, sin embargo este se redujo a 1.842 luego de dejar fuera a colegios particulares y los colegios de las comunas no geocodificadas (explicado más adelante). Por último, la información geográfica de 11 de estos establecimientos fue adquirida a través de Google Maps, ya que no se contaba con ella en la base original.

alcohol. La primera es para tiendas y negocios de compraventa. Las profesionales hacen referencia a consultas médicas o estudios de abogados, arquitectura, etc. Patentes industriales se refieren a la producción de alimentos, como pan, o producción de manufacturas. Finalmente las patentes de alcoholes son para botillerías, bares, restaurantes y otros locales de comida.

Como sólo buscamos identificar y utilizar puntos de oferta de alimento, se dejaron fuera del estudio las patentes profesionales ya que no son lugares en los que se encuentre esta información. Sí serán incluidas las patentes comerciales, industriales y de alcoholes. Se deberá tener cuidado con la inclusión de los últimos dos tipos de patentes, ya que incluyen locales que no contienen alguna oferta de alimentos, pero sí incluirán giros como panaderías o restaurantes diurnos los cuales sí son de nuestro interés²³. Es importante destacar que toda patente deberá contar con su fecha de apertura, ya que de lo contrario no se podrá tener la certeza de que un punto de oferta de alimento está ubicado alrededor de un colegio en un momento en específico, cuestión fundamental para construir la variable instrumental.

Para obtener esta información hubo que comunicarse con cada municipalidad de la región metropolitana para solicitar sus registros de patentes. En general la petición de información se hizo a través del sistema de Transparencia del Gobierno de Chile, pero en otras ocasiones hubo un contacto directo con personas de ciertas municipalidades para solicitar en detalle alguna información²⁴.

En el Anexo E se encuentra un detalle de todas las bases de datos y variables utilizadas en el estudio. Mientras que el Anexo F muestra una breve estadística descriptiva de las mismas.

4.- Metodología

Como ya se mencionó en la revisión de la literatura, existe un problema de endogeneidad a la hora de estimar el efecto de la obesidad sobre el rendimiento académico, explicado por una doble causalidad y por variables relevantes omitidas, en particular, factores psicológicos como ansiedad y autoestima. Debido a la limitación de datos psicológicos, genéticos o de panel, es que para poder resolver esta dificultad se decidió utilizar una variable instrumental. La literatura ha empleado instrumentos tales como: IMC de familiares, IMC del individuo rezagado o información genética relacionada a obesidad. Este trabajo es inédito en la literatura debido a que utiliza información geográfica geocodificada de patentes y colegios para construir la variable instrumental²⁵. Dicho esto, la variable instrumental deberá cumplir con las siguientes dos condiciones:

1. **Condición de Relevancia:** La variable instrumental debe estar correlacionada con la variable a instrumentar, en este caso IMC. Por un lado, la correlación puede ser débil, lo que acarrea una serie de problemas a la hora de realizar la estimación. O puede ser fuerte, lo que no presentaría problemas estadísticos a la hora de realizar la estimación. Stock y Yogo (2005) concluyen que basta un test F menor a 16,38, asintóticamente hablando, para que un único instrumento sea considerado débil²⁶.

²³La inclusión de este tipo de patentes debe tomarse con cuidado, debido a que incluyen locales que no contienen alguna oferta de alimentos. Revisar Anexo C para más información acerca de la distribución de patentes por tipo y otros temas relacionados.

²⁴ Ignacio Urrea fue quien dirigió estas operaciones.

²⁵ Se utilizó el programa ArcGis para geocodificar la información. Para mayores detalles revisar Anexo G.

²⁶ Stock y Stagner (1997) habían concluido con antelación que un instrumento era débil si es que este tenía un test F menor a 10. Para llegar a este nuevo valor del test, los autores extendieron el estudio de Cragg-Donald (1993).

2. **Condición de Exclusión:** La variable instrumental no debe estar correlacionada con el término de error, es decir, con factores no observables o no medibles por los investigadores. Sin embargo, ya que no se puede cuantificar esta correlación, la condición deberá ser justificada de manera teórica.

Como se comentó en la sección de datos, sólo ocuparemos aquellas patentes que permitan identificar un tipo de oferta de alimento, en particular comida chatarra. Michael Jacobson²⁷ definió este concepto como:

"Comida percibida como insalubre o con poco valor nutritivo, que contiene altos niveles de grasas, sal, carbohidratos y numerosos aditivos alimentarios. Al mismo tiempo, tiene carencia de proteínas, vitaminas y fibra, entre otros".

Basándose en esta definición, se entenderá por comida chatarra todos aquellos alimentos que cumplan con alguna de estas características o que sean consumidos masivamente por los habitantes chilenos, como pan y bebidas azucaradas. El listado completo de alimentos se encuentra en el Anexo G²⁸. Luego de definir lo que se entiende por comida chatarra, se procederá a obtener una distribución geográfica de colegios y de las patentes de comida chatarra por comuna de la Región Metropolitana, la cual puede ser vista en el gráfico 2. Sin embargo, existirán comunas para las cuales no se geo-codificó información de patentes, tema a tratar en la sección 4.5.

Antes de definir la variable instrumental, se deberá revisar tanto la lógica de la variable, como su viabilidad y área a la redonda alrededor de los colegios a utilizar, cuestiones que se analizarán a continuación:

4.1.- Lógica de la variable instrumental

Consideremos dos colegios A y B idénticos en todos los aspectos, excepto uno, la oferta de alimentos chatarra que tienen a su alrededor. Si la oferta de alimentos es superior para el colegio A que para el colegio B, sus estudiantes tendrán un mayor acceso a este tipo de alimentos y por ende una probabilidad de compra de alimentos chatarra más alta. Esto provocará finalmente que la probabilidad de tener un mayor peso al final del año para los alumnos del colegio A, sea más que para los del colegio B.

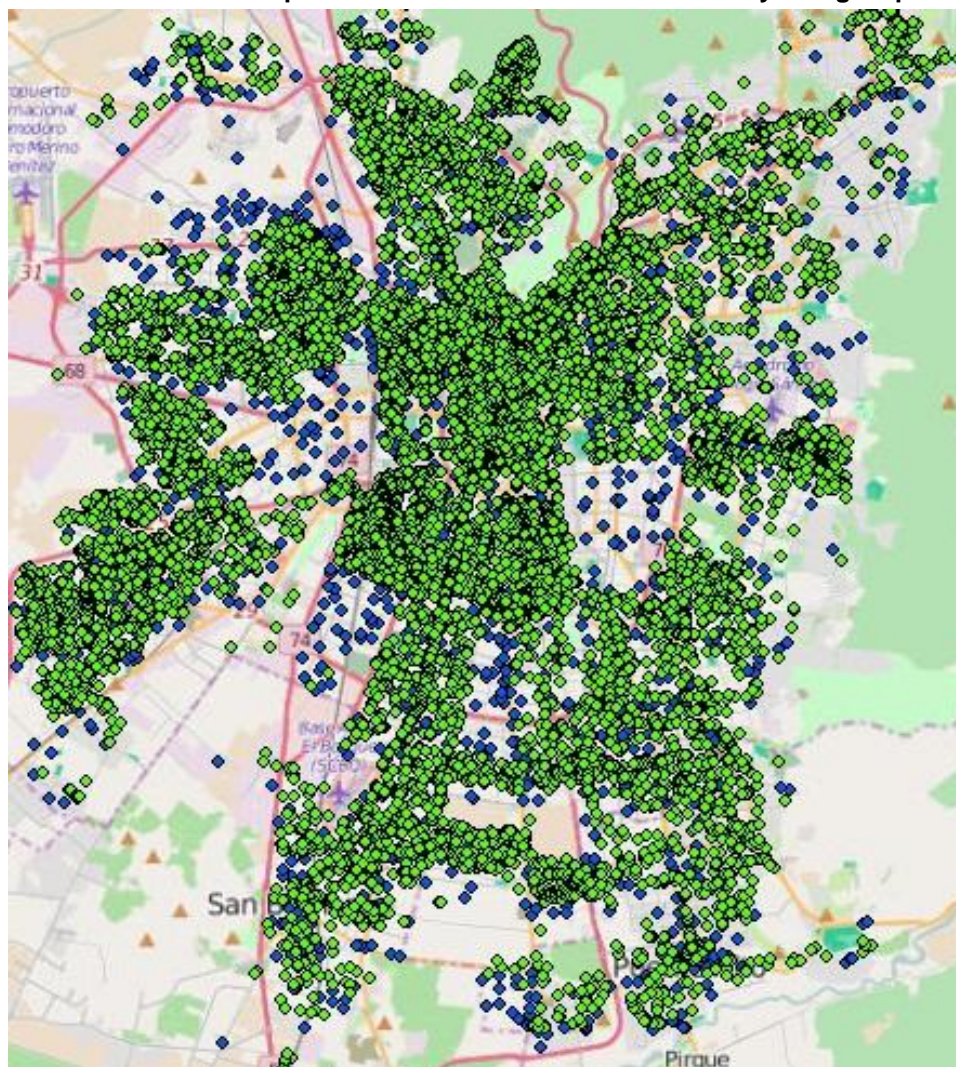
4.2.- Viabilidad variable instrumental

La relación anteriormente planteada cumple tanto con la condición de relevancia, como la de exclusión. Referente a la primera condición, efectivamente se ve afectado el peso de los individuos, Azar et al (2015) indican que cada un kilómetro de cercanía de un local de comida rápida a un colegio, aumenta un 5% las probabilidades de que sus alumnos presenten algún problema de peso (obesidad o sobrepeso). Además, la evidencia indica que un 95% de los chilenos no tiene una alimentación saludable (ENCA, 2014), siendo además el segundo país con mayor consumo de pan, el tercero con mayor consumo de bebidas azucaradas y el cuarto con mayor consumo de alimentos procesados y envasados, soportando el hecho de que los chilenos consumen elevadas proporciones de alimentos chatarra, lo que efectivamente aumenta el peso de los individuos, como recalcan Shanty et al (2004) y French et al (2000).

²⁷ Michael Jacobson es un doctor en microbiología del MIT quien fundó el Centro para la ciencia del interés público en 1971 y de finió este concepto en 1972. Fuente: <http://quees.la/comida-chatarra/>

²⁸ Para chequear la robustez de los resultados se consideró más de una definición de comida chatarra, ambas pueden verse en el Anexo B. Sin embargo sólo serán presentados los resultados para la 1era definición.

Grafico 2: Distribución espacial de locales de comida chatarra y colegios para RM



Nota: Elaboración propia en base al programa ArcGis. Verde: locales de comida chatarra; Azul: establecimientos educacionales.

La segunda condición exige que la variable instrumental no esté correlacionada con factores no observables, en nuestro caso: factores psicológicos, genéticos o habilidades de los individuos. Estas variables también son no observables para los dueños de los locales, por lo cual no determinan la instalación del local en cuestión. Además, si bien en algunos casos se hacen estudios sobre el público objetivo a atender, este no considera habilidades o el perfil psicológico (ansiedad, motivación) de los alumnos del establecimiento educacional más cercano. Por otro lado, existen dos motivos para utilizar la oferta de locales de comida rezagada y no así la oferta de locales de comida actual. El primero es evitar que locales de comida chatarra que fueron instalados en fechas posteriores a las mediciones del mapa nutricional sean considerados en el instrumento²⁹, esto ocurriría si se utiliza la oferta actual, lo que aumentaría el número de locales entorno a los colegios de una manera artificial. El segundo motivo es que el IMC de un individuo no depende

²⁹ Si bien se podría limpiar la variable instrumental para que no considerará estos locales al momento de considerar la oferta de locales contemporánea, las fechas de medición del mapa nutricional varía de establecimiento en establecimiento, por lo que sería agregar una complejidad innecesaria.

exclusivamente de su situación actual, sino que es una combinación de factores a través del tiempo (hábitos alimenticios, genética, horas de sueño, ejercicio, entre otros). Por lo mismo, instalar un local de comida muy cercano a la fecha de medición del IMC no debiese provocar grandes variaciones en el IMC de los individuos, ya que lo más probable es que no hayan comprado ahí o las veces que lo hayan hecho lo debiese llegar a afectar su IMC, es decir, considerar la oferta actual de locales podría también estar aumentando la influencia del número de locales entorno al colegio. Ninguno de estos problemas debiese ocurrir con la utilización del rezago en la medición de los locales, ya que estarían el suficiente tiempo establecidos para afectar el IMC de los individuos y sí o sí para la fecha de medición del mapa nutricional los locales que se contabilizan serían los correctos.

4.3.- Área a la redonda de los establecimientos educacionales

Un supuesto clave es definir el área a la redonda en que se medirá el número de locales de comida alrededor de un colegio. En una fase exploratoria se consideraron diferentes distancias, comenzando por 50 metros y aumentando la distancia hasta llegar a 1.000 metros. Conforme se iba aumentando esta área, se puede fácilmente apreciar que tanto la cantidad de locales de comida entorno a cada colegio, como el número de colegios que efectivamente tenían un local a la redonda aumentaban, esta situación se ve reflejada en el gráfico 3 y en las tablas 1 y 2.

Tabla 1: Promedio de locales por año y distancia

Distancia	2006	2007	2008	2009	2010	2011	2012	2013
1km	10,82	10,13	11,76	15,92	13,33	12,38	13,27	15,28
800m	7,48	7,09	8,17	11,03	9,26	8,59	9,10	10,42
500m	3,69	3,57	4,13	5,47	4,59	4,19	4,36	5,14
400m	2,85	2,83	3,23	4,17	3,49	3,21	3,31	3,90
300m	2,22	2,16	2,52	3,13	2,53	2,47	2,42	2,91
200m	1,78	1,65	1,93	2,39	1,93	1,96	1,87	2,09
150m	1,58	1,44	1,60	1,99	1,62	1,61	1,62	1,58
100m	1,40	1,32	1,36	1,63	1,36	1,34	1,40	1,41
50m	1,35	1,45	1,21	1,36	1,17	1,16	1,27	1,21

Fuente: elaboración propia en base a datos patentes municipales.

Tabla 2: Porcentaje de colegios con al menos un local de comida chatarra, por año y distancia.

Distancia	2006	2007	2008	2009	2010	2011	2012	2013
1km	93%	92%	92%	91%	90%	91%	91%	91%
800m	88%	86%	87%	86%	84%	85%	86%	87%
500m	71%	68%	69%	70%	68%	69%	71%	70%
400m	60%	56%	59%	60%	58%	57%	60%	58%
300m	44%	41%	44%	45%	45%	42%	45%	44%
200m	24%	23%	27%	27%	27%	25%	27%	26%
150m	15%	14%	18%	19%	18%	17%	17%	17%
100m	8%	7%	9%	10%	9%	8%	9%	8%
50m	2%	2%	2%	3%	3%	2%	2%	2%

Fuente: elaboración propia en base a datos patentes municipales y colegios.

En la tabla 1 se ve que pasar de 1km de distancia a 800m, reduce en un 30% aproximadamente el número de locales promedio en torno a los colegios que efectivamente siguen teniendo locales alrededor suyo. Si se continúa disminuyendo la distancia esta reducción se acrecienta mucho más, pasando a ser más del 50% al considerar 500m. Por otro lado, en la tabla 2 se puede observar que esta misma reducción de distancia, 1km a 800m, disminuye aproximadamente un 15% el número de colegios que tiene al menos un local a la redonda. Además, bajo los 500m de distancia, la cantidad de colegios sin locales de comida a su alrededor se reduce entre 35 y 40%, valores muy elevados considerando que estos colegios no estarían en la muestra que utiliza la variable instrumental.

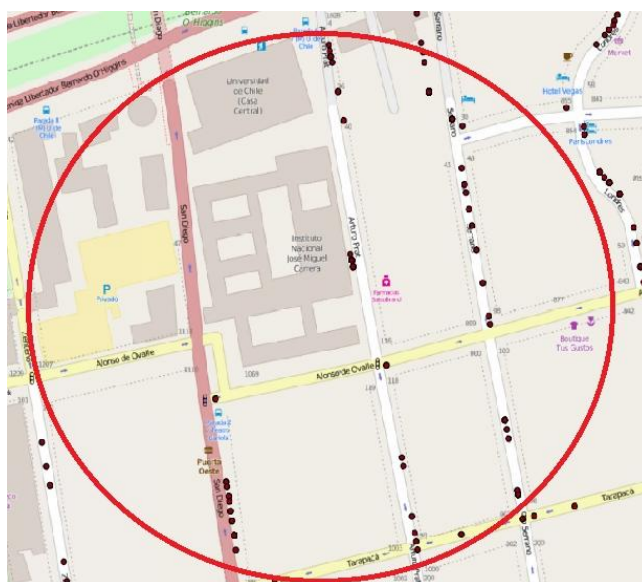
Debido a la rápida caída en términos de locales y de colegios, es que se decidió utilizar una distancia de un kilómetro a la redonda de los establecimientos educacionales, descartando utilizar distancias inferiores debido a que se dejaba de lado información relevante a la hora de instrumentalizar la variable endógena y se perdería fuerza en el instrumento, de todas formas, los resultados de las estimaciones son robustos para distancias inferiores. Por otro lado, es importante señalar que al aumentar la distancia entorno a los colegios, se está asumiendo que el área de influencia de los locales de comida sobre los alumnos del establecimiento crece, es decir, estoy diciendo que cualquiera local que se encuentre en ese rango afectará al individuo. Sin embargo, sabemos que los alumnos no irán hasta muy lejos para encontrar un local de comida chatarra si es que tienen uno más cercano (salvo excepciones), por lo que es muy probable que pasado cierto punto los individuos no se vean influenciados por los locales que se encuentran más alejados, por lo que de considerar distancias muy amplias se estaría agregando información que no sería útil. El caso extremo sería considerar un rango de toda la región metropolitana, a pesar de que todos los establecimientos tuviesen el mismo número de locales de comida, es claro que estos no influyen igual a los individuos ya que por ejemplo para la comuna de Maipú los locales de La Reina no lo influirán y viceversa. Por lo mismo el área de influencia no puede ser muy elevada, 1 kilómetro³⁰ parece razonable y ya ha sido utilizado por autores como Azar et al (2015).

³⁰ Se hizo el ejercicio de ver el número de locales para distancias mayores, verificando el punto señalado. Por ejemplo considerar dos kilómetros a la redonda aumenta en cuatro o cinco veces el número de locales a la redonda.

Con todo esto en mente, la variable instrumental corresponderá a la oferta de locales de comida chatarra entorno a un kilómetro a la redonda de cada colegio, corregida por la matrícula del establecimiento, medida en logaritmo y un periodo rezagado de la medición de IMC. Su representación gráfica puede ser vista en el gráfico 3 y su expresión matemática es la siguiente³¹:

$$Densidad_locale_{S_j,t-1} = \log\left(\frac{N^\circ \text{ de locales a 1km a la redonda del colegio } j, \text{ el período } "t - 1"}{Matrícula \text{ del colegio } "j", \text{ el período } "t-1"}\right)$$

Gráfico 3: Representación gráfica variable instrumental. Oferta de comida chatarra entorno a un kilómetro del Instituto Nacional.



Fuente: Elaboración propia en base a patentes municipales e información geográfica del Instituto Nacional.

4.4.- Modelos

Antes de presentar nuestro modelo, abordaremos brevemente la variable dependiente, las variables de control y la variable a instrumentalizar. Como se indicó en la sección de datos, se utilizará el Mapa Nutricional para alumnos de primero medio, sin embargo, las pruebas Simce no son aplicadas para este nivel. Para solucionar este problema es que se buscó el puntaje Simce más cercano a la medición de IMC, de esta forma la variable dependiente corresponderá al puntaje Simce de segundo medio. Cabe destacar que se aplica tanto la prueba de lectura como de matemáticas, razón por la cual se estimaran dos modelos diferentes, además, todos los puntajes han sido estandarizados.

La variable a instrumentalizar corresponderá al IMC de cada individuo, la creencia a priori es que se relacione negativamente con el rendimiento académico, aun cuando la literatura no haya llegado a un consenso en este aspecto.

³¹ En Anexo H se ve un ejemplo gráfico para una única comuna de su distribución espacial y se describe el proceso para identificar los locales entorno a esta.

Por último las variables de control se dividen en tres categorías: Individuales, a nivel de colegio y familiares. Con respecto a los controles individuales se tendrá: sexo, edad y puntajes SIMCE pasados del alumno. Sexo no tiene un signo definido en la literatura debido a la evidencia mixta que han arrojado las diferentes estimaciones. Así se ha encontrado que sólo a las mujeres les afecta negativamente la obesidad o que el sexo no tiene un impacto significativo en los resultados, independiente si la prueba es de matemáticas o lenguaje. La edad se incluyó en la estimación debido a que un mismo nivel de IMC, por ejemplo 20, representa de diferentes formas la obesidad para un niño de 8 años que otro de 16. Así, edad se incluirá en la estimación como el número de meses de vida al momento de ser encuestado el alumno. Los puntajes pasados fueron incluidos para verificar que efectivamente el IMC de los individuos tiene alguna relación con el rendimiento académico y que no es sólo la historia del individuo la que explica su propio rendimiento. Para poder incluir esta última variable de control se deberá combinar la información con bases pasadas de Simce.

Referente a los controles a nivel de colegio, se contará con el nivel socioeconómico del establecimiento (NSE) y su dependencia administrativa. El NSE corresponderá a cinco categorías: bajo, medio bajo, medio, medio alto y alto. Debido a un efecto par y acorde a lo que dice la literatura se espera que al ir aumentando en categoría socioeconómica, también aumente el puntaje del alumno. Por otro lado, la dependencia administrativa corresponderá a una dummy que toma valor uno para colegios particulares subvencionados y cero para colegios municipales³². Se espera que los colegios particulares subvencionados tengan en promedio un mejor rendimiento académico que los municipales, también por un efecto par.

Finalmente los controles a nivel familiar corresponden a: Educación de los padres e ingreso del hogar. Se dispondrá información tanto de la educación del padre como de la madre y se esperaría que si estos son más educados, el puntaje del individuo sea mayor. Así mismo con el ingreso del hogar, a mayor ingreso se esperaría un rendimiento académico mayor por parte del alumno.

Los controles familiares son agregados a partir de los cuestionarios de padres y apoderados, los cuales al no tener un 100% de respuesta disminuirán el número de observaciones de la muestra final, razón por la cual se considerará un modelo en cual se incluyen y otro en el cual no se incluyen estos controles en la estimación. Los modelos a estimar tienen la estructura de las siguientes dos ecuaciones y su listado completo es resumido en la Tabla 3.

$$SIMCE_{i,t} = \alpha + \beta X_{i,t,j} + \delta IMC_{i,j,t-1} + u_{i,t} \quad (1)$$

$$IMC_{i,j,t-1} = \gamma Z_{j,t-2} + e_{i,t-1} \quad (2)$$

Donde:

- $SIMCE_{i,t}$: Puntajes Simce del alumno "i" en el período "t". Puntajes Simce de matemáticas y lectura de segundo medio.
- $X_{i,t,j}$: *Controles individuales*: Edad, sexo y puntaje pasado, *Controles del colegio*: NSE y dependencia administrativa. *Controles familiares*: Educación de los padres e ingreso del hogar.
- $IMC_{i,j,t}$: Índice de masa corporal del alumno "i", que está en el colegio "j" en el período "t".
- $Z_{j,t-2}$: Logaritmo del número de locales de comida chatarra entorno al colegio "j", corregido por su matrícula, 1 periodos rezagado de la medición de IMC.
- $u_{i,t}$: Errores clusters por curso del individuo "i" en el periodo "t".

³² Se debe recalcar que debido a la utilización del Mapa Nutricional, no se dispone de información de colegios particulares paga dos.

Tabla 3: Modelos a Estimar

Modelo	Descripción
1	Educación media matemática sin controles familiares
2	Educación media matemática con controles familiares
3	Educación media lectura sin controles familiares
4	Educación media lectura con controles familiares

Fuente: Elaboración propia

Como ya se dijo, la literatura no sólo estima modelos lineales entre IMC y rendimiento académico, sino que también utiliza diferentes modelos no lineales para poder entender de mejor forma esta relación. Sabia (2007) y Ding (2009) se ocupan de esta no linealidad al utilizar variables dummy para identificar las diferentes categorías nutricionales de los alumnos, en particular aquellos que son obesos y que tienen sobrepeso. Luego, instrumentalizan cada una de estas categorías y encuentran que los efectos negativos de esta relación son para esos grupos y no otros. Sin embargo, para poder realizar este procedimiento es necesario contar con al menos el mismo número de instrumentos como categorías creadas, cuestión que no podemos realizar en nuestro estudio, por lo que se debe buscar otra alternativa para estimar modelos no lineales. Kaestner y Grossman (2009) utilizan otro enfoque al dividir en cuatro grupos el IMC de los individuos, de 0% a 5%, de 5% a 15%, de 85% a 95% y de 95% a 100%, sin embargo, el problema del número de instrumentos sigue manteniéndose, por lo cual tampoco es una alternativa viable.

Debido a que no se pueden utilizar las mismas metodologías de estos autores, es que se también se buscó en la literatura de obesidad y salarios como tratan la no linealidad³³. A grandes rasgos, la literatura utiliza tres metodologías: Regresiones cuantílicas, modelos semi-paramétricos y grados superiores de IMC. A continuación se analizará la viabilidad de aplicación de cada una de estas opciones:

Las regresiones por cuantiles analizan si es que para niveles altos o bajos de la variable dependiente (salarios, Simce), la variable endógena (IMC) afecta más o menos. Sin embargo, a nosotros nos interesa analizar la relación inversa, es decir, si un IMC más alto o bajo, influye más o menos en los puntajes, por lo mismo esta metodología fue descartada.

La inclusión de grados mayores de IMC hubiese sido una forma sencilla e ideal de incluir la no linealidad del IMC. Sin embargo, no fue posible obtener otro instrumento diferente al ya planteado, por lo que tampoco se utilizó esta metodología para estimar la no linealidad, ya que el único requisito de este procedimiento es tener el mismo número de instrumentos como de variables endógenas. Cabe destacar que se podría haber usado grados mayores del instrumento creado, sin embargo, esto generaría problemas de colinealidad.

Los modelos semi-paramétricos tienen la ventaja de que no suponen formas funcionales a priori, como sí lo harían aquellos modelos donde se incluyen grados mayores de IMC, como IMC^2 . Sin embargo, son intensivos en el uso de modelos computacionales y el costo de estimarlos puede ser mayor que el beneficio de obtener estos resultados. Por lo mismo en el presente trabajo se decidió no utilizar este tipo de modelos.

³³ Muchos aspectos metodológicos pueden ser extrapolados entre ambas literaturas

Tampoco fue posible estimar la no linealidad utilizando diferentes categorías de IMC al mismo tiempo o por sí solas, como obesidad y sobrepeso, debido a que en la primera opción se contaría con más variables endógenas que instrumentos, que en esencia es lo planteado por la literatura de obesidad y rendimiento académico. Mientras que en la segunda opción no se logró encontrar ninguna relación significativa.

Revisadas y comentadas las diversas formas de incluir la no linealidad, es que se decidió incluirla de la siguiente forma: En primer lugar se clasificó a los alumnos según IMC³⁴, de la siguiente forma:

$$IMC = \begin{cases} \text{Desnutrición} & IMC < 18,5 \\ \text{Normal} & 18,5 \leq IMC < 25 \\ \text{Sobrepeso} & 25 \leq IMC < 30 \\ \text{Obeso} & IMC \geq 30 \end{cases}$$

Posteriormente cada muestra fue restringida en cuatro muestras más pequeñas, según cada una de las categorías recién planteadas para finalmente estimar los modelos inicialmente planteados en donde sólo se tendría como variable endógena el IMC de los individuos. Cabe destacar que los modelos a estimar serán los mismos que los de la tabla 3 con la particularidad de que las muestras serán más acotadas.

4.5.- Construcción de las bases

Como una única base no cuenta con toda la información para realizar el estudio, fue necesario juntar diferentes bases para así poder llevar a cabo las estimaciones. A continuación se revisaran algunos aspectos de la información geográfica a utilizar, los cruces de información entre Mapa Nutricional y puntajes SIMCE y finalmente el proceso que se siguió para juntar en una única base esta información.

4.5.1.- Información Geográfica

En la sección de datos se indicó que sólo se utilizaría información geográfica de la RM y no de todo Chile. Si bien las bases del MN y Simce contienen información a nivel de país, existen limitaciones técnicas en el proceso de geo-codificación, razón suficiente para sólo utilizar información de la RM³⁵.

Dicho esto, nos centraremos en la RM la cual está compuesta por seis provincias: Chacabuco, Cordillera, Maipo, Melipilla, Santiago y Talagante, dentro de las cuales existen 52 comunas, distribuidas de la siguiente forma:

Tabla 4: Distribución de comunas según Provincia

Provincia	Nº de comunas
Chacabuco	3
Cordillera	3
Maipo	4
Melipilla	5
Santiago	32
Talagante	5
Total	52

Fuente: Subsecretaría de Desarrollo Regional y Administrativo.

³⁴ Clasificación según OMS.

³⁵ Para más detalles de estas limitaciones revisar Anexo G.

Sin embargo, no fue posible obtener información de las siguientes seis comunas: Alhué, Buín, Isla de Maipo, María Pinto, San Pedro y San Ramón. Además, luego de analizar la información disponible por comuna y siguiendo sugerencias de COES³⁶ se decidió acotar el estudio a la Provincia de Santiago, a la que se agregó las comunas de Puente Alto y San Bernardo.

Debido a que la variable instrumental identifica el número de locales que se encontraban alrededor del colegio un periodo previo a la medición del IMC de los individuos es de suma importancia identificar si un local estaba o no estaba en un determinado momento del tiempo. No tener claridad sobre este punto afectaría la medición de la variable instrumental, ya que se podría incluir a locales que no se encontraban en torno al colegio en ese periodo de tiempo. Por esta razón se excluyó las siguientes comunas: Estación Central, Macul, Renca y Lo Espejo³⁷, ya que no contaban con la información de las fechas de apertura de sus patentes. La muestra final contará con un total de 29 comunas.

Finalmente es importante recalcar que debido a que sólo se considera información geográfica de 29 comunas de Chile, la información tanto del Mapa Nutricional como de los puntajes Simce será acotada a estas comunas.

4.5.2.- Combinaciones entre Simce y Mapa Nutricional

Simce y Mapa Nutricional son aplicados a diferentes niveles educativos por lo que fue necesario combinar la información entre estas bases para así poder seguir a un individuo a través del tiempo y poder realizar el estudio. Sin embargo, no sólo se aplican a diferentes niveles educativos, sino que además las pruebas SIMCE no son aplicadas todos los años a los mismos niveles educativos, a diferencia del Mapa Nutricional, lo que dificulta aún más el seguimiento de un individuo.

Debido a que la causalidad que se quiere estudiar es desde obesidad hacia rendimiento académico, es que se identificó aquellas combinaciones en donde la medición del puntaje Simce sea posterior a la medición del mapa nutricional, de esta manera nos aseguramos que luego de corregir la endogeneidad efectivamente se analizara la causalidad que se busca³⁸. Se descartó de esta forma combinaciones en donde el puntaje Simce se midiera con anterioridad del IMC, porque se estaría analizando la dirección contraria, es decir, cómo cambios en los puntajes Simce afectan el IMC de los individuos. Además, se exigió que las bases tuviesen un puntaje Simce anterior, debido a que es uno de los controles utilizados en la estimación. En base a estos criterios será que la tabla 5 muestra las combinaciones que servirán para nuestro estudio.

Tabla 5: Combinaciones de Mapa Nutricional y Simce a utilizar en el estudio

Fuente Nutricional		Simce	
Fuente	Año	Año escolar	Año
Mapa N. Media	2012	2do medio	2013
Mapa N. Media	2011	2do medio	2012

Fuente: Elaboración propia en base a MN y Simce.

³⁶ Para revisar el proceso de geocodificación revisar Anexo G.

³⁷ Lo Espejo sólo tenía fecha de ingresos anteriores al periodo de análisis del estudio, por lo que no contaba con la información mínima necesaria. Además, el listado completo de todas las comunas utilizadas se encuentra en el Anexo D.

³⁸ Todas las combinaciones entre Simce y Mapa Nutricional pueden ser revisadas en el Anexo A, donde se debe destacar que sólo se dispone de información del Mapa Nutricional desde el año 2007 en adelante.

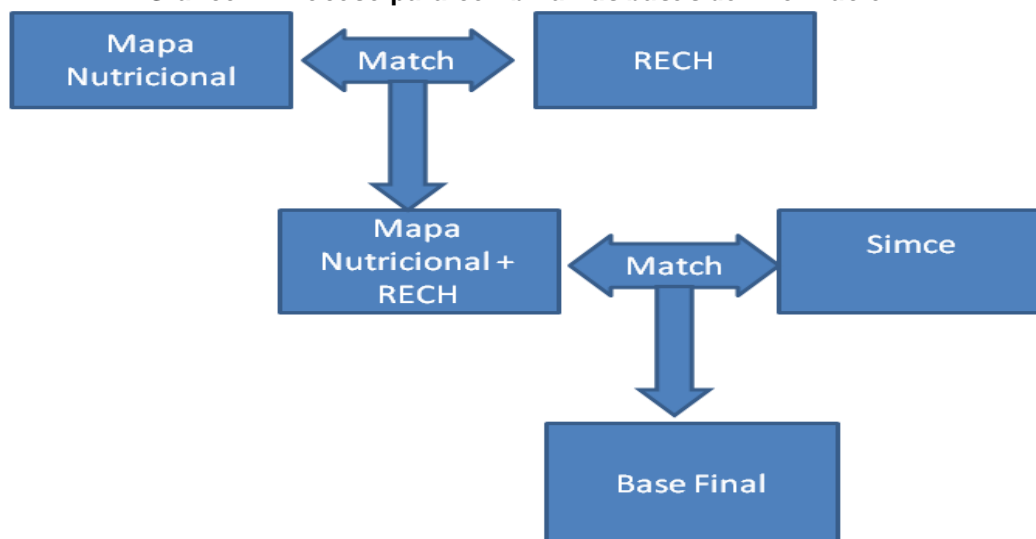
4.5.3.- Proceso de Pegado de las Bases

Para poder juntar las bases de Simce y del mapa nutricional es necesario contar con un identificador único por alumno que sea capaz de relacionar ambas bases. Si bien el mapa nutricional cuenta con un identificador que diferencia a los individuos dentro de esta base, no es un identificador válido para pegarlo con otras bases. Con esto en mente, se creó un identificador que cumpliera con estas dos características, ser único y relacionable a otras bases, este identificador se compone de la fecha de nacimiento, el género y el código del establecimiento al cual asiste el alumno, es decir: Id= "Fecha de Nacimiento + Sexo + Código RBD". No se utilizó el curso, ya que no se pudo obtener esta información, lo que finalmente repercutió en una mayor pérdida de información³⁹.

Una vez creado el identificador en las bases del mapa nutricional, se debería crear el mismo identificador en las bases SIMCE. Sin embargo no se cuenta con la fecha de nacimiento de los alumnos, por lo que fue necesario utilizar una tercera base de datos para así poder conectar estas dos bases de datos, estas fueron las bases RECH. De esta forma, se combinó en primera instancia las bases del mapa nutricional con las bases RECH y posteriormente estas fueron combinadas con las bases Simce. El gráfico 4 muestra de manera gráfica este proceso.

Descrito el proceso mediante el cual se combinarán los datos hay que destacar dos puntos: El primero hace referencia al filtro geográfico que se debe aplicar a la información, es decir, se deberá acotar la información de los estudiantes a aquellos que se encuentren dentro de alguna de las 29 comunas del estudio, quedando el resto fuera. El segundo punto es que debe garantizar que los identificadores sean únicos en cada base para que al cruzar la información entre las bases, no se estén asociando datos a individuos que no corresponden. Para esto hubo que eliminar los datos duplicados o recuperar los originales en el caso que fuese posible⁴⁰.

Gráfico 4: Proceso para combinar las bases de información.



Fuente: Elaboración propia.

³⁹ En colegios como el Instituto Nacional se pierde más información que en otros colegio debido a la gran cantidad de alumnos que existe por nivel educativo.

⁴⁰ Para un mayor detalle de este proceso revisar Anexo I.

Tanto el proceso de limpieza de información duplicada como el de posterior combinación de bases de datos disminuyen la muestra original y por ende la muestra final. La cantidad de observaciones que se pierde luego de cada proceso de limpieza o combinación de datos es resumida en las tablas 6 y 7, donde se debe recordar que se estudiarán dos combinaciones de datos.

Tabla 6: Número de datos a través del proceso de combinación de bases

Datos	RECH		Mapa Nutricional		Simce	
	2012.m	2011.m	2012.m	2011.m	2013.2m	2012.2m
Iniciales	103.786	107.179	65.410	47.462	53.324	56.456
Duplicados	21.942	22.385	10.731	7.954	331	872
Finales	81.844	84.794	54.679	39.508	52.993	55.584

Fuente: Elaboración propia en base a datos de RECH, Mapa Nutricional y Simce. Nomenclatura: M: media

Tabla 7: Número de datos a través del proceso de combinación de bases

Datos	RECH+MN		RECH+MN+Simce	
	2012.m	2011.m	2012.m	2011.m
Iniciales	54.679	39.508	53.474	38.131
No Pegados	1.205	1.377	19.461	14.227
Finales	53.474	38.131	34.013	23.904

Fuente: Elaboración propia en base a datos de RECH, Mapa Nutricional y Simce. MN es la base principal para realizar el proceso de combinación de datos, por lo que corresponderá al número inicial de datos. Esto se explica debido a que sólo para esos alumnos se tendrá información de peso y altura, por lo que tener más información disponible en las bases RECH y Simce no agregará un valor extra.

Al analizar las tablas se observa que más de la mitad de los datos iniciales se pierden luego de realizar todo el proceso de combinación de bases, disminuyendo la validez externa del estudio.

Por último y si bien con la información que se tiene hasta el momento no hay ningún problema para estimar la relación entre IMC y rendimiento académico, lo que será hecho, existe información que no se encuentra incluida en esta base de datos que sí puede ser obtenida de otras fuentes de información. Para esto se incorporará dos bases de datos adicionales, la primera base corresponde a los cuestionarios de padres de los alumnos que rinden el Simce, que entregará información acerca de controles familiares, como educación de los padres o ingreso del hogar. Y por otro lado, bases de puntajes Simce pasados, que servirá para controlar por la historia académica del alumno. Sin embargo, incluir estas bases de datos tiene asociado un costo, el cual queda plasmado en las tablas 8 y 9 donde se muestra la cantidad de información que se pierde al incluir estos nuevos controles.

Tabla 8: Número de datos a través del proceso de combinación de bases

Datos	RECH+MN+Simce+Simce Pasado	
	2012.m	2011.m
Fuente Actual	2012.m	2011.m
Fuente Pasado	2007.4b	2006.4b
Iniciales	34.013	23.904
No Pegados	9.518	7.521
Finales	22.459	16.383

Fuente: Elaboración propia en base a datos de RECH, Mapa Nutricional y Simce. Para mayor detalle acerca de los cruces de información pasada, revisar Anexo A.

Tabla 9: Número de datos a través del proceso de combinación de bases

	RECH+MN+Simce+Simce Pasado+ Controles Familiares	
Fuente Actual	2012.m	2011.m
Fuente Pasado	2007.4b	2006.4b
Datos		
Iniciales	22.459	16.383
No Pegados	10.138	9.466
Finales	12.321	6.917

Fuente: Elaboración propia en base a datos de RECH, Mapa Nutricional, Simce y encuesta de padres y apoderados.

Finalmente las muestras son reducidas en más de la mitad de las observaciones que se registraban en la tabla 7, por lo que al momento de estimar los modelos utilizando estos datos habrá que tener cuidado con las conclusiones que se pueden obtener. No obstante, controlar por más variables permitirá verificar la robustez de los resultados de las estimaciones.

5.- Resultados

Como se indicó en el apartado metodológico existen diferentes especificaciones a ser estimadas. En primer lugar se mostrarán los resultados de matemáticas y posteriormente los de lectura, del modelo lineal o principal. Recordemos que los controles familiares disminuyen el tamaño de la muestra, razón por lo cual se estimaron modelos con y sin estos controles. Finalmente se revisarán aquellos modelos que tratan la no linealidad en la relación a estimar.

5.1 Modelo Principal

La tabla 10 corresponde a las estimaciones de matemáticas vía MCO para ambas muestras, con y sin controles familiares.

Tabla 10: Resultados estimaciones MCO puntajes Simce matemáticas

Variables	2012	2011	2012	2011
IMC	-0,005*	-0,006*	-0,003	-0,007*
Sexo	-0,087*	-0,078*	-0,09*	-0,064*
Ptje_Mate2 (4to básico)	0,603*	0,619*	0,603*	0,617*
Dummy_NSE2	0,146*	0,066	0,152*	0,165
Dummy_NSE3	0,332*	0,204*	0,332*	0,305*
Dummy_NSE4	0,607*	0,495*	0,647*	0,615*
Dummy_NSE5	0,829*	0,963*	0,869*	1,375*
Dependencia_Adm	0,1*	0,156*	0,093*	0,156*
Edad	-0,005*	-0,008*	-0,005*	-0,007*
Educ_Padre			-0,001	-0,01
Educ_Madre			0,001	0,001
Ingreso_h2			0,036*	-0,036
Ingreso_h3			0,051*	-0,013
Ingreso_h4			0,02	0,018
Ingreso_h5			0,072*	-0,053
Constante	0,4962	1,165*	0,427*	0,958*
N° Observaciones	22.459	16.299	11.069	6.215

Fuente: Elaboración propia en base a datos Mapa Nutricional, Simce y RECH. * P-value<0,05. ** P-value<0,1 Los errores corresponden a clusters por curso.

Antes de analizar las estimaciones MCO, se destaca que incluir los controles familiares no cambia en demasía las estimaciones, de hecho se mantienen los signos y la significancia de casi todas las relaciones. La diferencia ocurre para la muestra del año 2011, donde la relación entre IMC y rendimiento académico (RA) ya no es significativa.

Dicho esto, vemos que existe un efecto negativo y significativo del IMC sobre el RA que va desde 0,005 hasta 0,007 desviaciones estándar (DE) para tres de los cuatro casos, es decir, entre 0,325 y 0,455 puntos del Simce. Dicho de otra forma, si tenemos un IMC de 22 y este aumenta a 23, el puntaje Simce del alumno en cuestión debiese reducirse máximo en 0,455 puntos en su prueba de matemática. La variable sexo también está relacionada negativa y significativamente con el RA, es decir, ser mujer reducirá los puntajes entre 0,06 y 0,09 DE, es decir, entre 3,9 y 5,85 puntos del Simce.

Por otro lado, al controlar por puntajes Simce de 4to básico se encontró una relación positiva y significativa y no cambia si es que se incluyen los controles familiares. Además, es una de las variables que más explica el desempeño actual del alumno, alcanzando valores de 0,67 DE, es decir, 43,55 puntos.

Al analizar el nivel socioeconómico del colegio se observa una relación positiva, significativa y creciente de este con los puntajes Simce, la cual comienza con valores entorno a las 0,01 DE y alcanza máximos de 0,9 DE. Esto tiene sentido al considerar que a mayor cantidad de recursos disponibles en los establecimientos, mayor será la infraestructura, profesores o materiales a los que pueden acceder estos establecimientos, para así potenciar el RA de sus alumnos.

La variable dependencia administrativa se relaciona positiva y significativamente con el puntaje, es decir, los colegios particulares subvencionados en promedio lo hacen mejor que los municipales⁴¹. Edad se relacionada negativamente y de manera significativa.

Con respecto a la educación de los padres, no se observa una relación significativa con el RA. Además, si bien la muestra del año 2012 tiene relaciones significativas en términos de los ingresos del hogar, no ocurre lo mismo con la muestra del año 2011. De hecho, no se registra un patrón único en sus signos, por lo que no es conclusiva esta evidencia.

Si bien estos resultados permiten entender en parte la relación entre obesidad y rendimiento académico, es de suma importancia corregir el problema de endogeneidad presente en la estimación, ya que de lo contrario los resultados serían sesgados. Por este motivo es que se utilizó la metodología de variable instrumental, donde al instrumentalizar la obesidad, en este caso el IMC de los individuos, se logra corregir el problema anteriormente mencionado y de esta forma estimar correctamente la relación planteada. En la tabla 11 se presentan los resultados de las estimaciones de matemáticas vía variable instrumental, con y sin controles familiares:

Tabla 11: Resultados e estimaciones IV puntajes Simce matemáticas

Variables	IV-2012	IV-2011	IV-CF-2012	IV-CF-2011
IMC	-0,402*	-0,282*	-0,449*	-0,49*
Log_Niños	0,106*	0,112*	0,096*	0,106*
F-test	20,86	16,43	8,6	5,87
N° Observaciones	20.737	15.076	10.188	5.921

Fuente: Elaboración propia en base a datos Mapa Nutricional, Simce y RECH. * p-value < 0,05. ** p-value < 0,1 Los errores corresponden a clusters por curso. El test F corresponde al test Kleibergen-Paap rk Wald F statistic. Log_Niños corresponde a la variable instrumental.

⁴¹ Recordemos que los colegios particulares pagados están fuera de la muestra debido a que MN no los encuesta.

A continuación analizamos la primera etapa de la estimación por variable instrumental, la cual tiene por objetivo verificar o dar indicios de la fuerza del instrumento. En los modelos sin controles familiares, ambos instrumentos son considerados fuertes, ya que presentan un test F superior a 16,38, pese a que el modelo para el año 2011 pase por muy poco la prueba. Sin embargo, al momento de incluir los controles familiares vemos que estos valores se reducen drásticamente, provocando que los instrumentos pasen a ser débiles. De todas formas e independiente del valor del test F y de los controles a utilizar, el instrumento se relaciona positiva y significativamente con el IMC del individuo. Así, considerando constante la matrícula de cada establecimiento, la apertura de un nuevo local de comida chatarra provoca un aumento promedio de un 10% en el IMC de los individuos.

Dicho esto, se analizará la segunda etapa o los resultados de estimar vía variable instrumental el efecto del IMC sobre el rendimiento académico, donde se puede ver que este efecto es significativo y negativo, corroborando nuestra hipótesis inicial. Además, refleja que esta relación se encontraba subestimada y entrega valores entre 0,28 y 0,49 DE, es decir, entre 18,2 y 31,9 puntos de la prueba Simce de matemáticas. Se debe destacar que al controlar por las variables familiares, el instrumento pierde fuerza y esto debería aumentar el valor de los coeficientes estimados vía VI⁴², producto del sesgo en las estimaciones de VI que esto produciría, cuestión que efectivamente ocurre para el año 2011. Sin embargo, esto no ocurre para el año 2012 y también se puede observar que el aumento en magnitud del parámetro estimado para el año 2011 no está tan alejado de este mismo parámetro para el año 2012. Esto nos permite entonces verificar la robusticidad del instrumento ante la presencia de más variables explicativas, las cuales no alteran el signo y la significancia de la relación. Por lo mismo, el hecho de que el instrumento pierda fuerza no es un gran problema.

Revisados los resultados asociados a matemáticas, se procederá a revisar en la tabla 12 y 13 los puntajes Simce de lectura, con y sin controles de factores familiares.

Tabla 12: Resultados estimaciones MCO puntajes Simce lectura

Variables	MCO-2012	MCO-2011	MCO 2012	MCO 2011
IMC	-0,003**	-0,005*	-0,001	-0,008*
Sexo	0,137*	0,142*	0,137*	0,175*
Ptje_Lect2 (4to.b)	0,592*	0,58*	0,597*	0,569*
Dummy_NSE2	0,089	0,121	0,031	0,142
Dummy_NSE3	0,201*	0,236*	0,139**	0,266*
Dummy_NSE4	0,375*	0,465*	0,329*	0,493*
Dummy_NSE5	0,507*	0,765*	0,467*	0,95*
Dependencia_Adm	0,047*	0,075*	0,026	0,088*
Edad	-0,003*	-0,008*	-0,003*	-0,007*
Educ_Padre			-0,002	-0,002
Educ_Madre			0,004	-0,001
Ingreso_h2			0,026	-0,024
Ingreso_h3			0,029	0,009
Ingreso_h4			-0,011	-0,005
Ingreso_h5			0,021	0,032
Constante	0,275*	1,063	0,21	1,001*
N° Observaciones	22.459	16.383	11.069	6.241

Fuente: Elaboración propia en base a datos Mapa Nutricional, Simce y RECH. * p-value<0,05. ** p-value<0,1 Los errores corresponden a clusters por curso.

⁴² Stock y Yogo (2005)

Al igual que en las estimaciones de matemáticas, se encuentra una relación negativa entre IMC y RA al estimar vía MCO en los modelos sin controles familiares. Al controlar por las variables familiares la relación negativa se mantiene, pero pasa a no ser significativa para la muestra del año 2012. El resto de las variables de control del modelo sin controles familiares son significativas, tienen el mismo signo y tienen magnitudes similares a sus pares de matemáticas, con la diferencia de que sexo cambia de signo. Es decir, las mujeres tendrán en promedio un mayor puntaje que los hombres, lo cual también es significativo. Por el lado de los modelos con controles familiares, conservan la misma dinámica y al igual que el caso de matemáticas las variables familiares son no significativas.

La tabla 13 presenta los resultados de las estimaciones de lectura vía variable instrumental, con y sin controles familiares:

Tabla 13: Resultados estimaciones IV puntajes Simce lectura

Variabes	IV-2012	IV-2011	IV-2012	IV -2011
IMC	-0,091	-0,305*	-0,085	-0,617*
Log_Niños	0,107*	0,113*	0,096*	0,104*
F-test	20,96	16,63	8,68	5,71
N° Observaciones	20.737	15.156	10.188	5.947

*Fuente: Elaboración propia en base a datos Mapa Nutricional, Simce y RECH. * P-value<0,05. ** P-value<0,1 Los errores corresponden a clusters por curso. El test F corresponde al test Kleibergen-Paap rk Wald F statistic. Log_Niños corresponde a la variable instrumental.*

La primera etapa de la estimación por variable instrumental, es muy parecida a la de matemáticas, ya que los test F varían muy poco entre sí, siendo en tres de los cuatro casos planteados ligeramente superiores para estos modelos. Además, al igual que antes, considerando constante la matrícula de cada establecimiento, la apertura de un nuevo local de comida chatarra provoca un aumento promedio de un 10% en el IMC de los individuos.

El efecto del IMC sobre el rendimiento es significativo para las muestras del año 2011, pero no así para el año 2012, además difiere bastante en magnitud entre los casos significativos y no. Sin embargo, el caso significativo es precisamente el que tiene la menor cantidad de observaciones, por lo que no es suficientemente conclusivo para indicar que existe una relación negativa y significativa entre ambas

Por último, se debe destacar que el efecto encontrado sobre el rendimiento académico, tanto en matemática como en lectura, es superior al de grandes políticas educativas, como la Ley SEP o la JEC, lo que sugeriría que sólo enfocándose en reducir el IMC de los alumnos aumentaría su desempeño. La implementación de la Subvención Escolar Preferencial (SEP) en los colegios particulares subvencionados aumentó entre 3 y 9 puntos las pruebas Simce de matemáticas y unos 5,5 puntos las pruebas de lectura, ambas para cuarto básico, según estudios realizados por el Centro de Estudios del Ministerio de Educación⁴³. Por otro lado, Bellei (2009) encuentra que la Jornada Escolar Completa tiene un efecto positivo entre 0,05 y 0,07 desviaciones estándar para la prueba de lectura y entre 0 y 0,12 desviaciones estándar para matemática, siendo todos los valores significativos y obteniendo los valores máximos para estudiantes que asisten a colegios rurales y públicos⁴⁴.

⁴³ (30 de Agosto de 2012) Serie Evidencias: Impacto de la Ley SEP en Simce: Una mirada a 4 años de su implementación

⁴⁴ No se entrega el valor de la desviación estándar, sin embargo, se nota que el impacto es menor que el encontrado en nuestro estudio.

Una de las razones que puede explicar este fenómeno es la calidad de la base de datos a utilizar. La muestra final pierde más de un 50% de las observaciones iniciales, por lo que el grupo que se está identificando puede ser uno en donde la obesidad afecte de una manera mucho mayor al que se encontraría si es que se realizase el mismo estudio para toda la población. Dicho de otra forma, la variable instrumental sólo afecta a un subgrupo de la población, o lo que es lo mismo esta sólo "trata" a un grupo llamado *compliers*, por lo que en la práctica se está obteniendo un efecto local promedio del tratamiento (LATE por sus siglas en inglés), el cual no necesariamente es extrapolable a toda la población y que de hecho permitiría entender los valores tan elevados en comparación a otras variables. Sin embargo, y pese a que existe la posibilidad de que el grupo que se ha encontrado sea muy específico, otra puede ser la explicación para estos resultados. Esta hace referencia a la fuerza de la variable instrumental, la cual no sería lo suficientemente fuerte, pese a cumplir con los test estadísticos. Esto afectaría los coeficientes estimados, en particular, los sobreestimaría, lo que justificaría los elevados resultados en comparación al resto de las políticas.

Independiente de cuál sea la razón, los resultados deben ser analizados con cuidado y sólo futuras investigaciones, que utilicen una base de datos más completa, permitirá llegar a una conclusión con respecto a este tema.

5.2 Modelos no lineales.

En la Tabla 14 se pueden revisar sólo los resultados de las estimaciones vía variable instrumental de los modelos no lineales para matemáticas. Se debe destacar que los puntajes de lectura no aparecen debido a que no presentaron resultados significativos. Además, no se mostrará la primera etapa de la estimación ya que se mantiene la lógica de los resultados del modelo principal.

Tabla 14: Resultados estimaciones no lineales - Simce matemáticas

Variables	IMC	Log_Niños	F-test	N° Observaciones
Controles Familiares- IV-2012				
Bajo Peso	-2,07	0,019	0,56	1.334
Normal	-0,861**	0,04*	4,17	7.289
Sobrepeso	-2,52	0,035	1,03	1.322
Obeso	-0,91	0,089	0,48	243
Sin Controles Familiares- IV-2012				
Bajo Peso	-3,55	0,007	0,135	2.648
Normal	-1,56**	0,026*	3,5	14.831
Sobrepeso	-3,65	0,02	0,76	2.728
Obeso	-0,71	0,06	0,465	530

*Fuente: Elaboración propia en base a datos Mapa Nutricional, Simce y RECH. * p-value<0,05. ** p-value<0,1 Los errores corresponden a clusters por curso. El test F corresponde al test Kleibergen-Paap rk Wald F statistic. Log_Niños corresponde a la variable instrumental.*

Luego de corregir la endogeneidad y estimar los modelos no lineales se ve una relación negativa y significativa entre IMC y RA, la cual se encuentra entre 0,86 y 1,56 DE para matemáticas, valores muy elevados. Dos aspectos deben ser destacados: El primero es que las relaciones son significativas al 10% de significancia y lo segundo es que se está empleando un instrumento débil. Esto repercutirá en los coeficientes estimados, ya que reflejarán en parte el sesgo de la estimación de variable instrumental, que en este caso en particular tenderá a aumentar

el valor estimado y por lo tanto sobreestimar esta corrección, razón por la cual se deben analizar con cuidado estos resultados⁴⁵.

Para finalizar esta sección destacaremos los siguientes puntos: En primera instancia, se encontró una relación negativa entre IMC y rendimiento académico tanto para los puntajes de matemáticas y de lectura, luego de corregir la endogeneidad inherente a la estimación. Sin embargo, el resultado de lectura debe ser tratado con más cuidado ya que sólo se registró este efecto para uno de los dos años estimados.

En segundo lugar, al momento de estimar los modelos no lineales se encontró que los extremos de la distribución de IMC (desnutrición y obesidad) no presentan una relación clara con el rendimiento académico de los individuos, cuestión que también ocurre con la condición de sobrepeso. Sin embargo, fue en la categoría IMC normal en la que se encontró una relación negativa y significativa. Esta evidencia permite postular las siguientes dos hipótesis:

- Por un lado, la penalización en rendimiento académico no se encontraría en individuos obesos o con sobrepeso, sino que se vería en aquellos individuos que pasan de tener un peso normal a ganar un par de kilos o incluso a cambiarse de categoría de IMC. No es claro el canal mediante el cual afecte el rendimiento académico y si bien diversas teorías pueden ser postuladas, el paper no tiene por objetivo esto, por lo que será un tema que se deberá seguir estudiando.
- Por otro lado, la relación entre IMC y rendimiento académico no responde necesariamente a las clasificaciones médicas de obesidad y sobrepeso. Es decir, el efecto negativo en rendimiento académico comienza a verse en niveles de IMC menores a los que definen que un alumno tenga sobrepeso, apoyando la hipótesis anterior.

Independiente de si estas hipótesis son o no correctas, es evidente que los modelos no lineales entregan información relevante al momento de estimar la relación entre IMC y rendimiento académico. El potencial de estos modelos aún no es alcanzado en su totalidad debido al problema de pérdida de información que se comentó en la sección de datos, cuestión que se espera mejorar en investigaciones futuras. Sin embargo, la relación negativa entre IMC y RA efectivamente es corroborada a través de ambos modelos.

6.- Conclusiones y Discusión

A lo largo del presente trabajo se estudió la relación entre rendimiento académico y obesidad. En primer lugar se tuvo que construir una única base de datos a partir del Mapa Nutricional, RECH y Simce, ya que estas por separado no contenían la información suficiente para realizar el estudio. Sin embargo, la construcción de esta base se vio afectada por la inexistencia de un identificador que permitiese relacionar todas estas bases. Para solucionar este problema es que creó un nuevo identificador, que si bien permitió reunir toda la información recayó en una pérdida importante de información.

Pese a que la literatura no ha alcanzado un consenso en termino de resultados, sí existe una concordancia con respecto al problema de endogeneidad presenta en la estimación. Nuestro trabajo es un aporte a la literatura debido a la utilización de una variable instrumental inédita en esta, la cual corresponde a la *densidad de locales de comida chatarra entorno a un colegio en un*

⁴⁵ Stock y Yogo (2005) investigan los problemas de los instrumentos débiles.

determinado momento del tiempo. La lógica detrás de esta variable es que una mayor oferta de alimentos chatarra entorno a un colegio, aumenta el acceso y por tanto la probabilidad de consumo de este tipo de alimento por parte de los alumnos de este establecimiento, recayendo finalmente en un aumento de IMC de estos. Lo interesante de este instrumento es que evidentemente está relacionada con la variable endógena de la estimación (obesidad) y que no se relaciona con factores no observables del individuo, ya que la instalación del local de comida no tiene que ver a priori con las habilidades o aspectos psicológicos de los alumnos, lo cual permitirá romper la endogeneidad y así estimar correctamente la relación.

Dicho esto, se estimaron modelos lineales entre IMC y rendimiento académico utilizando la metodología de variable instrumental. En primera instancia se encontró una relación negativa y significativa entre IMC y rendimiento académico en los puntajes Simce de matemáticas que toma valores entre 0,28 y 0,49 desviaciones estándar, es decir, entre 18,2 y 31,9 puntos de la prueba Simce. Con respecto a los puntajes de lectura no se encontró una relación significativa en dos de los cuatro casos y los casos significativos si bien mostraban una relación negativa entre IMC y rendimiento académico, diferían en casi el doble los coeficientes estimados, por lo que no son conclusivos los resultados. Sin embargo, los valores encontrados son mucho más elevados que los encontrados por el resto de la literatura de rendimiento académico, sea tanto en diferentes políticas educativas u otras variables, por lo cual deben ser tomados con cuidado. Explicaciones para este suceso hacen referencia a la fuerza del instrumento y a la calidad de la base de datos utilizada.

Posteriormente se estimaron modelos no lineales entre IMC y rendimiento académico encontrando que los extremos de la distribución de IMC (desnutrición y obesidad) y también la condición de sobrepeso no presentaban una clara relación con el rendimiento académico de los individuos, tanto para matemáticas como para lectura. Así fue sólo en la categoría de IMC normal en la que se encontró una relación negativa y significativa que osciló entre 0.8 y 1.6 desviaciones estándar, valores muy elevados con respecto a las estimaciones pasadas, probablemente porque el instrumento perdió fuerza y porque las muestras fueron restrictas. Independiente de este problema, plantearemos las siguientes dos teorías: La primera es que la penalización en rendimiento académico se da en aquellos individuos que pasan de tener un peso normal a tener sobrepeso o un par más de kilos, los que cambian notoriamente su aspecto físico. En segundo lugar, que las clasificaciones médicas de obesidad y sobrepeso no estarían capturando del todo la relación negativa entre IMC y rendimiento académico.

Otro punto en cuestión son las limitaciones y desafíos futuros de nuestro trabajo. Como se mencionó con anterioridad, debido a la inexistencia de un identificador que permita unir todas las bases de datos, se perdió mucha información, lo que le quita validez externa al estudio. A futuro se espera contar con este identificador lo que recaería en que la pérdida de información fuese mínima y por ende que el estudio obtenga resultados aún más representativos a nivel de la RM.

Para identificar los puntos de comida chatarra se utilizó información de patentes municipales, sin embargo más de alguna comuna presentó problemas para obtener la información necesaria para realizar nuestro trabajo, sea por falta de fechas de apertura, formato de calles, abreviaciones en calles o simplemente porque no se pudo adquirir la información. A futuro se utilizará información del Servicio de Impuestos Internos (SII), que además de ser más estandarizada debiese tener información de comunas a las cuales no pudimos geo-codificar.

Por otro lado, la medida de obesidad utilizada, IMC, no es la mejor para medir esta condición según diferentes autores, entre ellos Ivanovic (2013). El problema de esta medida es que no logra diferenciar musculatura de tejido graso, por lo que si bien ambas aumentan el peso del

individuo, el porqué lo aumentan es muy diferente. En este sentido se recomienda utilizar por ejemplo el tamaño de la circunferencia de la cintura o el porcentaje de grasa del cuerpo, información que no está disponible en el Mapa Nutricional de JUNAEB. El Simce de Educación Física sí dispone de esta primera medida sugerida, pero el número de observaciones con las que se cuenta no es muy elevado, por lo que a futuro este es un punto a seguir explorando, sea por la inclusión de una mejor medida de obesidad por parte de JUNAEB o una mayor exploración de las bases ya existentes.

Finalmente se harán recomendaciones de política para combatir la obesidad. En primer lugar se sugiere la implementación de un impuesto a las bebidas gaseosas al igual que cómo lo hizo México, país con los índices de obesidad infantil más elevados del mundo⁴⁶. En este país, a principios de 2014 se introdujo un impuesto a la venta de un 10% a las bebidas azucaradas, registrando un año después de su aplicación una reducción en las ventas de estas de un 12% y un aumento en las bebidas sin impuesto (agua embotellada) de un 4%⁴⁷. Si bien en este país ha dado resultado esta política, es necesario realizar los estudios pertinentes para la aplicación de este impuesto en nuestro país, pero considerando que somos el tercer país con mayor consumo de bebidas.

En segundo lugar, es claro que innumerables políticas contra la obesidad se han desarrollado e implementado en todo el mundo. Si bien no se recomendará una política en particular, sí se mencionaran los elementos que la política desarrollada debe tener para ser efectiva y que por lo tanto permita reducir los índices de obesidad y sobrepeso con el paso del tiempo. En palabras simples, las calorías consumidas deben disminuir y las calorías gastadas deben aumentar.

Para disminuir el consumo de calorías en primer lugar es necesario que la educación sanitaria o en materia de alimentos mejore, sea clara y que tenga una cobertura global de la población. La idea detrás de esta medida es que la población sepa qué es lo que debería consumir. Sin embargo, no basta con saber qué es lo que deberían comer, si es que al momento de adquirir un alimento no es clara la información nutricional de este, aspecto que también debiese ser regulado. Además, el acceso a los alimentos es desigual a través de la población, por lo que es esencial el rol del estado a través de JUNAEB para diseñar y entregar los alimentos necesarios para los alumnos de menores recursos.

Para aumentar el gasto de calorías se debe incentivar la actividad física de la población. Algunas medidas simples pueden ser tomadas por parte de la población, como ocupar medios de transporte no motorizados, como bicicleta o caminar y realizar actividad física por su cuenta, como salir a trotar o practicar algún deporte. En este sentido el Estado deberá desarrollar políticas para incentivar estos comportamientos, por un lado aumentando la infraestructura existente, como por ejemplo: creación de más ciclovías o centros deportivos municipales o también realizando eventos deportivos como corridas o cicletadas.

Con respecto a los establecimientos educacionales, es necesario que las clases de educación física sean más efectivas⁴⁸ y que su frecuencia aumente, cuestión que ya ha empezado a ocurrir desde el año 2013⁴⁹. Además se recomienda la existencia de una oferta de actividades

⁴⁶Noticiero, Televisa (2015) "México encabeza estadística en obesidad , en adultos y niños"

⁴⁷Medline Plus (2016) "Los impuestos sobre los refrescos redujeron el consumo en México"

⁴⁸ Disminuir lo más posible los tiempos "muertos" como por ejemplo el de duchas y cambio de ropa.

⁴⁹ Actualidad Deportes Noticia (ADN, 2013) "Schmidt: Este año se incrementará a 4 horas obligatorias de educación física para todos los niños" <http://www.adnradio.cl/noticias/nacional/schmidt-este-ano-se-incrementara-a-4-horas-obligatorias-de-educacion-fisica-para->

extra programáticas deportivas o que incentiven la actividad física que sea lo más variada posible por parte de los colegios, para así cubrir la mayor cantidad de gustos de los alumnos en cuestión.

Cabe destacar que la aplicación de la llamada *Ley Super 8* o 20.606⁵⁰ va en el sentido correcto para combatir los altos índices de obesidad y sobrepeso infantil. A grandes rasgos se implementará un nuevo etiquetado de alimentos que permitirá diferenciar a aquellos que tienen "excesos de" sal, grasa, entre otros, de aquellos que no. Estos alimentos no podrán ser comercializados al interior de los colegios, en específico en los quioscos y finalmente no se podrá realizar publicidad de estos alimentos a menores de 14 años.

Por último, la evidencia internacional y el ejemplo de programas desarrollados en otros países como Dinamarca⁵¹ indican que es necesario que la aplicación de las políticas sea de manera integral y que se enfoquen en la familia, más que en los alumnos. La razón es simple, los patrones de conducta y los hábitos de alimentación son aprendidos mayoritariamente de los padres o figuras paternas cercanas, por lo que si estos no cambian, es muy difícil que los cambios se materialicen en sus hijos.

Anexos

Anexo A: Combinaciones factibles entre Simce y Mapa Nutricional

A continuación se presenta una tabla con todas las combinaciones posibles entre las bases del Mapa Nutricional y las pruebas Simce a lo largo de los años. Como se indicó en la sección de Datos, el Simce es rendido por diferentes niveles educacionales en diferentes años, razón por la cual no se tiene la misma información para todos los años.

Tabla 15: Combinaciones factibles entre Simce y Mapa Nutricional

Mapa Nutricional / Simce	2006	2007	2008	2009	2010	2011	2012	2013	Match Posibles	Match Utilizados
Mapa Media 2013			4b						1	0
Mapa Media 2012		4b				8b		2m	3	1
Mapa Media 2011	4b						2m		2	1
Mapa Media 2009					2m				1	0

Nota: Elaboración propia según bases de datos disponibles para el estudio. Negro= Match utilizados; Gris= Match no utilizados debido a que el desfase temporal entre la medición de IMC y puntaje SIMCE es muy extensa; Negro no destacado= Matches no utilizados debido a que IMC fue medido después del puntaje, por lo que no se podría suponer a simple vista que ese es el IMC del individuo antes de rendir la prueba. 2b= Segundo básico; 4b=Cuarto básico; 6b= Sexto básico, 8b= Octavo básico y 2m= Segundo medio.

Anexo B: Definiciones de Comida Chatarra

Considerando la definición de comida chatarra propuesta por Jacobson, definiremos dos tipos. Es necesario aclarar que la forma de determinar si una patente es o no un local de comida chatarra es a través de su giro, actividad a la cual se dedica la empresa o negocio. Debido a que

todos-los-ninos/20130802/nota/1944116.aspx y La Tercera (2014) "Ministra de salud pide subir a seis las horas de educación física para frenar obesidad" <http://www.latercera.com/noticia/nacional/2014/09/680-595085-9-ministra-de-salud-pide-subir-a-seis-las-horas-de-educacion-fisica-para-frenar.shtml>

⁵⁰ Biblioteca de Congreso Nacional, Ley N°20.606 <http://www.leychile.cl/Navegar?idNorma=1041570>

⁵¹ BBC Mundo (2014) "El secreto danés para combatir la obesidad infantil" http://www.bbc.com/mundo/noticias/2014/11/141108_salud_obesidad_infantil_danese_secreto_lav

las municipalidades no tienen una descripción única para los giros, por ejemplo la venta de frituras puede estar representada en el giro como: fritura, ftas, fritas, etc, es que buscaremos palabras claves dentro de los diferentes giros en nuestro poder. Además, mientras mayor sea la glosa del giro de una patente, más pequeñas serán las descripciones de las actividades que realizan, por lo que muchas veces se ocuparan abreviaciones de las palabras claves a utilizar, cuestión que también fue considerado en nuestro estudio. De esta forma, las dos definiciones de comida chatarra serán:

Definición de Comida Chatarra 1:

Pizza, fritura, bebida, gaseosa, sopaipilla, empanada, pastel, asadera, masa dulce, helado, completos, hotdog, cecina, sándwich, amasandería, fuente de soda, confites, emparedado, cafetería, comida rápida, panadería, fiambrería, caramelos, abarrotes, gelatería, comida preparada, hipermercados, supermercado, minimarket, confitería, pastelería, pan, ave, ass, lomito, churrasco, hamburguesa, chorrillana, queque, rotisería, restaurant, tratoría, masas, bazar, buffer, torta, pollo asado, pollería, pulpería, tostaduría, papas fritas, tacos, bocadillos, muffins, comestibles, embutido, dulce, lácteo, expendio de refresco, snack, caramelo, chocolate, manjar, apanado, envasados, encurtido, aves faenadas, kiosko, horneados y comida china.

Definición de Comida Chatarra2:

Pizza, fritura, bebida, gaseosa, sopaipilla, empanada, helado, completo, hotdog, sándwich, fuente de soda, comida rápida, gelatería, comida preparada, comida china, ass, lomito, churrasco, hamburguesa, chorrillana, restaurant, pollo asado, papas frutas, tacos, refresco y snack.

La primera definición es más general que la segunda, esto es porque lo que se buscó con esta segunda definición es reflejar una oferta de alimentos chatarra relacionados lo más cercanamente a la comida rápida, por lo que a simple vista esperaríamos que los efectos estimados utilizando esta segunda definición sean mayores que los primeros. El punto en contra de la segunda definición es que al ser más acotada que la primera, identificará una menor cantidad de locales de comida, disminuyendo tanto el promedio de locales por colegio como el número de colegios con oferta de alimentos cercana a sus establecimientos.

Anexo C: Temas relacionados a patentes municipales

En la tabla a continuación se observa la distribución del tipo de patentes municipales. A partir de esta tabla se puede ver el total de patentes de cada tipo, su participación en el total de patentes y además cuántas de estas cumplen con la definición de Chatarra 1 o Chatarra 2, las cuales fueron revisadas en el anexo anterior. Cabe destacar que los porcentajes presentados son con respecto al total de cada categoría.

Tabla 16: Distribución de patentes comunales según tipo

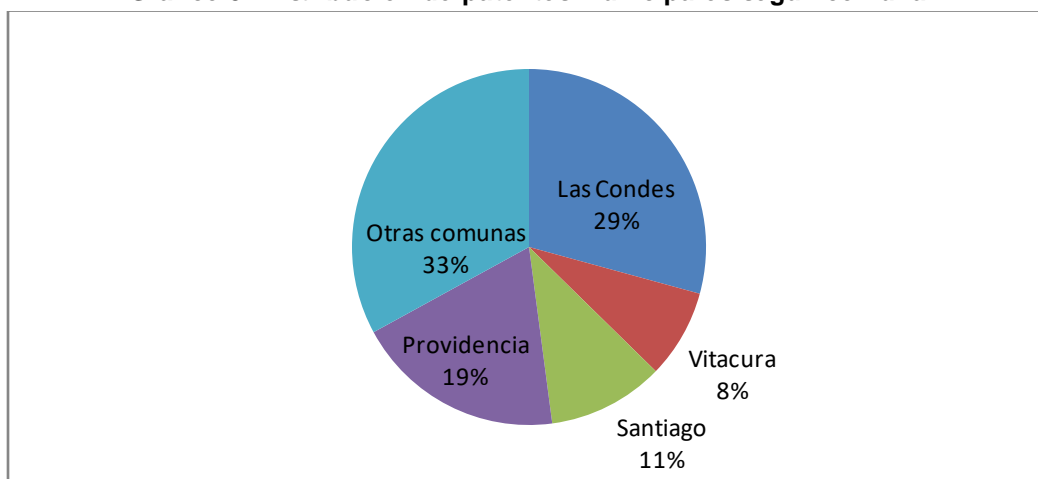
Tipo	Total	% del total	Chatarra_1	% del total	Chatarra_2	% del total
Alcohol	2.859	1,08%	2.226	5,63%	1.878	11,37%
Comercial	202.675	76,63%	27.273	68,98%	11.649	70,55%
Industrial	3.863	1,46%	691	1,75%	155	0,94%
Feria Libre	3.985	1,51%	1.223	3,09%	53	0,32%
Kiosco	382	0,14%	213	0,54%	134	0,81%
Microempresas	7.091	2,68%	3.409	8,62%	1.341	8,12%
Profesionales	18.507	7%	0	0	0	0

No aparece Glosa	8.785	3,32%	3.576	9,04%	890	5,39%
Otros	16.334	6,18%	927	2,34%	411	2,49%
	264.481		39.538		16.511	

Nota: Elaboración propia en base a patentes comunales. Tanto patentes comerciales como industriales incluyen las patentes definitivas y provisorias. Microempresas pueden ser de carácter familiar, comercial o industrial. En la categoría Otros podemos encontrar: patentes ambulantes, carros estacionados, domicilio postal tributario, servicios, MEF, etc. Fuente: Elaboración propia en base a información de patentes municipales.

No sólo es importante ver cómo se distribuyen las patentes por su tipo, sino también cómo se distribuyen a través de las comunas. Esto es lo que se busca representar a través del siguiente gráfico:

Gráfico 5: Distribución de patentes municipales según comuna



Fuente: Elaboración propia en base a información de patentes municipales.

Como se observa, cuatro comunas son las que concentran la mayoría de las patentes municipales. Debido a esto, es que se analizara la distribución de patentes en dos grupos de comuna. El primero hace referencia a las cuatro comunas con mayor concentración de patentes: Las Condes, Vitacura, Providencia y Santiago y el segundo grupo corresponderá al resto de las comunas.

En el primer gráfico se ve que son las patentes comerciales las que tienen mayor participación en cada una de las cuatro comunas en análisis. Cuestión que también ocurre para el segundo grupo, con la excepción de tres comunas para las cuales no se tiene una descripción del tipo de patente: Cerrillos, Cerro Navía y Pudahuel.

Gráfico 6: Distribución de patentes, Grupo I

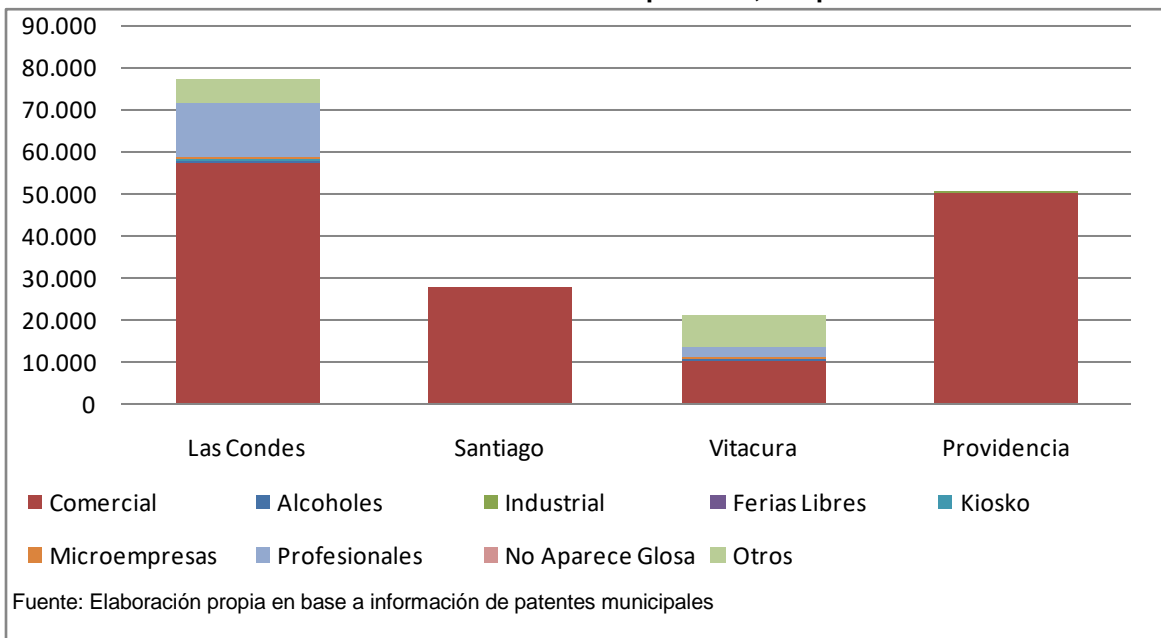
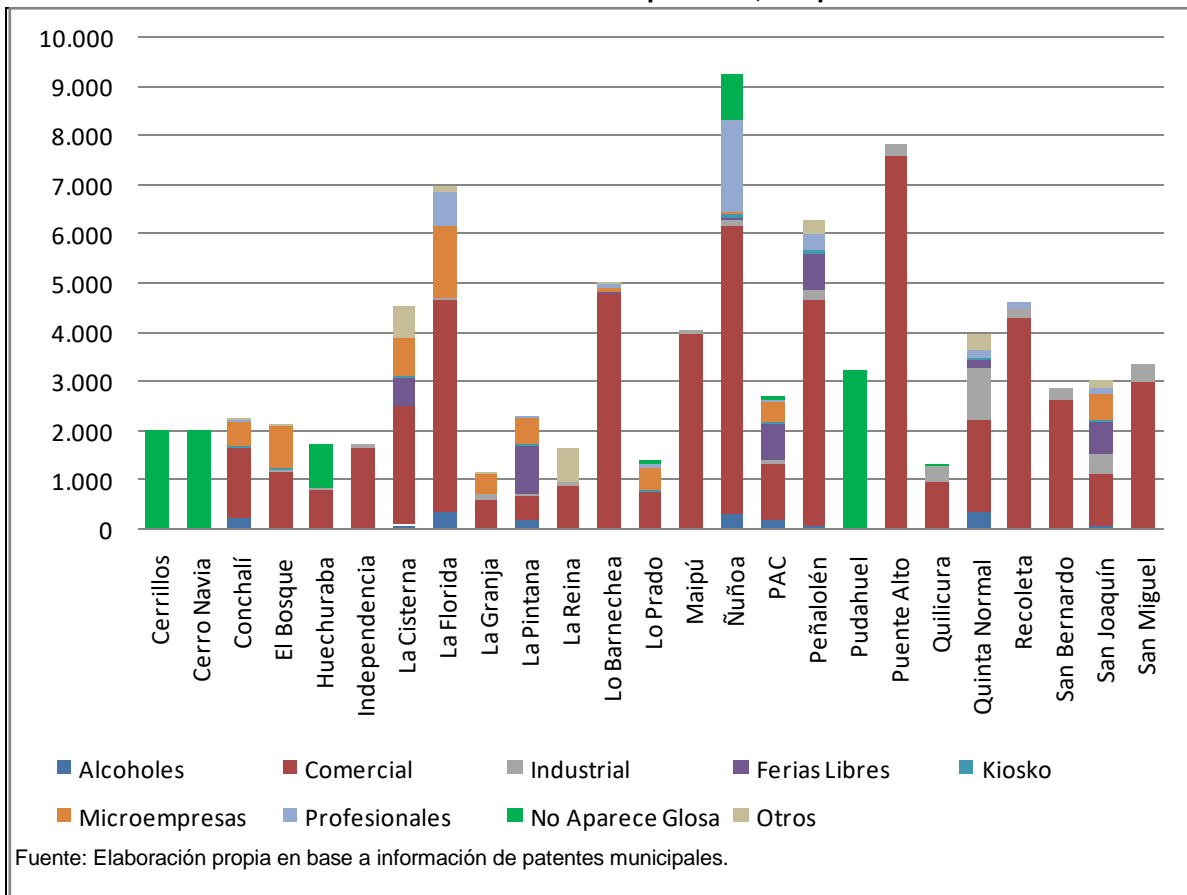


Gráfico 7: Distribución de patentes, Grupo II



Anexo D: Comunas de la Región Metropolitana utilizadas

A continuación se presenta un listado de todas las comunas de la Región Metropolitana indicando aquellas que son y no son utilizadas y el porqué de esta no utilización.

Tabla 17: Uso de comunas Región Metropolitana

Comuna	Provincia	Estado	Comuna	Provincia	Estado
Colina	Chacabuco	No utilizada: Fuera de Rango geocodificación	Las Condes	Santiago	Utilizada
Lampa	Chacabuco	No utilizada: Fuera de Rango geocodificación	Lo Barnechea	Santiago	Utilizada
Tiltil	Chacabuco	No utilizada: Fuera de Rango geocodificación	Lo Espejo	Santiago	No utilizada: Fecha de Otorgamiento
Pirque	Cordillera	No utilizada: Fuera de Rango geocodificación	Lo Prado	Santiago	Utilizada
Puente Alto	Cordillera	Utilizada	Macul	Santiago	No utilizada: Fecha de Otorgamiento
San José de Maipo	Cordillera	No utilizada: Fuera de Rango geocodificación	Maipú	Santiago	Utilizada
Buín	Maipo	No fue posible obtener información de la comuna	Ñuñoa	Santiago	Utilizada
Calera de Tango	Maipo	No utilizada: Fuera de Rango geocodificación	Pedro Aguirre Cerda	Santiago	Utilizada
Paine	Maipo	No utilizada: Fuera de Rango geocodificación	Peñalolén	Santiago	Utilizada
San Bernardo	Maipo	Utilizada	Providencia	Santiago	Utilizada
Alhué	Melipilla	No fue posible obtener información de la comuna	Pudahuel	Santiago	Utilizada
Curacaví	Melipilla	No utilizada: Fuera de Rango geocodificación	Quilicura	Santiago	Utilizada
María Pinto	Melipilla	No fue posible obtener información de la comuna	Quinta Normal	Santiago	Utilizada
Melipilla	Melipilla	No utilizada: Fuera de Rango geocodificación	Recoleta	Santiago	Utilizada
San Pedro	Melipilla	No fue posible obtener información de la comuna	Renca	Santiago	No utilizada: Fecha de Otorgamiento
Cerrillos	Santiago	Utilizada	San Joaquín	Santiago	Utilizada
Cerro Navia	Santiago	Utilizada	San Miguel	Santiago	Utilizada
Conchalí	Santiago	Utilizada	San Ramón	Santiago	No fue posible obtener información de la comuna
El Bosque	Santiago	Utilizada	Santiago	Santiago	Utilizada
Estación Central	Santiago	No utilizada: Fecha de Otorgamiento	Vitacura	Santiago	Utilizada
Huechuraba	Santiago	Utilizada	Isla de Maipo	Talagante	No fue posible obtener información de la comuna
Independencia	Santiago	Utilizada	Padre Hurtado	Talagante	No utilizada: Fuera de Rango geocodificación
La Cisterna	Santiago	Utilizada	Peñaflor	Talagante	No utilizada: Fuera de Rango geocodificación
La Florida	Santiago	Utilizada	El Monte	Talagante	No utilizada: Fuera de Rango geocodificación
La Granja	Santiago	Utilizada	Talagante	Talagante	No utilizada: Fuera de Rango geocodificación
La Pintana	Santiago	Utilizada			
La Reina	Santiago	Utilizada			

Fuente: Elaboración propia en base a patentes municipales.

Anexo E: Información de variables utilizadas según base de datos

Las siguientes tablas muestran la información que se tiene por cada una de las bases de datos utilizadas en el estudio:

Tabla 18: Bases SIMCE-RBD

Variable	Descripción
Grado	Grado en el que se rinde la evaluación
Rbd	Rol Base de datos del establecimiento
Nom_rbd	Nombre del establecimiento
Cod_reg_rbd	Código de región del establecimiento
Nom_reg_rbd	Nombre de región del establecimiento
Cod_pro_rbd	Código de provincia del establecimiento
Nom_pro_rbd	Nombre de provincia del establecimiento
Cod_com_rbd	Código de comuna del establecimiento
Nom_com_rbd	Nombre de comuna del establecimiento
Dependencia_adm	Código de dependencia administrativa. (0 Municipal, 1 Particular Subvencionado. Variable creada en función de cod_depe1, 5 categorías de dependencia en donde se eliminó los colegios particulares pagados).
NSE_rbd	Grupo socioeconómico del establecimiento. (Dummies por categoría: Baja, Media Baja, Media, Media Alta y Alta)
Letra_curso	Letra del curso
Cod_curso	Código del curso
Año	Año de la evaluación

Fuente: Elaboración propia en base a Simce

Tabla 19: Bases SIMCE- Alumnos

Variable	Descripción
Ptje_lect	Puntaje del estudiante en SIMCE de lectura
Ptje_mate	Puntaje del estudiante en SIMCE de matemática
Mrun	Identificador único MINEDUC (longitudinal)
Sexo	Género del estudiante. (0 Hombre, 1 Mujer)
Rbd	Rol Base de datos del establecimiento
Letra_curso	Letra del curso
Cod_curso	Código del curso
Idalumno	Identificador único Simce (transversal)
Año	Año de la evaluación

Fuente: Elaboración propia en base a Simce

Tabla 20: Bases SIMCE - Cuestionario de Padres

Variable	Descripción
Grado	Grado en el que se rinde la evaluación
Rbd	Rol Base de datos del establecimiento
Educ_padre*	Máximo nivel de educación alcanzado por el padre del alumno
Educ_madre*	Máximo nivel de educación alcanzado por la madre del alumno
Ingresos Hogar	Suma de los ingresos de todas las personas que aportan al hogar donde vive el estudiante en un mes normal. (5 categorías, 1=Menos de \$300.000; 2=Entre \$300.001 y \$600.000; 3= Entre \$600.001 y \$1.000.000; 4= Entre \$1.000.001 y \$1.600.000 ; 5= Más de \$1.600.000)
Idalumno	Identificador único Simce (transversal)
Letra_curso	Letra del curso
Cod_curso	Código del curso
Año	Año de la evaluación

Fuente: Elaboración propia en base a Simce *Revisar Tabla 22

Tabla 21: Conversión del nivel educacional en años de escolaridad

Nivel Educacional	Años de estudio
No estudió	0 años
1° Año de Educación Básica	1 año
2° Año de Educación Básica	2 años
3° Año de Educación Básica	3 años
4° Año de Educación Básica	4 años
5° Año de Educación Básica	5 años
6° Año de Educación Básica	6 años
7° Año de Educación Básica	7 años
8° Año de Educación Básica	8 años
1° Año de Educación Media	9 años
2° Año de Educación Media	10 años
3° Año de Educación Media	11 años
4° Año de Educación Media	12 años
4° o 5° Año de Educación Media Técnico Profesional	12 años
Educación Incompleta en un C.F.T o I.P	14 años
Titulado de un C.F.T o I.P	16 años
Educación Incompleta en una Universidad	15 años
Titulado de una Universidad	17 años
Grado de Magíster	19 años
Grado de Doctor	22 años

Fuente: Metodología de Construcción de Grupos Socioeconómicos SIMCE 2012

Tabla 22: Base Patentes Municipales

Variable	Descripción
Id	Identificador del local
Rol	Identificador único a nivel comuna de una propiedad o bien raíz
Nombre/Razón Social	Denominación legal con la que se conoce la empresa
Dirección	Lugar físico del local
Giro/Actividad Económica	Actividad a la que se dedica la empresa o negocio
Tipo de Patente	Comercial, Industrial o de Alcoholes
Fecha de Otorgamiento	Fecha en la que se entregó la primera patente
Fecha de Terminó	Fecha a la cual se vence o venció la patente
Valor	Costo de la patente
Comuna	Comuna de referencia
Rut	Identificación requerida por toda persona natural, jurídica o entidad que sea susceptible a ser sujeto ha impuesto.

Fuente: Elaboración propia en base a patentes Municipales

Tabla 23: Base Mapa Nutricional

Variable	Descripción
Año	Año de la evaluación
Grado	Grado en el que se rinde la evaluación
Código	Identificador único (Transversal)
Sexo	Género del estudiante (0= Hombre; 1= Mujer)
Edad	Número de meses de vida del alumno.
Peso	Medida de peso en kilogramos.
Talla	Medida de altura en centímetros.
IMC	$Peso / Talla^2$.
RBD	Rol Base de datos del establecimiento.
Región	Código de región del establecimiento.
Provincia	Código de provincia del establecimiento.
Comuna	Código de comuna del establecimiento.
Nombre_Comuna	Nombre de comuna del establecimiento.
Fecha_Med	Fecha en la que se midió y peso al alumno.
Fecha_Nac	Fecha en que nació el alumno.

Fuente: Elaboración propia en base a Mapa Nutricional

Tabla 24: Base Registro de Estudiantes Chilenos (RECH)

Variable	Descripción
Año	Año de la evaluación
Rbd	Rol base de datos del establecimiento
Nom_rbd	Nombre del RBD
Cod_reg_rbd	Código de región del establecimiento
Cod_com_rbd	Código de comuna del establecimiento
Nom_com_rbd	Nombre de comuna del establecimiento
Cod_depe	Código de dependencia administrativa (5 categorías, donde se eliminó los colegios particulares pagados)
Cod_ense*	Código del tipo de Enseñanza
Cod_grado	Código de grado medido
Let_cur	Letra del curso del alumno
Mrun	Identificador único MINEDUC (longitudinal)
Gen_alu	Género del alumno (0= Hombre; 1= Mujer)
Fec_nac_alu	Fecha de nacimiento del alumno
Cod_com_alu	Código de comuna donde reside el alumno
Nom_com_alu	Nombre de comuna donde reside el alumno
Prom_gral	Promedio general anual
Asistencia	Porcentaje anual de asistencia
Sit_fin	Situación de promoción al cierre del año escolar. (P= Promovido; R=Reprobado; Y: Retirado)
Sit_fin_r	Situación de promoción al cierre del año escolar, con indicador de traslado. (P= Promovido; R=Reprobado; Y: Retirado; T=Trasladado)

Fuente: Elaboración propia en base a RECH

Anexo F: Estadística Descriptiva

Tabla 25: Estadística Descriptiva 2013-2012						
Controles Individuales	Total		Hombre		Mujer	
IMC	21,98	3,37	21,84	3,41	22,1	3,33
Simce Matemática*	0,129	0,96	0,207	0,96	0,061	0,963
Simce Lectura	0,073	0,99	-0,035	1,02	0,169	0,97
Simce Pasado						
Matemática	0,252	0,92	0,316	0,93	0,19	0,9
Simce Pasado Lectura	0,21	0,94	0,152	0,95	0,25	0,93
Edad	181,63	8,52	181,83	8,56	181,45	8,47
N° de Observaciones	12.321		5.813		6.508	

Controles Colegio	Total		Hombre		Mujer	
Nivel Socioeconómico						
<i>Bajo</i>			41		39	
<i>Medio Bajo</i>			1.354		1.482	
<i>Medio</i>			2.847		3.021	
<i>Medio Alto</i>			1448		1.831	
<i>Alto</i>			123		135	
Dependencia_adm						
<i>Municipal</i>			1.552		1.406	
<i>Particular</i>						
<i>Subvencionado</i>			4.261		5.102	
Log_niños*	-4,71	1,08	-4,72	1,07	-4,69	1,07
N° de Observaciones	12.321		5.813		6.508	
* 11.358 = Hombres (5.373) + Mujeres (5.985)						

Controles Familiares	Total		Hombre		Mujer	
Educ Padres*	11,73	3,83	11,77	3,8	11,69	3,87
Educ Madres**	11,74	3,59	11,81	3,59	11,67	3,58
Ingresos Hogar						
<i>Bajo</i>			2.567		2.963	
<i>Medio Bajo</i>			1663		1882	
<i>Medio</i>			631		627	
<i>Medio Alto</i>			308		323	
<i>Alto</i>			467		506	
N° de Observaciones			5.636		6.301	
*11.379= Hombres (5.366) + Mujeres (5.633).						
**11.966 = Hombres (6.013) + Mujeres (6.333)						

Fuente: Elaboración propia en base a Mapa Nutricional, Simce, Rech. En paréntesis, desviaciones estándar. La tabla está hecha en función de la base que contiene todos los controles posibles, por eso mismo el bajo número de observaciones.

Tabla 26: Estadística Descriptiva 2012-2011			
Controles Individuales	Total	Hombre	Mujer
IMC	22,06 (3,32)	22,02 (3,3)	22,11 (3,32)
Simce Matemática*	0,057 (0,95)	0,12 (0,96)	-0,001 (0,93)
Simce Lectura	0,003 (0,99)	-0,125 (1,04)	0,116 (0,93)
Simce Pasado Matemática	0,194 (0,88)	0,268 (0,89)	0,128 (0,87)
Simce Pasado Lectura	0,138 (0,94)	0,078 (0,98)	0,191 (0,9)
Edad	181,6 (6,48)	181,66 (6,31)	181,53 (6,63)
N° de Observaciones	6.917	3.238	3.679
* 6.887 = Hombres (3.226) + Mujeres (3.661)			

Controles Colegio	Total	Hombre	Mujer
Nivel Socioeconómico			
<i>Bajo</i>		29	26
<i>Medio Bajo</i>		794	864
<i>Medio</i>		1.645	1.792
<i>Medio Alto</i>		758	960
<i>Alto</i>		12	37
Dependencia_adm			
<i>Municipal</i>		1.047	1.073
<i>Particular Subvencionado</i>		2.191	2.606
Log_niños*	-4,59 (1,06)	-4,64 (1,05)	-4,55 (1,07)
N° de Observaciones	6.917	3.238	3.679
* 6.602 = Hombres (3.085) + Mujeres (3.517)			

Controles Familiares	Total	Hombre	Mujer
Educ Padres*	12,7 (3,66)	12,77 (3,86)	12,91 (3,88)
Educ Madres**	12,84 (3,87)	12,59 (3,66)	12,8 (3,66)
Ingresos Hogar			
<i>Bajo</i>		1.235	1.345
<i>Medio Bajo</i>		813	894
<i>Medio</i>		296	399
<i>Medio Alto</i>		234	262
<i>Alto</i>		520	637
N° de Observaciones	6.635	3.098	3.537
* 6.420 = Hombres (2.998) + Mujeres (3.422).			
**6.640 = Hombres (3.105) + Mujeres (3.535)			

Fuente: Elaboración propia en base a Mapa Nutricional, Simce, Rech. En paréntesis, desviaciones estándar. La tabla está hecha en función de la base que contiene todos los controles posibles, por eso mismo el bajo número de observaciones.

Anexo G: Proceso y Resultados de Geocodificación

G.I.- Proceso de Geocodificación

Geocodificación es el proceso mediante el cual se transforma la descripción de una ubicación, como: par de coordenadas, dirección o un nombre de un lugar en una ubicación en la superficie de la tierra. En particular, tenemos una tabla con direcciones de comunas de la Región Metropolitana, correspondientes a locales señalados en las patentes municipales a las que les encontraremos sus coordenadas geográficas (latitud y longitud) y que localizaremos en el mapa de la Región metropolitana a través un Sistema de Información Geográfica (SIG). Un SIG es un conjunto de herramientas que integra y relaciona diversos componentes que permiten la organización, manipulación y análisis de datos procedentes del mundo real, en particular información geográficamente referenciada.

Para poder realizar esta labor utilizamos el programa ArcGis 10.2, en específico ArcMap. Existen tres pasos a seguir a la hora de geocodificar una dirección, los cuales están descritos en el *tutorial de geocodificación*⁵² de ArcGis y que serán descritos a continuación:

1.- Creación AddressLocator

Uno de los insumos necesarios para poder geocodificar es la creación de un "addresslocator" o un "localizador de direcciones" que básicamente es un algoritmo de búsqueda de información. A grandes rasgos sabemos que una dirección puede estar compuesta por alguno de los siguientes componentes: número, nombre de la calle, tipo de calle, ciudad, estado, código ZIP y además dependiendo del país de referencia, las direcciones pueden estar escritas de diferentes maneras. Por ejemplo es normal que en Chile uno indique primero el nombre de la calle y después la numeración del lugar, por ejemplo:

Diagonal Paraguay 257, Santiago, Región Metropolitana.

Por otro lado en Estados Unidos, se suele decir primero el número y luego el nombre de calle o muchas veces se indican manzanas o códigos ZIP. Por ejemplo:

320 Madison Street.

Uno de los problemas para crear un addresslocator, es que los estilos de localizadores de dirección están configurados por defecto para Estados Unidos, por lo que primeramente habría que cambiar la forma en que se escriben las direcciones chilenas para que coincida con uno de estos estilos. Una vez hecho esto, se requiere de un mapa con suficiente detalle para crear el addresslocator en base a este. Sin embargo, no pudimos conseguirnos la información desagregada a nivel de la Región Metropolitana para crear nuestro propio addresslocator. No obstante, COES nos facilitó dos address locators que tenían en su poder: PSAD56 y WGS84, dos sistemas de coordenadas diferentes específicos para Santiago. Este es el principal motivo para utilizar información sólo de la RM y no de todo Chile, ya que de lo contrario no podríamos geocodificar información fuera de esta región, porque no se contaría con las "calles" a las cuales hay que hacer referencia. Para nuestro estudio se decidió utilizar el addresslocator PSAD56, debido a que tenía un mayor porcentaje de match.

⁵² http://help.arcgis.com/es/arcgisdesktop/10.0/pdf/tutorial_geocoding.pdf

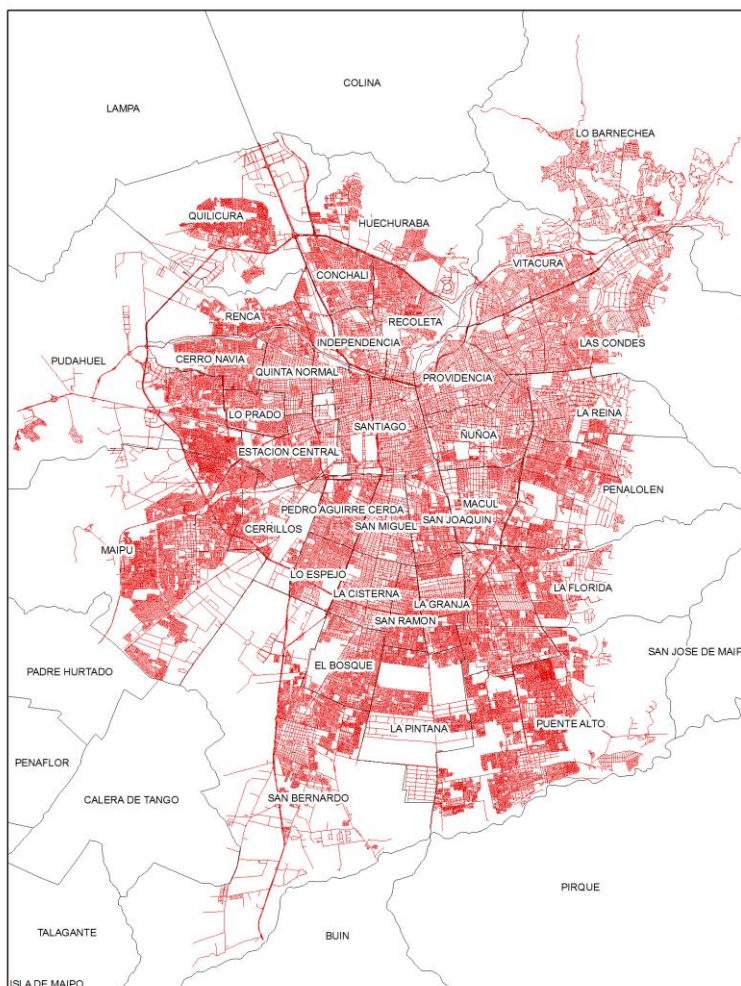
2.- Geocodificar

Con el addresslocator en nuestro poder, se procederá a geocodificar las direcciones de las patentes municipales. Para esto tuvimos que seguir el siguiente formato:

Numeración, Nombre Calle, Comuna

No será necesario recalcar la región en la que nos encontramos o el país, ya que sólo se realiza la búsqueda en Santiago de Chile, en particular para la Región Metropolitana. De hecho, el localizador con el que contamos cubrirá exclusivamente el área en rojo que podemos ver en el siguiente gráfico:

Gráfico 8: Zona de Cobertura de Localizador de Dirección



Fuente: COES.

Viendo el gráfico vemos que existen zonas que no están cubiertas por las calles enmarcadas en rojo, esto quiere decir que para algunas comunas como Colina o Lampa, no tendrán la información o el detalle suficiente a nivel de calles necesario para realizar la geocodificación, razón por lo cual se decidió excluir a este tipo de comunas de la muestra final.

Esto, sumado al hecho de que algunas comunas tenían patentes sin fechas de otorgamiento provocó que finalmente el estudio considerase un total de 29 comunas.

Una vez realizado el proceso de geocodificación existen tres resultados posibles, los cuáles se generan según el grado o nivel de concordancia que tiene una dirección con un punto o ubicación en el espacio en el cuál se está geocodificando. Así, se puede encontrar una nula concordancia, más de una o una exclusiva, entendiendo concordancia como las opciones en dónde puede estar la información que entregamos como datos:

- Matched: La dirección encontrada concuerda con un punto particular o dentro de las opciones disponibles es la con mayor precisión. De manera estándar se considera que sobre un 80% un punto se considera matcheado.
- Tied: La dirección encontrada tiene más de una concordancia, por lo mismo no se puede decidir dónde ubicar geoespacialmente la información.
- Unmatched: La dirección encontrada no concuerda con ningún punto en particular en la zona geoespacial o sus puntos porcentuales son muy bajos, por lo que no son matcheados.

3.- Volver a Geocodificar

Debido a que generalmente después de geocodificar la primera vez quedan muchos puntos Tied o Unmatched es frecuente que se vuelva a repetir el proceso de geocodificación hasta que el porcentaje total de direcciones matcheadas sea lo más alto posible. Algunos de estos errores ocurren por ejemplo porque las direcciones están mal escritas o porque contiene información extra que imposibilita encontrar la dirección o simplemente porque hay calles que se llaman igual dentro de una misma área espacial y entrega información de ellas dos.

G.II. - Resultados Geocodificación:

Como fue revisado en el anexo XX, el proceso de geocodificación entrega como output: Match, tied o unmatched. Con esto en mente se geocodificaron todas las patentes municipales en nuestra poder, incluidas las que no hacen referencia a una oferta de alimentos. El resultado de esta operación es presentado en las siguientes tablas, donde se puede observar que el porcentaje de logro de match por comuna varía, teniendo un promedio total de un 68% entre todas ellas.

Tabla 27: Resultados Geocodificación por Comuna

Comuna	Match	Tied	Unmatch	Total Matcheado	Total
Cerrillos	56,41%	3,87%	39,73%	1.999	3.544
Cerro Navia	83,38%	9,09%	7,53%	1.981	2.376
Conchalí	56,97%	2,45%	40,59%	2.236	3.925
Ñuñoa	66,49%	3,46%	30,05%	9.271	13.944
El Bosque	61,53%	16,17%	22,30%	2.116	3.439
Huechuraba	51,25%	4,13%	44,62%	1.701	3.319
Independencia	70,09%	2,87%	27,04%	1.711	2.441
La Cisterna	68,14%	11,03%	20,83%	4.517	6.629
La Florida	65,20%	11,31%	23,49%	6.966	10.684
La Granja	83,56%	5,84%	10,59%	1.144	1.369
La Pintana	62,26%	9,41%	28,32%	2.295	3.686
La Reina	50,55%	14,89%	34,56%	1.650	3.264
Las Condes	89,75%	6,07%	4,18%	77.358	86.194
Lo Barnechea	72,50%	8,36%	19,14%	5.023	6.928
Lo Prado	76,85%	4,08%	19,07%	1.394	1.814
Maipú	53,69%	8,26%	38,05%	4.054	7.551
PAC	61,16%	3,05%	35,79%	2.685	4.390
Peñalolén	53,86%	2,41%	43,73%	6.294	11.686
Providencia	76,38%	6,02%	17,60%	50.527	66.148
Pudahuel	76,44%	6,03%	17,53%	3.231	4.227
Puente Alto	59,82%	19,27%	20,92%	7.836	13.100
Quilicura	61,16%	12,68%	26,15%	1.326	2.168
Quinta Normal	82,07%	3,84%	14,08%	3.951	4.814
Recoleta	75,58%	6,74%	17,68%	4.608	6.097
San Bernardo	51,42%	20,92%	27,66%	2.846	5.535
San Joaquín	65,58%	0,69%	33,72%	3.020	4.605
San Miguel	87,76%	3,34%	8,90%	3.364	3.833
Santiago	82,98%	1,84%	15,18%	27.975	33.714
Vitacura	75,07%	4,89%	20,04%	21.376	28.473
	68,20%	7,35%	24,45%	264.455	349.897

Fuente: Elaboración propia en base a patentes municipales y geocodificación en ArcGis.

Algunas de las razones para no obtener una geocodificación perfecta es la sintaxis a utilizar en el addresslocator. Como indicábamos, se debe buscar el "número + dirección", lo que dejaría automáticamente fuera del proceso de geocodificación a lugares que entregan esta información de manera incorrecta o que simplemente no la tienen, como por ejemplo algunas ferias o terrenos rurales (Ej: Parcelas). Por otro lado, la dirección sólo debe poseer el nombre de la calle

en cuestión, no puede incluir ninguna especificación como: avenida, pasaje, block, departamento o alguna abreviación de estos términos u otros, lo que recaería en que tampoco pudiesen ser geocodificadas esas patentes. Los tildes tampoco pueden ser identificados por parte del addresslocator e incluso existen calles que no están registradas en este buscador, sea porque se crearon después de haberse creado el addresslocator o porque simplemente les cambiaron el nombre.

Esto explica el porqué los porcentajes de match no son perfecto. Además, como no fue posible limpiar las bases de datos de todas estas imperfecciones y como no todas las comunas tienen la misma cantidad de avenidas, blocks, departamentos, oficinas, etc, es normal que los porcentajes de match entre comunas varíe. Por último, del total de patentes municipales ya geocodificadas nos interesa conocer cuáles corresponden a comida chatarra y cuáles no, esto es lo que vemos en la siguiente tabla.

Tabla 28: Geocodificación comida chatarra por comuna

Comuna	Chatarra1	Chatarra2	Chatarra1	Chatarra2	Total
Cerrillos	15,58%	4,09%	552	145	3.544
Cerro Navia	37,37%	2,48%	888	59	2.376
Conchalí	23,59%	9,25%	926	363	3.925
Ñuñoa	9,95%	4,55%	1.388	635	13.944
El Bosque	31,52%	7,44%	1.084	256	3.439
Huechuraba	17,32%	11,96%	575	397	3.319
Independencia	15,53%	6,23%	379	152	2.441
La Cisterna	17,44%	4,90%	1.156	325	6.629
La Florida	18,93%	7,35%	2.023	785	10.684
La Granja	30,31%	6,72%	415	92	1.369
La Pintana	34,02%	13,16%	1.254	485	3.686
La Reina	7,51%	2,54%	245	83	3.264
Las Condes	4,33%	1,91%	3.729	1.650	86.194
Lo Barnechea	7,78%	2,86%	539	198	6.928
Lo Prado	38,48%	15,44%	698	280	1.814
Maipú	22,51%	14,49%	1.700	1.094	7.551
PAC	30,27%	14,74%	1.329	647	4.390
Peñalolén	17,75%	6,50%	2.074	760	11.686
Providencia	5,81%	2,96%	3.844	1.956	66.148
Pudahuel	37,76%	8,07%	1.596	341	4.227
Puente Alto	24,03%	7,53%	3.148	986	13.100
Quilicura	13,19%	3,55%	286	77	2.168
Quinta Normal	19,82%	11,20%	954	539	4.814
Recoleta	17,68%	6,89%	1.078	420	6.097
San Bernardo	20,05%	6,16%	1.110	341	5.535
San Joaquín	25,67%	11,70%	1.182	539	4.605
San Miguel	19,12%	13,57%	733	520	3.833
Santiago	10,95%	5,59%	3.691	1.886	33.714
Vitacura	3,39%	1,75%	964	499	28.473
Total	19,92%	7,43%	39.540	16.510	349.897

Fuente: Elaboración propia en base a patentes municipales y geocodificación en ArcGis.

Anexo H: Identificar locales de comida entorno a establecimientos educativos

Para cuantificar el número de locales de comida chatarra entorno a los colegios, fue necesario utilizar el toolbox "NearTable" de ArcGis. En primer lugar se debe entrenar las "unidades base" que corresponderán a los establecimientos educativos. Posteriormente, debemos entregar las "unidades cercanas" que son las patentes de comida chatarra. La herramienta funciona de la siguiente forma, tomando como centro a cada unidad base, se trazaran áreas a la redonda de estos puntos para cuantificar el número de unidades cercanas a estos puntos, obteniendo así el número de locales de comida entorno a la distancia utilizada.

Una representación gráfica de esto puede verse en la siguiente figura. Esta muestra las unidades base (puntos azules) y las unidades cercanas (puntos verdes) para la comuna de Ñuñoa. Además, las circunferencias entorno a estos puntos azules serán el área a la redonda que captura cada unidad base para cuantificar las unidades cercanas entorno a ellos. Se debe destacar que esto es sólo una representación gráfica de lo que se está realizando con el NearTable, por lo cual no se utilizan ni todos los colegios ni tampoco la misma distancia para los colegios, por eso es que las circunferencias tienen tamaños diferentes.

Gráfico 9: Representación Gráfica Variable Instrumental, Ñuñoa.



Fuente: Elaboración propia en base a ArcGis

Anexo I: Proceso de match de las bases

Los identificadores, id y mrun, deben ser únicos en cada base, ya que de lo contrario se podría estar asociando información a individuos que no son el mismo. Las bases del mapa nutricional sólo tendrán el identificador "id", por lo que simplemente se eliminarán los datos duplicados. Las bases RECH tienen ambos identificadores, id y mrun, donde los datos duplicados de "id" fueron eliminados. Sin embargo, se registraron "mrun" duplicados, antes de eliminarlos estos pueden ser filtrados para encontrar al alumno que efectivamente corresponden y de esta forma no eliminar el dato, sino que recuperar el original. Esto es posible debido a que las bases RECH contienen información de la situación académica del alumno al finalizar el año escolar, es decir, se podrá saber si es que el alumno: aprobó, repitió, se traslado de colegio o se retiro del sistema educativo. Esta información permitirá identificar por ejemplo a aquellos alumnos que se trasladaron de colegio y que aparecen dos veces en la base de datos, en el colegio desde el cual se trasladaron y su colegio actual, dato que será de nuestro interés. Aquellos "mrun" que no fueron capaces de ser asociados a un único alumno serán los que efectivamente se eliminarán de esta base de datos.

Finalmente las bases SIMCE también cuentan con casos de "mrun" duplicados. Antes de eliminar esta información se buscó el "mrun" duplicado en la base RECH, que en caso de ser único permitía diferenciar al "mrun" incorrecto del correcto. Para aquellos casos en los que no se pudo identificar los "mrun" reales, simplemente se eliminó la información.

Bibliografía

- Actualidad Deportes Noticia (2 de agosto de 2013) "Schmidt: Este año se incrementará a 4 horas obligatorias de educación física para todos los niños" <<http://www.adnradio.cl/noticias/nacional/schmidt-este-ano-se-incrementara-a-4-horas-obligatorias-de-educacion-fisica-para-todos-los-ninos/20130802/nota/1944116.aspx>>
- Atella, V. Noemice, P. & Vuri, D. (2008) "Are employers discriminating with respect to weight? European Evidence using Quantile Regression" *Economics and Human Biology*, Volume: 6, 305-329.
- Overseas Development Institute (2015) "The rising cost of a healthy diet" <<http://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/9579.pdf>>
- Averett, S. & Korenman, S. (1996) "The Economic Reality of the Beauty Myth" *Journal of Human Resources*. Volume: 31, 304-330.
- Azar, A. Franetovic, G. Martínez, M. & Santos, H. (2015) "Determinantes individuales, sociales y ambientales del sobrepeso y la obesidad adolescente en Chile" *Revista Médica de Chile*. Volumen: 143, 598-605.
- Baum, C. & Ford, W. (2004) "The wage effects of obesity: A longitudinal study" *Health Economics*, Volume:13, 885-899.
- BBC Mundo (10 de noviembre de 2014) "El secreto danés para combatir la obesidad infantil" <http://www.bbc.com/mundo/noticias/2014/11/141108_salud_obesidad_infantil_daneses_secreto_lav>
- Biblioteca del Congreso Nacional, Ley N°20.606 <<http://www.leychile.cl/Navegar?idNorma=1041570>>
- Campos, A. Sigulem, D. Moraes, D. Escrivao, A. & Fisberg, M. (1996) "Intelligent quotient of obese children and adolescents by the Weschler scale" *Revista de Salud Publica*, volumen: 30, 85-90.
- Cawley, J (2004) "The Impact of Obesity on Wages" *Journal of Human Resources*. Volume: 39, 451-474.
- Cawley, J. (2000) "Body Weight and women's labor market outcome" Working Paper. National Bureau of Economic Research, Cambridge.
- Cawley, J. (2014) "Overview of my research on the economics of obesity" Cornell University and National Bureau of Economic Research.
- Cawley, J. Rizzo, J. & Haas, K. (2007) "Occupation-specific absenteeism costs associated with obesity and morbid obesity". *Journal of Occupational and environmental medicine*.

- Childhood obesity and Academic Outcomes (2008). Institute for educational leadership and policy.
- Cragg, J. & Donald, S. (1993) "Testing Identifiability and Specification in Instrumental Variable Models" *Econometric Theory*, Volume:9, 222-240.
- Cutler, D., Glaeser, E. & Shapiro, J. (2003) "Why have Americans become more obese" *Journal of Economic Perspectives* 17, N°3: 93-118.
- Datar, A. & Sturm, R (2004) "Physical education in elementary school and body mass index: Evidence from the early childhood longitudinal study" *American Journal of Public Health*.
- Datar, A. & Sturm, R (2006) "Childhood overweight and elementary school outcomes" Volume: 30, 1449-1460.
- Datar, A. Sturm, R & Magnabosco, J. (2004) "Childhood overweight and academic performance: national study of kindergartners and first-graders" *Obesity Research*, Volume: 12, 58-68.
- Ding, W. Lehrer, S. Rosenquist, J. & Audrian-McGovern, J. (2009) "The Impact of poor health on academic performance: New evidence using genetic markers" *Journal of Health Economics*. Volume: 28, 578-597.
- Encuesta de Consumo Alimentario en Chile, ENCA (2014). Departamento de Nutrición, Escuela de Nutrición, Escuela de Salud Pública, Centro de Microdatos. <http://www.ipsuss.cl/site/artic/20141210/asocfile/20141210160211/enca_informe_final.pdf>
- Falkner, N. Neumark-Sztainer, D. Story, M. Jeffery, R. Beuhring, T. & Resnick, M. (2001) "Social, Educational and Psychological correlates of weight status in adolescents" *Obesity Research*, Volume 9: 32-42
- Finkelstein, E. Brown, D. Trogdon, G. Segel, E. & Ben-Joseph, R. (2007) "Age-Specific impact of obesity on prevalence and costs of diabetes and dyslipidemia" *Value in Health*, Volume:10, 45-51. "The medical cost of obesity are likely to increase with age.
- Finkelstein, E. Ruhm, C. & Kosa, K. (2005) "Economic causes and consequences of obesity" *Annual Review of Public Health*, Volume: 26, 239-257.
- Fletcher, J. & Lehrer, S. (2008) "Using the "Genetic Library" within families to examine the causal effect of health on academic achievement" *Society of Labor Economist*, New York.
- French, S. Harnack, L. & Jeffery, R. (2000) "Fast food restaurant use among women in the pound of prevention study: dietary, behavioral and demographic correlates". *International Journal of Obesity*, Volume:24, 1353-1359.

- Gable, S. Britt-Ranking, J & Jennifer, L. (2008) "Ecological Predictors and developmental outcomes of persistent childhood overweight" Contractor and Cooperator Report, N°42.
- Gable, S. Chang, Y. & Krull (2012) "Boy's and Girl's Weight Status and Match Performance from kindergarten entry through fifth grade: A mediated analysis" Child Development, Volume: 83, 1822-1839.
- Gortmaker, S. Must, A. Perrin, J. Sonol, A. & Dietz, W. (1993) "Social and economic consequences of overweight in adolescence and Young adulthood" The New England Journal of Medicine. Volume: 239, 1008-1012.
- Gregory, C. & Ruhm, C (2011) "Where does the Wage Penalty bite?" Chapter in National Bureau of Economic Research book Economic Aspects of Obesity, 315-347.
- Greve, J. (2008) "Obesity and labor market outcomes in Denmark" Economics and Human Biology, Volume: 6, 350-362.
- Institute for Health Metrics and Evaluation (IHME) <<http://www.healthdata.org/>>
- Ivanovic, D. Boris, P. Pérez, H. Olivares, M. Díaz, N. Urrutia, M. Almagía, A. Toro, T. Miller, P. Bosch, E. & Larraín, C. (2004) "Head size and intelligence, learning, nutritional status and brain development: Head, IQ, learning, nutrition and brain" Neuropsychologia.
- Ivanovic, D. Burrows, R. Correa-Burrows, P. Orellana, Y. (2014) "The relationship between unhealthy snacking at school and academic outcomes: a population study in Chilean schoolchildren" Article in Public Health Nutrition.
- Ivanovic, D. Burrows, R. Correa-Burrows, P. Orellana, Y. Almagía, A. & Lizana, P. (2014) "Scheduled Physical activity is associated with better academic performance in Chilean school-age children" Journal of Physical Activity and Health.
- Ivanovic, D. Burrows, R. Orellana, Y. Morales, G. Arancibia, V. Almagía, A. Ibaceta, C. & Lizana, P. (2011) "Prevalence of obesity and physical and eating habits of Chilean children attending to schools with high, medium and low academic achievement" Article in Pediatric Research.
- Ivanovic, D. Ibaceta, C. Correa, Pa. Orellana, Y. Calderon, P. Morales, G. Leyton, B. Almagia, A. Lizana, Pa. & Burrows, R. (2013) "Brain development and scholastic achievement in the Educational Quality Measurement System test in Chilean school-aged children" Article in Pediatric Research.
- Ivanovic, D. Leiva, B. Inzunza, N. Pérez, H. Castro, V. Jansana, J. Toro, T. Almagía, A. Díaz, A. Urrutia, M. & Cervilla, J. (2001) "The impact of malnutrition on brain development, intelligence and school work performance" Archivos Latinoamericanos de Nutrición.

- Ivanovic, D. Leave, B. Pérez, H. Almagía, A. Toro, T. Urrutia, M. Inzunza, N. & Bosch, E. (2002) "Nutritional status, brain development and scholastic achievement of Chilean high school graduates from high and low intellectual quotient and socio-economic status" British Journal of Nutrition.
- Ivanovic, D. Orellana, Y. Burrows, R. & Correa-Burrows, P. (2014) "Achievement in mathematics and language is linked to regular physical activity: a population study in Chilean youth" Article in Journal of Sports Sciences.
- Ivanovic, D. Rodríguez, M. Pérez, H. Alvear, J. Almagiá, A. Toro, T. Urrutia, M. Cruz, A. & Ivanovic, R. (2009) "Impact of nutritional status at the onset of elementary school on academic aptitude test achievement at the end of high school in a multicausal approach" The British Journal of Nutrition.
- Ivanovic, D. Rodríguez, M. Pérez, H. Alvear, J. Díaz, N. Leyton, B. Almagia, A. Toro, T. Urrutia, M. & Ivanovic, R. (2008) "Twelve-year follow-up study of the impact of nutritional status at the onset of elementary school on later educational situation of Chilean school-age children" European Journal of Clinical Nutrition.
- Kaestner, R. & Grossman, M. (2009) "Effects of weight on children's educational achievement" Economics of Education Review. Volume 28: 651-661.
- Krukowski, R. West, D. Philyaw, P. Bursac, Z. Phillips, M. & Raczynski, M. (2009) "Overweight children, weight-based teasing and academic performance". International Journey of Obesity, 274-280.
- La Tercera (28 de abril de 2014) "MINSAL: El 34% de los niños menores de seis años sufre obesidad o sobrepeso" <<http://papeldigital.info/lt/2014/04/28/01/paginas/010.pdf>>
- La Tercera (18 de abril de 2014) "Chile es el quinto país con más obesidad en OCDE" <<http://www.latercera.com/noticia/nacional/2014/04/680-574458-9-chile-es-el-quinto-pais-con-mas-obesidad-en-ocde.shtml>>
- La Tercera (29 de Mayo de 2015) "Chile, el país con menores desnutridos pero el segundo con más obesos de la región" <<http://www.latercera.com/noticia/tendencias/2015/05/659-631796-9-chile-el-pais-con-menos-ninos-desnutridos-pero-el-segundo-con-mas-obesos-en-la.shtml>>
- La Tercera (9 de Septiembre de 2014) "Ministra de salud pide subir a seis las horas de educación física para frenar la obesidad" <<http://www.latercera.com/noticia/nacional/2014/09/680-595085-9-ministra-de-salud-pide-subir-a-seis-las-horas-de-educacion-fisica-para-frenar.shtml>>
- Latinen, J. Power, C. Ek, E. Sovio, U. & Jarvelin, M (2002) "Unemployment and obesity among Young adults in a northern Finland 1966 birth cohort" International Journal of Obesity and related metabolic disorders. Volume: 26, 1329-1338

- Li, X. (1995) "A study of intelligence and personality in children with simple obesity" *International Journal of Obesity and related metabolic disorders*. Volume: 19, 355-357
- Mapa Nutricional, Junta Nacional de Auxilio Escolar y Becas (JUNAEB). <<http://www.junaeb.cl/mapa-nutricional>>.
- Medline Plus (8 de enero de 2016) "Los impuestos sobre los reflejos redujeron el consumo en México" <https://www.nlm.nih.gov/medlineplus/spanish/news/fullstory_156615.html>
- Mikkila, V. Lahti-Koski, M. Pietinen, P. Virtanen, S. & Rimpela, M. (2003) "Associates of obesity and weight dissatisfaction among Finnish adolescents" *Public Health Nutrition*, Volume: 6, 49-56.
- Ministerio de Salud "Encuesta Nacional de Salud, Año 2009 - 2010 (ENS)" <<http://web.minsal.cl/portal/url/item/bcb03d7bc28b64dfe040010165012d23.pdf>>
- Mo-Suwan, L. Lebel, L. Puetpaiboon, A. & Junjana C. (1999) "School performance and weight status of children and Young adolescents in a transitional society in Thailand" *International Journal of Obesity*, Volume: 23, 272-277.
- Must, A & Strauss, R. (1999) "Risks and consequences of childhood and adolescent obesity" *International Journal of Obesity*. Volume 23: 2-11
- Neckerman, K. Bader, M. Richards, C. Purciel, M. Quinn, J. Thomas, J. Warbelow, C. Weiss, C. Lovasi, G. & Rundle, A. (2010) "Disparities in the food environments of New York city public schools" *American Journal of Preventive Medicine*. Volume: 39, 195-202.
- Neumark-Sztainer, D. Falkner, N. Story, M. Perry, C. Hannan, P. & Mulert, S. (2002) "Weight-teasing among adolescents: correlations with weight status and disordered eating behaviours" *International Journal of Obesity*, Volume: 26, 123-131.
- Noticiero, Televisa (7 de enero de 2015) " México encabeza estadísticas en obesidad, en adultos y niños" <<http://noticieros.televisa.com/mexico/1501/mexico-encabeza-estadisticas-obesidad-adultos-ninos/>>
- Organización Mundial de la Salud (OMS) "Datos y cifras sobre obesidad infantil" <<http://www.who.int/end-childhood-obesity/facts/es/>>
- Organización Mundial de la Salud (OMS). "Centro de Prensa: Nota N°311, Enero 2015" <<http://who.int/mediacentre/factsheets/fs311/es/>>
- Pagan, J. & Davila, A (1997) "Obesity, occupational attainment, and earning" *Social Science Quarterly* Volume: 78, 757-770.
- Puhl, R. & Latner, J (2007) "Stigma, obesity and the health of the nation's children" *Psychological Bulletin* N°133: 557-580.

- Sabia, J. (2007) "The effect of body weight on adolescent academic performance" Southern Economic Journal, Volume: 73, 871-900.
- Sargent, J. Blanchflower, D. (1994) "Obesity and stature in adolescence and earning in Young adulthood" Archives of Pediatrics and Adolescent Medicine, Volume: 148.
- Serdula, M. Ivery, R. Coates, D. Freedman, D. & Williamson, T. (1993) "Do Obese Children become obese adults? A review of the literature" Preventive Medicine. Volume: 22 , 167-177.
- Shanty, A. Bowman, B. & Vinyard, T. (2004) "Fast food consumption of U.S. Adults: Impact on energy and nutrient intakes and overweight status" Journal of American Colleague of Nutrition Volume: 23, 163-168.
- Shimokawa, S. (2008) "The labor market impact of body weight in China: A semiparametric analysis. Applied Economics Volume 40: 949-968.
- Snell, E. Adam, E. & Duncan, G. (2007) "Sleep and the body mass index and overweight status of children and adolescents" Child Development, N°78: 309-323.
- Stock, J. & Staigner, D. (1997) & "Instrumental Variables Regression with weak instruments", Econometrica, Volume: 65, 557-586.
- Stock, J. & Yogo, M. (2005) "Testing for weak instruments in linear IV regression" Andrews DWK Identification and inference for econometric models: New York: Cambridge University Press: 2005, 80-108.
- Taras, H. & Datema, W (2005) "Obesity and student performance at school" Journal of School Health, Volume 75: N°8.
- Wada, R. & Tekin, E (2010) "Body composition and wages" Economics and Human Biology.
- Bellei, C. (2009) "Does lengthening the school day increase student's academic achievement? Results from a natural experiment in Chile. Economic of Education Review.
- "Impacto de la Ley Sep en SIMCE: Una Mirada a 4 años de su implementación" (30 de Agosto de 2012) Centro de Estudios Mineduc
- MIT "Treatment Effects" <<http://economics.mit.edu/files/32>>