



Nota Técnica

Índice de Ventas Diarias del Comercio Minorista (IVDCM) Fuentes, métodos y resultados^{1 2}

División Estadísticas
Banco Central de Chile

Resumen

El Índice de Ventas Diarias del Comercio Minorista (IVDCM) es un indicador nominal elaborado a partir de registros innominados de ventas con boleta electrónica proporcionados por el Servicio de Impuestos Internos (SII) e incluye el total de transacciones entre empresas y consumidores finales, independiente del medio de pago y canal de venta. Con el objetivo de obtener series de tiempo consistentes, el indicador se construye a partir de una muestra de empresas seleccionadas por su estabilidad de reporte. El IVDCM se actualiza dos veces al mes: el día 5 con datos de la primera quincena del mes anterior y el día 19 con información completa.

El contenido de esta publicación forma parte del proyecto *Estadísticas Experimentales* del Banco Central de Chile, por el cual se elaboran y publican estadísticas que se encuentran en proceso de desarrollo o consolidación –lo que podría reflejarse en revisiones importantes a las series publicadas–, y que se caracterizan por su carácter innovador, ya sea en las fuentes de información, los métodos estadísticos y/o el ámbito de estudio.

1. Antecedentes sobre la medición coyuntural del comercio minorista

La crisis sanitaria profundizó la necesidad de contar con información más oportuna de la actividad económica y de la demanda. La mayor exposición a cambios repentinos en el comportamiento de la población hace necesario contar hoy con información que logre desagregar los índices mensuales, de manera tal de evaluar correctamente los efectos sobre la actividad de *shocks*, tales como cuarentenas, o mejorar las proyecciones económicas.

Actualmente en Chile hay dos instituciones que publican información coyuntural para el seguimiento del comercio minorista: el Instituto Nacional de Estadísticas (INE) y la Cámara Nacional de Comercio (CNC).

El INE publica mensualmente el [Índice de Actividad del Comercio \(IAC\)](#), siendo uno de sus componentes la actividad minorista (división 47 del CIIU4.CL, 2012³). El IAC minorista es un indicador de volumen de tipo Laspeyres con ponderaciones fijas en el año 2013 y que utiliza información

¹ Contacto: contactocentral.bcentral.cl

² Esta nota técnica ha sido actualizada el 29 de julio de 2022.

³ Adaptación chilena del Clasificador Industrial Internacional Uniforme (CIIU), Rev. 4.

principalmente de encuestas a establecimientos⁴. Las ventas son deflactadas utilizando información de índices de precios tales como el IPC e IPP, dependiendo de la subactividad. Este indicador se publica con un desfase aproximado de 30 días respecto del mes de referencia.

Por su parte, la CNC publica el [Índice de ventas del Comercio Minorista](#) y el [Termómetro Semanal de las Ventas Minoristas semanales](#) construido a partir de información provista por las empresas asociadas. Los índices mensuales incluyen desagregaciones regionales y por establecimientos equivalentes o totales, mientras que su estadística semanal realiza un seguimiento de las ventas nominales agregadas. Este último reporte comenzó a ser difundido en medio de la crisis sanitaria con un desfase de 7 días.

2. Objetivo del IVDCM y definiciones generales

a) Objetivo

El objetivo del IVDCM es servir de complemento al índice mensual calculado por el INE, proporcionando información nominal, relativa a una muestra de empresas, para el agregado del comercio minorista, con periodicidad diaria y frecuencia de publicación quincenal.

b) Fuente de información

La fuente de información básica utilizada corresponde al registro de ventas diarias del SII que contiene la venta de bienes y servicios realizadas por unidades legales que operan en el territorio nacional y que están autorizadas a emitir boletas electrónicas. Estos datos son complementados, cuando corresponde, con información de transacciones individuales de boletas electrónicas emitidas por firmas nacionales. Ambas fuentes básicas de información son recibidas en el BCCh en virtud de un convenio de colaboración vigente entre ambas instituciones⁵.

c) Selección y composición de la muestra de firmas

Para seleccionar las empresas pertenecientes al comercio minorista se utilizó el directorio de empresas gestionado por el BCCh para la elaboración de las cuentas nacionales, el cual se nutre principalmente de la actividad económica reportada por los contribuyentes en el formulario 29 del SII (declaraciones de IVA).

El criterio utilizado para la selección de la muestra de empresas fue definido de acuerdo con dos grandes objetivos: (1) el poder contar con una serie temporal con suficiente estabilidad y (2) proveer de un periodo de tiempo que permitiese comparaciones interanuales relevantes para el análisis.

Como resultado de la aplicación de estos criterios, la muestra obtenida quedó conformada por empresas con presencia mensual entre enero 2018 y septiembre 2019, previo a los *shocks* relacionados con el estallido social y la crisis sanitaria.

Por otra parte, cabe destacar que, dado que la obligatoriedad de emisión de boletas electrónicas para todo tipo de empresas se inició en marzo de 2021, la muestra del IVDCM considera un número reducido de firmas, sin embargo, éstas cubren diversos tamaños y subramas al interior de la actividad minorista.

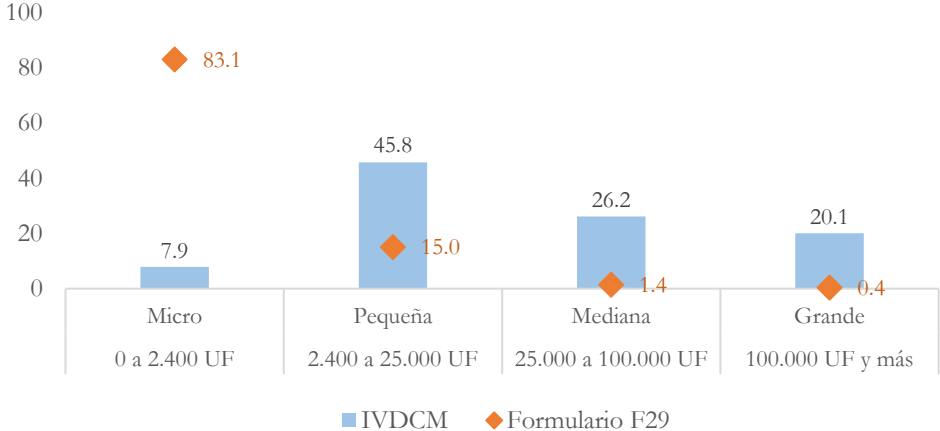
El gráfico 1 muestra la participación por tamaño de empresa, de las firmas seleccionadas para la elaboración del IVDCM versus la distribución que presenta el total de empresas minoristas declarantes

⁴ Ver: [Documento metodológico del Índice mensual de actividad del Comercio](#)

⁵ “Convenio de Intercambio de Información y Colaboración entre el Banco Central de Chile y el Servicio de Impuestos Internos”, firmado el 31 de marzo de 2021.

de IVA, para el año 2019. Del gráfico se desprende que el IVDCM cuenta con una mayor presencia de empresas “grandes”, lo cual se debe a que fueron éstas quienes tempranamente se registraron voluntariamente como emisores de boletas electrónicas.

Gráfico 1: Participación por tamaño de empresa
(Número de empresas como porcentaje del total)

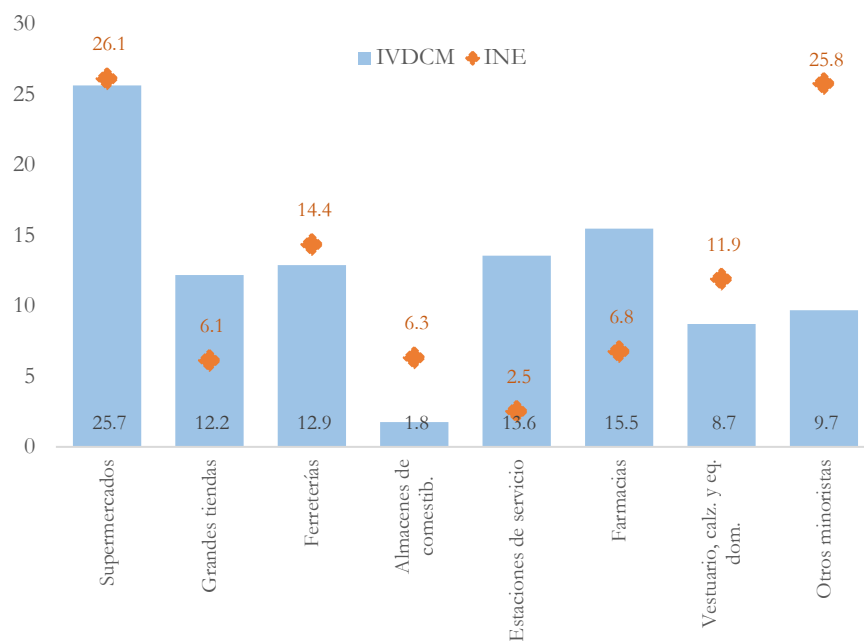


Fuente: Banco Central de Chile en base a información del SII

El hecho de contar con una mayor presencia de grandes empresas tiene un impacto en la importancia relativa de las subactividades minoristas que componen el indicador. Como se muestra en el Gráfico 2, las *Grandes Tiendas* y *Farmacias* tienen una mayor ponderación en el IVDCM, comparado con los ponderadores que se utilizan en la elaboración del IAC del INE. Estas dos subactividades comparten como característica el ser industrias relativamente concentradas, esto es, con un reducido número de participantes. En contraposición, los *Otros minoristas* –donde se clasifican empresas unipersonales–, y los *Almacenes de comestible*, tienen una menor ponderación en el IVDCM.

En adición a la razón antes mencionada, los ponderadores pueden diferir por definiciones metodológicas. El IAC utiliza los ponderadores de Cuentas Nacionales, los que a su vez se estiman a partir del nivel de valor agregado de cada subactividad. El IVDCM, por su parte, se define como un indicador de ventas nominales brutas, más cercano al concepto de gasto de consumo familiar. A modo de ejemplo, esta diferencia metodológica se refleja principalmente en la subactividad de *Estaciones de servicio*, donde la participación en términos de valor agregado está muy por debajo de lo que representan sus ventas brutas en el IVDCM.

Gráfico 2: Ponderadores por subactividad minorista: Contraste IVDCM e IAC-INE (Div. 47)
(porcentaje del total)



Fuente: Banco Central de Chile en base a información del SII

Con respecto a la representatividad de la muestra, esta cubre poco más de un 25% de las ventas totales del comercio minorista de acuerdo con información registrada en el formulario 29 (Cuadro 1).

Cuadro 1: Número de firmas y representatividad de la muestra IVDCM

Subactividad	Número de firmas	Representatividad (%)
Supermercados	27	21.1
Grandes tiendas	29	21.3
Ferreterías	180	39.8
Almacenes de comestibles	189	6.6
Estaciones de servicio	510	56.5
Farmacias	170	55.9
Vestuario, calz. y eq. domést.	282	27.3
Otros minoristas	732	23.5
Comercio Minorista	2.119	25.2

d) Otras definiciones

Si bien se cuenta con información de ocho subactividades minoristas, de acuerdo con lo expuesto en el cuadro 1, la difusión del índice se realizará a nivel agregado del comercio minorista. Esta decisión tiene su fundamento en que aún no se cuenta con una cobertura suficiente en algunas subactividades, en las cuales existen emisores importantes que quedan fuera de la muestra al no presentar una desagregación diaria consistente y estable a través del tiempo.

3. Metodología

a) Depuración de ventas por empresa

La base de datos, por su naturaleza administrativa y frecuencia diaria, está sujeta a diversos errores de reporte, lo que se refleja en la presencia de datos atípicos y faltantes que deben ser solucionados para asegurar estabilidad y la interpretabilidad de las series temporales.

En primer lugar, para detectar datos atípicos, se dispuso del monto de venta y el número de transacciones diarias. Para cada firma, se analiza la distribución que forman ambas variables en un espacio bidimensional, marcando como atípicas las observaciones aisladas que no forman parte de regiones densas. Esto se lleva a cabo mediante el algoritmo de aprendizaje no supervisado *DBSCAN*⁶.

Posteriormente, tanto los datos faltantes originales como las observaciones detectadas como atípicas son imputadas respetando la estacionalidad intra semanal propia de cada firma. Para esto, cada serie temporal es descompuesta de forma aditiva para obtener su componente estacional utilizando el algoritmo *STL decomposition*⁷. Luego, la serie desestacionalizada es interpolada linealmente en los días faltantes para añadir a continuación el componente estacional que fue detectado previamente.

Finalmente, en caso de que una firma presente datos faltantes en el cálculo de una nueva semana coyuntural, se utiliza la información de las firmas pertenecientes a su “vecindario” (métodos pares). Esta decisión se basa en la necesidad de reflejar correctamente cambios drásticos de nivel en las series de ventas, cuestión que fue particularmente relevante durante la emergencia sanitaria. Así, para imputar un dato faltante correspondiente al día t de una empresa en particular i ($Ventas_{i,t}$), se multiplica a su nivel de 7 días atrás ($Ventas_{i,t-7}$) la mediana de los ratios de variación semanal de las n empresas presentes y pertenecientes a la misma subactividad j y estrato de ventas k .

$$Ventas_{i,t} = Ventas_{i,t-7} * Mediana \left(\frac{Ventas_{n,j,k,t}}{Ventas_{n,j,k,t-7}} \right)$$

b) Cálculo del índice

Los agregados de ventas de cada día y_t son obtenidos mediante la sumatoria de las ventas de las n empresas pertenecientes a la muestra del comercio minorista, siendo N el total de firmas.

$$y_t = \sum_{n=1}^N Ventas_{n,t}$$

Una vez obtenidas las ventas totales, se calcula el índice diario con base promedio enero 2018 = 100.

$$IVDCM_t = \frac{y_t}{\left(\frac{\sum_{t=1}^{31} y_t^{enero.2018}}{31} \right)} * 100$$

⁶ Para mayor detalle, ver anexo 1.

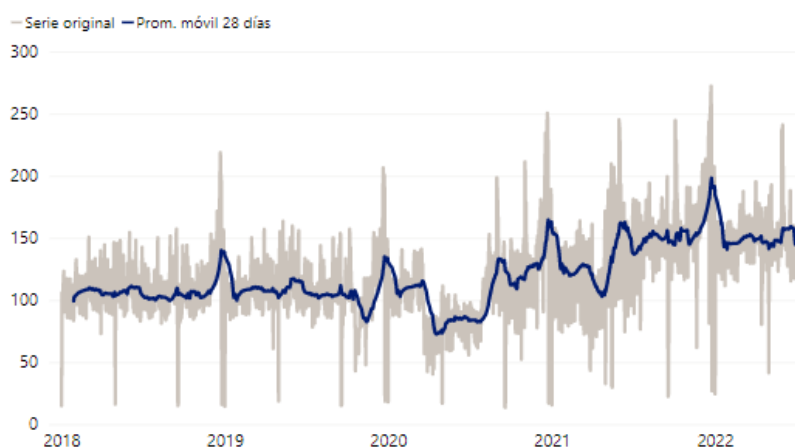
⁷ R. B. Cleveland, W. S. Cleveland, J.E. McRae, and I. Terpenning (1990) STL: A Seasonal-Trend Decomposition Procedure Based on LOESS. Journal of Official Statistics, 6, 3-73.

4. Productos estadísticos

a) Serie diaria

El producto final consiste en un índice diario agregado del comercio minorista, que comienza el 1 enero de 2018 hasta el último dato disponible de acuerdo con el calendario de actualización de datos indicado más adelante (Gráfico 3). Cabe destacar que los puntos altos de la serie corresponden a campañas navideñas o eventos particulares del comercio, mientras que los puntos mínimos responden, en general, a feriados irrenunciables.

Gráfico 3: Índice de ventas diarias del comercio minorista
(Índice enero 2018=100 y promedio móvil de 28 días)



Fuente: Banco Central de Chile en base a información de boletas electrónicas del SII.

5. Calendario de actualización de datos y difusión

Los datos se publicarán con frecuencia quincenal de la siguiente forma:

- Durante la primera semana del mes en curso (día 5 o hábil posterior) se difundirá el indicador incluyendo ventas hasta la primera quincena del mes inmediatamente anterior.
- Durante la tercera semana del mes en curso (día 19 o hábil posterior) se difundirá el indicador conteniendo información completa para el mes inmediatamente anterior.

En cuanto a las revisiones del indicador, al operar dentro del marco provisto por la iniciativa de *Estadísticas Experimentales* del Banco Central de Chile, se deja abierta la posibilidad de corregir las series cada vez que se considere necesario.

La difusión se realizará en el capítulo de “Estadísticas Experimentales” de la Base de Datos Estadísticos (BDE) del Banco Central. Adicionalmente, se presentarán gráficos interactivos en una sección especial de la página web institucional con miras a impulsar el uso de este producto estadístico.

Anexo 1. Algoritmo DBSCAN para la detección de datos atípicos

Contexto

Los algoritmos de agrupamiento o *clustering* son utilizados para realizar tareas de clasificación no supervisada sobre bases de datos que contienen múltiples dimensiones. Sin embargo, su rápida y efectiva aplicación a una colección de varias bases de datos masivas requiere de:

- El menor conocimiento experto posible para determinar los parámetros, dado que frecuentemente no se conoce a priori, por ejemplo, el número de *clusters* o grupos homogéneos de observaciones de una base de datos particular.
- La posibilidad de detectar *clusters* de forma arbitraria, debido a que se pueden presentar diversas formas en las n dimensiones disponibles (esféricas, elongadas, lineales, etc.).
- Ser eficiente computacionalmente dado el contexto de datos masivos.

Los algoritmos tradicionales de *clustering*, ya sean de tipo particional como *K-Means* o jerárquico, no ofrecen garantías para el cumplimiento de estos requerimientos, puesto que son altamente dependientes del criterio experto (por ejemplo, sobre el número de *clusters* presentes) y/o no son flexibles en cuanto a la forma de los *clusters* a identificar. El algoritmo que aquí se describe, llamado DBSCAN (*"Density-Based Spatial Clustering of Applications with Noise"*), lo logra mediante una aproximación basada en densidades, con mínima parametrización. Adicionalmente, este modelo ha sido diseñado para detectar la presencia de observaciones atípicas, que no pertenecen a ninguno de los grupos "válidos".

Algoritmo DBSCAN

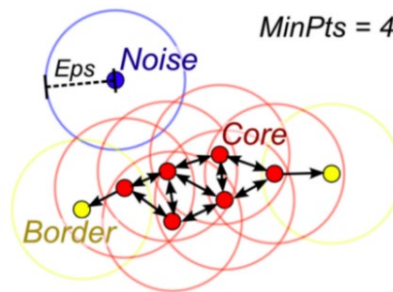
El modelo que propone DBSCAN utiliza una estimación simple del nivel de densidad mínimo requerido para determinar un *cluster* válido, mediante un parámetro que actúa como umbral para el número de "vecinos" u observaciones necesarias, llamado *"Minimum points"* (*minPts*), dentro de un radio *"epsilon"* (ϵ) normalmente medido en distancia *Euclidean*.

Haciendo uso de estos dos parámetros, las observaciones son clasificadas en tres categorías:

- *"Core"*: centroides de cada *cluster*, presentan un vecindario denso, con más de *minPts* dentro del radio ϵ .
- *"Border"*: observaciones que son parte del radio de un centroide (son parte de una región densa), pero no cumplen con tener más de *minPts* dentro de su propio radio ϵ .
- *"Noise"*: observaciones que no son parte de las clasificaciones anteriores, es decir, no son parte de regiones densas.

La categorización que lleva a cabo DBSCAN se ilustra en la Figura 1 en un espacio de 2 dimensiones. Con un parámetro *minPts=4*, los puntos de color rojo son clasificados como *"core"*, los de color amarillo como *"border"*, mientras el punto azul es el único atípico detectado (*"noise"*).

Figura 1: Ilustración algoritmo DBSCAN



Fuente: Vibhor Agarwal, 2019.

Con respecto al proceso computacional, este se puede resumir de forma conceptual en los siguientes pasos:

1. Computar vecinos de cada observación e identificar puntos “core”.
2. Unir observaciones “core” para formar clusters.
3. Para cada observación no clasificada como “core”, hacer:
 - 3.1. De ser posible, agregar a un cluster cercano (punto “border”)
 - 3.2. Si no, clasificar como “noise”

El resultado final entregado por el algoritmo corresponde a la asignación de cada observación a alguno de los *clusters* o grupos válidos detectados, más una categoría adicional para las observaciones atípicas.

Referencias

- Ester, M., H. P. Kriegel, J. Sander, and X. Xu, A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In: Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, Portland, OR, AAAI Press, pp. 226-231. 1996.
- Schubert, E., Sander, J., Ester, M., Kriegel, H. P., & Xu, X. DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems (TODS)*, 42(3), 19. 2017.
- Agarwal, V. Let Us Cluster Datapoints Using DBSCAN. Medium (<http://www.medium.com>). 2019.
Consulta web: Enero de 2022.