

# Los nuevos métodos de recogida de información del Índice de Precios de Consumo

**Cristina de la Fuente**

*Subdirectora Adjunta de Estadísticas Coyunturales. INE*

El Índice de Precios de Consumo (IPC) mide la evolución de los precios de bienes y servicios de consumo adquiridos por los hogares residentes en España. Este índice, por lo tanto, cobra especial importancia en tiempos en los que la inflación es elevada.

En 2022 la inflación media anual fue la más alta de los últimos 36 años.

Como es bien sabido, los usos que se hace del IPC son muy diversos. Además de su utilización como medida de la inflación, se utiliza en los ámbitos económico, jurídico y social, ya que tradicionalmente se utiliza para la revisión de contratos de arrendamiento de inmuebles, como elemento de referencia en la negociación salarial, en la fijación de las pensiones, en la actualización de las primas de seguros y otros tipos de contrato, y como deflactor en la Contabilidad Nacional, por mencionar algunos ejemplos.

Para que la medición de la evolución de los precios sea lo más precisa posible, el IPC debe adaptar sus elementos estructurales a los cam-

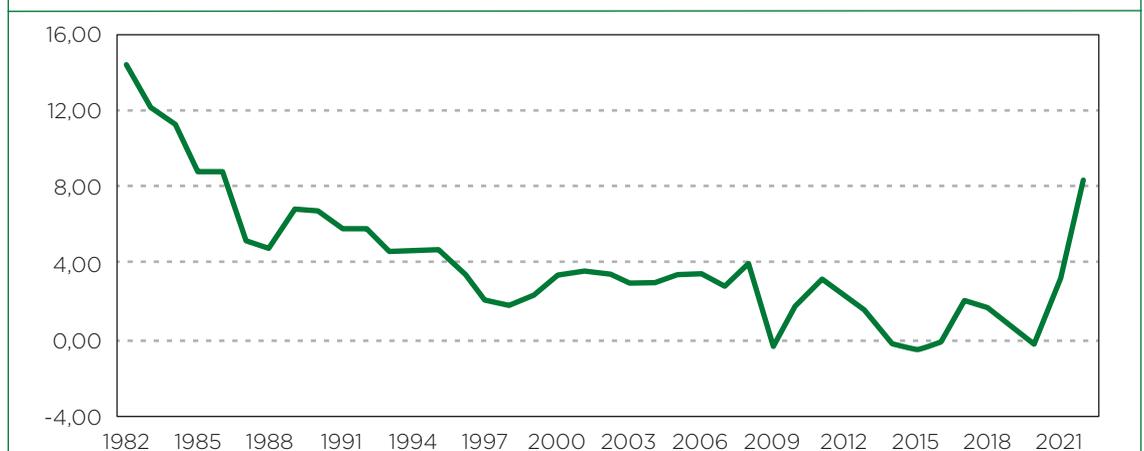
bios que se vayan produciendo en la economía. Esta capacidad de adaptación se consigue con la actualización de sus ponderaciones y de su cesta de productos que se realiza cada año. Con ello se garantiza un alto grado de representatividad de este indicador.

Otro elemento básico que determina la calidad del IPC, y sobre el que pocas veces se habla, es el proceso de recolección de los precios y las características de los productos de la cesta de la compra. En los últimos años, gran parte de los trabajos de los responsables de la encuesta en el INE se han encaminado a la mejora de este proceso mediante la automatización de este y la utilización de bases de datos disponibles en las empresas distribuidoras y en las diferentes páginas web de los diversos informantes. Estos son algunos de los avances conseguidos.

## Hacia la recogida informatizada de precios en el IPC

Actualmente, en el IPC base 2021, se recogen mensualmente unos 210.000 precios, la mayor parte de ellos en los casi 35.000 establecimien-

**GRÁFICO 1.** IPC. ÍNDICE GENERAL. Variaciones medias anuales. Porcentaje



**Fuente:** Instituto Nacional de Estadística.

tos incluidos en la muestra, distribuidos en los 177 municipios seleccionados.

Como la mayor parte de los artículos se recogen mediante visita del personal del INE a los establecimientos de la muestra, el INE trabaja permanentemente para desarrollar nuevos métodos que permitan reducir los costes de recogida, aumentando la precisión de sus estimaciones y, sobre todo, reduciendo los costes y la carga de respuesta a sus informantes.

La visita personal a los establecimientos garantiza que el precio recogido corresponde al de venta al público y que el producto para el cual se recoge el precio es el mismo a lo largo del tiempo. Está recogida en el establecimiento, además de la toma de los precios, conlleva la recopilación de información relevante para su correcto seguimiento, como, por ejemplo, el cambio de alguna de sus características o la necesidad de su sustitución por haber desaparecido del mercado.

Por este motivo, durante el año 2023, se está trabajando para poder implantar la recogida informatizada, utilizando dispositivos electrónicos en vez de hacer las correspondientes anotaciones en un cuestionario en papel. Con esta medida se conseguirá que esta información se obtenga con mayor rapidez, así como una reducción los potenciales errores de grabación y por lo tanto se tendrá una mayor eficiencia en el tratamiento de toda la información recogida.

### La implantación de *Scanner data*

Siguiendo también esta línea de trabajo, en enero de 2020 el INE comenzó a utilizar las bases de datos de algunas empresas, con información sobre las ventas de sus productos, para sustituir la recogida de los precios en esos establecimientos. Este método, denominado *scanner data*, consiste en la obtención y utilización para el cálculo del IPC de la información de las ventas en la línea de caja de los establecimientos, con información del número de unidades vendidas y los ingresos de cada uno de esos productos, codificados según la clasificación propia de cada empresa y con el código EAN (los 13 dígitos del código de barras del artículo).

El método *scanner data* no es exclusivo del INE, sino que ya se usa en varios países de nuestro entorno, como Francia, Holanda e Italia

entre otros. Se trata de una forma precisa, integral y eficaz de medir la evolución de precios, por lo tanto, esta implantación se ha realizado bajo el apoyo de Eurostat, que promueve su utilización en su papel de armonizador de los IPC de los estados miembros de la UE.

La información que se ha incorporado al IPC con *scanner data* es la de productos de gran consumo (alimentos envasados, bebidas, productos de limpieza y mantenimiento del hogar, productos de parafarmacia, alimentos y productos para mascotas y artículos de cuidado personal). Gracias a esta implantación, se ha eliminado la recogida de precios de estos artículos en aquellos establecimientos de las cadenas que aportan sus bases de datos.

La forma de obtención de la información no es el único cambio que conlleva el uso del *scanner data*. Como la naturaleza de los datos es diferente (unidades vendidas e ingresos frente al precio) y el volumen de datos a tratar es muy superior, el procedimiento de cálculo no puede ser el mismo que el utilizado en los artículos con recogida tradicional.

La primera tarea que hubo que llevar a cabo para el correcto tratamiento de los datos de *scanner data* es la correspondencia de clasificaciones, que es imprescindible para asociar automáticamente la variedad de los artículos recogida con la clasificación ECOICOP utilizada en IPC.

Al igual que en la recogida tradicional de IPC, se necesita seleccionar, dentro de cada parcela ECOICOP, una muestra de los artículos representativos que se empleará para medir la evolución de precios de dicha parcela. En el caso de *scanner data* se realiza con la información de los ingresos obtenidos el año anterior, actualizándola cada año para incorporar los artículos que han empezado a consumirse durante ese año y excluyendo los que dejan de hacerlo.

Dentro de los artículos seleccionados hay que identificar las variedades más vendidas de cada uno para que formen parte de la cesta. Esta selección se realiza de manera dinámica cada mes dentro de cada una de las cadenas, lo que evita tener que estar imputando hasta final de año aquellas variedades que dejan de venderse y poder así incorporar otras nuevas de mayor consumo. La exclusión de variedades se realiza cuando estas dejan de consumirse durante tres

meses seguidos (para evitar desechar aquellas variedades cuyo consumo ha disminuido o desaparecido de forma puntual). Este método combina la representatividad de la cesta con la comparabilidad a lo largo del tiempo, ya que calcula las variaciones mensuales con las variedades que están en la cesta en ambos meses.

Otra diferencia con el método de cálculo del IPC en la recogida tradicional reside en el cálculo de los índices elementales: se tienen en cuenta los valores unitarios de cada variedad incluida en la cesta de *scanner data*. Este valor unitario, calculado como ingresos de la variedad dividido para las unidades vendidas, no tiene por qué coincidir con el precio unitario de la variedad, ya que no se corresponde con una única transacción, sino con todas las que se han producido en las tres semanas de las que se recoge información y en las que ha podido tener distintos precios.

El índice elemental en los artículos de *scanner data*, igual que en los artículos de recogida tradicional, es el nivel más bajo de agregación en el que se calcula un índice sin utilización de ponderaciones. A diferencia del cálculo del IPC tradicional, en el que el índice elemental es el cociente de los precios medios del mes actual y el de diciembre del año anterior, el IPC con artículos de *scanner data* se calcula a partir de las tasas de variación mensuales de las medias geométricas de los valores unitarios. Es decir, el índice elemental del mes  $m$  es el índice elemental del mes anterior multiplicado por la tasa de variación del mes actual, ya que solo entre cada par de meses consecutivos el contenido de la cesta de variedades de cada artículo es homogéneo.

Finalmente, la agregación del IPC de recogida tradicional con el IPC de *scanner data* se realiza a nivel de subclase-provincia, al ser el

mínimo nivel de agregación común entre ambas cestas. Para realizar esta agregación se dispone de pesos a nivel de subclase-provincia distinguiendo además por tipo de recogida.

El tamaño de la muestra de *scanner data* es de 493 bienes de consumo, frente a los 462 bienes y servicios de consumo mediante recogida directa de precios.

### Desarrollo del método *Web scraping*

No toda la recogida tradicional se realiza en persona en los establecimientos seleccionados en la muestra. El sistema de recogida de precios es diferente según las características del artículo.

Cuando la gestión se optimiza si la recogida se realiza de manera centralizada, los precios de los artículos se recogen desde los servicios centrales del INE. Por ejemplo, son artículos de recogida centralizada aquellos que tienen los mismos precios en todo el país, los que sus precios están sujetos a tarifas publicadas en el BOE o los que son comercializados por pocas empresas (telefonía, seguros, electricidad, gas, etc.).

Aprovechando esta recogida centralizada de ciertos artículos se está trabajando también en la obtención de datos a través de *web scraping*, es decir, obtención masiva de precios a través de internet. Conceptualmente es muy similar a la metodología tradicional, en cuanto al uso de precios de oferta, pero es necesario desarrollar herramientas específicas para la obtención de los datos. Estas herramientas tienen por objetivo localizar, extraer y organizar el contenido de las bases de datos de las páginas web de las empresas que ofertan los productos incluidos en la cesta de artículos del IPC, que tienen estructuras muy diferentes entre sí. Actualmente se recogen de esta manera los precios de apartamentos turísticos y se está trabajando en otras parcelas del mismo sector económico.

Estas son las líneas principales en las que se está trabajando en el ámbito del IPC. Las líneas directrices, como se ve, están perfectamente establecidas, pero todavía queda mucho camino por recorrer para que los nuevos métodos de recolección de información descritos se extiendan a la mayor parte de la cobertura de este indicador. Esta es la tarea por hacer en los próximos años. ●

#### Para saber más...

- Base de datos IPC: [https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica\\_C&cid=1254736176802&menu=ultiDatos&idp=1254735976607](https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176802&menu=ultiDatos&idp=1254735976607)
- Principales cambios IPC base 2021: [https://www.ine.es/metodologia/t25/principales\\_caracteristicas\\_base\\_2021.pdf](https://www.ine.es/metodologia/t25/principales_caracteristicas_base_2021.pdf)
- Metodología scanner data: [https://www.ine.es/metodologia/t25/ipc\\_scanner\\_data.pdf](https://www.ine.es/metodologia/t25/ipc_scanner_data.pdf)