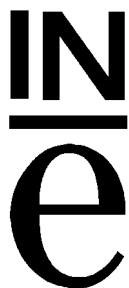


INSTITUTO NACIONAL DE ESTADÍSTICA



Principales novedades metodológicas del Índice de Precios de Consumo Base 2021

Madrid, enero de 2022

1. Introducción

La base 2016 del Índice de Precios de Consumo (IPC) se caracterizó por ser la primera que abrió las puertas hacia una nueva forma de concebir este indicador. Con su implantación se incorporó por primera vez en el IPC información procedente de las bases de datos de los supermercados al cálculo del índice (lo que se denomina *scanner data*). Asimismo, en esta misma línea de aprovechamiento de la información disponible en distintas fuentes, se inició el desarrollo de aplicaciones para la obtención de datos de Internet (web scraping).

Con estos cambios se dio uno de los pasos más decisivos en la concepción de este indicador en los últimos años. Con ello, sin duda, la precisión y la eficiencia del IPC aumenta progresivamente, a medida que se va incorporando información de más empresas.

En esta línea de trabajo, el IPC base 2021, es la materialización de muchos de los desarrollos iniciados en la base anterior, que renuevan el indicador y lo orientan hacia una nueva concepción de su producción, que se apoya cada vez más en el uso de bases de datos de las empresas y el aprovechamiento de los avances tecnológicos.

Los cambios introducidos en la nueva base 2021 se pueden agrupar en dos grandes bloques:

A. Cambios relacionados con la estructura

- Actualización de la cesta de la compra
- Actualización de la estructura de ponderaciones

B. Tratamientos metodológicos

- Introducción de nuevo método de tratamiento del vestido
- Ajustes en los criterios de las ofertas
- Cambio en la estimación de la falta de precio

Además, como se ha dicho, a lo largo de la base 2021 se continuará trabajando en el proceso para automatizar la recolección de información. En este proceso se incluyen los siguientes métodos.

C. Nuevos métodos para la obtención de datos

- Web scraping
- Scanner data
- Recogida informatizada

2. Principales novedades IPC base 2021

A. CAMBIOS RELACIONADOS CON LA ESTRUCTURA

En este bloque se incluyen las revisiones que afectan a elementos estructurales de la encuesta. El objetivo de las mismas es mantener el IPC actualizado, de forma que represente adecuadamente los hábitos de consumo de los hogares.

Para ello, se actualizan los dos aspectos básicos que reflejan el comportamiento de los hogares frente al consumo: el contenido de la muestra de bienes y servicios que componen la denominada cesta de la compra, y la estructura de ponderaciones, o importancia relativa de los mismos, basada en el gasto que los hogares realizan en cada uno de ellos.

La **revisión de la cesta de la compra**, incluye el análisis del contenido de la misma, desde el punto de vista de la pertinencia de que los artículos que la componen figuren o no en ella, en función de la importancia del gasto relativo de cada uno. De esta forma, se excluyen aquellos que han perdido relevancia en el conjunto del consumo en los últimos años, y se incorporan los que han empezado a figurar como importantes dentro de los patrones de consumo de los hogares.

Como consecuencia de estos ajustes, **la cesta de la compra del IPC base 2021 pasa a tener 955 artículos** (de los cuales 462 son de recogida tradicional y el resto se recogen mediante scanner data), frente a los 977 de la base anterior (con 480 de recogida tradicional).

Asimismo, el número de subclases (mayor nivel de desagregación para las que se publican índices) se reduce desde las 221 de la base 2016 a las 199 en la base 2021.

La siguiente tabla contiene la relación de subclases que dejan de publicarse en la nueva base.

Subclases que desaparecen en el IPC, base 2021

Código	Subclase
01143	Leche en conserva
01152	Margarina y otras grasas vegetales
03132	Artículos de mercería
03141	Limpieza de artículos de vestir
03220	Reparación y alquiler de calzado
04323	Servicios de mantenimiento de los sistemas de calefacción
05121	Alfombras y moquetas
05323	Planchas
05402	Cubertería
05511	Grandes herramientas y equipos con motor
06121	Pruebas de embarazo y anticonceptivos no orales
07224	Lubricantes
07362	Servicios de mudanza y guardamuebles
08109	Otros servicios postales
09111	Equipos para la recepción, registro y reproducción de sonido
09113	Dispositivos audiovisuales portátiles
09121	Cámaras
09132	Accesorios para equipos de procesamiento de información
09149	Otros soportes
09222	Grandes bienes duraderos para ocio en interiores
09322	Equipos de acampada y ocio al aire libre
09331	Productos de jardinería
09522	Revistas y publicaciones periódicas
12520	Seguro relacionado con la vivienda

Por su parte, las subclases nuevas en el IPC, base 2021 son las siguientes:

Subclases nuevas. IPC, base 2021

Código	Subclase
06129	Otros productos médicos n.c.o.p.
07213	Accesorios para vehículos personales

El otro elemento que se actualiza es la **estructura de ponderaciones**. En este caso, se trata de obtener una nueva estructura, que se adapte a las pautas de consumo de los hogares, a partir de la información de gasto proporcionada por la Encuesta de Presupuestos Familiares (EPF).

Aunque la estructura de ponderaciones del IPC se revisa cada año, es en el proceso de cambio de base cuando se acomete al mayor nivel de desagregación funcional y geográfico. La actualización se realiza a niveles de 7 dígitos, con información procedente de la EPF, así como del gasto en consumo final de los hogares de la Contabilidad Nacional, y utilizando estadísticas disponibles en los diferentes sectores de actividad de la economía.

La siguiente tabla muestra las ponderaciones de los doce grandes grupos en el conjunto nacional utilizadas en el año 2021, y su correspondencia con las que entran en vigor en el año 2022:

Ponderaciones de grupos (tanto por cien)

Grupo	2021	2022	Variación (%)
01. Alimentos y Bebidas no alcohólicas	23,6	22,6	-4,2
02. Bebidas alcohólicas y Tabaco	3,2	3,1	-2,5
03. Vestido y calzado	6,4	6,0	-5,5
04. Vivienda	13,6	14,2	4,9
05. Menaje	5,9	5,8	-2,4
06. Medicina	3,9	4,4	11,1
07. Transporte	12,4	13,0	4,2
08. Comunicaciones	3,7	3,6	-4,6
09. Ocio y Cultura	6,8	6,4	-6,2
10. Enseñanza	1,7	1,6	-4,1
11. Hoteles, cafés y restaurantes	11,6	13,0	12,0
12. Otros bienes y servicios	7,1	6,3	-11,1
TOTAL	100	100	:

B. TRATAMIENTOS METODOLÓGICOS

B.1. Cambio del tratamiento del vestido

En el IPC algunos productos reciben tratamientos específicos diferentes de los aplicados a la mayoría de los artículos de la cesta, debido a sus especiales características. El vestido es uno de ellos.

Hay dos elementos que definen las características del mercado del vestido y determinan las oscilaciones de sus precios: por un lado, los **periodos de rebajas**, y por otro, y estrechamente ligado al anterior, las marcadas **temporadas** en las que aparecen y desaparecen los diferentes tipos de prendas de los establecimientos. Ambos factores no solo determinan la tendencia de los precios sino que exigen la aplicación de una metodología específica que permita la medición precisa de la evolución de los precios.

Las rebajas en el vestido

En el año 2002 se incluyó en el IPC la recogida de los precios con oferta o descuento y rebajados, lo que supuso la implantación de un nuevo patrón en la tendencia de este indicador. En el caso de los precios con oferta esto supuso una mayor oscilación de los mismos en el corto plazo, ya que las políticas de descuentos de las compañías comercializadoras no siguen un patrón estacional necesariamente.

Las campañas de rebajas del vestido y del calzado, en cambio, en el momento de la implantación en el IPC sí que tenían un patrón estacional muy marcado, ya que a comienzos del año 2000 el calendario de inicio y finalización de las mismas estaba regulado por ley.

En la actualidad, sin embargo, los periodos de rebajas ya no están regulados, lo que hace que a la marcada estacionalidad de las rebajas tradicionales de invierno y de verano se le hayan ido añadiendo otros periodos con reducciones de precios. Así, por ejemplo, en numerosas ocasiones se aprecia la prolongación en el tiempo de los periodos tradicionales de rebajas, ya sea anticipándose o alargando dicho periodo. Además, a lo largo del año se producen reducciones de precios en cualquier momento, desligado del periodo tradicional.

Este comportamiento del mercado del vestido no está adecuadamente recogido en el IPC, que se ciñe al esquema de rebajas tradicional, obviando comportamientos de los precios en periodos intermedios.

La temporada en el vestido

El otro elemento diferenciador del vestido frente al resto de artículos de la cesta de la compra es la estacionalidad. A lo largo del año, hay dos temporadas claramente marcadas: primavera-verano y otoño-invierno. El inicio de ambas supone la introducción en el mercado de prendas propias de cada estación, lo que exige una metodología específica para una estimación precisa de la evolución de los precios.

El principal problema de la temporalidad en la práctica es que el calendario de entrada de cada temporada difiere en años consecutivos. Así, en un establecimiento de la muestra pueden estar disponibles ya los artículos de la temporada de primavera-verano cuando se visita a principios del mes de marzo en un año determinado, mientras que ese mismo mes del año siguiente, pueden no encontrarse aún en esa misma semana sino en la siguiente, en la que ya no se visita. Esto supone un problema desde la óptica de la medición del IPC, ya que la comparación interanual de los precios (que es la forma de analizar su evolución en este tipo de artículos) se ve truncada por ese motivo, ya que los precios de la nueva temporada entrarían en el IPC en abril, en lugar de en marzo, como el año anterior.

Tratamiento actual hasta diciembre de 2021

Las variaciones de precios de los artículos de vestido y calzado se deben únicamente al cambio de temporada o a las rebajas. Por tanto, para facilitar la interpretación de la evolución de los precios interanual, el proceso aplicado consistía en replicar un mismo esquema cada año, tratando de que el número de precios rebajados o en temporada en cada provincia, sea

similar un año y el siguiente, dejando pendientes para el mes siguiente, o estimando si aún no se recogieron, las ofertas y las variaciones de temporada, lo que estabiliza la tasa interanual.

Evidentemente, este tratamiento suponía una modificación consciente del momento de inclusión en el IPC de los precios recogidos en campo en aras de dotar de estabilidad a la tasa anual.

Cambio en el tratamiento del vestido en la base 2021

Debido a que, como se ha comentado anteriormente, las ofertas son cada vez más frecuentes en estos artículos, para una correcta representación de la realidad del mercado, los precios de la nueva temporada y con rebaja deberían incorporarse en el momento en que se recojan en el establecimiento, al igual que ocurre en el resto de productos donde también se aplican descuentos. Como se ha dicho, esto conlleva la ruptura del esquema entrada-salida de temporada y de rebajas en años consecutivos, y con ello una evolución de las tasas de variación anuales más oscilantes.

Este cambio de tratamiento es necesario para que el IPC recoja de forma más precisa la tendencia de los precios en el corto plazo, sin embargo su incorporación va a provocar una ruptura en la serie, por lo que las tasas anuales a lo largo del año 2022 vendrán influidas en parte por este cambio introducido.

B.2. Tratamiento de las ofertas

Umbrales de disminución de precios

Desde que se comenzaron a recoger los precios con descuento en el IPC, en la base 2001, se definió un umbral mínimo para las disminuciones de los precios, y se situó en el 50%. Sin embargo, unos años después esta limitación se eliminó, ya que se consideró que cualquier descuento, sea de la magnitud que sea, si es representativo, se debe tener en cuenta.

La consecuencia de no establecer umbrales de disminuciones de precios es que las subidas correspondientes a bajadas de más del 50% son proporcionalmente mayores que la disminución, con lo que las repercusiones en los agregados del IPC son mucho mayores que las de los descensos (por ejemplo: una bajada de un 50%, se corresponde con una subida del 100% cuando el precio vuelve al nivel inicial, pero con una disminución del 70%, el aumento es de un 233%). Esto sucede en unas pocas ocasiones, pero cuando ocurre los agregados del IPC se ven afectados de forma muy sensible ante un comportamiento puntual de un precio.

Por la experiencia acumulada, en el cambio de base 2021 se restablecerán los límites establecidos originalmente.

B.3. Tratamiento de la falta de precio

La falta de precio no es uno de los problemas que más afectan al IPC. Cuando ésta se produce, el precio se estima con la información que proporciona la recogida de los precios en los demás establecimientos donde el producto buscado sí que tiene precio. Esta estimación se realiza, como mucho, durante dos meses consecutivos, ya que en el tercer mes el producto ausente debe haber sido sustituido.

El método de estimación del precio consiste en aplicar la tasa de variación media del resto de precios recogidos en la provincia, siempre que hayan variado más de la mitad de los mismos. Si no han variado más de la mitad de los precios, entonces se considera que la estimación más precisa es la repetición del mismo.

Este método, a pesar de tener una base metodológica consistente, llega en numerosas ocasiones a plantear problemas debido a que el número de precios de cada artículo puede ser escaso para que se pueda aplicar el criterio. Por ello, se ha considerado necesario ampliar la cobertura de precios con los que se estime.

Cambios en el tratamiento de la falta de precio en la base 2021

El cambio introducido consiste en utilizar la variación de precios del producto para el conjunto nacional para estimar la falta de precio.

C. NUEVOS MÉTODOS PARA LA OBTENCIÓN DE DATOS

A lo largo de los últimos años la información utilizada por el IPC proveniente de bases de datos de empresas o bien recopilada por métodos telemáticos ha ido en aumento. Esto evita las visitas a los establecimientos para recolectar los precios y las principales características de los productos de la muestra.

Los procedimientos más utilizados en la mayoría de los países de nuestro entorno son *web scraping*, basado en la extracción de la información de las páginas web de las empresas más representativas de cada sector, y *scanner data*, consistente en solicitar a la empresa las bases de datos de ventas de los productos en todos sus establecimientos.

A estos dos métodos se unirá en los próximos meses la recogida de los precios en los establecimientos mediante dispositivos informáticos, lo que dotará al IPC de un mayor dinamismo, y aumentará el grado de precisión en el proceso de recogida y grabación de la información.

La base 2021, irá extendiendo paulatinamente la utilización de estos tres métodos de registro de la información a la parte de la cesta de la compra en la que sea factible.

C.1. Implantación de web scraping

Se denomina *web scraping* al proceso mediante el cual se localiza y se recopila información de forma automática de la web. Para ello, es preciso programar diferentes aplicaciones

informáticas que se adapten a los diversos diseños y particularidades de cada una de las páginas web, que ofertan los productos incluidos en la cesta de la compra del IPC, así como sus principales características y sus precios de venta al público. Dichas aplicaciones tienen como objetivo localizar, extraer y organizar el contenido de las bases de datos que utilizan las empresas.

Se trata, por tanto, de una herramienta muy útil para la captación de los precios de manera automática y que puede suplir, siempre que se considere factible, la recolección de precios en los establecimientos físicos por la recopilación de la información de la web.

A lo largo de la base 2021 se trabajará en la incorporación de este método en aquellos sectores donde su aplicación optimice los resultados. A continuación se describen los principales aspectos metodológicos para su implantación.

Metodología para la adaptación de web scraping al IPC

El proceso de implantación de web scraping en el IPC lleva acarreadas un conjunto de tareas cuyo objetivo es adaptar las características de las bases de datos así obtenidas a la estructura de cálculo del IPC. Las fases del proceso de adaptación son comunes en todos los casos, sin embargo para cada página web de la que se extraiga la información el trabajo dentro de cada fase debe ser único, dada la exclusividad de cada una de ellas.

Las fases para la implantación son las siguientes:

- **Programación de la extracción de la información.** El motor de extracción de la información debe programarse para que ésta se obtenga de forma automática con una frecuencia que habitualmente será semanal. Su diseño debe tener en cuenta las características de cada página web, la clasificación y jerarquización de los productos y la información disponible de cada uno.

Según las características del producto y del mercado donde este se comercializa se puede optar por dos tipos de extracción:

- **Extracción completa.** Se trata de obtener toda la información de la página web del informante, sin discriminar los productos por sus características. De esta forma, se dispone de toda la información sobre precios y características de los productos y no solo las de aquellos contenidos en la muestra limitada del IPC.
 - **Extracción selectiva.** Se trata de localizar y extraer únicamente aquellos productos cuyas características se adapten a los de la muestra del IPC. Es, por tanto, una equiparación plena de la recogida en web a la recogida en establecimientos.
- **Adaptación de la clasificación a la ECOICOP**

Este es uno de los temas cruciales en cualquier método de obtención automatizada de la información. Cada compañía comercializadora tiene su propia clasificación, lo que hace necesario un trabajo meticuloso de asignación de los productos a las parcelas propias de la ECOICOP. Esta adaptación puede ser directa, si cada producto se

relaciona con una parcela concreta, o puede requerir un análisis en detalle del contenido de ambas clasificaciones antes de establecer la correspondencia.

Por otra parte, aunque normalmente la clasificación no se ve modificada con demasiada frecuencia, es posible que con el paso del tiempo la empresa decida realizar algún cambio en la misma, por lo que uno de los controles de calidad de la información obtenida por este método debe enfocarse en la comprobación de las clasificaciones.

- **Control de calidad de la información**

Se diseñará un sistema de control de la calidad de la información obtenida. En esencia, este debe basarse en la comprobación del número de registros, elementos comunes entre dos periodos consecutivos, análisis de los elementos no comunes, comprobación de la estabilidad de la clasificación utilizada, etc.

- **Cálculo de precios medios e índices**

El proceso de cálculo se realiza para cada página web de forma independiente. Esto significa que la integración con el IPC no se hace incorporando los precios recolectados por web scraping al cálculo del indicador, sino que se calcularán previamente los precios medios para cada producto, posteriormente los respectivos índices y sus agregaciones, y serán estos los que se incorporen a su respectiva agregación de la ECOICOP, que puede ser a diferente nivel según cada caso.

C.2. Extensión de scanner data

Se denomina *Scanner Data* al proceso de obtención y utilización para el cálculo del IPC de la información procedente de las bases de datos de las empresas, que contienen el registro de las ventas realizadas de cada uno de sus productos en la línea de caja de todos sus establecimientos. Habitualmente, esta información consiste en el número de unidades vendidas y los ingresos para cada uno de los productos comercializados, codificados según la clasificación propia de cada empresa.

El método *scanner data* ya está siendo utilizado en varios países de nuestro entorno, ya que se trata de una alternativa más eficiente, precisa y completa de medir la evolución de los precios. Por ello, la oficina de estadística europea, EUROSTAT, promueve su utilización en el ámbito de la armonización de los índices de precios de los estados miembros de la UE.

A lo largo de los últimos tres años el INE, basándose en la experiencia de otros países, ha desarrollado el modelo metodológico más adecuado para el tratamiento de la información proveniente de las cadenas de supermercados e hipermercados y su posible integración en el cálculo del IPC. Una vez finalizada la fase de diseño metodológico se ha iniciado el proceso de implantación, cuyo punto de partida fue enero de 2020, cuando comenzó la incorporación de manera gradual de las diferentes cadenas de establecimientos, en la medida en que se va disponiendo de la información.

De momento la información que se incorporará al IPC será la relativa a 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).

La consecuencia inmediata de la implantación de *Scanner Data* es la eliminación de la recogida de precios en aquellos establecimientos de las cadenas incorporadas. Pero no es el único cambio. Debido a la naturaleza de la información contenida en las bases de datos (fundamentalmente, número de unidades vendidas por producto e ingresos obtenidos) el procedimiento de cálculo no puede ser el mismo que el utilizado hasta ahora en el IPC.

Próximas incorporaciones al procedimiento scanner data

Las etapas para extender *Scanner Data* en el IPC serán las siguientes:

- **Contacto e incorporación de otras compañías de gran consumo**

Se continuarán los trabajos de contacto y captación de las principales compañías comercializadoras de gran consumo. Se incluyen, las más importantes en cuanto a cuota de mercado a nivel nacional y se iniciarán los trabajos de identificación de las cadenas más importantes a nivel regional.

Este proceso de incorporación de las cadenas a nivel nacional y regional es constante. Por ello, no es imprescindible que finalice cuando termine el cambio de base, sino que permanecerá abierto hasta conseguir las cadenas más representativas del mercado.

- **Estudio de viabilidad de aplicación de scanner data a otros productos.**

Además de productos de gran consumo, se va a establecer una relación de los productos susceptibles de ser incorporados a este proyecto. Se estudiará cada sector y se decidirá cuáles son sus principales distribuidores, para realizar los contactos.

- **Estudio de nuevas fórmulas de cálculo de scanner data**

En los países europeos que llevan más tiempo utilizando el scanner data en el cálculo del IPC se ha planteado la utilización de fórmulas de cálculo alternativas, diferentes de la que habitualmente se usan en el IPC, basadas en índices tipo Laspeyres. Se estudiarán estas nuevas fórmulas (métodos multilaterales) y las ventajas e inconvenientes de su utilización.

C.3. Implantación de la recogida informatizada

Dentro del proyecto de modernización de los instrumentos y procedimientos de obtención de la información para el IPC, la utilización de dispositivos electrónicos para la recogida de los precios en los establecimientos tiene una importancia capital.

Es preciso señalar que el proceso de encuesta a los establecimientos no solo implica la anotación de los precios de los productos de la muestra, sino que está acompañado del registro de una información adicional absolutamente relevante para el correcto seguimiento de los mismos, como es cualquier alteración que se haya producido en el producto desde la anterior visita, o la propuesta de un cambio de producto o establecimiento si estos han dejado de ser representativos o han desaparecido del mercado.

En la actualidad, toda esta información se anota en los cuestionarios en papel, que posteriormente debe ser grabada en las dependencias del INE, para su procesamiento e inicio del proceso de validación, control y cálculo de los índices. La implantación de la recogida informatizada facilita y automatiza todo este proceso, lo cual redundará en una mayor rapidez en disponer de la información grabada (el proceso actual de anotación en papel y posterior grabación se reduce a la grabación de la información en el momento en que ésta es observada), una disminución de los potenciales errores de grabación y una mayor eficiencia en el tratamiento de la información adicional recogida.