

INSTITUTO NACIONAL DE ESTADÍSTICA



Metodología de cálculo del *scanner data* en el IPC e IPCA

Subdirección General de Precios y Presupuestos Familiares

Febrero 2020

1. Introducción

La mayor parte de los precios utilizados en el cálculo del Índice de Precios de Consumo (IPC) se recogen mediante visita personal a los establecimientos. Este sistema garantiza que el precio recogido es el de venta al público, y que el producto para el cual se recoge el precio es el mismo a lo largo del tiempo. En ambos casos es el propio encuestador quien certifica su cumplimiento.

La muestra del IPC está diseñada para que éste represente a la mayor parte de la población, y cubra los aspectos más relevantes de su comportamiento como consumidores. Así, los criterios utilizados para la selección de los municipios se basan en el tamaño poblacional y su ubicación geográfica dentro de cada provincia. A su vez, en cada uno de los municipios se definen las zonas comerciales más representativas, en la cuales se eligen los establecimientos con mayor afluencia de público.

Por su parte, para la determinación del número de precios que se recogen de cada artículo se tiene en cuenta su variabilidad (son necesarios más precios cuanto mayor sea la frecuencia de variación) y la importancia o ponderación del artículo (cuanto más relevante sea un artículo en la cesta de la compra, será necesario dedicar más recursos para asegurar la precisión de la medición).

De esta forma, en el Índice de Precios de Consumo (IPC) base 2016, el número de precios que se procesan mensualmente es aproximadamente 220.000. La mayor parte de los mismos se recogen en los alrededor de 33.000 establecimientos incluidos en la muestra, distribuidos entre los 177 municipios seleccionados.

El sistema de recogida de precios es diferente según las características del artículo. Así, artículos cuyos precios son los mismos en todo el país o su gestión se optimiza si se realiza de forma centralizada, se recogen en los servicios centrales del INE. Sin embargo, para la mayor parte de artículos de la cesta de la compra la recolección de precios se realiza mediante la visita del personal del INE a los establecimientos de la muestra seleccionada.

Este sistema de selección de la muestra y los procedimientos de recogida de la información garantiza un indicador con un alto grado de calidad y precisión, a la altura de los mejores IPC de los países de nuestro entorno. Sin embargo, el INE trabaja constantemente para desarrollar nuevos métodos que permitan producir sus estadísticas de forma más eficiente, reduciendo los costes de recogida, aumentando la precisión de sus estimaciones y, sobre todo, tratando de disminuir la carga que conlleva responder a las mismas por parte de los informantes.

En esta línea de trabajo, en los últimos años el INE ha implantado en muchas de sus estadísticas la utilización de nuevos métodos y técnicas para la obtención de la información, basados en la explotación de registros administrativos y en el uso de dispositivos electrónicos de recogida.

En el caso del IPC, la recogida de los precios en los establecimientos mediante dispositivos electrónicos será una realidad en el corto plazo. En cuanto a la obtención

de la información a partir de registros de las empresas, el INE ha trabajado en los últimos años en el desarrollo del método denominado *scanner data*, que consiste en utilizar las bases de datos de las empresas, con información sobre ventas de sus productos, como sustitutos de la recogida de los precios en los establecimientos.

Este documento presenta las principales características de este método, y detalla el procedimiento de cálculo de índices a partir de la información obtenida de las bases de datos, así como su integración en el IPC.

2. Qué es *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. Como no podía ser de otra manera, España se ha sumado a la propuesta, para lo cual el INE inició un proyecto piloto con el objetivo de evaluar todos los aspectos sobre la posible implantación de *scanner data* en el cálculo del IPC y del IPCA.

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 será enero de 2020 y se realizará de manera gradual en la medida en que se vaya disponiendo de la información de las diferentes cadenas de establecimientos.

En esta primera fase 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 que aportan sus bases de datos. 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.

Desde sus inicios, hubo que tener en cuenta dos grandes cuestiones para abordar la adaptación de este nuevo sistema de obtención de información a los métodos habituales del IPC:

A. Obtención de la información

Una de las razones fundamentales por las que *scanner data* es un método en auge en la mayoría de los países de nuestro entorno es que la información necesaria para llevarlo a cabo no requiere una elaboración específica por parte de las empresas informantes.

La información básica para el desarrollo del proyecto es, para cada producto codificado por la empresa y en cada establecimiento, la siguiente: ingresos o valores unitarios, cantidades, denominación del producto, descripción (si existe algún campo donde se distinga), código (EAN¹ o el interno que utilice la empresa) y clasificación interna.

El formato de las bases de datos, el sistema de transmisión de las mismas y los demás aspectos relacionados con la disposición de la información los debe decidir la empresa, para que la carga sea la menor posible. Evidentemente, esto condiciona los trabajos que debe abordar el INE, ya que deben diseñarse procesos para adaptarse a los diferentes contenidos, formatos y sistemas de acceso de las bases de datos de cada cadena que participe en el proyecto. Por tanto, existirán tantos procedimientos de obtención y gestión de las bases de datos como empresas informantes.

B. Utilización de la información

Uno de los principales dilemas que se plantan para utilizar este tipo de información en el IPC es de tipo conceptual: la incorporación de datos sobre ventas de las compañías exige cambios en los métodos y las definiciones utilizados tradicionalmente en la metodología de cálculo de este indicador. Los principales retos conceptuales son los siguientes:

- **Diferencias entre los conceptos precio y valor unitario**

El IPC mide, por definición, la evolución de los precios de los bienes y servicios adquiridos por los hogares. Se recoge, por tanto, el precio de venta al público en cada establecimiento. La utilización de *scanner data*, sin embargo, cambia esta filosofía ya que exige que para cada código EAN, se utilice su valor unitario (total de ingresos / total de unidades vendidas), pero no el precio propiamente dicho.

En realidad, el valor unitario no se corresponde con una única transacción real sino que representa a todas las realizadas a lo largo de un periodo de tiempo fijado. Esto supone un cambio en la definición del IPC y en los distintos tratamientos aplicados, como los de descuentos y ofertas. Sin embargo, teniendo en cuenta el objetivo del IPC y a la vista de los resultados que se obtendrán con la utilización de las bases de datos, se ofrecerá una visión global de la evolución de los precios del mercado.

- **Diferencias entre producto y la gama completa de variedades**

La utilización de las bases de datos permite disponer de la información de todas las variedades vendidas de un producto. Por ejemplo, supongamos que para el aceite de oliva, en un establecimiento, pueda haber en torno a 250 variedades (250 EAN distintos). Esto difiere del procedimiento habitual del IPC que, por su concepción, realiza el seguimiento de precios de una única variedad en cada establecimiento.

¹ EAN: European Article Number (13 dígitos del código de barras del artículo)

- **Volumen de información**

Con *scanner data* el volumen de datos es significativamente superior al que se obtiene con la recogida tradicional de precios. Por ello, además de los requisitos técnicos para el tratamiento de dicha información también hay que introducir novedades metodológicas, para determinar las variedades que cada mes deben formar parte de la muestra.

Estas cuestiones conceptuales no suponen escollos insalvables que impidan la utilización de *scanner data* en el IPC, pero conllevan la necesidad de desarrollar nuevos sistemas de gestión de grandes bases de datos, diseños metodológicos para el tratamiento de la información y nuevos métodos de validación de los resultados.

El siguiente apartado se centra en estos nuevos métodos destinados a adaptar la información recibida de las empresas al cálculo del IPC.

3. Proceso de cálculo de los índices de precios a partir de la información generada por el método *scanner data*

- **Correspondencia de clasificaciones**

El gran reto de *scanner data* es obtener índices de precios compatibles con los del IPC a partir de la información contenida en las bases de datos de las empresas, ya que dichos índices deberán integrarse con las agrupaciones funcionales en que está estructurado el IPC, basadas en la ECOICOP.

Por tanto, el primer paso necesario para el uso de *scanner data* es buscar la **correspondencia entre la clasificación de productos utilizada internamente por la empresa colaboradora y la clasificación de artículos ECOICOP utilizada en el IPC**. Esta asignación es imprescindible para poder asociar de forma automática las variedades con los artículos a los que hacen referencia únicamente a través de la familia de productos en la que está encuadrado.

- **Selección de artículos**

Al igual que en el IPC de recogida tradicional, se debe seleccionar una muestra de artículos representativos dentro de cada parcela ECOICOP, cuya evolución de precios representará la evolución de la parcela.

En el caso de *scanner data* esta cesta de artículos se realizará en función de los ingresos del año anterior, descartando aquellos artículos cuyo consumo no es representativo. Esta cesta de artículos permanecerá fija durante un año entero, y se revisará a finales de cada año, pudiéndose incluir artículos que han empezado a consumirse durante el año y excluyendo aquellos que dejaron de hacerlo.

Cada cadena de establecimientos incluida en el cálculo del IPC tendrá su cesta de artículos *scanner data* representativa, en función de los artículos que se venden en dicha cadena.

- **Selección de variedades**

Aunque uno de los principios que subyace en *scanner data* es el aprovechamiento máximo de toda la información disponible en las bases de datos de las empresas, la práctica indica que parte de dicha información puede no ser útil para el propósito del proyecto, y puede llegar incluso a ser una rémora en el proceso de cálculo. Se trata de las variedades cuyas ventas no son significativas y/o son inestables en el tiempo, lo que las hace poco representativas de la evolución global de los precios del artículo, requiriendo métodos de corrección para que su entrada o salida del mercado no condicione la evolución del índice.

Por ello, es necesario seleccionar las variedades más vendidas de cada artículo para que formen parte de la muestra. Esta selección se puede realizar de dos maneras diferentes, o bien de **forma estática**, es decir, se realiza la selección una vez al año, con la información del año anterior (como para la selección de los artículos) y se mantiene estable a lo largo del año, o bien se puede establecer una **muestra dinámica**, donde mes a mes se selecciona la muestra de variedades que entrarán en el cálculo del IPC.

El método estático se basa en el mantenimiento de una cesta de variedades fija. Esto tiene como consecuencia que para las variedades que dejan de comercializarse a lo largo del año, deben realizarse imputaciones de su valor unitario hasta el inicio del año siguiente, cuando se realizará la nueva selección de variedades. Asimismo, aunque las variedades fuesen estables a lo largo del año, con el método estático pueden darse casos en los que se mantengan en la muestra variedades que han dejado de ser representativas en gasto y, por el contrario, no incorporando otras nuevas de mayor consumo. Se estaría desaprovechando así el gran potencial de *scanner data*.

En cambio, el método dinámico, en el que las variedades seleccionadas para el cálculo puede variar entre dos meses consecutivos, elimina de forma automática aquellos productos que han dejado de ser representativos e incorpora los nuevos, siempre que las magnitudes de los ingresos sean significativas. Y combina la representatividad de la cesta con la comparabilidad a lo largo del tiempo, ya que calcula las variaciones mensuales con las variedades en ambos meses.

Como ya hemos señalado anteriormente, cada artículo está compuesto por numerosas variedades cuya identificación es su respectivo código EAN. El proceso para seleccionar qué productos forman parte del cálculo y cuáles deben desaparecer cada mes, se realiza de forma independiente en cada uno de los artículos, para cada uno de los formatos de establecimiento (hipermercado, supermercado grande, supermercado pequeño, etc.) que tenga la cadena y en cada provincia.

- **Fórmula de cálculo del IPC de scanner data**

A continuación se describe el procedimiento de cálculo del IPC de *scanner data*, desde la selección de variedades de cada producto hasta su índice agregado. Hasta la obtención de los índices agregados, este proceso se realiza para cada una de las variedades, formato de establecimiento y provincia.

1) Selección de variedades comunes

A partir de los datos agregados de las tres primeras semanas completas del mes de referencia m se seleccionan las variedades comunes entre el mes m y el mes anterior $m-1$.

Nota: A lo largo de este apartado, cuando se haga mención al mes m se refiere a esas tres primeras semanas completas de dicho mes.

2) Cálculo de los porcentajes de ingresos de cada variedad dentro del artículo

Para cada una de las variedades seleccionadas como se ha descrito en el apartado anterior, se calcula el peso, en tanto por ciento, que tiene cada variedad dentro del artículo al que pertenece, tanto para el mes m como para $m-1$:

$$\text{Peso variedad}_{i,f,p}^{m,t} = \frac{\text{Ingreso variedad}_{i,f,p}^{m,t}}{\sum_{i \in j} \text{Ingreso variedad}_{i,f,p}^{m,t}} * 100 ,$$

donde:

$\text{Ingreso variedad}_{i,f,p}^{m,t}$ son los ingresos, en euros, de la variedad i -ésima del artículo scanner data j , del formato de establecimiento f y la provincia p , en el mes m del año t .

3) Selección de las variedades que se utilizarán para el cálculo de la evolución de precios del artículo scanner data

Para asegurar que las variedades que van a formar parte del cálculo son representativas del consumo, entre las variedades comunes en los dos meses, únicamente se utilizarán para el cálculo aquellas que tengan un porcentaje significativo de ventas. Además, las variedades seleccionadas deben cubrir un alto porcentaje de las ventas totales del artículo.

Así, para la selección de variedades, se utilizará el siguiente algoritmo:

$$\frac{\text{Peso variedad}_{i,f,p}^{m-1,t} + \text{Peso variedad}_{i,f,p}^{m,t}}{2} > \frac{1}{N_{f,p} * \delta} * 100 ,$$

donde:

$\text{Peso variedad}_{i,f,p}^{m-1,t}$ es el peso de la variedad i -ésima, común entre m y $m-1$, del artículo scanner data j , del formato de establecimiento f y la provincia p , en el mes $m-1$ del año t ,

$\text{Peso variedad}_{i,f,p}^{m,t}$ es el peso de la variedad i -ésima, común entre m y $m-1$, del artículo scanner data j , del formato de establecimiento f y la provincia p , en el mes m del año t ,

$N_{f,p}$ es el número de variedades del artículo scanner data j , del formato de establecimiento f y la provincia p , comunes entre los meses m y $m-1$ del año t , y

$\delta = 1,25$, que hace que el umbral determinado por la fórmula seleccione las variedades comunes de modo que se cubra un porcentaje de ventas superior al 80%.

Además, a las variedades seleccionadas de este modo se le añadirán las que formaron parte de la muestra del mes anterior pero no han sido seleccionadas en el actual, ya sea porque en el mes m no se han vendido o sí se han vendido, pero no lo suficiente para superar el umbral.

Si pasados tres meses, la variedad sigue sin superar el umbral de gasto de selección o desaparece, la variedad se eliminará definitivamente de la cesta. Es decir, una vez que una variedad se selecciona un mes porque ha tenido un gasto significativo, se mantendrá en la muestra durante tres meses consecutivos aunque no tenga consumo. De esta forma, se trata de evitar desechar variedades cuyo consumo haya disminuido o desaparecido de forma puntual.

4) Cálculo de valor unitario (VU) de cada variedad

Para cada variedad incluida en la muestra, se calcula el valor unitario como el cociente entre los ingresos y el número de unidades vendidas:

$$VU \text{ variedad}_{i,f,p}^{m,t} = \frac{\text{Ingresos variedad}_{i,f,p}^{m,t}}{\text{Unidades vendidas variedad}_{i,f,p}^{m,t}},$$

donde:

$VU \text{ variedad}_{i,f,p}^{m,t}$ es el valor unitario, en euros, de la variedad i -ésima del artículo scanner data j , del formato de establecimiento f y la provincia p , en el mes m del año t ,

$\text{Ingresos variedad}_{i,f,p}^{m,t}$ son los ingresos, en euros, de la variedad i -ésima del artículo scanner data j , del formato de establecimiento f y la provincia p , en el mes m del año t , y

$\text{Unidades vendidas variedad}_{i,f,p}^{m,t}$ son las unidades vendidas de la variedad i -ésima del artículo scanner data j , del formato de establecimiento f y la provincia p , en el mes m del año t .

Es importante señalar que este valor unitario **no tiene por qué coincidir con el precio de adquisición del producto**, ya que no se corresponde con una única transacción, sino con todas las que se han producido a lo largo de las tres semanas en las que dicho producto ha podido tener distintos precios.

5) Cálculo de la variación mensual del artículo

A partir de los valores unitarios de las variedades que componen la cesta del mes m , se calculará la variación mensual del artículo, del siguiente modo:

- a. Se calcula la **media geométrica de los valores unitarios del mes m** del año t , de las variedades seleccionadas en el mes m del año t :

$$\overline{VU}_{j,f,p}^{m,t} = \sqrt[n_{j,f,p}^{m,t}]{\prod_{i=1}^{n_{j,f,p}^{m,t}} VU \text{ variedad}_{i,f,p}^{m,t}}$$

, donde:

$\overline{VU}_{j,f,p}^{m,t}$ es la media del valor unitario del artículo scanner data j , del formato de establecimiento f y la provincia p , en el mes m del año t ,

$VU \text{ variedad}_{i,f,p}^{m,t}$ es el valor unitario de la variedad i -ésima del artículo scanner data j , del formato de establecimiento f y la provincia p , en el mes m del año t , y

$n_{j,f,p}^{m,t}$ es el número de variedades seleccionadas del artículo scanner data j , del formato de establecimiento f y la provincia p , en el mes m del año t .

b. Se calcula la **media geométrica de los valores unitarios del mes (m-1)** del año t , de las variedades seleccionadas en el mes m del año t .

$$\overline{VU}_{j,f,p}^{(m-1),t} = \sqrt[n_{j,f,p}^{m,t}]{\prod_{i=1}^{n_{j,f,p}^{m,t}} VU \text{ variedad}_{i,f,p}^{(m-1),t}}$$

, donde:

$\overline{VU}_{j,f,p}^{(m-1),t}$ es la media del valor unitario del artículo scanner data j , del formato de establecimiento f y la provincia p , en el mes $m-1$ del año t ,

$VU \text{ variedad}_{i,f,p}^{(m-1),t}$ es el valor unitario de la variedad i -ésima del artículo scanner data j , del formato de establecimiento f y la provincia p , en el mes $m-1$ del año t , y

$n_{j,f,p}^{m,t}$ es el número de variedades seleccionadas del artículo scanner data j , del formato de establecimiento f y la provincia p , en el mes m del año t .

Como se puede ver en la formulación, las variedades que intervienen en las medias geométricas de los dos meses son las mismas, ya que así la variación del artículo únicamente reflejará variación de precios y no variación debida a la diferente composición en la cesta de variedades.

c. Se calcula la **variación mensual del artículo**, como la variación de los valores unitarios medios:

$$\Delta_{j,f,p}^{m,t} = \left(\frac{\overline{VU}_{j,f,p}^{m,t}}{\overline{VU}_{j,f,p}^{(m-1),t}} - 1 \right) * 100$$

, donde:

$\Delta_{j,f,p}^{m,t}$ es la variación del artículo scanner data j , del formato de establecimiento f y la provincia p , en el mes m del año t , y

$\overline{VU}_{j,f,p}^{m,t}$ es el valor unitario del artículo scanner data j , del formato de establecimiento f y la provincia p , en el mes m del año t .

$\overline{VU}_{j,f,p}^{m-1,t}$ es el valor unitario del artículo scanner data j , del formato de establecimiento f y la provincia p , en el mes $m-1$ del año t .

Cabe reseñar que la media de valores unitarios del mes $m-1$, utilizada para los cálculos de la variación mensual del mes m , será distinta de la que se utilizó para calcular la variación mensual del mes $m-1$ (ya que en esta última, las variedades seleccionadas eran las comunes entre $m-1$ y $m-2$)

6) Cálculo de índices elementales referidos a diciembre del año anterior

En el ámbito de IPC, un agregado elemental es el componente de consumo de más bajo nivel de agregación para el cual se obtienen índices, y en cuyo cálculo no intervienen ponderaciones. A los índices de estos agregados se les denomina **índices elementales**.

En el IPC se calcula un índice elemental para cada artículo de la cesta de la compra en cada una de las provincias, por lo que el agregado elemental es el *artículo-provincia*.

De manera análoga, en el caso de *scanner data* un agregado elemental representa a un artículo *scanner data*, en un formato de establecimiento determinado y en una provincia concreta, es decir, *artículo-formato-provincia*. Por tanto, la fórmula de cálculo de un índice elemental es la siguiente:

$$dic(t-1)I_{j,f,p}^{m,t} = dic(t-1)I_{j,f,p}^{(m-1),t} * \left(1 + \frac{\Delta_{j,f,p}^{m,t}}{100} \right)$$

, donde:

$dic(t-1)I_{j,f,p}^{m,t}$ es el índice del artículo scanner data j , del formato de establecimiento f y la provincia p , en el mes m del año t , referido a diciembre del año $t-1$,

$dic(t-1)I_{j,f,p}^{m-1,t}$ es el índice del artículo scanner data j , del formato de establecimiento f y la provincia p , en el mes $m-1$ del año t , referido a diciembre del año $t-1$, y

$\Delta_{j,f,p}^{m,t}$ es la variación mensual del artículo scanner data j , del formato de establecimiento f y la provincia p en el mes m del año t .

Nota: a diferencia del cálculo del IPC tradicional, en el que el índice elemental se puede expresar como un cociente de los precios medios del mes actual y el de diciembre de t , el IPC con artículos de *scanner data*, se calcula a partir de las tasas de variación mensuales debido a que, como ya se ha dicho, solo entre cada par de meses consecutivos el contenido de la cesta de variedades de cada artículo es homogéneo.

7) Cálculo de índices agregados referidos a diciembre del año anterior

Cálculo de ponderaciones

Las ponderaciones de cada artículo *scanner data* representan la relación entre el gasto realizado en dicho artículo y el gasto total realizado en todos los artículos vendidos por la cadena que forman parte del índice.

Por tanto, las ponderaciones en vigor durante el año t se calcularán en base a las ventas anuales de cada artículo, referidas al año anterior, teniendo en cuenta todas las

semanas del año (no las tres semanas mensuales que se utilizan para los cálculos de las variaciones mensuales del IPC *scanner data*), de la siguiente manera:

$$W_j^t = \frac{\text{ingresos del artículo } j \text{ en el año } (t - 1)}{\text{ingresos totales en el año } (t - 1)}$$

Estas ponderaciones son diferentes para cada formato de establecimiento en cada una de las agregaciones geográficas (provincias, comunidades autónomas y conjunto nacional) y a partir de ellas se obtienen las ponderaciones de las distintas agregaciones funcionales. Así, la ponderación del agregado funcional A se obtiene como suma de las ponderaciones de los artículos *scanner data* que componen dicha agregación:

$$W_A^t = \sum_{j \in A} W_j^t$$

Agregaciones funcionales dentro de una provincia

El índice, referido a diciembre del año anterior, de una agregación funcional S (5 dígitos ECOICOP) en una provincia p , se obtiene como agregación de los índices elementales de los artículos pertenecientes a dicha agregación, en cada uno de los formatos, con las ponderaciones vigentes en el año t :

$$dic(t-1)I_{S,p}^{m,t} = \sum_{j \in S} \sum_{f=1}^3 dic(t-1)I_{j,f,p}^{m,t} * W_{j,f,p}^t$$

donde:

$dic(t-1)I_{j,f,p}^{m,t}$ es el índice referido a diciembre de $(t-1)$ del artículo j , del formato de establecimiento f y la provincia p , en el mes m del año t , referido a diciembre del año $(t-1)$,

$W_{j,f,p}^t$ es la ponderación del artículo j , del formato de establecimiento f y la provincia p , en vigor durante el año t , y

S son las subclases ECOICOP a 5 dígitos.

Con estas agregaciones podemos obtener los índices de cualquier artículo *scanner data*, agregando los formatos de establecimiento de la cadena.

Del mismo modo, podemos obtener los índices de una subclase ECOICOP en un formato determinado, agregando los índices de los artículos de ese formato, que pertenecen a esa subclase:

$$dic(t-1)I_{f,S,p}^{m,t} = \sum_{j \in S} dic(t-1)I_{j,f,p}^{m,t} * W_{j,f,p}^t$$

Agregaciones geográficas de una agregación funcional

De la misma forma que en el caso anterior, el cálculo del índice de una agregación geográfica R superior a la provincia, para una agrupación funcional determinada A , se calcula de la siguiente manera:

$$dic(t-1)I_{A,R}^{m,t} = \sum_{p \in R} dic(t-1)I_{A,p}^{m,t} * W_{A,p}^t, \text{ donde:}$$

$dic(t-1)I_{A,p}^{m,t}$ es el índice referido a diciembre de $(t-1)$ de la agrupación funcional A en la provincia p , y

$W_{A,p}^t$ es la ponderación de la agregación funcional A en la provincia p , en vigor durante el año t .

- **Integración del IPC de recogida tradicional con el IPC de *scanner data***

La integración del IPC de *scanner data* con el IPC de recogida tradicional se realizará a nivel de subclase/provincia de IPC, ya que es la mínima agregación común entre la cesta de recogida tradicional y la cesta *scanner data*. Para la agregación de estos conjuntos de índices es preciso disponer de los pesos relativos de ambos.

Cálculo de ponderaciones

A partir de las ponderaciones del IPC vigentes cada año, se debe calcular una nueva estructura en la que cada cadena de supermercados tendrá su propio peso en cada una de las provincias.

Este peso será igual en todas las subclases en las que se va a incorporar el *scanner data* y en sus respectivos agregados elementales. La parte de la ponderación de la subclase/provincia que corresponda a la cadena se repartirá entre los agregados elementales *scanner data* y el resto de la ponderación entre los agregados elementales de recogida tradicional, proporcionalmente.

Por tanto, las ponderaciones del IPC recogida tradicional + *scanner data* se calculan de la siguiente manera:

- **Para cada artículo j de la cesta *scanner data*, dentro de la subclase S :**

$${}_{SC}\tilde{W}_{j,f,p}^t = {}_{RT}W_{S,p}^t * coef_p^C * \frac{{}_{SC}W_{j,f,p}^t}{\sum_{j \in S} \sum_f {}_{SC}W_{j,f,p}^t}, \text{ donde}$$

${}_{SC}\tilde{W}_{j,f,p}^t$ es la ponderación del artículo *scanner data* j , del formato de establecimiento f y la provincia p , en vigor durante el año t , dentro del IPC de recogida tradicional+*scanner data*,

${}_{RT}W_{S,p}^t$ es la ponderación de la subclase S del IPC de recogida tradicional, en la provincia p , en vigor durante el año t ,

$coef_p^C$ es el peso, en tanto por uno, de la cadena C frente a la recogida tradicional, en la provincia p , y

${}_{SC}W_{j,f,p}^t$ es la ponderación del artículo *scanner data* j , del formato de establecimiento f y la provincia p , en vigor durante el año t .

- Para cada artículo k de la cesta de recogida tradicional, dentro de la subclase S :

$${}_{RT}\tilde{W}_{k,p}^t = {}_{RT}W_{S,p}^t * (1 - coef_p^C) * \frac{{}_{RC}W_{k,p}^t}{\sum_{k \in S} {}_{RC}W_{k,p}^t} = {}_{RC}W_{k,p}^t * (1 - coef_p^C),$$

donde

${}_{RT}\tilde{W}_{k,p}^t$ es la ponderación del artículo de recogida tradicional k , en la provincia p , en vigor durante el año t , dentro del IPC de recogida tradicional+scanner data,

${}_{RT}W_{S,p}^t$ es la ponderación de la subclase S del IPC de recogida tradicional, en la provincia p , en vigor durante el año t ,

$coef_p^C$ es el peso, en tanto por uno, de la cadena C frente a la recogida tradicional, en la provincia p , y

${}_{RC}W_{k,p}^t$ es la ponderación del artículo de recogida tradicional k , en la provincia p , en vigor durante el año t .

De esta manera se obtiene una estructura de ponderaciones en la que la ponderación de la subclase S en la provincia p será la suma de la ponderación de recogida tradicional y la ponderación de *scanner data*:

$${}_{RC+SC}\tilde{W}_{i,p}^t = \sum_{j \in S} \sum_f {}_{SC}\tilde{W}_{j,f,p}^t + \sum_{k \in S} {}_{RT}\tilde{W}_{k,p}^t$$

Cálculo de índices agregados referidos a diciembre de $(t-1)$

Con esta nueva estructura de ponderaciones y con los índices elementales, tanto de *scanner data* como de recogida tradicional, se obtienen índices agregados funcionales referidos a diciembre de $(t-1)$.

El índice referido a diciembre del año anterior, de una agregación funcional A en una provincia p en el mes m del año t se calculará de la siguiente manera:

$$dic(t-1)I_{A,p}^{m,t} = \sum_{j \in A} \sum_f dic(t-1)I_{j,f,p}^{m,t} * {}_{SC}\tilde{W}_{j,f,p}^t + \sum_{k \in A} dic(t-1)I_{k,p}^{m,t} * {}_{RT}\tilde{W}_{k,p}^t,$$

donde

$dic(t-1)I_{j,f,p}^{m,t}$ es el índice del artículo scanner data j , del formato de establecimiento f y la provincia p , del mes m del año t , referido a diciembre del año $(t-1)$,

${}_{SC}\tilde{W}_{j,f,p}^t$ es la ponderación del artículo scanner data j , del formato de establecimiento f y la provincia p , en vigor durante el año t , dentro del IPC de recogida tradicional+scanner data,

$dic(t-1)I_{k,p}^{m,t}$ es el índice del artículo de recogida tradicional k , en la provincia p , del mes m del año t , referido a diciembre del año $(t-1)$,

${}_{RT}\tilde{W}_{k,p}^t$ es la ponderación del artículo de recogida tradicional k , en la provincia p , en vigor durante el año t , dentro del IPC de recogida tradicional+*scanner data*.

Cálculo de índices en base 2016

Los índices referidos a diciembre de $(t-1)$ se encadenan a los correspondientes índices en base 2016. Así el índice en base 2016, de una agregación funcional A en una provincia p en el mes m del año t será:

$${}_{16}I_{A,p}^{m,t} = {}_{16}I_{A,p}^{dic(t-1)} * \frac{dic(t-1)I_{A,p}^{m,t}}{100}$$

Para $t=2019$, los ${}_{16}I_{A,p}^{m,t}$ son los calculados únicamente con recogida tradicional. Para $t=2020$, el ${}_{16}I_{A,p}^{dic19}$ será, por tanto, el calculado con recogida tradicional y los $dic(t-1)I_{A,p}^{m,t}$ los calculados como unión del IPC de recogida tradicional y el IPC de *scanner data*.

Los índices para las distintas agregaciones geográficas se calcularán de manera análoga.