

Nota acerca del IRPH. Un análisis desde la ESTADÍSTICA

Informe realizado por JUAN ETXEBERRIA MURGIIONDO¹.

El presente análisis versa, desde el punto de vista estadístico, sobre el índice conocido como IRPH y definido como *“la media simple de los tipos de interés medios ponderados por los principales de las operaciones de préstamos con garantía hipotecaria de plazo igual o superior a tres años para adquisición de vivienda libre que hayan sido iniciadas o renovadas en el mes que se refiere el índice por el conjunto de entidades”*². En el mismo se concluye que:

1.- Este índice es manifiestamente mejorable como indicador de tendencia central y para ello, al menos, se debería efectuar una ponderación de las entidades en función del número de operaciones y de la cuantía de las mismas, que se efectúan en el período utilizado para el cálculo de la media ponderada de las operaciones efectuadas por la propia entidad.

2.- Otro aspecto que contribuiría a que el citado índice fuera un indicador más “robusto” de la tendencia central de los índices de las entidades, consistiría en eliminar las consecuencias del sesgo y desviación al valor central de la distribución que los valores atípicos tienen en el resultado del índice como consecuencia de la fórmula utilizada para el cálculo del mismo.

3.- Derivado de la fórmula utilizada se demuestra la posibilidad y capacidad de las entidades que aportan los datos para el cálculo del índice, para incidir, influir, manipular y condicionar el resultado final tanto de cada una de ellas de forma individual como de forma grupal, en el caso de que las mismas decidieran hacerlo de forma coordinada. Y, evidentemente, las entidades saben y son conscientes de esa potencialidad. Cada entidad de forma autónoma, es decir sin necesidad de ponerse de acuerdo con otras, puede prever de forma precisa y exacta las consecuencias que los incrementos (positivos o negativos) en los

¹ Acerca del autor: Juan Etxeberria Murgiondo es **Catedrático Acreditado de Estadística Aplicada** (fecha de acreditación por la ANECA: 19 de Enero de 2015). **Licenciado en Ciencias Exactas (especialidad Estadística)** el año 1979 y **Doctor en Ciencias de la Educación**, 1987. **Profesor titular de Estadística Aplicada** adscrito al departamento de Métodos de Investigación y Diagnóstico en Educación de la Universidad del País Vasco – Euskal Herriko Unibertsitatea. Se ha especializado en los Métodos de Regresión y Análisis Multivariantes. Es autor, entre otros, de los libros: *Programación y análisis estadísticos básicos con SPSS/PC+*, *Análisis de datos y textos*, *R Commander: Gestión y análisis de datos*, *Regresión Múltiple*, *Análisis Descriptivo de Datos en Educación* y *Análisis Inferencial de Datos en Educación* y *Estatistika Aplikatua: teoria eta praktika*. Así mismo, es autor de numerosos artículos en el ámbito de la Estadística Aplicada. Es director de la colección “Cuadernos de Estadística” y de la colección “Manuales de Metodología de Investigación Educativa” ambos de la Editorial La Muralla.

² El informe es igualmente aplicable a los ya desaparecidos IRPH Cajas e IRPH Bancos, pues su método de cálculo era idéntico, simplemente limitando su alcance a las cajas y a los bancos.

Esta definición junto con otros aspectos relacionados con el IRPH son el resultado del análisis de diferentes fuentes de información. De todas ellas, quiero destacar el documento de fecha 2 de Junio de 2015 en el que el Banco de España contesta al interrogatorio planteado por los abogados Maite Ortiz y José María Erauskin en el marco del procedimiento ordinario 81/2014 del Juzgado de lo Mercantil número 1 de Burgos.

tipos de interés de las operaciones realizadas en un período de tiempo determinado van a tener en el resultado del cálculo posterior del IRPH.

4.- El hecho de incluir las comisiones que las entidades aplican a los clientes en la información de los tipos medios ponderados que se envía al Banco de España para que efectúe el cálculo del nuevo índice tiene un efecto inflacionista a la vez que multiplicador (por valores superiores a la unidad) de dicha comisión, pues en el cálculo del posterior índice se incluirá el efecto de las comisiones, que se incrementará con la nueva comisión que la entidad vaya a aplicar al cliente en la siguiente operación.

El IRPH como Medida de Tendencia Central.

La Estadística provee un conjunto de herramientas que posibilitan “observar, ver y describir” la realidad que, a veces, se esconde detrás de amplios conjunto de datos. Es más, aporta reglas de observación y tratamiento de datos para que, mediante índices muy definidos, se pueda representar, de la forma más fehaciente posible, la distribución de gran número de datos que, se supone, están bien medidos (sin errores de medida). Entre estos conjuntos de indicadores, cuando se trata de una única variable, se encuentran las medidas de centralización y las de variabilidad (medidas de tendencia central y de dispersión). Así mismo están las medidas de asimetría y curtosis que obviaré en este análisis.

El objetivo de las medidas de tendencia central es reflejar y visualizar mediante un único indicador el valor central de un amplio conjunto de datos. Entre dichas medidas destaca, por su amplio uso, la Media Aritmética (en términos coloquiales media), aunque no debemos desdeñar otros índices como la mediana o variantes de la Media Aritmética.

La media es una medida adecuada cuando los valores a considerar en el cálculo de la misma representan al mismo número de sujetos (la mayor parte de las ocasiones cada valor es el referente de un elemento unitario). Así, en un conjunto de estudiantes, la media (aritmética) de las edades puede considerarse como un indicador válido de la tendencia central de todas las edades. Si queremos calcular la media de edad de las personas que un día están en un centro escolar deberíamos anotar la edad de todas y cada una de esas personas y calcular la media aritmética de las mismas.

¡Sin embargo no siempre es así!

Una mala utilización del procedimiento de cálculo de la media puede inducir a cometer errores que distorsionen la “visualización” del verdadero valor central del colectivo que queremos describir. Así, si la media de edad del profesorado es de 43 años, si la del alumnado es de 13 y la del Personal de Administración y Servicios es de 37, es un Error Estadístico, presentar como medida de tendencia central de la edad de todo el colectivo la media de los valores 43, 13 y 37: Media = 31 años. Este error es debido a que son muchas las personas que conforman el subconjunto del alumnado y menos las del profesorado y el PAS. Este error, habitual por otra parte, ha servido como motivo de muchas chanzas sobre el mal uso, abuso y mentiras de la Estadística *“Son hamburguesas de pollo y caballo. Tiene mezclada la carne de un caballo con la de un pollo”*.

Para subsanar este posible error, en el caso que conozcamos la media de las edades y el número de personas de cada uno de los diferentes subconjuntos, la Estadística posibilita abreviar dicho cálculo. Se trata de la denominada Media Ponderada. La lógica de la fórmula estadística que se utiliza en su cálculo es exactamente la misma que la de la media aritmética. Simplemente se tiene en cuenta el tamaño de cada subconjunto y se procede al cálculo. Esta media, que se utiliza desde tiempos ignotos en el comercio, construcción,.... para el cálculo de los precios de las mezclas de productos de diferente calidad, las resistencias de las argamasas,... posibilita obtener un indicador que refleja mejor la tendencia central de la distribución.

Este es el **primer error** (en el orden en que redacto este documento) que está detrás del procedimiento del cálculo del IRPH: suponer que cada una de las entidades tiene el mismo peso, es decir que tienen el mismo número de elementos (operaciones). Si calculamos la media de sus medias, estamos desvirtuando el objetivo de las medidas de tendencia central, al menos si es éste el objetivo del IRPH. Para corregir este error sería suficiente con el cálculo de la MEDIA PONDERADA de las mismas teniendo en cuenta el número de operaciones (y el volumen o cantidad de las mismas) que realiza cada entidad.

LA IMPORTANCIA DE LOS DATOS ATÍPICOS: OUTLIERS.

Un dato se considera atípico cuando su diferencia con respecto de la media es grande (habitualmente una diferencia se considera grande cuando supera dos desviaciones típicas). Desde el punto de vista estadístico, es ampliamente conocida la importancia diferencial que los datos atípicos (*outliers*) tienen en el cálculo de la media, así como las repercusiones que la existencia de dichos datos puede tener en que la media no “represente” el valor central de la distribución y, por tanto, no cumpla con el requisito exigido a toda medida de tendencia central: representar la centralidad. Así, en el caso de la media simple (media aritmética) de las personas que están en el aula, incluir la edad del profesor alterará de forma sustancial la edad media del colectivo de alumnos. También la importancia de la edad del profesor es mucho mayor que la de una persona del alumnado, en el sentido de que si eliminamos del conjunto de personas al profesor, el resultado de la media va a tener una modificación muy superior que en el caso de que eliminemos a una persona del colectivo de alumnos. En este caso, obviamente, estoy tratando la edad del profesor como valor atípico, porque se diferencia mucho de (su distancia a) la edad media de los alumnos.

La existencia de valores atípicos impele a utilizar indicadores de tendencia central diferentes a la media aritmética. La denominada Estadística Exploratoria definida, en sus inicios, por Tukey, J.W. (1977) con el nombre de Exploratory Data Analysis, ha dedicado importantes esfuerzos a detectar la existencia de estos valores, a corregir las posibles desviaciones que los datos atípicos producen en las medidas de tendencia central y a la definición de nuevos índices que posibilitan su detección y, otros, a minimizar la importancia de este tipo de datos.

Entre otras medidas, se trabaja con la denominada MEDIA RECORTADA al X%, que se obtiene calculando la media de las puntuaciones que quedan después de haber eliminado del

conjunto de datos el X% de los valores mayores y el X% de los valores menores³. Se considera que es una medida de tendencia central más “robusta” en el sentido de que se minimiza la importancia de los valores extremos de la distribución, pues se eliminan en el porcentaje que se estima conveniente. También es posible considerar la MEDIANA como medida más robusta que la Media, la Mediana viene definida como el valor que ocupa la posición central de todos los datos, una vez que éstos han sido ordenados (de menor a mayor o de mayor a menor). Los valores extremos de la distribución afectan muy poco (o nada) al resultado del estadístico.

Como indicador de la importancia que los datos atípicos tienen en Estadística, baste decir que en los procedimientos de Análisis Multivariantes se incluye la opción de detectar estos casos. Así en el caso de la regresión múltiple lineal, podemos calcular los Estadísticos de influencia (DfBetas), y los Diagnósticos por Casos en los que posibilita la definición de un valor como atípico en función del tamaño del residuo (considerado en desviaciones típicas),... A continuación es el investigador quien, tras el análisis de los datos detectados, tiene la opción de decidir la forma de abordar el análisis y el tratamiento de los mismos.

Considero que este es el **segundo error** en el cálculo del IRPH: obviar la existencia de datos atípicos y la distorsión que ellos producen en los resultados y seguir utilizando la media aritmética como indicador de centralidad, dejando al margen el uso de índices de tendencia central más robustos. Este error ya se ha corregido en la fórmula que utiliza el EURIBOR donde se trabaja con medias recortadas lo que, sin duda, posibilita que el resultado obtenido responda “mejor” a lo que se pretende conseguir: un indicador de la tendencia central de los créditos concedidos/renovados en determinado período de tiempo.

INFLUENCIABILIDAD

El director del centro escolar antes aludido sabe con precisión y exactitud que dentro de un año la media de edad de los alumnos de un aula (en el caso de que permanezcan las mismas personas) será exactamente de un año más. Es una obviedad.

Si el director de este centro escolar quiere que el año siguiente la edad de las personas que conforman el aula (25 alumnos y un profesor) sea exactamente la misma que la de este año, lo tiene muy fácil. Cambia al profesor por otro que hoy tenga 26 años menos y logrará la aparente paradoja: dentro de un año la EDAD MEDIA de las personas del aula será exactamente igual a la de hoy. Con absoluta certeza. También este director puede calcular con exactitud y precisión, cuál va a ser la repercusión que va a tener en la media de edad de las personas del aula cambiar un profesor por otro de mayor o menor edad. El cambio que se produce en la edad de cada persona que conforma el colectivo condiciona el valor medio del mismo. Si cambio un profesor por otro de mayor altura, aumenta la media de la altura de las personas del aula: si el nuevo profesor es de menor altura, la media de las alturas disminuye. Y además, de forma muy sencilla, se puede cuantificar de forma precisa el incremento (positivo o negativo) que cada cambio conlleva.

³ Hay autores que al hablar de media recortada al X%, reparten ese porcentaje en dos mitades, la mitad en cada extremo. Sin embargo la práctica habitual, y así lo calcula SPSS, es la de eliminar el X% de los valores de cada extremo.

Si para el cálculo de la EDAD MEDIA utilizáramos el procedimiento que, de forma errónea (desde el punto de vista estadístico) se utiliza en el proceso de cálculo del IRPH, esto es, dando la misma importancia al colectivo de alumnos (25 personas) y al colectivo de profesores (una persona), bastaría que el director del centro cambiara al profesor por uno que fuera sólo 2 años menor. Como consecuencia, sólo con cambiar al profesor por otro con 2 años menos, lograría la misma aparente paradoja anterior. Dentro de un año la EDAD MEDIA de las personas del aula será exactamente igual a la de hoy. Con absoluta certeza. Con precisión. Esta es una propiedad bien conocida y estudiada por la Estadística Descriptiva, con las consecuentes posibilidades de manipulación y margen de maniobra que otorga a quien ostenta la capacidad de cambiar los datos. En ambos casos se desvirtúa el objetivo de esta medida de centralización: representar el valor central de la distribución.

La casuística anterior puede ejemplificar la situación que se produce en el cálculo del IRPH. Dado que el cálculo del mismo se realiza calculando la MEDIA SIMPLE (no ponderada) de los datos remitidos por el conjunto de Entidades que operan en el mercado, cada entidad que opera en dicho mercado sabe con precisión y exactitud la repercusión al alza o a la baja que el incremento positivo o negativo de los tipos de interés de su entidad va a tener en el cálculo del próximo IRPH: cada unidad de cambio en la entidad, automáticamente supone un cambio del $(1/n) \%$ donde n es el número de entidades cuyos datos se tienen en cuenta para el cálculo del Índice. La demostración matemática de la repercusión e influencia es trivial y se deriva de la propia fórmula de cálculo del valor de IRPH:

$$IRPH = \sum_{j=1}^{j=n} \frac{i_j}{n}$$

Donde i_j : tipo de interés de la entidad j

n : número total de entidades

Esta capacidad que tienen las entidades de influenciar y, en consecuencia, manipular el resultado final, no se puede corregir trabajando con la MEDIA PONDERADA. Podemos observar que las fórmulas (que se indican a continuación) son equivalentes. Si conociera cuál es la ponderación (el peso) de cada entidad en el cálculo de la media ponderada, al igual que en el caso de trabajar con la media aritmética, una entidad podría conocer con precisión y exactitud y, por tanto, podría predecir el impacto que los cambios en los tipos de interés que se producen en el indicador de su entidad puede tener en el cálculo del IRPH.

La fórmula de la media ponderada es:

$$IRPH_{ponderado} = \sum_{j=1}^{j=r} \frac{n_j * i_j}{n}$$

Donde n_j : Ponderación o peso de la entidad j . En las r entidades cuyos datos se utilizan

i_j : Tipo medio (ponderado) de los préstamos concedidos

n : Suma de las ponderaciones

De forma inmediata surge la siguiente pregunta: ¿Cómo es posible obtener un índice “justo” en el sentido que refleje las variaciones del mercado pero en el que cada entidad no pueda manipular de forma directa el resultado de dicho índice?

Como antes he indicado son muchas y variadas las formas e índices que la Estadística utiliza para corregir los efectos directos que los valores (o cambios) de un elemento de la muestra tienen en el resultado final. El uso de la Mediana como indicador es una de las formas. El uso de las medias recortadas al X% es otra. Como antes he indicado, los análisis multivariantes posibilitan detectar dichos valores atípicos (uni o multivariantes), siendo conscientes de que su detección y posterior diagnóstico de la razón por la que dichos datos son atípico es el primer paso para un correcto análisis que posibilite visualizar de una forma más real la complejidad del conjunto de datos objeto de estudio. Y, por supuesto, trabajando con la media ponderada (en la que se tuviera en cuenta la cuantía de las operaciones efectuadas por cada entidad que aporta los datos), aunque no se eliminara la capacidad de influencia de las entidades, al menos se obtendría un indicador más preciso de la tendencia central.

En otros contextos, muy próximos y relacionados con la economía, es habitual utilizar como índice de referencia uno que estuviera utilizando criterios y/o indicadores externos a las entidades que luego van a formar parte interesada, e incluso ser una de las partes del contrato que efectúan utilizando como criterio de “coste” el resultado de un índice obtenido con base en las operaciones efectuadas por ellas mismas. Como es el caso. Así se anularía la capacidad de influencia que tienen dichas entidades en el índice de referencia. Baste como ejemplo la amplia utilización del IPC como indicador del Incremento del Coste de la Vida que es utilizado, como punto de partida, en variados contextos para determinar o referenciar incrementos salariales (positivos o negativos), los precios de muchos productos, la planificación de presupuestos,...

Es preciso reseñar que (en Estadística) la robustez de un estadístico es un indicador de la menor posibilidad de ser influenciados por cada uno de los datos. En el caso del IRPH, dado que para su cálculo se utiliza la media, y es un indicador muy poco robusto (en el sentido estadístico), es evidente y trivial que cada una de las entidades, independientemente de su tamaño o volumen tiene una capacidad de influencia exactamente cifrada en el $(\text{incremento}/n) \%$, pudiendo ser este incremento positivo o negativo.

Considero pertinente destacar que todo lo anterior es válido tanto en el supuesto de que las entidades actúen de forma independiente, como si las entidades (un subconjunto del total) se pusieran de acuerdo. En el caso de que una entidad actuase de forma independiente su influenciabilidad es precisa: $1/n$ (y así lo explicita el Banco de España en las contestaciones a las preguntas 4, 5 y 6). En el caso de que varias actuaran de forma conjunta este potencial se multiplicaría y, en este caso, la Estadística y los indicadores de tendencia central que la misma utiliza se convertirían en opacos en su utilidad para visualizar el valor central de la distribución.

LAS COMISIONES EN EL CÁLCULO

El Banco de España confirma la inclusión de las comisiones en el índice que las entidades le referencian para realizar el cálculo del IRPH. Así, en la contestación a la primera parte de la pregunta 13 que se realiza al Banco de España formulada así: *“Es cierto que las medias facilitadas por cada caja de ahorros incluían además comisiones y, por tanto no representarían de manera fiel el precio del dinero?”,* el Banco responde que *“El tipo declarado es la Tasa Anual Equivalente (TAE) que efectivamente incluye comisiones”*.

El problema surge en el punto en que la Tasa Anual Equivalente tiene en cuenta las comisiones que las entidades añaden al valor del IRPH en un momento dado. Es como si en el peso de determinado producto incluyéramos el del envoltorio, y para determinar el peso medio de los productos en un momento dado tomáramos con referencia los pesos del producto junto con los envoltorios. En esta circunstancia para aumentar el peso del producto nos bastaría utilizar un envoltorio más pesado. Es evidente que existe un sesgo al alza en el peso derivado del hecho de tomar como indicador la suma del peso del producto y el del envoltorio. Esto se agrava si tenemos en cuenta que en las operaciones futuras las entidades vuelven a envolver el producto (ya con envoltorio) en un segundo envoltorio, lo que vuelve a incrementar el precio del producto. Y esto es así porque una vez publicado el IRPH, a la hora de realizar las operaciones futuras, cada entidad tienen la potestad de añadir las comisiones, lo que, indudablemente, introduce un efecto multiplicador de las mismas, y de nuevo son las entidades quienes pueden establecer (o no) el incremento derivado de las citadas comisiones. Dado que este hecho se repite en los sucesivos procesos de cálculo del IRPH, se origina una serie geométrica de razón variable, pero siempre mayor que 1, que incrementa el importe del interés del préstamo, lo que repercute de forma automática (al alza) el valor del siguiente IRPH.

Es obvio el sesgo al alza que de esta práctica se deriva, y también muy sencilla la forma de evitarlo. Bastaría con eliminar de las medias que cada entidad facilita la parte proporcional a las comisiones, lo que en el ejemplo antedicho he denominado envoltorio. Es decir para hacer el cálculo del peso medio de los productos, tomar como datos reales los pesos de los mismos sin tener en cuenta el de los envases en los que se encuentran.


COROLARIO

Hasta aquí el análisis del IRPH realizado desde el punto de vista estadístico en el que, como ha quedado demostrado, se observa que es un indicador que no es válido como medida de tendencia central y como tal es **manifiestamente mejorable** utilizando, por ejemplo, para su cálculo la media ponderada (en función del peso de las entidades) de las medias ponderadas de las operaciones efectuadas por cada entidad.

Un segundo aspecto que robustecería el indicador es la eliminación de los valores atípicos, utilizando para ello cualquiera de las técnicas existentes de entre las que, por su sencillez, destaca la utilización de las medias recortadas al X%.

La mejora de los aspectos anteriores conllevaría a disminuir la capacidad que las entidades que aportan los datos tienen en este momento, con el procedimiento utilizado, para incidir, influir, manipular y condicionar de forma individual o coordinada en el valor resultante posterior del IRPH.

Y, por último, el hecho de incluir las comisiones que las entidades aplican a sus clientes en sus operaciones en el indicador que envían al Banco de España y sirve de base para el cálculo del IRPH, introduce un sesgo al alza en dicho cálculo, que también es necesario corregir.



Fdo: Juan Etxeberria Murgiondo

En Donostia, el 28 de Junio de 2016