

Pliego de prescripciones técnicas para el “Análisis de la movilidad durante la pandemia por Covid-19 (Estudio de movilidad EM-4) a partir de la base de datos Luca Transit”

1. Justificación y condiciones generales del servicio

El INE puso en marcha, a finales de 2019, un proyecto de medición de la movilidad a partir de datos de posicionamiento de teléfonos móviles (EM-1) con un objetivo muy específico: se trataba de construir matrices de movilidad cotidiana (residencia-trabajo) para delimitar la forma de las áreas metropolitanas de las ciudades. El trabajo se basaba en datos agregados (recuentos totales de flujos origen-destino) de los tres principales operadores de telefonía móvil (OTM) del país.

Este proyecto formaba parte de las investigaciones del Censo de Población y Viviendas 2021, que se basará en registros administrativos; se ha tratado de estudiar una fuente alternativa para completar la información sobre movilidad cotidiana, la cual no se encuentra en los diferentes registros.

El EM-1 consistía en la publicación de tres tipos de matrices:

- Matriz 1, de movilidad cotidiana
- Matriz 2, de población de día y de noche
- Matriz 3, de población estacional

El brote de covid-19 y el consiguiente estado de alarma propiciaron el aprovechamiento del estudio mencionado anteriormente, para el que los OTM ya habían proporcionado datos y que estaba en su última fase, para así medir la movilidad durante el periodo de confinamiento, el proyecto EM-2.

La gran ventaja de seguir esta aproximación es que los programas ya estaban desarrollados y el método ya estaba validado y rodado. Los resultados se podían obtener de forma casi inmediata y se podían comparar con la movilidad previa al brote, de la cual se acababan de recibir datos.

El proyecto EM-2 supuso la producción de una “matriz 1” diariamente mientras se mantuvo el estado de alarma de primavera de 2020. Este proyecto EM-2 dio lugar a un indicador formó parte de la lista de indicadores para el seguimiento de la desescalada. En concreto se tomaba, para cada área de residencia, el porcentaje de personas que salen de ella (en horario laboral).

Posteriormente el INE puso en marcha una nueva operación estadística con la finalidad de dar continuidad a la serie mencionada anteriormente y poder ofrecer información de la movilidad durante lo que entonces se llamaba “la vuelta a la normalidad”. Este nuevo proyecto, llamado EM-3, pretendía dar indicadores homogéneos con los que se habían ofrecido durante el estado de alarma con lo que no se propuso ninguna modificación metodológica, que siempre haría más difícil de interpretar la serie.

Se trataba de medir la movilidad de la población durante el periodo posterior al confinamiento de primavera de 2020, dado que esta información sobre movilidad de la población es un importante indicador para el análisis de la evolución tanto del fenómeno desde el punto de vista de salud (la movilidad de la población es un vector de aumento

de contagios) como económico (la movilidad es un indicador muy directo sobre la actividad económica).

Situándonos en el contexto de la situación de la pandemia de Covid-19 en el momento en el que se idea el proyecto EM-3 (entre mayo y junio de 2020), el objetivo que se pretendía alcanzar era medir la movilidad cotidiana (en horario de estudio o trabajo) a partir de la terminación del estado de alarma, suponiendo que se iba a producir una “vuelta a la normalidad”. Así, se previó (de forma optimista, podríamos decir ahora) que seis meses de medición del fenómeno eran más que suficientes: el proyecto EM-3 consistió en la producción de información de movilidad cotidiana para los meses de junio a diciembre de 2020.

Pero en la situación actual (diciembre de 2020) la segunda ola de la pandemia golpea con fuerza en toda Europa y no parece previsible que nos vaya a abandonar a corto plazo. Los responsables de salud pública de organismos nacionales e internacionales, con las lógicas incertidumbres que lo rodean, hablan de que el brote de coronavirus puede afectar a la población durante todo 2021.

En este contexto de enorme incertidumbre se plantea la necesidad de poner en marcha un nuevo estudio de movilidad (EM-4) que permita monitorizar la movilidad de la población, de manera homogénea a como hemos venido haciendo, por un tiempo que es difícil de estimar pero que sin duda será como mínimo de varios meses.

Para realizar este estudio se propone la elaboración de una serie de tablas con información agregada y anonimizada obtenida a partir del posicionamiento de los teléfonos móviles contenidos en la base de datos Luca Transit, que incorpora información de los usuarios de Movistar, del operador Telefónica Soluciones.

2. Especificaciones detalladas del servicio

Al igual que en el estudio EM-3, y para permitir la comparación con él, se debe entregar al INE dos tipos de matrices de datos que se describe a continuación.

Para la construcción de estas tablas, la unidad de observación geográfica es el “**área de movilidad INE**”. Estas áreas han sido elaboradas por el INE. Para la determinación de las áreas se considera la población empadronada en ellas a 1 de enero de 2019, de forma que cada área supere un umbral aproximado de 5.000 empadronados (en algunos casos no se podrá respetar ese umbral, pero en todo caso se formarán áreas en las que resida un mínimo de 1.000 personas).

Con esta delimitación se pretende garantizar la confidencialidad de la información y que el número de usuarios de Movistar en cada área sea suficiente para el análisis de movilidad entre las diferentes áreas. Las áreas, en términos generales, se han construido de la siguiente manera:

- Dentro de cada provincia, los municipios menores de 5.000 habitantes, se agrupan geográficamente para alcanzar al menos ese umbral de población.
- Cada uno de los municipios de entre 5.000 y 50.000 habitantes constituye un área.
- Los municipios de más de 50.000 habitantes se desagregan en distritos o barrios, siempre con un umbral superior a esos 5.000 habitantes. En el caso más extremo, el

municipio de Madrid, se divide en 128 barrios, todos ellos superando el umbral de 5.000 habitantes contemplado anteriormente.

El número total de áreas estará por debajo de 3.215. El tamaño medio de cada área rondará los 15.000 habitantes.

El INE entregará a la empresa adjudicataria la lista de áreas de movilidad, con la información necesaria para su delimitación, tanto alfanumérica como geométrica, en ficheros fácilmente tratables (formatos Excel, Shape).

Delimitadas estas áreas, los dos tipos de matrices a construir son:

MATRIZ Nº1 DE MOVILIDAD COTIDIANA (PERNOCTACIÓN-DESTINO COTIDIANO)

Se trata de una matriz de origen-destino formada por $N \times N$ áreas, en donde N será aproximadamente 3.215 áreas. En la casilla "fila A-columna B" de esta matriz figurará el número de teléfonos que, de media, se desplazan cotidianamente del área de pernoctación (A) al área de destino cotidiano (B). Además se tienen que entregar dos vectores de totales, por origen y destino de tamaño $N \times 1$.

Para la construcción de esta matriz, en primer lugar se debe intentar determinar, **para todos los teléfonos móviles de Movistar residentes en España**, su área de movilidad en la que se entiende que reside el usuario del teléfono móvil, que se denominará "área de pernoctación".

El algoritmo de determinación del **área de pernoctación** será el siguiente:

El área de pernoctación de cada teléfono móvil es aquella en donde el teléfono móvil se encuentra durante más tiempo entre las 22:00 del día anterior y las 06:00 durante el día observado. Se puede estudiar alguna otra forma de asignar área de pernoctación propuesta por el contratista. De todos modos, el método finalmente adoptado debe ser aprobado previamente por el INE.

El universo de teléfonos, es por tanto el de aquellos para los cuales se ha logrado asignar un área de pernoctación. Solo para ellos, asignada el área de pernoctación, se intentará determinar el **área de destino cotidiano** de los teléfonos móviles de Movistar que recoge la base de datos Luca Transit.

El método es parecido al de asignación de área de pernoctación pero, por la naturaleza del procedimiento que se diseña, no siempre será posible determinar área de destino cotidiano para un teléfono móvil. Para determinar **área de destino** se deberán tomar todas las áreas (incluida la de pernoctación) en las que se encuentra el terminal durante el horario 10:00 a 16:00 siempre y cuando se encuentre a ese terminal en esa área al menos durante dos horas al día. Si hay más de un área, se tomará aquella más frecuente (en la que se está más tiempo), que no sea la de pernoctación. Si solo el área de pernoctación cumple esta condición, entonces el área de pernoctación será también la de destino. Al igual que en el caso anterior, si la empresa adjudicataria propone una definición alternativa deberá ser aprobada previamente por el INE.

Tal y como se ha dicho anteriormente, este método no siempre va a poder obtener un área de destino, pues en algunos casos no se podrá determinar que exista un área a la que se acude de forma recurrente. Por tanto no es necesario asignar un área de destino a todos los teléfonos móviles, pero sin embargo sí que es necesario que todos los teléfonos móviles tengan asociada un área de pernoctación.

Por cada día de observación se deberá enviar al INE tres ficheros:

- La matriz "NXN", en realidad es un fichero con las siguientes variables:

- * Identificador de área de pernoctación
- * Identificador de área de destino
- * Número de teléfonos móviles con esas áreas de pernoctación y destino

Por razones de confidencialidad el operador puede considerar no incluir en este fichero valores inferiores a un cierto umbral (que debe especificar). En tal caso, si el número de movimientos observado para una celda pernoctación-destino de la matriz es inferior a ese umbral ese dato vendrá marcado con un carácter ".", "#", u otro. En todo caso, estos valores no proporcionados sí figurarán en el cómputo de vectores de totales tanto de orígenes como de destinos. Solo debe incluirse la pareja pernoctación-destino si hay dato numérico o bien esta marca de confidencialidad.

- El vector de orígenes de tamaño N contiene para cada área de movilidad el número total de teléfonos móviles detectados en ella durante la noche observada, es decir, los que pernoctan en ella. La suma de estos N valores constituye el universo de móviles observado el día en cuestión.

-El vector de destinos contiene cuantos teléfonos se cuentan en total en cada área de destino cotidiano según el algoritmo propuesto, independientemente de su área de origen (de pernoctación).

MATRIZ Nº 3 DE POBLACIÓN ESTACIONAL (RESIDENCIA-PERNOCTACIÓN)

(Nota: No hay una matriz nº 2 en este proyecto pero se mantiene esta nomenclatura porque si existe tal matriz en el estudio EM-1)

Se trata de construir una matriz N X N donde N es el número de áreas de movilidad. Cada matriz corresponde a un día concreto. Además se entregan los vectores de totales por origen (es decir, residencia) y destino (es decir, pernoctación).

Para cada día elegido, en la celda fila A-columna B figuraría el número de teléfonos móviles que, "residiendo" en el área A, pernoctan en el área B durante ese día, según las siguientes especificaciones:

Como en la matriz 1, solo se consideran los números de teléfonos móviles con residencia en España. No se incorporan los terminales del extranjero en "roaming".

Para determinar el universo de móviles de la fecha en cuestión, se parte, como en la matriz 1, de aquellos terminales para los cuales se puede determinar un área de pernoctación.

Para el día elegido, se determina para cada terminal el área **donde pernocta** (donde ese teléfono móvil se encuentra más frecuentemente durante las horas 22:00 de la noche anterior a 06:00 en esa fecha). La franja horaria se puede modificar por acuerdo entre las partes.

Para cada uno de los teléfonos a los que se ha podido localizar en el día elegido, y por tanto, se le ha asignado área de pernoctación, se debe encontrar su **área de residencia**.

En este caso se tomará como referencia fechas anteriores a la fecha objeto de estudio en cada caso (normalmente, será el área donde se localice el teléfono en horario nocturno durante un periodo de entre dos y tres meses previos al día de referencia). El método finalmente elegido se debe acordar entre el INE y la empresa adjudicataria.

El contenido de la matriz 3 es por tanto, tres ficheros por cada día observado:

- el fichero "NXN" de orígenes-destinos (o residencia-pernoctación), con las siguientes variables:

- * Identificador de área de residencia
- * Identificador de área de pernoctación
- * Número de teléfonos móviles con esas áreas de residencia y pernoctación

Por razones de confidencialidad el operador puede considerar no incluir en este fichero valores inferiores a un cierto umbral (que debe especificar). En tal caso, si el número de movimientos observado para una celda residencia-destino de la matriz es inferior a ese umbral ese dato vendrá marcado con un carácter ".", "#", u otro. En todo caso, estos valores no proporcionados sí figurarán en el cómputo de vectores de totales tanto de orígenes como de destinos. Solo debe incluirse la pareja residencia-pernoctación si hay dato numérico o bien esta marca de confidencialidad.

- El vector de destinos contiene cuantos teléfonos se cuentan pernoctando en total en cada área de movilidad según el algoritmo propuesto. La suma de estos N valores constituye el universo de móviles observado el día en cuestión.

- El vector de orígenes (áreas de residencia) de tamaño N contiene cuántos de los teléfonos móviles detectados durante la noche observada, tienen asignada cada una de las áreas de residencia según el método anterior.

Se proponen aquí otros elementos básicos de la metodología necesaria para producir los datos requeridos.

DETERMINACIÓN DE LAS ÁREAS DE MOVILIDAD

El INE proporcionaría la capa de delimitación geográfica para las aproximadamente 3.215 áreas en que se dividiría el territorio nacional.

Se determina el área de movilidad en la que se encuentra un teléfono en cada momento mediante intersección entre éstas y la información sobre localización que utilice la base de datos Luca Transit (pueden ser áreas de Voronoi asociadas a la distribución de antenas u otras delimitaciones alternativas más precisas). El método exacto a utilizar debe ser acordado entre el INE y la empresa adjudicataria.

TELÉFONOS Y POBLACIÓN

El estudio debería realizarse exclusivamente con teléfonos móviles de población residente en España.

La empresa adjudicataria deberá reflejar en las matrices solicitadas información sobre número de teléfonos de Movistar que se mueven (según los datos disponibles en su

base de datos Luca Transit), y no de personas: una persona que porte dos teléfonos debería contarse como dos teléfonos a menos que se disponga de información metodológica precisa que permita discernir este caso.

La información solicitada, además, se deberá referir solo a teléfonos móviles y no a todos los dispositivos susceptibles de tener una tarjeta SIM. Se debe acordar con el INE la metodología empleada para contabilizar solo estos terminales.

Teniendo en cuenta la zonificación planteada de España en aproximadamente 3.215 áreas, es de esperar que la base de datos Luca Transit disponga de información en todas estas áreas. Si hubiese algún caso en donde esto no fuese así, deberá aportarse justificación a este respecto y será necesaria la aceptación por el INE de la delimitación alternativa que se proponga, si fuera el caso.

3. Referencia temporal y calendario de los productos a entregar

Se contempla que el contrato tenga una duración de un año.

Desde el inicio del contrato, El INE establecerá mensualmente un calendario de matrices a entregar, tanto de tipo 1 como de tipo 3. El calendario correspondiente a cada mes se entregará por parte del INE con una antelación mínima de una semana. El calendario mensual establecerá las fechas de referencia para matrices tipo 1 y 3.

Calendario matrices tipo 1:

Las matrices de tipo 1 serán entregadas por el contratista al INE de forma individual con un retraso no superior a 72 horas desde la finalización de la fecha de referencia de los datos.

El número de matrices de tipo 1 a suministrar por parte del contratista cada semana estará entre un mínimo de cero y un máximo de cuatro matrices.

Calendario matrices tipo 3:

Las matrices de tipo 3 se entregarán, también conforme se vayan generando, con un retraso no superior a 20 días desde la finalización de la fecha de referencia de los datos.

El número de matrices de tipo 3 a solicitar será de un mínimo de 2 y un máximo de 8 a lo largo del periodo del contrato.

Número total de matrices a entregar:

El INE se compromete a solicitar un mínimo de 52 matrices de tipo 1 y dos matrices de tipo 3 y un máximo de 110 matrices de tipo 1 y 8 de tipo 3 a lo largo del contrato.

El contrato contempla así un precio fijo por estas primeras 52+2 matrices y un precio por cada una de las matrices de tipo 1 o 3 que se soliciten adicionalmente hasta los máximos prefijados.

4. Controles de calidad

Los datos suministrados serán sometidos a un control de calidad exhaustivo, con especial atención a la consistencia entre la información reflejada por cada una de las matrices y la cifra de población existente en cada una de las áreas.

Si en este proceso se detectan valores llamativos o claramente erróneos, inconsistencias internas o defectos en la cobertura, los ficheros respectivos serán devueltos a la empresa suministradora, quien será informada de las deficiencias encontradas y deberá informar de lo sucedido y subsanar dichas incidencias en un plazo máximo de un mes.

En Madrid,

El Subdirector General de Estadísticas Sociodemográficas

Fdo. Antonio Argüeso Jiménez