**5.3-4. Revisar y validar. Depurar e imputar**

5.3.1. Ejecutar detección y tratamiento de errores (input)

**30260 - Estadística del Padrón Continuo**

|  |
| --- |
| Si no es aplicable marcar |

|  |
| --- |
| Input:  **A1: Datos provisionales**   * Fichero de microdatos obtenidos sin depurar E403E.CIF[*aaaa*].SELTODOS.F3[*aa*] (output de 4.3.7) * Ficheros necesariospara la depuración:   Lugar de Nacimiento\_[*aaaa*] (valores válidos) (output de 4.3.8)  Ficheros con las sustituciones (de valores validados) realizadas en cifras definitivas el año anterior (output de 5.3.1 A2 del año anterior):  SUSSEXO[*aaaa*-1]  SUSfnac[*aaaa*-1]  SUSnpromu[*aaaa*-1]  Base Padronal.   * Reglas de detección y tratamiento de errores (outputde 2.5.10). * Reglas de cálculo de variables derivadas (output de 2.2.4).   **A2. Datos definitivos.**   * Ficheros provinciales E403E.FEX02.A[*aaaa*].P[*xx*] con microdatos sin depurar (output de 4.3.7). * Fichero E377E.CALL.TRAM.P01-52.F[*aa*]1231.G[*aa*]03dd con el callejero (output de 4.3.8). * Fichero Pobcon[*aa*] con población contada sin depurar (output de 4.3.7). * Fichero de Alegaciones sin depurar(output de 4.3.7) * Listado estándar de países (output de 4.3.8) * Relación de municipios y sus códigos (output de 4.3.8) * Fichero DEPav.a00\_conerrorescorregidos con correcciones realizadas previamente en el Avance(output de 5.3.1 A1). * Fichero Historia.sas7bdat con la HistoriaPadronal (output de 4.3.7). * Fichero MNP.MNP\_nacimientos\_F141016 con nacimientos registrados en MNP (output de 4.3.8). * Tablaspadron.pcvi26dni\_mi del Ministerio del Interior con DNIs (output de 4.3.8) * Fichero MNP90.sas7bdat (CRUCE.MNP90) con el cruce que se realizó en 2009 entre los mayores de 90 años y las defunciones de MNP (output de 4.3.8) * Ficheros E403E.FEX.SUS.MUNN.A[*aaaa*] con correspondencias de códigos de lugares de nacimiento (output de 4.3.8) * Ficheros E403E.FEX.SUS.MUNIVALI.A[*aaaa*] con códigos válidos de lugares de nacimiento (output de 4.3.8). * Ficheros E403E.FEX.SUS.NACI.A[*aaaa*] con correspondencias de códigos de nacionalidades (output de 4.3.8). * Ficheros E403E.FEX.SUS.NACIVALI.A[*aaaa*] con códigos válidos de nacionalidades (output de 4.3.8). * Reglas de detección y tratamiento de errores (output de 2.5.10). * Reglas de cálculo de variables derivadas (output de 2.2.4). |

|  |
| --- |
| Output:  **A1: Datos provisionales**   * Fichero con los datos depurados de población del Avance A00. * Correcciones del Avance DEPav.a00\_conerrorescorregidos * Informe de depuración e imputación en fichero   Resumen Imputaciones REALIZADAS.txt   * Informe de depuración e imputación en fichero   Valores\_Variables\_Avance[*aaaa*].xls  **A2: Datos definitivos**   * Ficheros provincialesE403E.FEX41.A[*aaaa*].P[*xx*] (234 posiciones) con microdatos depurados de la Estadística del Padrón Continuo.  Ficheros E403E.FEX11.RESULTA.A[*aaaa*].Pxxcon valores brutos y depurados.  * Ficheros E403E.FEX14.RESULT.A[aaaa].Pxx con valores brutos y depurados. |

|  |
| --- |
| Proceso:  **A1. Depuración Datos Avance**  En este proceso se depuran los datos de población correspondientes al avance que se obtienen en marzo.Se realiza en marzo-abril.  El proceso es el siguiente:  [FASE-1. Creación de la estructura de carpetas y de ficheros que vamos a necesitar en la depuración](#Fase1).   * 1. Crear estructura de carpetas y copiar los ficheros especificados en el input según la documentación especificada.   2. Copiar fichero E403E.CIF[*aaaa*].SELTODOS.F3[*aa*]y renombrarlo como AVANCE\_[*aaaa*]\_L60.   FASE-2.  2.1- Ejecución secuencial de la regla de detección "DetCompara" con las siguientes parametrizaciones:  Variable a depurar: Sexo  Valores válidos: '1', '6' y contenidos en SUSSEXO[*aaaa*-1].  Variable a depurar: Fecha de nacimiento  Valores válidos: Valores conducentes a edades en [1, 100] y aquéllos tomados de SUSfnac[*aaaa*-1]  Variable a depurar: Nacionalidad  Valores válidos: Valores tomados de ¿?  Variables a depurar: Provincia y Municipio  Valores válidos: Valores tomados de SUSnpromu[*aaaa*-1]  2.2.- Ejecución secuencial de la regla de tratamiento "TratCompara" con las siguientes parametrizaciones:  Variable a tratar: Sexo.  Origen de los valores válidos: SUSSEXO[*aaaa*-1].  Reglas complementarias: Si quedan valores inválidos tras la imputación, se imputa a '6' o, a juicio del experto, buscando en la Base Padronal.  Variable a tratar: Fecha de nacimiento.  Origen de los valores válidos: Lugar de Nacimiento\_[*aaaa*].  Reglas complementarias: Se imputan manualmente fechas específicas señaladas por el experto (véase documentación).  Variable a tratar: Nacionalidad.  Origen de los valores válidos: ¿?  Reglas complementarias: Si quedan valores inválidos tras la imputación, se imputa a '108' (nacionalidad española).  Variables a tratar: Provincia y Municipio de Nacimiento  Origen de los valores válidos: Lugar de Nacimiento\_[*aaaa*].  Reglas complementarias: Para valores no encontrados en el conjunto anterior, la provincia para extranjeros se imputa por el país de nacionalidad y para el resto, su municipio se imputa por el municipio de residencia.  Se aplican estos dos pasos (2.1 y 2.2) hasta que no queden valores inválidos.  FASE 3.- Se integran los conjuntos de datos de registros inicialmente válidos e inválidos y posteriormente se dividen por provincias y por nacionalidad (española y extranjera).  La aplicación de estas fases se realiza mediante la ejecución del programa “00-DEPURACIÓN AVANCE\_[*aaaa*]” sobre el fichero AVANCE\_[*aaaa*]\_L60.Se generan automáticamente los ficheros  Resumen Imputaciones REALIZADAS.txt  Valores\_Variables\_Avance[*aaaa*].xls.  Se aplica de modo integrado en este programa la reglade cálculo de variables derivadas “DerivarEdad”.  Se obtiene el fichero final A00, con los microdatos que quedan validados automáticamente tras ejecutar las tareas anteriores.  En el documento Depuracion Avance [*aaaa*].doc se describe todo el proceso detalladamente.  **A2. Depuración Datos definitivos.**  Actividad “Depurar Datos Geográficos”  Esta actividad se realiza mediante la ejecución del procedimiento FEX (en el HOST; véase documentación) y consiste en:   1. Depuración del código de unidad poblacional (CUN):   Se ejecutan de modo integrado las reglas "DetCompara" y "TratCompara" con las siguientes parametrizaciones:  Variables a depurar: Cpro, Cmun, cun  Valores válidos: Valores contenidos en el tramero (E377E.CALL.TRAM.P01-52.F[*aa*]1231.G[*aa*]03dd).  Variable a tratar: Mismas a depurar.  Origen de los valores válidos: Tramero (E377E.CALL.TRAM.P01-52.F[*aa*]1231.G[*aa*]03dd).  Reglas complementarias: La comparación de las siete variables comprende numerosos casos diferentes con las correspondientes acciones de tratamientos automáticos y manuales (ver documentación).   1. Depuración de seccionado:   Se ejecutan de modo integrado las reglas "DetCompara" y "TratCompara" con las siguientes parametrizaciones:  Variables a depurar: dist, secc  Valores válidos: Valores contenidos en el tramero (E377E.CALL.TRAM.P01-52.F[*aa*]1231.G[*aa*]03dd).  Variable a tratar: Mismas a depurar.  Origen de los valores válidos: Tramero (E377E.CALL.TRAM.P01-52.F[*aa*]1231.G[*aa*]03dd).  Reglas complementarias: La comparación de las variables comprende numerosos casos diferentes con las correspondientes acciones de tratamientos automáticos y manuales (ver documentación).   1. Añadir Código Postal   Se añade el código postal al fichero de partida, obtenido a través del callejero.   1. El procedimiento genera los ficheros E403E.FEX11.RESULTA.A[*aaaa*].Pxx y E403E.FEX14.RESULT.A[*aaaa*].Pxx.   Actividad “Depurar variable”   1. Ejecución del programa correspondiente que implementa las reglas de detección "DetCompara" y "TratCompara" para la variable respectiva sobre el fichero de población contada Pobcon[*aa*] y sobre el fichero de alegaciones tabla-60, con la parametrización:   Variables a depurar: Nombre.  Valores válidos: Valores contenidos en  (a) DEPav.a00\_conerrorescorregidos  (b) Historia Padronal;  (c) Nacimientos registrados en MNP(ver tarea 4.3.8);  (d) Tablas Ministerio del Interior (ver tarea 4.3.8);  (e) Fichero MNP90.  Variables a tratar: Mismas a depurar.  Origen de los valores válidos: Valores contenidos en  (a) DEPav.a00\_conerrorescorregidos  (b) Historia Padronal;  (c) Nacimientos registrados en MNP(ver tarea 4.3.8);  (d) Tablas Ministerio del Interior (ver tarea 4.3.8);  (e) Fichero MNP90.  Reglas complementarias: Breve descripción.  Programa: Nombre.   1. Generación del fichero con los registros de Pobcon[*aa*] y de tabla-60 que tienen errores en las variablesespecificadas y sus correspondientes valores imputados.   En este proceso se obtienen los registros (unidades estadísticas) definitivos con las variablesdepuradas.  El proceso para la depuración del conjunto de datos definitivos consiste en:   1. Ejecutar la actividad “Depurar Datos Geográficos”sobre el ficheroE403E.FEX02.A[*aaaa*].P[*xx*] para obtener E403E.FEX15.TOTAL.A[*aaaa*].P[*xx*], que contienen código de distrito, sección y unidad poblacional depurados, mediante la ejecución parcial del procedimiento FEX. Este subproceso se ejecuta a partir de abril sin fecha fija. 2. Ejecutar la actividad “Depurar Datos Geográficos” sobre el ficheroE403E.FEX30.CVPCFEX.F[*aaaa*] (no disponible hasta la reunión del Consejo de Empadronamiento, en noviembre)para obtener E403E.FEX36.TOTAL.A[*aaaa*] fichero de alegaciones con código de distrito, sección y unidad poblacional depurados, mediante la ejecución parcial del procedimiento FEX. Este subproceso se ejecuta a partir de abril sin fecha fija. 3. Ejecutar la actividad “Depurar variable” con la siguiente parametrización:   Variable a depurar: Sexo;  Reglas complementarias: Si quedan valores sin imputar, se imputa el valor '6';  Programa: Depuracionsexo.sas;  Se obtiene el ficheroSUSSEXO[*aa*]\_\_CON\_alegaciones, que contiene las variables niden y sexo. Este proceso se realiza en el mes de octubre.   1. Ejecutar la actividad “Depurar variable” con la siguiente parametrización:   Variable a depurar: Fecha de nacimiento;  Reglas complementarias: Si quedan valores conducentes a edades fuera del intervalo [1, 100], se imputa la fecha '19500101';  Programa: "Depuracion FECHA de NACIMIENTO.sas";  Se obtiene el fichero SUSFNAC[*aa*]\_\_\_CON\_alegaciones, que contiene las variables niden, fnac1 y fnac2. Este proceso se realiza en el mes de octubre.   1. Ejecutar la actividad “Depurar variable” con la siguiente parametrización:   Variable a depurar: Provincia y municipio de nacimiento.  Reglas complementarias: Si quedan valores no válidos, se imputa por provincia y municipio de residencia.  Programa: "Depuracion LUGAR de NACIMIENTO.sas".  Se obtiene el fichero SUSNPROMU[*aa*]\_\_\_CON\_alegaciones que contiene las variables lnac y niden. Este proceso se realiza en el mes de octubre.   1. Ejecutar la actividad “Depurar variable” con la siguiente parametrización:   Variable a depurar: Nacionalidad.  Reglas complementarias: Si quedan valores no válidos, se imputa por la nacionalidad española.  Programa:¿?.  Se obtiene el fichero SUSNACIA[*aaaa*] que contiene las variables naci y niden. Este proceso se realiza en el mes de octubre.   1. Últimas depuraciones. Procedimiento FEX41 (SGTIC). Este proceso se realiza a partir del mes de octubre, una vez concluidos los anteriores. 2. Se genera el fichero completo como unión de E403E.FEX15.TOTAL.A[*aaaa*] y E403E.FEX36.TOTAL.A[*aaaa*] 3. Se sustituyen las variables depuradas en los pasos anteriores. 4. Se ejecuta la regla de detección "DetCompara"sobre las variables Sexo, Provincia y Municipio de Nacimiento, Fecha de Nacimiento y Nacionalidad para comprobar que no ha quedado ningún valor inválido. Se ejecuta la regla "TratCompara" para obtener unas últimas depuraciones para algunos registros que no se han podido identificar en las fases anteriores (alegaciones con NIDEN=0000000000). 5. Se generan manualmente los ficheros E403E.FEX.SUS.SECDEF.A[*aaaa*] y E403E.FEX.SUS.CUNDEF.A[*aaaa*] con últimas correcciones de secciones censales y CUN. 6. Los ficheros provinciales completos obtenidos se llaman E403E.FEX41.A[*aaaa*].P[*xx*].   Los conjuntos de datos contenidos en los ficheros provinciales completos quedan validados automáticamente cuando se ejecutan las tareas anteriores. |

|  |
| --- |
| Herramientas:  **A1. Datos provisionales**   * Programa SAS 00-DEPURACION AVANCE\_2015.sas (output de 3.2.5).   **A2. Datos definitivos**   * Procedimiento FEX (programación elaborada por la SGTIC) (output de 3.2.5) * Excel * Programa SAS Depuracionsexo.sas (output de 3.2.5) * Programa SAS Depuracion FECHA de NACIMIENTO.sas (output de 3.2.5) * Programa SAS Depuracion LUGAR de NACIMIENTO.sas(output de 3.2.5) |

|  |
| --- |
| Documentación:  **A1. Datos provisionales**   * Depuracion Avance [*aaaa*].doc   **A2. Datos definitivos**   * Procedimiento de obtención del fichero de explotación estadística.docx * DEPURACIÓN Variables demograficas.doc * Programa SAS Depuracionsexo.sas * Programa SAS Depuracion FECHA de NACIMIENTO.sas * Programa SAS Depuracion LUGAR de NACIMIENTO.sas |

|  |
| --- |
| Unidad responsable:  **A1. Datos provisionales**  - Unidad promotora.  - SGTIC  **A2. Datos definitivos**  - Unidad promotora.  - SGTIC. |