

Censo Nacional 2011

Costa Rica

Criterios, Procedimientos, Reglas de Consistencia y
otros procesos especiales a ser aplicados durante el
proceso de captura y codificación de las Boletas
Censales

14 de Julio de 2011

Criterios de Procesamiento para el Censo Nacional 2011 de Costa Rica

| | | |
|-----------|--|-----------|
| 1 | INTRODUCCIÓN | 4 |
| 2 | REGLAS DE IMPUTACIÓN AUTOMÁTICA DURANTE LA CAPTURA | 4 |
| 3 | REGLAS DE INTEGRIDAD APLICABLES SOBRE EL CONTENIDO DE LAS BOLETAS..... | 4 |
| 4 | REGLAS ESPECIALES PARA EL PROCESAMIENTO..... | 6 |
| 5. | REGLAS DE CONSISTENCIA ESTRUCTURAL..... | 7 |
| 5.1. | ALCANCE DEL MÓDULO CONSISTENCIA ESTRUCTURAL..... | 7 |
| 5.2. | LISTA DE REGLAS DE CONSISTENCIA | 8 |
| 6. | CRITERIOS DE TRABAJO PARA LAS TAREAS DE VERIFICACIÓN INTERACTIVAS..... | 11 |
| 6.1. | REGLAS A SEGUIR EN LA ETAPA “VERIFICACIÓN DE REGLAS” | 11 |
| 6.2. | REGLAS A SEGUIR EN LA ETAPA “VERIFICACIÓN” DE NÚMEROS, MARCAS Y TEXTO, Y EN EL CONTROL DE CALIDAD..... | 12 |
| 6.2.1. | ¿CÓMO TRABAJAR CON LOS CAMPOS DE TIPO MARCA? | 12 |
| 6.2.2. | ¿CÓMO TRABAJAR CON LOS CAMPOS DE TIPO NÚMEROS? | 14 |
| 6.2.3. | ¿CÓMO TRABAJAR CON LOS CAMPOS DE TIPO TEXTO? | 15 |
| 7. | REGLAS PARA LA CODIFICACIÓN DE VARIABLES ABIERTAS | 16 |
| 7.1. | REGLAS PARA LA CODIFICACIÓN AUTOMÁTICA..... | 16 |
| 7.2. | REGLAS PARA LA CODIFICACIÓN ASISTIDA..... | 16 |
| 7.3. | REGLAS PARA LA ACTUALIZACIÓN DE DICCIONARIOS DE REFERENCIA | 17 |
| 8. | DEFINICIÓN DEL NIVEL DE ERROR ACEPTABLE..... | 18 |
| 8.1. | TASA DE ERROR MÁXIMA ACEPTABLE..... | 18 |
| 8.2. | FORMA DE CÁLCULO DE LA “TASA DE ERROR” | 18 |

Criterios de Procesamiento para el Censo Nacional 2011 de Costa Rica

9. FLUJO DE PROCESAMIENTO Y PROCEDIMIENTOS DE ENTREGA Y ACEPTACIÓN 22

9.1. TRABAJO POR LOTE INEC 22

9.2. SECUENCIA DE TRABAJO 22

9.3. PROCEDIMIENTOS A SEGUIR PARA LA DEVOLUCIÓN Y ACEPTACIÓN DE LOTES PROCESADOS..... 22

Criterios de Procesamiento para el Censo Nacional 2011 de Costa Rica

1 Introducción

El presente documento expresa el acuerdo establecido entre el INEC y el Consocio ITS/Active Software relativo a todos los criterios, procedimientos y reglas que serán aplicadas en el proceso de captura y codificación del material del Censo Nacional de Población y Vivienda de Costa Rica 2011, así como también la definición de las tasas de error aceptables para el procesamiento.

2 Reglas de Imputación automática durante la captura

| | Regla o situación | Acción a realizar |
|---|--|---|
| 1 | El campo Número de Persona es imputado automáticamente. | No se reconoce el número de persona escrito en la boleta, y se asigna un número consecutivo de Persona por cada hogar. Esto implica que la vivienda/hogar debe ser consistente estructuralmente. Esta imputación se realiza al final del proceso, una vez que ya se ha decidido cómo se forman las boletas. |
| 2 | Si el campo Número de Hogar está vacío y se trata de una vivienda desocupada (o sin datos de ocupación), y que no contiene ningún campo lleno en las partes Hogar ni de Personas, entonces se imputa el valor 0 (cero) al número de Hogar. | Esta operación se realiza en forma previa a la parte interactiva de la Consistencia Estructural, previo a la evaluación de otras reglas. Para este caso la generación del CSPro no va a generar ningún registro de hogar ni de persona. |
| 3 | Si el campo Número de Hogar está vacío y se trata de una vivienda ocupada con moradores presentes, y que contiene algún campo lleno en las partes Hogar, entonces se imputa el valor 1 (uno) al número de Hogar. | Esta operación se realiza al inicio de la Consistencia Estructural, previo a la evaluación de otras reglas. |
| 4 | Si el campo Número de Hogar está vacío y se trata de una vivienda ocupada con moradores ausentes, y que no contiene ningún campo lleno en las partes Hogar, entonces se imputa el valor 1 (uno) al número de Hogar. | Esta operación se realiza al inicio de la Consistencia Estructural, previo a la evaluación de otras reglas. |

3 Reglas de Integridad aplicables sobre el contenido de las boletas

A continuación se definen las reglas aplicables a la tarea del Control de Integridad de los datos. En esta etapa el software analiza que una serie de reglas de consistencia se cumplan. Se asume en este instante que la captura ya fue realizada correctamente, y se desea detectar eventuales problemas de error en la escritura de la información en la boleta en papel. Un funcionario de INEC podrá analizar esos casos y tomar la decisión de dejar la inconsistencia o de modificar el dato.

Criterios de Procesamiento para el Censo Nacional 2011 de Costa Rica

| | Regla o situación | Acción a realizar |
|---|--|---|
| 1 | Se define que una Persona existe si al menos un campo de información está lleno. | Crear ese registro de persona si cualquier campo de sus preguntas 1 a 31 de "Persona" contiene información. |
| 2 | Los campos p30 y p31 deberían ser inferiores a 12. | Enviar a Integridad INEC para verificación y eventual corrección. Se debe dejar la información que está escrita. |
| 3 | La AE la Boleta debería ser igual a la del C2 | Algunas veces los censistas equivocan la escritura del número de AE en la boleta, invierten dos cifras u otros errores. Si el operador detecta que el censista escribió mal en el papel entonces podrá corregirlo en el dato. De lo contrario, se deberá dejar el AE escrito en el papel, ya que puede tratarse de una boleta que se entreveró y cambió de bolsa de UGM, pero es correcta. |
| 4 | La UGM de la Boleta debería ser igual a la del C2 | Algunas veces los censistas equivocan la escritura del número de UGM en la boleta, invierten dos cifras u otros errores. Si el operador detecta que el censista escribió mal en el papel entonces podrá corregirlo en el dato. De lo contrario, se deberá dejar la UGM escrita en el papel, ya que puede tratarse de una boleta que se entreveró y cambió de bolsa de UGM, pero es correcta |
| 5 | La cantidad de Hijos Vivos es \leq a la cantidad de Hijos Nacidos | Verificar. Ante la duda se deja la información tal como está. |
| 6 | Una página de Persona debería contener información en al menos los campos Parentesco y Sexo. | Esto es para detectar si alguna página de Persona no contiene algún campo con basura, pero que en realidad la persona es vacía. El operador puede decidir si mantener o eliminar a esa persona. |
| 7 | En el campo P16 hay información inconsistente, ya que hay doble marca. | Analizar y registrar la marca que corresponda, dando prioridad a la marca del mayor nivel de estudios. |
| 8 | En la pregunta H9 hay marca doble | Analizar y registrar la marca que corresponda. Si hay combinación de las opciones 1 a 3 entonces quedará como válida la opción 1. Si hay combinación de la opción 4 con alguna de las opciones 1 a 3, entonces quedará 99 indicando que es una doble marca, |
| 9 | En la pregunta P5 o P19 o P29 hay una marca doble. | Analizar y registrar la marca que corresponda. Se puede dejar marca doble (99) o se puede corregir según corresponda. |

Criterios de Procesamiento para el Censo Nacional 2011 de Costa Rica

4 Reglas especiales para el procesamiento

A continuación se definen reglas especiales que serán aplicadas durante el procesamiento.

| | Regla o situación | Acción a realizar |
|---|--|--|
| 1 | El literal no está normalizado (es decir que contiene caracteres especiales, contiene minúsculas o tildes, no respeta la separación entre palabras) | Se intentará normalizar el literal, tal que quede en mayúsculas, sin tildes, sin caracteres especiales y respetando la separación entre palabras. Esta regla no es obligatoria, sino que se aplicará en función de las posibilidades técnicas de interpretación de esos literales. |
| 2 | El literal contiene faltas de ortografía | Se intentará corregir esas faltas y obtener un literal idéntico en contenido, pero correcto ortográficamente. |
| 3 | Número de Vivienda será capturado sobre 3 dígitos. | Se extenderá la zona de captura de información de este campo en dos dígitos, uno a la derecha y otro a la izquierda. |
| 4 | En las preguntas 5 y 19 de Persona se asignará automáticamente el literal del país para el caso en que esté marcada la opción 3 y un óvalo de alguno de los países | Se imputará el literal del país en el campo "Otro país", de acuerdo con el campo de marca. Esto solo aplica si el literal está vacío. |
| 5 | Para el caso anterior se codificará automáticamente el País. | Se asignará automáticamente el código de país que ha sido marcado en la pregunta 5 o 19. |
| 6 | En caso de presencia de una "doble marca", el sistema asignará el valor numérico 99. | En caso de duda sobre la presencia real de doble marca, el sistema lo enviará a una revisión previa. Se imputa este valor automáticamente. |
| 7 | Campos 24 y 27 expandirlos a 60 caracteres | El reconocimiento se hace normal, pero si hay textos que deben expandirse o que fueron "comprimidos" dentro de la zona normalmente prevista, podrán ser entonces expandidos, pero esta operación no será obligatoria ni sistemática. |
| 8 | Para pregunta p4 "año", el campo contiene menos dígitos que los 4 previstos para el año. | Se deja tal cual viene, sin intentar extrapolar a un año a cuatro dígitos. El valor que se lea será un numérico, independientemente de la posición que tenga dentro del campo. O sea que un 18.. es igual a .18. e igual a ..18, siendo en todos los casos el valor numérico dieciocho. |

Criterios de Procesamiento para el Censo Nacional 2011 de Costa Rica

5. Reglas de Consistencia Estructural

5.1. Alcance del módulo Consistencia Estructural

Estas reglas son aplicadas en la última etapa del proceso, en el módulo “Consistencia Estructural”, una vez que todos los datos ya han sido capturados y están disponibles.

La Consistencia Estructural verifica que la construcción de números, identificadores clave y números consecutivos son consistentes dentro de su UGM, es decir que estas boletas se ordenan siguiendo la pareja AE/UGM. Esta consistencia es indispensable para poder generar archivos de salida consistentes en formato CSPro. El propósito de este módulo es de detectar problemas de consistencia estructural y de corregir esos errores, incluso cuando los mismos fueron escritos por el censista en la boleta papel. Es decir, se trata de una actividad de modificación de la información original, por lo cual la misma solo puede ser realizada por INEC y bajo su responsabilidad.

Tal como puede observarse en la tabla de definición de Reglas de Consistencia, el software busca la detección de inconsistencias apoyándose en la hipótesis de que el resto de los datos es correcto. Por el contrario, si se presentan simultáneamente varios errores o inconsistencias, o si hay un único error suficientemente completo y específico, el software no podrá garantizar la detección de inconsistencia de manera absoluta. El uso de este software por parte de INEC permite entonces de obtener una significativa mejora en la calidad de los datos censales, y de corregir la enorme mayoría de inconsistencias estructurales, pero no se debe asumir que su funcionamiento puede obtener un resultado consistente en todos los casos, debido a la complejidad y multiplicidad de casuísticas. Por esta razón no podrá ser asumido como causa de rechazo de un Lote el hecho de que, habiendo sido capturada correctamente la información de las boletas, persiste de todas formas alguna inconsistencia estructural en dicho Lote.

Si en el lote se han capturado Números de AE y UGM que no son válidos (no en base de datos), el lote si podrá ser devuelto, según se especifica en Regla 2 del punto 5.2.

Criterios de Procesamiento para el Censo Nacional 2011 de Costa Rica

5.2. Lista de Reglas de Consistencia

En la siguiente tabla se presenta la lista completa de Reglas de Consistencia que serán evaluadas.

| Nro. Regla | Inconsistencia observada | Acción del Software y de INEC | Conjunto de valores que ha sido observado en este caso (en color rojo se pone en evidencia el error constatado) | | | | | | Acción manual que se sugiere sea realizada por los operadores autorizados de INEC |
|------------|--|---------------------------------------|--|-----------------------------|-------------------------------|----------------------------|---------------------|---|--|
| | | | Número AE y UGM (Carpeta) | Ejemplo Núm. Orden Vivienda | Ejemplo Número Hogar | Campo Continua | Cantidad de Hogares | Cantidad de Personas del Hogar | |
| 1 | Identificadores Boleta están Vacíos | Bloqueante | Vacío | Vacío | Vacío (salvo si no hay Hogar) | Vacío (salvo si no aplica) | --- | --- | Digitar valor correcto mirando el cartapacio en la que están |
| 2 | Número de AE/UGM de una Boleta es erróneo | Bloqueante | No válido (no existe en la BD de AE/UGM del censo) | | | | | | Analizar cuál debería ser el número correcto y corregir. Es un problema bloqueante. No se puede seguir si eso no se corrige. |
| 3 | Número de AE/UGM de una Boleta es válido pero diferente a los esperados en este Cartapacio | No bloqueante. INEC decide qué hacer. | Diferente al de la Carpeta | | | | | | Analizar cuál debería ser el número correcto. Es posible dejar el mismo sin cambiarlo. No es un problema bloqueante. |
| 4 | Vivienda Repetida caso A | Bloqueante | OK | 1ra: 8 2da: 8 | 1 1 | vacío vacío | 1 1 | Inferior a 7 Inferior a 7 (mirando el campo H5) | Segunda vivienda debería ser otro, podría ser el siguiente o si no el último + 1. Corregir |
| 5 | Vivienda Repetida caso B | Bloqueante | OK | 1ra: 8 2da: 8 | 1 1 | vacío SI | 1 1 | Inferior a 7 Superior a 6 (mirando H5 de la primer boleta) | Segunda vivienda debería ser 9 . Corregir |
| 6 | Vivienda con error en su número | Bloqueante | OK | 1ra: 8 2da: 9 | 1 1 | vacío SI | 1 ---- | Superior a 6 (mirando campo H5) Vacío | Segunda vivienda debería ser 8 . Corregir |

Criterios de Procesamiento para el Censo Nacional 2011 de Costa Rica

| Nro. Regla | Inconsistencia observada | Acción del Software y de INEC | Conjunto de valores que ha sido observado en este caso (en color rojo se pone en evidencia el error constatado) | | | | | | Acción manual que se sugiere sea realizada por los operadores autorizados de INEC |
|------------|---|-------------------------------|--|-----------------------------|----------------------|-----------------------|-----------------------|---|---|
| | | | Número AE y UGM (Carpetá) | Ejemplo Núm. Orden Vivienda | Ejemplo Número Hogar | Campo Continua | Cantidad de Hogares | Cantidad de Personas del Hogar | |
| 7 | Hogar repetido | Bloqueante | OK | 1ra: 8 2da: 8 | 1 1 | vacío vacío | 1ra: 2 2da: existe | Inferior a 7 Inferior a 7 (mirando el campo H5 Total personas) | Segunda boleta Núm. de Hogar debería ser 2. Corregir |
| 8 | Hogar con error en su número | Bloqueante | OK | 1ra: 8 2da: 8 | 1 2 | vacío SI | 1ra: 1 2da: vacío | Superior a 6 (mirando el campo H5) 2da boleta: vacío | Segunda boleta debería ser hogar 1. Corregir |
| 9 | Campo Continuación | Bloqueante | OK | 8 8 | 1 1 | Vacío Vacío | 1ra: 1 2da: vacío | Superior a 6 (mirando el campo H5 de la primer boleta) 2da boleta: vacío | Llenar el campo Continuación de la segunda boleta del hogar. N.B.: esta regla solo aplica si todas las páginas de persona están consecutivas. Si se han utilizado 3 boletas para un hogar de 12 personas (cuando debían ser 2 boletas), el sistema no podrá detectarlo. |
| 10 | Una vivienda con 2 boletas contiene información de Vivienda en las dos boletas (puede ser porque hay 2 hogares en la vivienda o porque hay 1 hogar con más de 6 personas) | Bloqueante. | OK | OK | Ok | OK | OK | OK | Colocar la información pertinente únicamente en la boleta 1, analizando qué información debe ser colocada. El resto se desecha. |
| 11 | Un hogar con dos boletas contiene información de Hogar en las dos boletas (puede ser cuando el hogar tiene más de 6 personas) | Bloqueante. | OK | OK | OK | OK | OK | OK | Colocar la información pertinente únicamente en la primera boleta de cada hogar, analizando qué información debe ser colocada. El resto se desecha. |

Criterios de Procesamiento para el Censo Nacional 2011 de Costa Rica

| Nro. Regla | Inconsistencia observada | Acción del Software y de INEC | Conjunto de valores que ha sido observado en este caso (en color rojo se pone en evidencia el error constatado) | | | | | | Acción manual que se sugiere sea realizada por los operadores autorizados de INEC |
|------------|---|--|--|----------------------------------|----------------------|----------------|---------------------|--------------------------------|---|
| | | | Número AE y UGM (Carpeta) | Ejemplo Núm. Orden Vivienda | Ejemplo Número Hogar | Campo Continua | Cantidad de Hogares | Cantidad de Personas del Hogar | |
| 12 | Cartapacio realizado por dos empadronadores y construido uniendo sus cartapacios físicos diferentes, pero tal que se utilizó el número de AE de solo uno de ellos. Puede ser uno de los casos de "barrido". Los números de Vivienda se repiten. | No bloqueante | Todos iguales | Repetidos entre las dos carpetas | | | | | Mantener el número de AE que se usó en terreno, y reasignar los números de vivienda que estén repetidos. |
| 13 | Una vivienda tiene al menos una persona pero no tiene información de Hogar y no está marcada la opción 14 de la V1. | Bloqueante. INEC decide qué hacer. Se archivó CSPro solo guardará la información de la primer boleta de la vivienda. | OK | OK | OK | OK | OK | OK | No es aceptable que una persona no pertenezca a un Hogar. Se debe resolver esta inconsistencia, ya sea borrando la persona (porque en realidad era una persona borrada y una Vivienda sin moradores, o agregando información del hogar porque el empadronador olvidó registrar información del hogar. |
| 14 | Una vivienda ha sido transcripta o ha sido duplicada, tal que hay dos boletas completas e iguales para la misma vivienda. | Bloqueante | OK | OK | OK | OK | OK | OK | Una de las dos debe ser eliminada. La forma de marcar la que será eliminada es de asignarle un número de vivienda superior a 900. Durante los procesos de crítica e imputación estos registros de vivienda completos deberán ser eliminados. |

Criterios de Procesamiento para el Censo Nacional 2011 de Costa Rica

6. Criterios de trabajo para las tareas de Verificación interactivas

6.1. Reglas a seguir en la etapa “Verificación de Reglas”

A continuación se definen las reglas de consistencia de datos que el sistema verifica en forma automática.

Si los datos reconocidos no cumplen estas reglas, los mismos serán mostrados al operador del puesto “Verificación de Reglas” para su verificación y corrección si aplica.

| | Regla o situación | Acción a realizar |
|----|--|--|
| 1 | La edad de la persona no es coherente con relación a su fecha de nacimiento. | Verificar si está bien reconocido por el sistema. En todos los casos se debe dejar el valor que ha sido escrito en el papel. |
| 2 | El Numero de Hogar debería estar comprendido entre 0 y 4 | Se supone que si hay más de 4 hogares es altamente probable que se trate de un error de interpretación, pero de todas formas, es posible. El usuario debe ingresar el valor que se muestra en pantalla a pesar que dichos valores sean incorrectos y no cumplan la regla. Si es número mayor a 9, se ingresará 9. |
| 3 | Líneas de teléfono - número entre 5 y 20. | El usuario debe ingresar el valor que se muestra en pantalla. Valor máximo 99. |
| 4 | Total de personas (página 3 – campo 5) | El usuario debe ingresar el valor que se muestra en pantalla a pesar que dichos valores sean incorrectos y no cumplan la regla. |
| 5 | Total de persona colectiva (página 3 – campo 6) | El usuario debe ingresar el valor que se muestra en pantalla a pesar que dichos valores sean incorrectos y no cumplan la regla. |
| 8 | El Total de personas de la Vivienda debe ser superior o igual al Total de personas del Hogar | El usuario debe ingresar el valor que se muestra en pantalla a pesar que dichos valores sean incorrectos y no cumplan la regla. |
| 9 | Cantidad de personas fuera del país. más de 9 | El usuario debe ingresar el valor que se muestra en pantalla a pesar que dichos valores sean incorrectos y no cumplan la regla. Si es número mayor a 9, se ingresará 9. |
| 10 | El Año de Llegada al país debe ser superior o igual al año de nacimiento | El usuario debe ingresar el valor que se muestra en pantalla a pesar que dichos valores sean incorrectos y no cumplan la regla. |
| 11 | Día - entre 1 y 31 | El usuario debe ingresar el valor que se muestra en pantalla a pesar que dichos valores sean incorrectos y no cumplan la regla. |
| 12 | Mes - entre 1 y 12 | El usuario debe ingresar el valor que se muestra en pantalla a pesar que dichos valores sean incorrectos y no cumplan la regla. |
| 13 | Año - entre 1910 – 2011 | El usuario debe ingresar el valor que se muestra en pantalla a pesar que dichos valores sean incorrectos y no cumplan la regla. |
| 14 | Edad entre 1 - 100 | El usuario debe ingresar el valor que se muestra en pantalla a pesar que dichos valores sean incorrectos y no cumplan la regla. Si es número mayor a 999, se ingresará 999 y se deberá anotar el |

Criterios de Procesamiento para el Censo Nacional 2011 de Costa Rica

| | Regla o situación | Acción a realizar |
|----|--|---|
| | | número de lote, número de carpeta y número de formulario. También debe anotar el valor real que se debería ingresar y en cual campo. |
| 15 | Año de Llegada entre 1910 y 2011 | El usuario debe ingresar el valor que se muestra en pantalla a pesar que dichos valores sean incorrectos y no cumplan la regla. |
| 16 | Hijos entre 1 y 12 | El usuario debe ingresar el valor que se muestra en pantalla a pesar que dichos valores sean incorrectos y no cumplan la regla. Si es número mayor a 99, se ingresará 99. |
| 17 | Totales de Hijos Vivos <= total de hijos | El usuario debe ingresar el valor que se muestra en pantalla a pesar que dichos valores sean incorrectos y no cumplan la regla. |
| 18 | Página vacía - si sexo y parentesco vacío - entonces se sospecha que todo debería estar vacío. | El usuario debe ingresar el valor que se muestra en pantalla a pesar que dichos valores sean incorrectos y no cumplan la regla. O sea que se deja cualquier campo que esté claramente lleno o escrito, así sea uno solo en toda la Persona. |
| | | |

6.2. Reglas a seguir en la etapa “Verificación” de números, marcas y texto, y en el control de calidad

6.2.1. ¿Cómo trabajar con los campos de tipo MARCA?

| | Regla o situación | Acción a realizar |
|---|---|--|
| 1 | Si un óvalo de una pregunta está claramente marcado entonces se debe registrar en el sistema. | Registrar el número que corresponde con dicho óvalo. Este registro debe hacerse incluso en el caso en que pudiera pensarse que esa respuesta no es lógica. En este instante solo debemos registrar fielmente lo que está escrito, sin poner en cuestión su coherencia. Existe una excepción a esta regla, es para el campo P16. Ver regla número 3. |
| 2 | Si en una misma pregunta de marcas hay más de un óvalo marcados claramente | Se debe ingresar el valor 99. Eso se denomina “doble marca”. Cuidado: muchas veces es una “falsa doble marca”, ya que algún óvalo fue marcado pero luego borrado. La práctica debe permitir hacer la diferencia en la “apariencia” que presenta una verdadera marca frente a cómo se ve una marca borrada. Ver más detalles en el siguiente punto. |
| 3 | Caso particular de “doble marca” para el campo P16 “último grado que aprobó”. Hay casos en que se han marcado varios óvalos, pero el que vale es el nivel más alto. | Se decide que se adoptará como buena únicamente la marca en el óvalo de mayor nivel de estudio, por lo cual se desestima la doble marca para este caso particular. |

Criterios de Procesamiento para el Censo Nacional 2011 de Costa Rica

| | Regla o situación | Acción a realizar |
|----|---|--|
| 4 | Si en una pregunta no hay ningún óvalo marcado | No se debe ingresar ningún valor, o sea que debe quedar en "blanco". |
| 5 | En el óvalo hay una traza visible con apariencia de ser una marca borrada | Observar dos cosas: 1) mirar cómo son el resto de las marcas de esta página y de esta boleta y 2) analizar si este óvalo dudoso es consistente o inconsistente con el contexto. Si este óvalo no es consistente con el contexto y esta marquita es diferente al patrón de marquitas hechas, entonces podemos deducir que es una marquita borrada, por lo que será en "blanco". Por el contrario, si es una marquita consistente y posible, será adoptada como "negro". Finalmente, si esto fuese inconsistente con el contexto, incluyendo los pases y flujos de la boleta, pero la traza visible es suficientemente oscura y abundante, entonces será de todas formas considerada como "negro". Existe un grado de incertidumbre en la decisión, que tendrá un fuerte componente subjetivo. |
| 6 | En el óvalo hay una traza visible de escritura no conforme, pero visible. Podría ser una rayita horizontal, un check, una cruz. | Si es consistente con el contexto esta marquita será rescatada como siendo "negro", pero solo en el caso en que sea claramente visible y de tamaño suficiente. Si es muy pequeña, apenas visible y claramente diferente del resto del patrón de marcas de esa página será considerada como "blanca". Existe un grado de incertidumbre en la decisión, que tendrá un fuerte componente subjetivo. |
| 7 | En el óvalo hay una traza apenas visible de escritura, bajo forma de un puntito, una suciedad u otra forma no estándar | Analizar el contexto, incluyendo los pases y flujos de la boleta, para tratar de determinar si podría tratarse de una marquita o no. Si este óvalo no es consistente con el contexto entonces la decisión será "blanco". Por el contrario, si es consistente y es suficientemente visible como para poder representar una marca, será considerado "negro". Existe un grado de incertidumbre en la decisión, que tendrá un fuerte componente subjetivo. |
| 8 | Hay escritura visible fuera del óvalo previsto. | Será ignorada. No se puede deducir que se trate de una marca. |
| 9 | En el óvalo hay una traza de escritura claramente visible que dibuja el contorno del óvalo, pero sin rellenarlo. | Al igual que en los casos anteriores, se analizará el contexto, siendo que se intentará considerar como una marca normal. |
| 10 | Existe una aparente doble marca, tal que una de ellas es claramente una marca y la otra(s) es dudosa | Se asumirá como únicamente buena la marca más fuerte, y no se señalará como doble marca. |

Criterios de Procesamiento para el Censo Nacional 2011 de Costa Rica

6.2.2. ¿Cómo trabajar con los campos de tipo NUMEROS?

| | Regla o situación | Acción a realizar |
|----|--|--|
| 1 | Se lee claramente un valor | Se debe digitar exactamente el valor que se lee. |
| 2 | No se entiende bien cuál es el valor escrito. | Analizar el contexto para tratar de interpretar lo que se ha escrito. Comparar este número con otros similares escritos por la misma persona en esta boleta. |
| 3 | Los números escritos no son muy bien definidos y parecen ser totalmente inconsistentes: por ejemplo, la edad de un Jefe de Familia es 05, siendo que probablemente se trata de 65, ya que el 6 está mal escrito. | Tratar de interpretar lo que realmente quiso escribir el empadronador, mirando números similares en la boleta y analizando el contexto. |
| 4 | Los números escritos sobresalen del espacio de escritura previsto. | Tratar de interpretar lo que está escrito y digitarlo completamente. |
| 5 | El campo Número de Vivienda tiene más de 2 dígitos, ya sea escritos dentro de los 2 rectángulos o escritos por fuera. | Esto sucede cuando hay más de 99 viviendas en la UGM. Se debe digitar el número completo que está escrito, ya sea que tenga dígitos extra a la izquierda o a la derecha. El resultado podrá entonces ser un número de 3 dígitos como máximo. |
| 6 | Pregunta 7 del Bloque III, la cantidad de personas es mayor a 9. | Debe ingresarse el valor 9. |
| 7 | Pregunta 19 del Bloque II, la cantidad de teléfonos celulares es mayor a 99. | Debe ingresarse el valor 99. |
| 8 | Pregunta 1 y 5 del Bloque III. Cantidad mayor a 99. | Debe ingresarse el valor 99. |
| 9 | Bloque I, Hogar mayor a 9, ya sea porque se escribe al lado o porque se escriben dos cifras dentro del único rectángulo. | Debe ingresarse el valor 9. |
| 10 | El campo años o año (preguntas 3 y 4 de Persona) no han completado todos los dígitos previstos. | Se debe interpretar como un número común, sin tener para nada en cuenta la posición de los dígitos dentro del campo. Es decir que en ningún caso hay que completar con ceros a la izquierda ni a la derecha, sino simplemente escribir el número que se lee. |
| 11 | El campo numérico contiene información inconsistente o extranjera, como por ejemplo está marcado (marca) o texto, en lugar de cifras. | Es una inconsistencia que no puede ser resuelta con interpretación, por lo que se colocará el valor máximo que acepta el campo (puede ser "9", o "99", o "999", o "9999"). |

Criterios de Procesamiento para el Censo Nacional 2011 de Costa Rica

6.2.3. ¿Cómo trabajar con los campos de de tipo TEXTO?

| | Regla o situación | Acción a realizar |
|---|--|---|
| 1 | Respetar el texto que ha escrito el empadronador. | Copiar fielmente el texto que ha escrito el empadronador. |
| 2 | El texto no está normalizado | Intentar normalizar: <ul style="list-style-type: none"> - Todo en mayúsculas - No aceptar caracteres especiales ni tildes - Corregir las faltas de ortografía simples y fáciles de corregir. Si es muy complicado, dejar como está. No es obligatorio corregir. - Si hay abreviaciones, intentar expandirlas, pero sólo si las mismas son fáciles y evidentes. De lo contrario dejar el texto como está. Por ejemplo si dice DR PEDIATRA se puede expandir a DOCTOR PEDIATRA. |
| 3 | Dos palabras están pegadas (o sea, no separadas por un espacio en blanco). A veces esto sucede cuando la primer palabra termina en el primer renglón y la segunda palabra comienza en renglón siguiente. | Se deben separar, introduciendo un espacio en blanco entre ambas. |
| 4 | El texto no respeta los lugares (rectángulos blancos) previstos de escritura | Tratar de todas formas de interpretar el texto completo y digitarlo. Se dispone de hasta 60 caracteres para escribir. Pueden haber casos incluso que se escribe fuera del espacio de la pregunta. Si es simple se debe intentar digitar todo eso escrito. |
| 5 | El texto escrito está muy borroso. Esto puede suceder porque se trata de un texto que fue borrado, o porque el empadronador no utilizó el lápiz adecuado. | Hay que hacer un estudio para determinar si se trata de texto borrado o de texto escrito con lápiz inadecuado. Se aconseja comparar esa escritura con el contexto de la boleta: si todos los demás textos son bien negros y precisos y este texto es borroso, muy probablemente sea un texto borrado. En dicho caso ese texto borrado será dejado enteramente en blanco. |

Criterios de Procesamiento para el Censo Nacional 2011 de Costa Rica

7. Reglas para la Codificación de Variables Abiertas

7.1. Reglas para la codificación Automática

El sistema de procesamiento y de codificación dispone de un módulo de “Codificación Automática”. Este módulo analiza los literales de las variables abiertas definidas e intenta determinar cuál es el código numérico que le corresponde. Para realizar esta tarea el software se apoya en un referencial oficial denominado “diccionario por campo”: se trata de archivos de parejas literal/código numérico.

Se asume por definición que todo literal presente en este diccionario tiene asociado un código correcto.

Este archivo o diccionario es entregado por INEC al Consorcio al inicio del proyecto, y luego es actualizado exclusivamente por parte de representantes de INEC.

Existen cuatro tipos de diccionarios:

- De Ocupación
- De Rama de Actividad
- De Cantón
- De País
- Diccionario de parejas Ocupación/Rama de Actividad

7.2. Reglas para la codificación Asistida

| | Regla o situación | Acción a realizar |
|---|--|---|
| 1 | Uno y solo uno de los dos campos 24 y 27 está vacío, mientras que el otro contiene un literal. | Se enviará a pantalla de Codificación Asistida doble, a los efectos que el operador decida si es posible asignar algún literal y código al campo que está vacío. Se debe aceptar que un literal vacío tenga código numérico. Esto aplica en todos los casos, incluso cuando haya sido posible codificar automáticamente el campo que contiene un literal. |
| 2 | Para el caso de marca de “otro país” para las preguntas p5 y 019, se codificará automáticamente el País. | Se asignará automáticamente el código de país que ha sido marcado en la pregunta 5 o 19. |
| 3 | Para el caso de que en las preguntas 5 y 19 y 29, el subcampo “Cantones”, vinieran literales de países, autorizar a que sean codificados en ese lugar. | Aceptar codificar un nombre de un País que haya sido escrito en el campo Cantón. Solo será codificado, pero el resultado quedará dentro de la variable Cantón. Será responsabilidad posterior de INEC la extracción de esos literales/códigos y llevarlos a la variable correspondiente |
| 4 | Para el caso de los literales de “Otro país” podría haber literales de Cantones. En preguntas p5 y p19. | Aceptar codificar un nombre de Cantón que haya sido escrito en el campo País. Solo será codificado, pero el resultado quedará dentro de la variable País. Será responsabilidad posterior de INEC la extracción de esos literales/códigos y llevarlos a la variable correspondiente. Esta forma de trabajo deberá ser evaluada en función del eventual impacto negativo que tenga para la codificación |

Criterios de Procesamiento para el Censo Nacional 2011 de Costa Rica

| | Regla o situación | Acción a realizar |
|---|---|---|
| | | automática de los países. |
| 5 | Mejorar la información contextual para la codificación asistida de Rama de Actividad y Ocupación. | En pantalla de codificación asistida, para el caso de codificación de P24 y p27, se presentará en pantalla además los valores de la P3, p16, p17 y p28. Adicionalmente el operador podrá solicitar manualmente la visualización de las preguntas p25 y p26, presentando en pantalla la imagen de la página que está siendo analizada. |
| 6 | Manejo de casos que no sean codificables | Se utilizará el código 4 (cuatro) para cuando definitivamente no sea posible asignar un código en Ocupación o Rama de Actividad. |
| 7 | Manejo de casos que no sean pertinentes | Se utilizará el código 89 para cuando el literal no aplica para ese contexto. Para ello se incluirá también en las tablas de codificación automática los literales que sean definidos como “no aplicables”. |
| 8 | El literal ha sido mal reconocido por el sistema. | El operador deberá corregirlo y digitarlo correctamente. |

7.3. Reglas para la Actualización de Diccionarios de referencia

Los Diccionarios de referencia para la Codificación son archivos incrementales. La versión inicial de estos archivos es entregada por INEC al Consorcio al inicio del procesamiento, y luego los responsables de INEC harán una actualización incremental de los mismos a medida que el sistema informático va identificando nuevos literales.

Se asume que la actualización de estos diccionarios será realizada por INEC dentro del procesamiento, siguiendo el flujo y ritmo normal de procesamiento del material censal, procurando que no se produzcan atrasos en la actualización de los diccionarios. Dicho de otra manera, INEC deberá analizar y tomar decisiones sobre todos los nuevos literales que aparezcan cada día, con el fin de actualizar los diccionarios de referencia al finalizar el día y sin que queden literales pendientes de ser analizados. Podrían surgir excepciones a este principio metodológico, pero la tendencia general debe ser compatible con esos principios antes enunciados.

El Consorcio proporcionará a los responsables de INEC el software, la capacitación a su uso, el o los puestos de trabajo necesarios para esta tarea.

Criterios de Procesamiento para el Censo Nacional 2011 de Costa Rica

8. Definición del nivel de error aceptable

8.1. Tasa de error máxima aceptable

Se establece que INEC aceptará los lotes INEC procesados cuando el nivel de calidad de los mismos sea compatible con un cierto valor de calidad establecido de común acuerdo. El nivel de calidad se define con relación a una “tasa de error aceptable” como siendo el máximo porcentaje de error que será tolerado por INEC en cada Lote procesado por el Consorcio, desagregado por tipo de variable. Si el error observado en un Lote INEC es inferior o igual a la máxima tasa de error aceptada, entonces se asume que dicho lote posee un nivel de calidad aceptable. Ese nivel de calidad de un Lote INEC se mide de la siguiente forma:

| Tipo de Variable | Porcentaje máximo de error aceptado | Unidad de medida |
|-----------------------|-------------------------------------|---|
| Marcas | 0,5% (cero coma cinco) | A nivel de campo |
| Númericas | 1% (uno) | A nivel de campo (a exclusión del campo V20 “teléfono”) |
| Literales alfabéticos | 2% (dos) | A nivel de campo |

8.2. Forma de cálculo de la “tasa de error”

La tasa de error es el cociente de la cantidad total de casos de error sobre el universo de ese tipo de variable aplicado a la totalidad del Lote.

El “**caso de error**” se define como aquel donde el resultado final capturado es de un cierto valor y razonablemente se debería haber interpretado otro valor diferente a partir de la lectura de la variable analizada. Es decir que los errores se contabilizarán únicamente cuando se ha interpretado mal algo que estaba correctamente escrito.

Para el caso de los campos de números y texto no se contabilizarán como “caso de error” a aquellos casos en que la grafía original sea tan confusa o que haya sido sobrescrita tal que genere una duda razonable respecto al valor que realmente se quiso escribir.

Para el caso de los campos de marca, no se contabilizarán como “caso de error” aquellos en que los óvalos hayan sido mal borrados, ni tampoco aquellos casos en que la indicación de marca sobre un óvalo sea tan tenue o pequeña que dificulte su interpretación.

El universo de una variable se calcula incluyendo todas las variables que han sido analizadas en el proceso de captura. A continuación se detalla la forma de cálculo de error aceptable para cada tipo de variable:

Criterios de Procesamiento para el Censo Nacional 2011 de Costa Rica

| Tipo de Variable | Cálculo del error | Cálculo del Universo |
|-------------------------|---|---|
| Marcas | Suma de todos los campos con error del Lote. Se contabiliza 1 error por cada campo erróneo. | Suma de todos los campos de marca de una boleta. |
| Numéricas | Suma de todos los campos con error del lote (excluyendo el campo V20 – “teléfono”). Se contabiliza 1 error por cada campo erróneo. | Suma de todos los campos numéricos de una boleta. |
| Literales alfanuméricos | Suma de todos los campos erróneos del lote, teniendo en cuenta que los literales pueden ser normalizados en el proceso de captura (es decir, no se hará comparación string a string). Se contabilizará 1 error por cada campo erróneo, pero solo si eso ha ocasionado una interpretación de la “información” incorrecta en la codificación. (Suma de los campos con información que resultaron erróneos. Se estaría contabilizando 1 error cuando se ha capturado una frase que cambia completamente el sentido de lo escrito en la boleta, y cuando en la reconstrucción de textos equivocadamente se escribió una frase que cambia completamente el sentido de lo escrito en la boleta). | Suma de todos los campos escritos en las boletas del lote (Suma de todos los campos de texto con información en la boleta) |

Para efectos de la revisión de marcas, números y texto, el INEC implementará el siguiente procedimiento :

1. Registrar en la plantilla de CSpro cada caso de acuerdo al siguiente diccionario:

a) Para los campos de marca se utilizarán los siguientes valores:

- 1) No registró la marca.
- 2) Sin marca y registró marca.
- 3) Doble marca y no registró 99.
- 4) Una sola marca y el sistema asignó 99.
- 5) Sin marca y asignó 99.
- 6) El número asignado no coincide.

Criterios de Procesamiento para el Censo Nacional 2011 de Costa Rica

- 7) El campo es blanco.
- 8) El campo de marca es correcto.

b) Para reconocimiento numérico se utilizarán los siguientes valores:

- 1) El número capturado no coincide con el número escrito.
- 2) Existe número escrito y no hay número capturado.
- 3) El campo numérico es blanco.
- 4) El campo numérico es correcto.

c) Para reconocimiento alfabético se utilizarán los siguientes errores:

- 1) El texto capturado no coincide con el texto escrito.
- 2) Existe texto escrito y no hay texto capturado.
- 3) El campo alfabético es blanco.
- 4) El campo alfabético es correcto.

2. Generar las estadísticas de error a través del CSpro:

Las estadísticas de error corresponden al porcentaje de campos con error según el tipo de marca. Las fórmulas para el cálculo de los mismos se detallan a continuación:

- La fórmula de cálculo del error para los campos de marca es:

$$\left(\frac{\text{Suma de campos erróneos de marca}}{\text{Total de campos válidos de marca}} \right) * 100$$

Los campos erróneos de marca son aquellos que tienen valores de 1 a 6 según el diccionario citado en el punto 4 de la metodología. El total de campos válidos de marca corresponde a la suma de los valores 1 a 8 indicados en la metodología en el punto 4a.

- La fórmula de cálculo del error para los campos numéricos es:

Criterios de Procesamiento para el Censo Nacional 2011 de Costa Rica

$$\left(\frac{\text{Suma de campos numéricos erróneos}}{\text{Total de campos numéricos válidos}} \right) * 100$$

Los numéricos erróneos son aquellos que tienen valores 1 y 2 según el diccionario citado en el punto 4 de la metodología. El total de campos numéricos válidos corresponde a la suma de los valores 1 a 4 indicados en la metodología en el punto 4b.

- La fórmula de cálculo del error para los campos alfabéticos es:

$$\left(\frac{\text{Suma de campos alfabéticos erróneos}}{\text{Total de campos alfabéticos válidos}} \right) * 100$$

Los campos alfabéticos erróneos son aquellos que tienen valores 1 y 2 según el diccionario citado en el punto 4 de la metodología. El total de campos alfabéticos válidos corresponde a la suma de los valores 1, 2 y 4 indicados en la metodología en el punto 4c.

3. Aceptar o rechazar el lote.

- El lote se rechazará si el porcentaje de error de campos de marca es mayor al 0,5% o si el porcentaje de campos numéricos con error es mayor al 1% o si el porcentaje de campos de texto es mayor al 2%.
- De ser rechazado el lote el proceso de revisión de la muestra se hará una vez más y el lote se rechazará si el porcentaje de campos de marca es mayor al 0,25% o si el porcentaje de campos numéricos con error es mayor al 0,5% o si el porcentaje de campos de texto es mayor al 1%.
- Si el lote se rechaza una tercera vez los criterios a tomar en consideración serán definidos por el INEC.

Criterios de Procesamiento para el Censo Nacional 2011 de Costa Rica

9. Flujo de procesamiento y procedimientos de entrega y aceptación

9.1. Trabajo por Lote INEC

Se establece el acuerdo de que la unidad de intercambio y evaluación es el Lote INEC. El Consorcio ha entregado a INEC una base de datos conteniendo la descripción completa y detallada de todos los Lotes, su localización y su contenido (Cajas, Cartapacios y UGM), tal como existen físicamente en la Bodega Censal. Siendo que en el transcurso del proceso está previsto la recepción de más material censal (los “pendientes”), la base de datos del contenido de la bodega será actualizada, y el Consorcio entregará en cada caso la versión actualizada, de tal forma que ambas partes estén utilizando el mismo referencial respecto del material que existe en la bodega.

9.2. Secuencia de trabajo

Se establece el acuerdo que el Consorcio va a entregar a INEC lotes procesados según una secuencia de números de lote INEC definida por INEC, de común acuerdo con el Consorcio.

El Consorcio hará su mejor esfuerzo para respetar este orden de entrega de Lotes a INEC, aunque excepcionalmente podría no respetarse por causa de que algún elemento de un cierto Lote tenga problemas que impliquen postergar su culminación.

9.3. Procedimientos a seguir para la Devolución y Aceptación de Lotes Procesados

Se establece el acuerdo que para la devolución de Lotes Procesados por parte del Consorcio a INEC se seguirá el siguiente procedimiento:

- 1) El Consorcio entrega a INEC uno o varios Lotes Procesados, entregando para cada Lote un documento de Devolución. Dicho documento indica el número de Lote y la lista de cajas que están siendo devueltas, así como la localización de dicho Lote en la Bodega Central. Acompañando cada Lote se entrega un DVD conteniendo la totalidad de los datos y de las imágenes correspondientes al lote.
- 2) INEC hará una revisión superficial del material recibido, verificando que las cajas devueltas corresponden a lo declarado en el documento de Devolución. Si corresponde entonces un responsable de INEC firmará la recepción provisoria del lote, y se aceptará la fecha de la recepción provisoria indicada en el documento, o de ser diferente, se pondrá separadamente la fecha del día.
- 3) A partir de ese momento comienza el período de 5 días hábiles dentro de los cuales INEC podrá emitir una notificación de rechazo del Lote. Una vez terminado ese período y sin haber recibido una notificación contraria, el Lote quedará automática y definitivamente Aceptado.
- 4) En caso de existir una notificación de rechazo del Lote, INEC entregará un documento al responsable del Consorcio, donde se consignará la fecha de dicha notificación. Asimismo, INEC dispone del acceso al módulo de Gestión de Lotes INEC, para registrar en el mismo la aceptación o rechazo de un lote INEC por parte de un operador autorizado de INEC.

Criterios de Procesamiento para el Censo Nacional 2011 de Costa Rica

- 5) En todos los casos, ya sea de aceptación o rechazo, INEC acumulará el conjunto de la información relativa a su evaluación y entregará una única respuesta, incluyendo la totalidad de las observaciones. Se considerará que luego de entregado dicho documento no habrá por parte de INEC ninguna otra comunicación con nuevos casos. Esto podría ser negociable entre las partes.
- 6) En todos los casos ese Lote de material físico (las Cajas y su contenido) será mantenido en la posición fija asignada en la Bodega Central, y estará a disposición de INEC durante el período de evaluación establecido. En caso de requerirlo, INEC deberá disponer de sus propios recursos para extraer cajas de dicho Lote, con fines de análisis. Una vez culminado dicho análisis las cajas que eventualmente INEC haya extraído deberán ser nuevamente colocadas en su sitio.