

**SERVICIO DE CODIFICACION Y LIMPIEZA DE DATOS DE LA
ENCUESTA NACIONAL DE CONSUMO FAMILIAR DE
ALIMENTOS – ENCOFA
AL
CENTRO NACIONAL DE ALIMENTACION Y NUTRICION
DEL
INSTITUTO NACIONAL DE SALUD**

**INFORME SOBRE EL 100%
DE LA LIMPIEZA DE LA BASE DE
DATOS ENCOFA**

Lizette Ganoza Morón
Noviembre del 2007

INFORME SOBRE EL 100 POR CIENTO DE LA BASE DE DATOS LIMPIA DE LA ENCUESTA DE CONSUMO FAMILIAR - ENCOFA 2006.

Lizette Ganoza Morón

Como parte de la Consultoría del Servicio de Codificación y Limpieza de datos de la Encuesta Nacional de Consumo Familiar de Alimentos – ENCOFA se hace entrega del III Producto: Informe sobre el 100% de la base de datos limpia de la ENCOFA.

La Encuesta Nacional de Consumo Familiar de Alimentos 2006 se encuentra en un archivo denominado Base ENCOFA 2006.xls, la cual esta conformada por diversas hojas de cálculo. Estas son:

- 100: contiene información acerca del Jefe de Familia, identificación del formulario, identificación y dirección de la vivienda, características y servicios básicos del hogar además del nombre del encuestador y supervisor.
- 200: identificación del formulario, características de los miembros del hogar, datos de altitud y hemoglobina.
- 300: identificación del formulario e identificación de los comensales
- 400: identificación del formulario y consumo familiar de alimentos que consta de 2 hojas de cálculo: 400 - 1 y 400 - 2
- 500: identificación del formulario y frecuencia de consumo semicuantitativa de alimentos ricos en vitamina A y hierro.
- 600: identificación del formulario y antropometría de los miembros del hogar

Siendo el objeto de esta consultoría el consumo familiar de alimentos se hará referencia a lo encontrado en este.

Antes de ello es necesario mencionar que se ha encontrado errores en los códigos de alimentos empleados en la pregunta 411, en otros casos ha habido alimentos que no concordaban con la encuesta por lo que han tenido que ser eliminados, y en algunos casos se ha tenido que añadir alimentos que se encontraban en la encuesta y no se encontraban en la base de datos digitalizada.

Existe un listado de 330 alimentos que no presentaban nombre del alimento ni descripción de la preparación, en estos casos ha sido necesario remitirse a las encuestas para poder realizar las correcciones respectivas en la base de datos por lo que actualmente existe mayor número de alimentos codificados.

Se ha efectuado la revisión de todas las encuestas que se encuentran en la base de datos entregada por el personal a cargo de la ENCOFA para verificar su ámbito, vivienda, el número de hogar al que pertenecen así como el día en que se realizó la encuesta. Se ha encontrado diversos errores los cuales son mencionados en el presente informe.

Después de realizar una primera etapa de limpieza de la información presentada en el Consumo Familiar de Alimentos que consistió en emplear rangos de cantidades para cada alimento, se decidió efectuar la verificación de cada uno de los datos presentados por 16 conglomerados que equivaldría a un 10% de la información total. Mediante esta revisión se ha encontrado errores en cantidades presentadas para algún alimento, en el código asignado a alimentos así como también datos que no

habían sido digitados y formaban parte de la encuesta. De los 16 conglomerados revisados 15 presentaron algunos de los tipos de errores mencionados.

Las personas encargadas de la planificación y análisis de las encuestas de ENCOFA deben tener en cuenta que el proceso de limpieza demanda un tiempo prolongado en especial si se tiene en cuenta la cantidad de información que presenta esta base de datos por lo que el análisis de la información no puede ser realizado en paralelo.

También se hace mención en las siguientes hojas acerca de la dificultad de cruzar la información de la descripción de la preparación con los alimentos que la conforman, esto hubiese sido interesante realizar pero no ha sido factible, además de que no forma parte de esta consultoría.

Limpieza de Datos

La información de la Encuesta Nacional de Consumo Familiar se encuentra ingresada en formato Excel en las hojas de cálculo denominadas: 400-1.xls. y 400-2.xls.

En el archivo de limpieza de datos se ha tratado de diferenciar los errores realizados tanto por el digitador como por el encuestador asignando el color azul a aquellos errores realizados por el digitador mientras que el color fucsia corresponde a los errores efectuados por el encuestador.

En el proceso de limpieza de datos se tiene lo siguiente:

1. La pregunta 401 (Número de orden) según el manual de crítica de la encuesta nacional de consumo familiar de alimentos (ENCOFA) presenta un rango de 1 hasta 30. En el transcurso de la limpieza de datos se hizo la verificación de esta información y se encontró que 14 alimentos presentaban un rango mayor de 30, al preguntar al encargado de las encuestas sobre esto mencionó que entonces deberían de haber ampliado el rango, por lo que se puede decir que no ha habido consistencia en el ingreso de la información.
2. Se ha encontrado errores en la pregunta 401 tanto por parte del encuestador como por parte del digitador en estos casos se encontró que no se siguió debidamente el orden correlativo de los números.
3. Se ha tratado de realizar el cruce de información de la descripción de la preparación con la descripción de los ingredientes pero existe el problema de que solo la primera línea de cada preparación tiene la información concerniente a la preparación por lo que no ha sido posible efectuarlo ya que al tratar de hacerlo ha ocurrido que los demás alimentos de una misma preparación se mezclaban con los otros alimentos de otras preparaciones.
4. Para realizar el tipo de cruce de información mencionado en el acápite 3, se debería de haber empleado para el ingreso de los datos de la encuesta un adecuado programa de ingreso con sus respectivos rangos y consistencias en el cual se repetía el número de preparación a la que pertenecía determinado alimento.
5. El tiempo de comida solo ha sido digitado en la primera preparación y en el primer alimento de esta preparación más no así en todos los alimentos y

preparaciones que abarcaban este tiempo de comida, si se quisiera realizar un análisis de la información obtenida por ámbito y por tiempo de comida esto no se podría realizar ya que habría mezcla de los datos de las diferentes preparaciones y tiempos de comida. Para llevar a cabo este tipo de análisis se tendría que colocar toda la información faltante a cada alimento que forma la base de datos de la encuesta ENCOFA lo cual demandaría tiempo y no es el objetivo de esta consultoría.

6. Por las razones expuestas en los acápites que preceden a este, no es factible realizar este tipo de limpieza de los datos por ende no se puede realizar un análisis a ese nivel.
7. En la pregunta 409 Componente de enlace se ha corregido errores realizados tanto por el digitador como por el encuestador. En el caso del digitador ha ocurrido que no digito la información o esto lo realizó en forma errónea y en el caso del encuestador error en el enlace asignado.
8. En el formato de la encuesta (P. 409) se menciona que el componente de enlace esta constituido por el número de orden/número de componente (401/406), aquí existe un error ya que el número de componente es el 404 y no el 406.
9. Con respecto a la descripción del alimento se ha encontrado errores por parte del digitador posiblemente por que al digitar erróneamente el código del ingrediente (P. 411) aparecía la descripción del alimento, esta situación también puede ocurrir si se digitaba erróneamente el alimento aparecía el código correspondiente en la columna 411.
10. Con respecto a la pregunta 411(código del alimento) se ha encontrado que 330 alimentos no solo no presentaban esta información sino que tampoco existía una descripción del alimento por lo que ha sido necesario verificar cada uno de estos datos en las encuestas (fuente original).
11. La mayoría de los casos que no presentaban descripción del alimento correspondían a especies como ajinomoto, sibarita, palillo entre otras, hierbas como manzanilla, anís, etc., y en algunos casos esto correspondía a vitaminas pastillas, como uña de gato, magnesol, vitamina E entre otras.
12. La revisión de los alimentos mencionados en los acápites 10 y 11 se tuvo que realizar ya que era imposible saber con anticipación de que alimento se trataba al no presentar información en la mayoría de los casos en la pregunta 405 (descripción de la preparación).
13. Es necesario mencionar que cuando se realiza el ingreso de la información esto debe ser realizado con toda la información que se encuentra en la encuesta aún si la cantidad consumida sea cero, lastimosamente este no es el caso de la encuestas digitadas de la ENCOFA en la que muchos alimentos no han sido ingresados por tener un consumo de cero. Cabría saber si hubo alguna disposición por parte del personal encargado en que estos casos no tenía que ser digitado y si hubiese sido así debería haberse empleado el mismo criterio para todas las encuestas.

14. La mayoría de los alimentos que conforman el grupo de los 330 alimentos que no presentaban descripción del alimento (P. 410) y código del alimento (P. 411) no tienen un código asignado por parte de CENAN.
15. Algunos alimentos han sido asemejados a otros por lo que en algunos casos estos no coinciden con la información proporcionada en la encuesta hubiese sido mejor asignar un código nuevo a cada uno de los alimentos, e incrementar progresivamente la base de códigos que poseen.
16. El crear nuevos códigos hubiese facilitado en cierta medida la asignación de códigos del IIN teniendo en cuenta que los códigos de esta base de datos es más amplia. Se tiene por ejemplo:
 1. Jurel seco salado cuando se hacia referencia a la caballa seco salado
 2. Jurel seco salado cuando se hacia referencia al bonito seco salado.
17. En el caso de las vitaminas pastilla se ha optado por colocar como código de Instituto de Investigación Nutricional 999999 para que posteriormente no se han tomados en cuenta al momento de realizar los análisis.
18. Para asignarles los códigos proporcionados por el Instituto de Investigación Nutricional (IIN) a los alimentos que conforman la base de datos de ENCOFA se ha encontrado algunos problemas debido a que la codificación empleada por el IIN es para algunos alimentos más específica que la información empleada por el CENAN. Esta información es detallada a continuación:
 - a) En el caso de habas frescas, los códigos proporcionados por el IIN presentan dos alternativas habas frescas sin cáscara sin vaina o habas frescas sin vaina con cáscara, por lo que ha sido necesario revisar 113 datos a fin de poder decidir cuál de estos códigos debía de ser el empleado.
 - b) En el caso de habas sin cáscara se ha efectuado la revisión de 10 casos a fin de diferenciar si los datos proporcionados correspondían a habas frescas o habas secas y en el caso de habas frescas efectuar la diferencia mencionada en el acápite anterior.
 - c) En el caso del pollo, los códigos empleados por IIN no solamente son específicos para cada parte del ave sino también implican la información de si este alimento fue consumido con piel o sin piel, lo que no ocurre con la codificación empleada por CENAN. De los 474 datos de encuestas que presentan como consumo alguna parte del pollo solo 1 encuesta menciona que el consumo de la pechuga de pollo fue sin piel. En los demás casos se ha optado por colocar el código de la parte del ave con piel. Cabe mencionar que en un principio se conversó sobre la posibilidad de que el consumo en la ciudad de Lima podría realizarse sin piel pero esto puede o no puede ocurrir por lo que finalmente se optó por colocar el código para todas las partes del pollo con piel a excepción de aquella encuesta que hacía mención a esta observación.

- d) Existen 28 datos de encuestas que presentan información de pollo entero con menudencia, en estos casos se debería de haber optado por pesar el pollo y la menudencia por separado y más aún optar por pesar cada parte del pollo y de la menudencia por separado. Al no contar con esta especificidad se ha colocado un código que implica al pollo entero con menudencia.
- e) Existen 173 datos de encuesta que presentan información de pollo entero sin menudencia, como se menciono anteriormente lo ideal hubiese sido pesar cada parte del pollo por separado. Al no contar con esta especificidad se ha colocado un código que implica al pollo entero sin menudencia.

19. Existen algunos alimentos en las encuestas que pueden con llevar a errores al momento de asignarles los códigos proporcionados por el IIN, así se tiene:

- a) Se ha revisado los datos presentados por 38 encuestas con respecto al maíz morado, si se tiene en cuenta que este alimento se usa para obtener la chicha morada por lo que el maíz finalmente es quitado de la preparación, no debería ocurrir que el peso final sea igual al peso bruto. En algunos casos este error ha ocurrido por una deficiente digitación y en otros por la persona que realizó la encuesta.

De acuerdo al formato de registro de esta información se tiene 2 caminos:

- Colocar la misma cantidad para peso bruto y residuo obteniéndose al final cero (la opción menos empleada), ó
- Colocar la misma cantidad para peso bruto y para peso neto y luego en la pregunta 417 (situación del ingrediente) colocar dos (quitado todo) para que la cantidad final sea cero.

Cualquiera que hubiese sido el camino elegido lo importante era uniformizar el criterio de recojo de información para así evitar errores como los que se ha podido observar en los datos de la ENCOFA. Particularmente hubiese optado por la primera opción ya que evita el problema de olvido de colocar información en la p 417 opción 2 (quitado todo).

- a) Se ha revisado los datos presentados por 6 encuestas con respecto al café en polvo, en estos casos se hace mención de que era café en polvo con azúcar de acuerdo al código empleado por el encuestador pero al momento de revisar la encuesta era café y el azúcar se encontraba aparte. En este caso se solicitó al IIN crear y buscar información para café con azúcar que finalmente no fue utilizado.
- a) Existen 4 encuestas que incluyen waffer pero no se especifica si es de chocolate o vainilla. En estos casos se ha optado por emplear el código de waffer de vainilla dado que es más habitual su consumo.
- a) En el caso del limón se ha empleado el código de limón entero (644); si bien es cierto para peso bruto correspondía este código hay que tener en cuenta que el objetivo principal de estas encuestas es obtener el

consumo de las personas para posteriormente evaluar su dieta por lo tanto lo que consumen es el jugo del limón y el código a emplear en la pregunta 411 (código del alimento por CENAN) es 552.

20. Se ha encontrado errores en la pregunta 411 (código asignado por CENAN) algunos por parte del encuestador y otros por parte del digitador. Por mencionar algunos ejemplos se tiene:
 - a) 1085 café en polvo con azúcar soluble cuando en realidad se trataba del 1585 pollo encuentro.
 - b) 1048 sal cuando en realidad era 1028 agua
 - c) 1928 sopa de fideos con plátano cuando en realidad era 1028 agua.
 - d) 644 limón cuando en realidad era 552 jugo de limón.
 - e) 1986 Sopa de verduras cuando en realidad era 1085 café.
 - f) 786 ají fresco sin clasificar cuando en realidad era 733 espárrago, entre otros.
21. Se ha realizado la revisión de algunos alimentos cuyos códigos empleados en la pregunta 411 (código del alimento) son muy amplios como:
 - a) Pescado magro, en este caso se refería al pescado perico.
 - b) Grasas de animales promedio, en este caso hubo error del código empleado porque el alimento era jamonada
22. Se ha efectuado la revisión de datos con respecto a la pregunta 414 (residuo) comparando con la pregunta 413 (peso bruto) en los cuales el residuo era mayor al peso bruto por lo que se ha realizado las verificaciones y correcciones respectivas (8 casos).
23. En el caso de infusión de hierbas, cebada, los encuestadores han colocado la cantidad utilizada como residuo (P. 414) y para realizar el cálculo de peso neto se procedió a restar del peso bruto el residuo, al presentar estos datos solo información en el residuo han llevado a confusión por lo que sería necesario unificar criterios para el llenado de información de la encuesta ya que debería de utilizarse la misma metodología en todos los alimentos.
24. La manera más adecuada de introducir la información de infusiones cebada hubiese sido colocar la misma cantidad para peso bruto y residuo obteniéndose al final en cantidad consumida cero.
25. Con respecto a la pregunta 412 (Procedencia del ingrediente) se ha encontrado 27 casos que presentaban en esta pregunta cero como respuesta. Siendo las alternativas de 1 a 4 se ha procedido a revisar y corregir esta información.
26. Con respecto a la pregunta 416 que hace referencia a la condición al pesar se ha efectuado la revisión de 128 datos que presentaban como información cero cuando las alternativas para el llenado de información en esta pregunta van de

- 1 a 3. en estos casos ha habido encuestas en que el encuestador se había olvidado de llenar esta información.
27. Para la pregunta 416 se ha realizado un listado de aquellos alimentos (822) que presentaban como alternativa 2 (cocido) encontrándose que 58 alimentos presentaban como información la alternativa 1 y no 2 como figuraba en la base de datos.
 28. En el listado obtenido con la alternativa dos (cocido) para la P. 416 se ha observado algunas incongruencias como la gaseosa, aquí hubo error ya que la respuesta era uno.
 29. Se ha tenido que verificar la información proporcionada en la pregunta 415 (peso neto) sobre el peso del arroz porque algunos datos hacen referencia a que fueron pesados cocidos entonces era necesario cerciorar si el peso al que se hacia referencia era peso crudo o cocido,
 30. Es necesario mencionar que la utilidad de la información proporcionada con esta alternativa (P416) sirve para verificar si se ha utilizado factor de corrección, ya sea de bruto a neto o de cocido a crudo, sin embargo en el formato de datos no se digita la información correspondiente al factor lo cual habría servido como información para verificar el dato final obtenido en algunos casos.
 31. También es necesario acotar con respecto a la pregunta 416 que puede con llevar a que la persona que revisa la encuesta piense que el peso del alimento registrado es cocido cuando la respuesta es 2 y por tanto el código del alimento tiene que ser de alimento cocido.
 32. Se ha efectuado un listado de datos con respecto a la pregunta 416 que presentan como información el código tres (remojado). En este caso ha sido necesario verificar la información proporcionada de 134 alimentos a fin de realizar las correcciones respectivas de acuerdo a la información proporcionada en la encuesta. Se ha encontrado algunos errores como:
 - a) En este listado se ha encontrado pescados frescos por lo que ha sido necesario verificar esta información.
 - b) En este listado también aparece el agua por lo que ha sido necesario corregir la información. (error encuestador)
 33. Se ha corregido el código empleado para los diferentes tipos de panes en esta pregunta (P.416) encontrándose 1902 alimentos que presentaban como respuesta el número 1 cuando lo correcto es el número 2 que hace referencia a cocido.
 34. También se ha efectuado la corrección para esta pregunta (P.416) de aquellos panes (n= 198) cuyo código empleado es el de pan sin especificar su tipo (Código 450) que tuvieron como respuesta el número 1 cuando se debió de emplear el número 2. Esta corrección ha sido efectuada tanto en la base de datos como en las encuestas.

35. Dada la magnitud del error en el código empleado en la respuesta a la pregunta 416 en el caso de los panes, estaría indicando que no ha existido un adecuado entendimiento por parte del personal que iba a realizar las encuestas al momento de efectuar la capacitación por lo que siempre es necesario emplear retroalimentación (feedback).
36. Se ha realizado la corrección en las encuestas con respecto a los panes así como también de los errores presentados con respecto al código empleado para el alimento o en la cantidad de alimento consumido. Estas correcciones han sido señaladas en las diferentes encuestas con lapicero de tinta roja.
37. Se ha revisado la información proporcionada en el caso del té en especial hojas secas que al final presentan datos de cantidad consumida. En estos casos se ha optado por colocar tanto en peso bruto como en residuo la misma cantidad obteniendo al final como cantidad consumida cero.
38. Cuando en el caso del té existía información en situación del ingrediente (P.417) quitado todo (opción 2) esto se ha mantenido y se ha realizado la corrección respectiva en la cantidad consumida que debía de ser cero.
39. Se ha efectuado la revisión del listado obtenido cuando la respuesta a la pregunta 417 (situación del ingrediente) era 2 (quitado todo) y la información en la pregunta 422 (cantidad consumida) era mayor que cero. Este listado estaba conformado por 41 alimentos cuya información ha sido corregida tanto en la base de datos como en las encuestas.
40. Se ha efectuado la revisión y corrección del listado obtenido con datos negativos a la pregunta 422 (cantidad consumida). Este listado estaba constituido por 18 alimentos.
41. Se ha efectuado la revisión de los datos proporcionados en la pregunta 415 (peso neto) mediante la resta del residuo al peso neto. Se ha encontrado algunos errores como resultados negativos o cantidades mayores al peso bruto. Estos datos han sido consultados en las encuestas para efectuar las correcciones respectivas.
42. Los datos de la pregunta 422 (cantidad de ingrediente consumido) en algunos casos se ha obtenido mediante una fórmula en la base digitalizada y en los otros casos han sido digitados tal y como se encuentran en la encuesta, existiendo en esta última alternativa algunos errores efectuados por parte del encuestador al realizar los cálculos respectivos y otros errores por parte del digitador.
43. Si se hubiese empleado en todos los datos de alimentos la fórmula para calcular la cantidad de ingrediente consumido (P.422) se hubiese podido minimizar los errores por digitación y también por que no decirlo los errores de cálculo efectuados por los encuestadores.
44. Como mencione anteriormente es necesario y muy importante uniformizar el manejo de la información.
45. Existen algunos alimentos que no se encuentran en la base de datos pero que sí se encuentran en las encuestas en especial cuando se trataba de enlaces a

los cuales se les ha añadido algún tipo de ingrediente. Esta información ha tenido que ser añadida a la base de datos.

46. Después de realizar varias modalidades de limpieza de las encuestas se decidió efectuar una revisión del 10% de conglomerados (16 conglomerados) que conforman toda la base de datos. Cada conglomerado está constituido por 12 o 13 encuestas. De estos 1 conglomerado no ha presentado errores mientras que los demás presentaban algún tipo de error que no ha sido posible encontrar en la limpieza de encuestas mediante rangos o usando criterios lógicos. Algunos de estos errores correspondían a error en la cantidad digitada, datos que no aparecían en la encuesta (acápite 43), errores de código de alimento.
47. Cabe recalcar que es necesario realizar una adecuada supervisión del personal que realiza la digitación así como también del personal que participa como encuestador; de esta manera se trataría de minimizar el número de errores y mejorar la calidad de la información.
48. Se ha procedido a eliminar 15 registros de la base datos por no encontrarse esta información en las encuestas a la que se suponía pertenecían.
49. Existe 13 encuestas que presentan error en el día de la semana digitado para la encuesta estas son:
 - Conglomerado: 002, ámbito: 2, vivienda: 128, hogar: 1, día: **miércoles**
 - Conglomerado: 012, ámbito: 2, vivienda: 001, hogar: 1, día: **miércoles**
 - Conglomerado: 020, ámbito: 2, vivienda: 026, hogar: 1, día: **lunes**
 - Conglomerado: 020, ámbito: 2, vivienda: 052, hogar: 1, día: **viernes**
 - Conglomerado: 030, ámbito: 2, vivienda: 021, hogar: 1, día: **lunes**
 - Conglomerado: 052, ámbito: 3, vivienda: 011, hogar: 1, día: **domingo**
 - Conglomerado: 058, ámbito: 3, vivienda: 079, hogar: 1, día: **viernes**
 - Conglomerado: 061, ámbito: 4, vivienda: 016, hogar: 1, día: **sábado**
 - Conglomerado: 119, ámbito: 1, vivienda: 001, hogar: 1, día: **miércoles**
 - Conglomerado: 131, ámbito: 1, vivienda: 071, hogar: 1, día: **miércoles**
 - Conglomerado: 154, ámbito: 1, vivienda: 005, hogar: 1, día: **jueves**
 - Conglomerado: 155, ámbito: 1, vivienda: 014, hogar: 1, día: **miércoles**
 - Conglomerado: 130 ámbito: 1, vivienda: 063, hogar: 1, día: **miércoles**
50. Existe 1 encuesta que presentó error en el día de la semana asignado por parte del encuestador, esta es:
 - Conglomerado: 015, ámbito: 2, vivienda: 013, hogar: 1, día: **viernes**
51. En el conglomerado 44 la encuesta 25 del ámbito 3 ha tenido que ser arreglada ya que la información de los alimentos no correspondía a la información de la encuesta, se tuvo que correr toda la información e incorporar dos alimentos para que la encuesta quede correctamente corregida.

52. Se ha efectuado la revisión y corrección de los datos proporcionados para el conglomerado: 117, ámbito: 1, Vivienda: 002, y hogar: 1 ya que esta encuesta presentaba 9 datos de alimentos con información missing.
53. En el conglomerado 157, ámbito: 1, vivienda: 41, hogar 1 se tiene que la familia ha consumido agua mineral pero no se ha colocado la cantidad que ha sido consumida en este caso para efectos de análisis se va a optar por eliminar esta información.
54. En el segundo archivo de datos: 400-2.xls se encuentran las encuestas que faltaban ingresar del conglomerado 117, 118, 130 y 135. Asimismo se encuentran las encuestas de los conglomerados 122, 137, 138, 139 y 140.
55. Los datos que se encuentran en la base de datos de la encuesta codificada con el conglomerado 48, ámbito 3, vivienda 18 en realidad pertenecen a la encuesta codificada con el conglomerado 122, ámbito 1, vivienda 010, hogar 1.
56. Al comparar la identificación de la encuesta digitada con el formato de la encuesta se ha encontrado 29 encuestas con errores en el número de hogar digitado, así se tiene:
 - a) Conglomerado 37, ámbito 3, vivienda 023 hogar 9
 - b) Conglomerado 37, ámbito 3, vivienda 040 hogar 17
 - c) Conglomerado 37, ámbito 3, vivienda 050 hogar 25
 - d) Conglomerado 37, ámbito 3, vivienda 068 hogar 33
 - e) Conglomerado 37, ámbito 3, vivienda 080 hogar 41
 - f) Conglomerado 37, ámbito 3, vivienda 099 hogar 45
 - g) Conglomerado 37, ámbito 3, vivienda 114 hogar 53
 - h) Conglomerado 37, ámbito 3, vivienda 131 hogar 61
 - i) Conglomerado 37, ámbito 3, vivienda 148 hogar 69
 - j) Conglomerado 37, ámbito 3, vivienda 155 hogar 60
 - k) Conglomerado 37, ámbito 3, vivienda 180 hogar 77
 - l) Conglomerado 37, ámbito 3, vivienda 241 hogar 101
 - m) Conglomerado 62, ámbito 4, vivienda 002 hogar 2
 - n) Conglomerado 62, ámbito 4, vivienda 004 hogar 4
 - o) Conglomerado 62, ámbito 4, vivienda 005 hogar 5
 - p) Conglomerado 62, ámbito 4, vivienda 006 hogar 6
 - q) Conglomerado 62, ámbito 4, vivienda 007 hogar 7
 - r) Conglomerado 62, ámbito 4, vivienda 008 hogar 8
 - s) Conglomerado 62, ámbito 4, vivienda 009 hogar 9
 - t) Conglomerado 62, ámbito 4, vivienda 010 hogar 10

- u) Conglomerado 62, ámbito 4, vivienda 011 hogar 11
- v) Conglomerado 62, ámbito 4, vivienda 012 hogar 12
- w) Conglomerado 62, ámbito 4, vivienda 013 hogar 13
- x) Conglomerado 124, ámbito 1, vivienda 006 hogar 6
- y) Conglomerado 124, ámbito 1, vivienda 008 hogar 8
- z) Conglomerado 124, ámbito 1, vivienda 025 hogar 28
- aa) Conglomerado 124, ámbito 1, vivienda 052 hogar 56
- bb) Conglomerado 153, ámbito 1, vivienda 071 hogar 27
- cc) Conglomerado 156, ámbito 1, vivienda 058 hogar 2

57. Al comparar la identificación de la encuesta digitada con el formato de la encuesta se ha encontrado errores en el número de vivienda digitado en 4 encuestas, así se tiene

- 1. Conglomerado 19, ámbito 2, vivienda 103 hogar 1
- 2. Conglomerado 53, ámbito 3, vivienda 045 hogar 1
- 3. Conglomerado 73, ámbito 4, vivienda 008 hogar 1
- 4. Conglomerado 147, ámbito 1, vivienda 094 hogar 1

1. Al comparar la identificación de la encuesta digitada con el formato de la encuesta se ha encontrado 14 encuestas con error en el ámbito, así se tiene:

- 1. Conglomerado 54, ámbito 2, vivienda 003 hogar 1
- 2. Conglomerado 54, ámbito 2, vivienda 005 hogar 1
- 3. Conglomerado 54, ámbito 2, vivienda 008 hogar 2
- 4. Conglomerado 54, ámbito 2, vivienda 014 hogar 1
- 5. Conglomerado 54, ámbito 2, vivienda 016 hogar 1
- 6. Conglomerado 54, ámbito 2, vivienda 019 hogar 1
- 7. Conglomerado 54, ámbito 2, vivienda 022 hogar 1
- 8. Conglomerado 54, ámbito 2, vivienda 034 hogar 1
- 9. Conglomerado 54, ámbito 2, vivienda 039 hogar 1
- 10. Conglomerado 54, ámbito 2, vivienda 044 hogar 1
- 11. Conglomerado 54, ámbito 2, vivienda 050 hogar 1
- 12. Conglomerado 54, ámbito 2, vivienda 052 hogar 2
- 13. Conglomerado 56, ámbito 3, vivienda 023 hogar 1
- 14. Conglomerado 86, ámbito 4, vivienda 024 hogar 2

Los dos últimos errores (conglomerado 56 y 86) en el ámbito corresponden al encuestador

1. En el conglomerado 081 existe el problema de que en las encuestas el ámbito señalado en ellas es el tres (3) pero en la base de datos figura con ámbito 4. Cabe señalar que en el sobre donde se guarda este conglomerado han tachado el ámbito 3 y han colocado con lapicero rojo cuatro (4) pero esta modificación no ha sido realizada en las encuestas por lo que en una primera intención se optó por corregir el ámbito de acuerdo a la información que presentan las encuestas, posteriormente se revisó la base de datos y teniendo en cuenta que desde el conglomerado 061 hasta el conglomerado 097 pertenecen al ámbito 4 se decidió dejar está información con el número de ámbito 4 tal como se mostraba en la base de datos.
2. El sobre que contiene las encuestas que conforman el conglomerado 130 refiere que esta conformado por 11 encuestas y con la encuesta identificada como la vivienda 063 que se encuentra en el archivo 400-2.xls son 12 encuestas.
1. El conglomerado 139, ámbito 1 según la información escrita en el sobre donde se encuentra guardadas las encuestas menciona que esta conformado por 12 encuestas. Al momento de realizar el listado de las mismas se encontró que eran 13 encuestas. Estas son:
 64. Vivienda: 002 hogar: 1 día: sábado
 65. Vivienda: 002 hogar: 3 día: sábado
 66. Vivienda: 002 hogar: 4 día: sábado
 67. Vivienda: 008 hogar: 1 día: martes
 68. Vivienda: 009 hogar: 1 día: martes
 69. Vivienda: 010 hogar: 1 día: sábado
 70. Vivienda: 012 hogar: 1 día: sábado
 71. Vivienda: 013 hogar: 1 día: sábado
 72. Vivienda: 022 hogar: 1 día: sábado
 73. Vivienda: 025 hogar: 1 día: sábado
 74. Vivienda: 026 hogar: 1 día: martes
 75. Vivienda: 026 hogar: 2 día: martes
 76. Vivienda: 030 hogar: 1 día: martes
2. La encuesta del conglomerado 139 identificada como la vivienda 009 no se encuentra en los archivos que conforman la base de datos de Consumo Familiar.
3. La encuesta identificada como: conglomerado 010, ámbito 2, vivienda 027, hogar 1, día jueves no se encuentra ingresada en la base de datos.
4. En la base de datos la encuesta del conglomerado 132, ámbito 1 vivienda 059, día viernes se encontraba doblemente ingresada. Al ser comparada la información se encontró que un grupo de datos sí correspondía a la encuesta

identificada con esta codificación pero la otra era completamente diferente. Posteriormente se pudo ubicar que estos datos correspondían a la encuesta 147, ámbito1, vivienda 072, hogar 1, día jueves.

5. Al realizar la modificación del número de vivienda 094 del conglomerado 147, ámbito 1, hogar 1, se obtuvo una doble digitación de esta encuesta en la base de datos pero existía la diferencia en el día de semana (domingo en vez de sábado). Se optó por comparar los datos y se encontró que eran diferentes pero una parte de los datos eran iguales a los presentados por la encuesta identificada como: conglomerado 147, ámbito 1, vivienda 094, hogar 2. Estos datos todavía no han sido descartados de la base de datos pero la mejor opción sería eliminarlos ya que se desconocen a que encuesta pertenecen.
1. La encuesta identificada como: conglomerado 068, ámbito 4, vivienda 012, hogar 1, día martes se encuentra en la base de datos proporcionado por el personal a cargo de la ENCOFA, pero los datos no coinciden con la encuesta. Estos datos no han sido eliminados y faltaría ingresar toda la información de esta encuesta.
2. En la base de datos se encuentra información de la encuesta identificada como: conglomerado 105, ámbito 5, vivienda 040, hogar 1 día sábado, pero esta encuesta no existe físicamente. Esta información no ha sido eliminada de la base de datos.
3. Si se tiene en cuenta la encuesta mencionada en el acápite anterior, el conglomerado 105 estaría conformado por 12 encuestas y no 11 encuestas, como se informa en el sobre que contiene las encuestas de este conglomerado.
4. Al efectuar la revisión del código de identificación de las encuestas se encontró que había error en la vivienda digitada, se realizó la modificación respectiva y se obtuvo al final que 2 encuestas presentaban la misma identificación esto es: conglomerado 073, ámbito 4, vivienda 008, hogar 1, siendo lo único que las diferencia el día de semana, una corresponde al domingo y la otra al día sábado. Al hacer la consulta respectiva mencionaron que no podían existir 2 encuestas que solo se diferencien por el día de la semana. Posteriormente se observó que en la primera hoja donde se encuentra los nombres de los miembros del hogar existe una nota en la encuesta del día sábado en la que se escribe que corresponde a la vivienda 014; de acuerdo a esta información se ha optado por realizar esta modificación.
5. Es preciso mencionar que el colocar una nota en la que se mencione que se realiza una modificación en la identificación de una encuesta tal como se hizo para la encuesta del conglomerado 073 no es la metodología más adecuada de dar a conocer esta información. Lo ideal hubiese sido señalar con lapicero de tinta de color esta modificación en cada una de las páginas que identifican la encuesta y así se hubiese evitado el trabajo especificado en el acápite 70 de este informe.
6. Las modificaciones que se han realizado en la identificación de la encuesta solo han sido anotados en el formato que corresponde a consumo familiar.

7. Las modificaciones que se han realizado en la identificación de la encuesta han sido realizadas en las bases de datos referente al consumo familiar
8. Los nuevos archivos denominados 4001.xls y 4002.xls contienen todas las modificaciones expresadas en este informe para la Encuesta de Consumo Familiar.
9. Dada la importancia de conocer no solo el número de individuo para poder evaluar el consumo familiar sino también las características de los mismos, se ha procedido a realizar las modificaciones de identificación de la encuesta en las hojas de cálculo:
 - 200 por contener información con respecto a edad y sexo de los individuos que conforman la familia
 - 300: por contener información sobre la identificación de los comensales.

Para lo cual se ha creado los archivos 201-2.xls. y el 301-2.xls..

Estas modificaciones no forman parte de la consultoría pero se ha realizado a fin de facilitar el trabajo de análisis de consumo para la institución que lo realice.

10. Se recomienda realizar las modificaciones efectuada en la identificación de la encuesta en todas las hojas de cálculo que conforman la ENCOFA para evitar descoordinación entre todos los formatos de la encuesta.
11. Es necesario reincidir que en la base de datos de la ENCOFA se ha encontrado errores de digitación que hubiesen sido de menor magnitud si hubiese existido una adecuada supervisión de los digitadores y por supuesto un adecuado programa de ingreso de la información; asimismo también se ha encontrado errores por parte de los encuestadores que podrían haber sido evitados si hubiese existido una adecuada supervisión.
12. La información de la ENCOFA ha sido ingresada en programa Excel se recomienda emplear otro tipo de programa que permita el ingreso de todos los datos así se evitaría la creación de 2 archivos (hojas de cálculo) para un mismo tipo de información. Este no es un adecuado programa de ingreso para la cantidad de información que se obtenía en las encuestas y el manejo de las mismas (consistencia de los datos ingresados).
13. Se espera que este informe sirva de ayuda para mejorar la calidad de la información obtenida teniendo en cuenta que el CENAN es el organismo público a nivel nacional que efectúa este tipo de trabajo de investigación y de ahí parten las directrices para mejorar la calidad de la alimentación, los alimentos que conforman la canasta familiar a nivel local, regional, entre otros.