

APÉNDICE 3

Organización y Administración del Procesamiento de Datos

CONTENIDO

Manejo de la Base de Datos	A3.4
Control de Documentos	A3.4
Como Instalar y Utilizar los Programas de Entrada de Datos y Análisis	A3.6
Paquete EPI INFO para Encuestas para las Metas de Mitad de Década	A3.6
Instalación de Programas	A3.6
Ejecución de los Programas	A3.6
Entrada de Datos	A3.7
Notas sobre los módulos específicos	A3.8
Chequeos de Inconsistencias	A3.8
Frecuencias de las Variables	A3.8
Análisis de los Datos	A3.8
Respaldo de los Datos	A3.9
Entrada a EPI INFO	A3.9
Salida al Subdirectorío DOS	A3.9
Ejecución de CSAMPLE para Obtener Errores Estándar para Encuestas por Conglomerados	A3.9
Notas al Programador	A3.11
Remoción de Módulos	A3.11
Adición de Módulos	A3.11
Cambio de Fraseo de las Preguntas o las Categorías de Respuesta	A3.11
Cambio del Análisis/Chequeos de Consistencia	A3.12
Consolidación de Bases de Datos de Varios Computadores	A3.12

MANEJO DE LA BASE DE DATOS

Sugerimos que usted realice la entrada de datos y la edición (crítica) de datos en lotes pequeños, generalmente completando la entrada de datos para un conglomerado de una vez. Los datos deben almacenarse en el disco duro. Después que se ha entrado un lote, los datos se pasan por un programa de chequeo de consistencias y los resultados se imprimen.

Los errores que se deben a digitación incorrecta deben ser corregidos inmediatamente. El listado de errores debe entregarse a la supervisora para decidir los cursos de acción (por ejemplo, cambio de los datos, consulta con la entrevistadora o dejar el error sin resolver). Una vez que se han hecho las correcciones de los errores en los datos, el programa de inconsistencias se corre de nuevo para verificar las correcciones. Cuando no hay más errores que pueden corregirse en la oficina, los datos se respaldan en un disquete.

☛ Rotule sus disquetes con los nombres de los archivos, la región y la fecha de los cambios más recientes.

Cuando la base completa de datos está lista para análisis, el programa de chequeo debe usarse de nuevo como un chequeo final antes de que se produzcan frecuencias para cada pregunta.

CONTROL DE DOCUMENTOS

Las operaciones de procesamiento de datos a menudo se tornan caóticas debido al volumen de datos y a los múltiples procedimientos que tienen que ejecutarse. Un buen control de documentos puede aliviar algunos de estos problemas.

Una muestra de una forma (formulario) para el control de documentos se presenta en la Figura A3.1. Esta forma contiene una fila para cada lote o grupo de datos.

1. Registre el número de cuestionarios que retornan a la oficina del terreno para cada conglomerado. Verifique estas cifras contra los totales del trabajo de campo.
2. Cuando un lote de cuestionarios se entrega para entrada de datos, registre la fecha.
3. Cuando un lote ha sido digitado, registre la fecha, el número de cuestionarios y el número de errores.
4. Anote la fecha en la que se hacen las correcciones.
5. Cada vez que se producen listados de errores para un lote, anote el número de documentos con errores y la fecha.
6. Cuando la base de datos principal se actualiza con un lote, anote la fecha y el número de registros agregados.

Los datos para un cuestionario se entran para cada hogar en el orden en el cual se recolectan: los datos de hogares del Panel de Información de Hogar, los datos para todos los niños del Listado de Madres y Niños, los Módulos de Agua y Saneamiento, y Yodización, los datos del Módulo de Educación para cada niño en edad de ingreso a la escuela, el Módulo de Toxoide Tetánico para cada madre, el Módulo de Enfermedad Respiratoria Aguda para cada madre, y por último, los módulos de salud para cada niño. Una vez que se entran todos los datos para un hogar, el digitador pasa al hogar siguiente.

La entrada de datos puede ser efectuada por varios digitadores. Al final de la digitación de cada conglomerado se produce un informe de errores con una descripción de cualquier rubro inconsistente.

Las correcciones, anotadas en las hojas de errores, se utilizan para corregir los registros correspondientes en los disquetes. Cuando no hay más errores que se puedan corregir, los datos deben respaldarse en disquetes.

CÓMO INSTALAR Y UTILIZAR LOS PROGRAMAS DE ENTRADA Y ANÁLISIS DE DATOS

PAQUETE EPI INFO PARA ENCUESTAS DE METAS DE MITAD DE DÉCADA: INSTALACIÓN DE LOS PROGRAMAS

Todos los archivos para entrada de datos y análisis deben ser copiados en un directorio separado del disco duro. Para instalar el paquete, primero cree un nuevo directorio en el disco duro desde DOS:

MD C:\MIDDECAD

Luego copie todos los archivos de la disquetera A al disco duro:

COPY A:*. * C:\MIDDECAD

Se supone que EPI INFO v.6 ya está instalado en el disco duro. Si no, siga los procedimientos descritos en el manual de EPI INFO para la instalación. El directorio \EPI6 debe ser incluido en el PATH del computador. Usted debe verificar que así es, con el comando PATH en DOS. Si \EPI6 no está incluido en el PATH, modifique su archivo AUTOEXEC.BAT para incluirlo.

EJECUCIÓN DEL PROGRAMA

El paquete debe ejecutarse siempre del directorio \MIDDECAD. Cambiense primero a este directorio con los comandos:

**C:
CD \MIDDECAD**

Entonces para correr los programas, entre:

MENU

Esto trae un menu de selecciones que incluye:

1. Entrada de datos
2. Chequeos de inconsistencia
3. Frecuencias de las variables
4. Análisis de los datos
- B. Respaldo de los datos
- Q. Regreso al sub-directorio DOS

Entrada de Datos

Escoja la opción 1 del menú bien sea para entrar nuevos datos o para editar datos ya existentes. La selección de esta opción llama la primera pantalla de entrada de datos para la digitación del Panel de Información de Hogar. Se puede entrar inmediatamente un nuevo cuestionario: Para editar un cuestionario existente, teclee <CTRL F> y entre los números de identificación del CONGLOMERADO y HOGAR del cuestionario deseado. Presione entonces <F3> y <ENTER> para capturar este cuestionario para edición. El oprimir <CTRL N> de nuevo permitirá el ingreso de nuevos cuestionarios. Para detalles adicionales sobre cómo recorrer las pantallas de entrada de datos en EPI INFO, consulte el manual de entrada de datos de EPI INFO.

La entrada de datos está organizada en varios módulos, incluyendo:

1. Información del hogar (ocurrencia única)
2. Listado de las madres y niños (ocurrencias múltiples)
3. Agua y saneamiento (ocurrencia única)
4. Yodización de la sal (ocurrencia única)
5. Educación (ocurrencias múltiples)
6. Toxoide tetánico (ocurrencias múltiples)
7. Infección respiratoria aguda (ocurrencias múltiples)
8. Salud del niño (ocurrencias múltiples)

Esta organización le permite al personal de entrada de datos completar toda la entrada de datos para una página del cuestionario antes de pasar a la página siguiente, eliminando así el tener que estar pasando páginas.

El ingreso a cada módulo se controla desde la parte inferior de la pantalla. Para empezar la entrada de datos del Listado de Madres y Niños, simplemente presione <ENTER> al frente de GoToChild. Ésto llamará una pantalla para ingresar la primera línea del listado. Cuando se entran los datos y se graban en el disco, se repite la pantalla puesto que se permiten múltiples líneas. Después de ingresar todas las líneas presione <F10> (**RETURN**) y reaparecerá la pantalla de hogar. Presione <ENTER> en frente de GoToWater para entrar al Módulo de Agua y Saneamiento. Como solamente se permite una ocurrencia del Módulo de Agua y Saneamiento por hogar, la pantalla del hogar reaparece inmediatamente una vez se salva el registro de agua. Los mismos procedimientos se aplican para los módulos restantes. Cuando no hay datos para un módulo, éste puede saltarse colocando el cursor al final del módulo y presionando <ENTER> o simplemente presionando <F10> inmediatamente después que se ingresa al módulo. Cuando se ha terminado de entrar los datos para todos los módulos, presione <ENTER> en frente de Finish para terminar la entrada de datos.

Los números del conglomerado y el hogar se concatenan para formar una HOUSEID (identificación de hogar) la cual sirve como clave para concatenar todos los archivos. La HOUSEID debe identificar cada hogar unívocamente y no debe estar repetida. Además, los números de línea de la madre o el niño sirven como clave para concatenar la información del individuo con la línea correspondiente en el Listado de

Niños. Es importante que este número de línea sea digitado correctamente debido a que muchos de los indicadores se basan en la conexión de los datos del listado y los módulos específicos.

Notas sobre los módulos específicos:

Panel de Información de Hogar: La fecha de la entrevista tiene que ingresarse porque todas las edades y fechas de nacimiento tienen que verificarse contra la fecha de la entrevista.

Formulario para el Listado de Madres y Niños: Los números de línea para madres o guardianes deben terminar en un 0 (i.e., 1-0, 2-0, etc.). La fecha de nacimiento no se ingresa para ellos. Los números de línea para los niños deben terminar en un número del 1 al 9 (i.e., 1-1, 1-2, etc.)

Toxoide tetánico: El número de línea de la madre (terminado en 0) debe ingresarse aquí para ser utilizado para concatenar la madre a su hijo más joven. (La clave para concatenar los niños se basa en el número de la línea más 1, ya que las fechas de inmunización necesitan compararse contra la fecha de nacimiento del niño más joven). La información para la pregunta 6 no necesita entrarse, puesto que la respuesta a la pregunta 6 es simplemente la suma de las preguntas 2 y 4.

Enfermedad Respiratoria Aguda (ERA): Debe ingresarse el número de línea de la madre (terminado en 0) si bien la clave para la concatenación se basa en el número de la línea más 1.

Niños menores de cinco años: Como todos los módulos para niños menores de cinco años (diarrea, vitamina A, lactancia, inmunización y antropometría) se llenan para un niño antes de continuar con el siguiente niño, ellos se tratan como un solo módulo para efectos de entrada de datos. La información sobre la fecha de la entrevista y la fecha de nacimiento del niño se deben reentrar en este módulo debido a que EPI INFO calcula los índices antropométricos en el momento de la entrada de datos.

Chequeos de Consistencia

Los datos para cada conglomerado deben ser chequeados para consistencia interna tan pronto como todos los datos han sido ingresados para ese conglomerado. Escoja la opción 2 para ejecutar el chequeo de consistencia. Se le preguntará si usted desea enviar los resultados a la impresora o a un archivo para futura revisión. También se le preguntará sobre el rango de conglomerados a chequear. Si usted sólo está interesado en el conglomerado 101, teclee **101** como el conglomerado inicial y como el final. Si desea correr los chequeos de consistencia para el archivo completo, teclee **000** como el conglomerado inicial y **999** como el conglomerado final.

Los errores de consistencia se envían a un archivo llamado CHECK.TXT o a la impresora directamente. Cada mensaje de error muestra: (1) el número de identificación del hogar, (2) el módulo de cuestionario que contenía el error, (3) una descripción del error y (4) el número individual, si es del caso.

Frecuencias de las variables

Cuando todos los datos han sido ingresados, será de utilidad examinar un listado de frecuencias de todas las variables en el archivo de datos. El seleccionar la opción 3 producirá un listado de frecuencias para todas las variables con la excepción de HOUSEID y CHILDDID (identificación del niño).

Análisis de los Datos

Seleccione la opción 4 para analizar los datos. Esta opción ejecuta el conjunto de comandos que contiene el programa INDICATR.PGM, el cual produce los resultados para los indicadores, de acuerdo con las especificaciones establecidas en el Apéndice 4. Usted puede obtener un listado de los archivos .PGM simplemente en pantalla o enviando el archivo a su impresora. La salida del programa se limita a mostrar los indicadores específicos para la muestra total. El programa se puede modificar para mirar solamente a subgrupos poblacionales y después hacer la corrida para ese subgrupo. Las instrucciones para modificar el programa se encuentran en la sección "Notas para el Programador" más abajo.

El programa de análisis no produce los errores estándar para los indicadores. Estos se pueden calcular utilizando la opción CSAMPLE en EPI INFO v.6. Véase más abajo las instrucciones de como usar CSAMPLE.

Respaldo de los Datos

Los datos deben ser respaldados en disquetes diariamente, utilizando disquetes diferentes cada día de la semana. El seleccionar la opción B copiará los archivos apropiados al disquete. Este procedimiento asume que el tamaño del archivo de datos no excederá la capacidad de un solo disquete. En caso que los archivos de datos lleguen a ser demasiado grandes para un solo disquete, el administrador de los datos necesitará encontrar un sistema alternativo para hacer el respaldo y deberá modificar el archivo MENU.BAT según el caso. Algunos sistemas que se pueden considerar son PKZIP, BACKUP (DOS), o PCTOOLS.

Entrada a EPI INFO

En caso que hubiese necesidad de trabajar directamente con el paquete EPI INFO, la opción E lo llevará a la pantalla de iniciación de EPI INFO v.6. Todos los programas y utilidades de EPI INFO se pueden ejecutar desde esta pantalla. Cuando salga de EPI INFO, usted regresará al menú. (La opción E realmente no aparece en el listado del menú, debido a que el personal de entrada de datos no necesitará entrar regularmente a EPI INFO directamente. La opción es funcional, sin embargo.)

Regrese al Subdirectorío de DOS

Seleccione la opción Q para salir del menú del sistema.

EJECUCIÓN DE CSAMPLE PARA OBTENER LOS ERRORES ESTÁNDAR PARA ENCUESTAS DE CONGLOMERADOS

Para estimar los errores estándar correctos para los indicadores en EPI INFO, se requiere tener en disco una base de datos que contenga las variables generadas para el indicador de interés. Por ejemplo, supongamos que estamos interesados en obtener los errores estándar para el indicador 11.1, el porcentaje de menores de 5 años con bajo peso para la edad. Para crear tal base de datos, agregue dos líneas al archivo INDICATR.PGM después de la línea "report waz.rpt":

```
ROUTE TEMP.REC  
WRITE "RECFILE" CLUSTER LOW2WAZ
```

y corra el análisis. Esto creará un archivo temporal llamado TEMP.REC el cual contiene solamente la variable del conglomerado y la variable relacionada con si el niño tiene o nó, bajo peso para la talla.

Seleccione entonces la opción CSAMPLE entre las opciones del menu de PROGRAMS en EPI INFO v.6. Teclee **TEMP** como el nombre del archivo. Aparecerá una pantalla de opciones. Bajo Main, teclee **LOW2WAZ**, la variable para la cual usted necesita los errores estándar. Especifique la variable de estratificación solamente si el diseño muestral fue estratificado. Especifique la variable de la UPM (en el capítulo 4, el término “áreas menores” se refiere a estas Unidades Primarias de Muestreo), a la cual generalmente se le da el nombre **CLUSTER**. No se requiere ponderación a menos que así lo exija el diseño muestral. Especifique una variable de Crosstab solamente si usted necesita examinar errores estándar en subgrupos poblacionales. Los resultados se pueden ver inmediatamente en la pantalla, enviados a la impresora o guardados en un archivo para uso futuro. Si los datos no están ya ordenados por el programa **CLUSTER**, entonces utilice **SORT**. Seleccione **TABLES** y el análisis será ejecutado.

La salida produce el error estándar para el indicador y su intervalo (límites superior e inferior con una confianza del 95 por ciento.) El programa también calcula el efecto de diseño obtenido en la encuesta para la variable de interés. Por ejemplo, al analizar peso para la talla, la salida podría lucir como el ejemplo que se muestra en la Figura A3.2, en el cual el porcentaje (redondeado a la décima más cercana) con bajo peso para la edad es 21.0 con un intervalo de confianza de (19.3, 22.7).

Figura A3.2 Ejemplo de la salida CSAMPLE al analizar peso para la edad

ANÁLISIS DE DISEÑO MUESTRAL COMPLEJO CON CSAMPLE

Analysis of LOW2WAZ
 LOW2WAZ

		Total
0		
Obs		2168
Percent	V	79.009
SE%		0.887
LCL%		77.271
UCL%		80.746
1		
Obs		576
Percent	V	20.991
SE%		0.887
LCL%		19.254
UCL%		22.729
Total Obs		2744
Design eff.		1.300

LCL = lower confidence limit
 UCL = upper confidence limit

Sample Design Included:

 Sampling Weights--None
 Primary Sampling Units from CLUSTER
 Stratification--None

0 records with missing values

NOTAS AL PROGRAMADOR

En cada país en donde se utilice el paquete se requerirán algunas modificaciones al programa. Por ejemplo, la definición de “agua segura y conveniente” depende del país, y por lo tanto, el programa de análisis INDICATR.PGM debe editarse para que refleje la definición utilizada en un país específico. La definición de edad de ingreso escolar varía de país a país. Existen también módulos alternativos para vitamina A al igual que para cobertura de inmunización. Las modificaciones apropiadas para un país particular tienen que ser hechas por un programador familiarizado con EPI INFO. Lo mejor es que el programador entienda la estructura de la base de datos antes de hacer alguna modificación. Esto es especialmente cierto si va a haber muchos cambios a los cuestionarios.

Cualquier cambio a los archivos .QES o .CHK requiere que el correspondiente archivo .REC sea recreado. Esto se puede lograr seleccionando ENTER en EPI INFO v.6 y escogiendo la opción 2, “Crear un nuevo archivo de datos del archivo .QES.” Como los datos se encuentran en el archivo .REC, es importante que todos los cambios a los archivos .QES y .CHK se hagan antes de que empiece la entrada de datos. Si es necesario hacer alguna modificación a estos archivos después de que se han entrado datos, consulte el manual de EPI INFO en “Revisión de la Estructura de un Archivo de Datos (Selección 3 en el Menú),” en el capítulo 8, “Entrada de Datos.”

Remoción de Módulos

Cualquiera de los Módulos del Cuestionario Modelo que un país decida no implementar puede ser removido fácilmente. Si el módulo a remover es uno de los listados atrás en la página A3.6, una simple eliminación en el archivo HOUSEHLD.CHK impedirá utilizar ENTER para ese módulo. La variable para ese módulo en HOUSEHLD.QES también deberá eliminarse.

Si solamente se quiere borrar una parte del módulo de entrada de datos (por ejemplo, las preguntas de lactancia en los Módulos para Niños Menores de Cinco Años), entonces se deben eliminar las líneas correspondientes a esas preguntas en los archivos .QES y .CHK. Los módulos alternativos de vitamina A e inmunización que no se implementen también deben ser removidos de esta manera. Debe agregarse o eliminarse del archivo .QES un número apropiado de líneas para garantizar que los saltos de página en la pantalla quedan en puntos lógicos.

Adición de Módulos

Si se van a agregar módulos adicionales al cuestionario, será necesario desarrollar nuevos archivos .QES y .CHK. Los archivos existentes pueden servir como plantillas sobre cómo hacer esto. En la mayoría de los casos, las preguntas adicionales se pueden agregar a los módulos existentes, en cuyo caso será suficiente modificar los archivos .QES y .CHK.

Cambio del Fraseo de las Preguntas o las Categorías de Respuesta

Los cambios al fraseo de las preguntas (incluyendo la traducción a idiomas diferentes) simplemente requieren que las pantallas que se definen en los archivos .QES se modifiquen para mostrar el fraseo apropiado. Los nombres originales de las variables en los archivos .QES, .CHK y .PGM no tienen que cambiarse. Sin embargo, si se cambian los nombres de las variables, los cambios deben hacerse en todos estos archivos.

Si se cambian las categorías de las respuestas que son aceptables para las preguntas, tienen que hacerse los cambios apropiados en el archivo .CHK (para permitir todas las respuestas válidas y rechazar las respuestas inválidas) y en el archivo .PGM file (para garantizar que los indicadores se calculan utilizando las categorías correctas). El paquete no asigna explícitamente rótulos a las respuestas por lo que generalmente no hay necesidad de entrar rótulos para las respuestas.

Cambio del Análisis/Chequeos de Consistencia

Tanto el chequeo de consistencia como el cálculo de los índices se logra a través de los archivos .PGM. Estos archivos se pueden editar para remover referencias a módulos que no se implementan y para especificar las definiciones de ciertos indicadores (e.g., agua segura y conveniente, edad de ingreso escolar y frecuencia de uso de la cápsula de vitamina A). A menudo se desean los indicadores para otros subgrupos específicos, tales como región o sexo.

- ☞ Obtención de las estimaciones para subgrupos: Para calcular cada indicador para subconjuntos específicos de los datos, la modificación más fácil sería seleccionar apenas aquellas observaciones en el subconjunto y correr de nuevo el programa de análisis. Por ejemplo, para calcular los indicadores para hombres únicamente, seleccione las observaciones de hombres:

```
SELECT SEX = 1.
```

Como la instrucción "select" tiene que ser ejecutada cada vez que se lee una nueva base de datos, INDICTR.PGM incluye al final una subrutina ":SELECTSEX" para realizar el análisis para cualquier subconjunto de datos que se requiera. Edite esta subrutina y ejecute de nuevo el programa revisado para calcular los indicadores para niños únicamente. Para niñas únicamente, revise la subrutina con la línea:

```
SELECT SEX = 2
```

y corra de nuevo el archivo INDICATR.PGM. La instrucción "select" también se puede revisar para solamente considerar observaciones en categoría específicas de REGION, grupo de EDAD o URBANO/RURAL. Por supuesto que el programador debe asegurarse que la variable de interés está siempre disponible generando el correspondiente comando RELATE antes de la selección.

- ☞ Se pueden crear archivos .PGM adicionales para otros propósitos—por ejemplo, para calcular la duración mediana de lactancia o para examinar la distribución por edad de los niños. El examen de los archivos .PGM existentes puede ser útil como punto de partida para nuevos análisis.
- ☞ Todos los programas de análisis en EPI INFO se han escrito asumiendo que la muestra en el país ha sido diseñada para que sea autoponderada (i.e., todas las ponderaciones son 1.0). Sin embargo, se incluye una subrutina para hacer el análisis ponderado tan descomplicado como sea posible. El programador sólo necesita editar la subrutina ":WEIGHT" al final del archivo INDICATR.PGM, para incluir la asignación de ponderaciones como se describe en el capítulo 7.

Concatenación de Bases de Datos de Varios Computadores

En algunos casos será necesario entrar datos de cuestionarios en varios computadores para después concatenar los datos para el análisis. Esto puede lograrse con el programa MERGE en EPI INFO. Los archivos .REC de un computador deben ser copiados en otro pero en un subdirectorio diferente. El programa MERGE, al utilizar la opción CONCATENATE, creará un nuevo archivo con datos de ambos archivos. El proceso se puede repetir si se usan más de dos computadores. El programa MERGE debe completarse en forma separada para cada módulo.