

Curso a distancia de Epi.info versión 6.04 (II)

P Arias Bohigas, J C Fernández Merino

NOTA PREVIA

En el CD ROM distribuido con el número anterior existe un problema con la instalación de las tutorías. Para poder instalarlas es necesario copiar los archivos `INSTALAR.BAT` y `CURSUNED.EXE`, que están en la carpeta `EPIINFO\TUTORIAS`, en un disquete y proceder a la instalación desde allí en lugar de hacerlo desde el CD. Lamentamos los problemas y quebraderos de cabeza que esto os haya podido ocasionar. Recuerda, si tienes algún problema ponte en contacto con nosotros usando el correo-e (o) o con la dirección de la revista.

UN BREVE RECORDATORIO

En el número anterior revisamos algunos conceptos relacionados con el mundo de las bases de datos: como utilizar `ANALYSIS` para abrir una tabla mediante la orden `READ`, como realizar listados usando `LIST` y como ojear los contenidos de la tabla mediante la orden `BROWSE` (tecla F4). También vimos como obtener distribuciones de frecuencias (`FREQ`), tablas (`TABLES`), comparaciones de medias (`MEANS`) y aprendimos a realizar una regresión lineal simple. Finalmente pudimos ver como hacer gráficos con `EpiInfo`.

En este capítulo podrás usar de nuevo estas órdenes al mismo tiempo que aprendes algunas cosas más relacionadas con `ANALYSIS`. Para conocer más a fondo todo lo que repasaremos en este capítulo usa las tutorías `ANALYSIS` de datos (I) y (II) del menú `ANALYSIS`. Además, en todos los ejemplos que usemos trabajaremos con la tabla `ACCIDEN.REC` que se encuentra en la carpeta `C:\TUTOREPI` (debes leerla usando `READ C:\TUTOREPI\ACCIDEN.REC`).

CURSO EPIINFO VERSIÓN 6.04: MÓDULO ANALYSIS (II)

Trabajar sólo con una parte de la tabla

En muchas ocasiones desearás obtener una distribución de frecuencias o una tabla de los registros que

cumplen una determinada condición en tu tabla. Por ejemplo, puedes desear obtener la distribución de frecuencias de turnos de trabajo (`TURNO`) de las mujeres (`M`) incluidas en la tabla `ACCIDEN`, olvidándote momentáneamente de los datos de los hombres. Para ello deberás usar la orden `SELECT`.

La orden `SELECT` tiene una sintaxis un poco especial:

`SELECT <VARIABLE> <condición>` donde `<variable>` es el nombre de cualquier variable o campo de la tabla y `<condición>` es la condición que deben cumplir los registros con los que deseas trabajar, por ejemplo: `SELECT SEXO="M"`, te permitirá trabajar sólo con los registros de las personas de sexo femenino (`M`). Como condiciones se pueden usar la igualdad (`=`), mayor que (`>`), menor que (`<`), distinto de (`<>`), mayor o igual que (`>=`), etc. Hay que recordar que, excepto los números, los valores a la derecha de los signos de comparación, deben ir entre comillas. Así, `BARCELONA` debe escribirse entre comillas, por ejemplo para seleccionar las que trabajan en esta ciudad, `13/04/1959` debe ir entre comillas para seleccionar los nacidos después de esta fecha, pero `40` no necesita comillas si queremos trabajar con las personas cuya edad sea mayor o igual que 40 años.

MUY IMPORTANTE, recuerda que los criterios de selección se acumulan, es decir si ya has usado una orden `select` y usas otra, esta última se acumulará a la primera. Para evitarlo debes eliminar cualquier selección previa, lo que puedes hacer escribiendo `SELECT`, sin más.

A veces, querrás seleccionar los registros que cumplen criterios en más de una variable, por ejemplo las personas que trabajan en turno de tarde (`TURNO="T"`) y tienen más de 45 años (`EDAD>45`). Para esto debes usar lo que se llaman operadores lógicos (`AND`, `OR`). La orden anterior se escribiría: `SELECT TURNO="T" AND EDAD>45`. Debes tener cuidado por que no es lo mismo que `SELECT TURNO="T" OR EDAD >45`. Prueba las dos condiciones y mira donde está la diferencia (OJO, acuérdate de quitar la primera selección antes de incluir la segunda).

Si te fijas en la parte superior de la pantalla de `ANALYSIS`, verás que allí aparecen los criterios de selección activos en cada momento. Con los criterios

de selección activo, podrás realizar cualquier distribución de frecuencias, tablas, comparación de medias, regresión lineal o gráfico que desees.

Como siempre puedes encontrar información sobre la orden SELECT pulsando F1 (la tecla de ayuda).

¿Dónde están mis resultados?

Cuando se van obteniendo resultados, se pueden consultar los anteriores usando las teclas de desplazamiento <av pag> y <re pag>. Sin embargo, cuando sales de ANALYSIS, todos estos resultados se pierden. Para conservar una copia de los resultados puedes utilizar dos técnicas, una es enviar los resultados a una impresora, para lo cual se usa la orden ROUTE PRINTER (ROUTE es la orden y PRINTER es el objeto), la otra técnica consiste en enviar los resultados a un archivo mediante ROUTE <archivo>, donde <archivo> es cualquier nombre de archivo válido. La última opción es la más aconsejable porque permite utilizar después un procesador de textos para editar e imprimir los resultados, evitando los problemas que EpiInfo tiene para gestionar impresoras en redes y en entorno Windows.

Es importante que sepas que sólo los resultados de las órdenes que uses después de la orden ROUTE serán enviados al destino elegido, no así los resultados anteriores.

Consulta la ayuda de la orden ROUTE para saber más sobre como utilizarla.

¿Cómo adaptar ANALYSIS a mis preferencias?

Los resultados que ves en pantalla (o en impresora) se obtienen con la configuración que tiene ANALYSIS por defecto, pero a veces tu puedes desear que se vean de otra forma. Por ejemplo, puedes desear que al hacer una tabla, ANALYSIS muestre los porcentajes de fila y columna, o puedes desear que trabaje incluyendo los valores nulos, etc. Existe una orden que te permite configurar algunas cosas de ANALYSIS a tu gusto, se trata de la orden SET.

Para empezar escribe SET y pulsa la tecla ENTER (intro o return). Verás que aparece un listado de las cosas (parámetros) que puedes configurar y el estado en que está en ese momento. Vamos a ver algunos de ellos:

STATISTICS, normalmente está ON (activo), es decir cuando se realiza una tabla, se presentan los resultados estadísticos asociados.

PERCENTS, normalmente está OFF (inactivo), es decir, en las tablas no se incluyen los porcentajes de fila y columna.

IGNORE, normalmente está ON, es decir ANALYSIS no incluye los valores nulos (missing) al hacer una tabla o una distribución de frecuencias.

CRITERIA, normalmente activo, lo que significa que los criterios de selección existentes en el momento de hacer cualquier tabla se imprimirán al inicio de la misma.

Si, por ejemplo, deseas cambiar el estado de IGNORE, sólo debes escribir SET IGNORE=OFF. Con esto habrás puesto el parámetro IGNORE inactivo y a partir de ese momento ANALYSIS incluirá los valores missing en las tablas.

¿Cómo crear programas para automatizar tareas?

A veces ocurre que de una misma tabla se quiere obtener una serie de resultados periódicamente incluyendo la información más actualizada. Por ejemplo, en una tabla con los datos de casos de una enfermedad atendidos, queremos obtener todos los meses los mismos gráficos, tablas y comparaciones de media. Aunque podrías escribir cada vez todas las órdenes, es más cómodo escribirlas en un programa que ejecutarás de principio a fin con una sola orden.

Un programa de ANALYSIS es un conjunto de órdenes de ANALYSIS escritas una debajo de otra. Las órdenes son las mismas que podemos usar desde el teclado y se pueden escribir en cualquier procesador de texto que permita guardar el archivo con un formato "no documento", es decir como un archivo de tipo ASCII (por ejemplo puedes usar WORDPAD).

A continuación escribimos un programa que lee la tabla ACCIDEN.REC, hace una distribución de frecuencias de SEXO, compara las medias de edad por turnos y hace un gráfico de tarta de los turnos de trabajo de las mujeres. Al finalizar el programa cierra el programa ANALYSIS.

```
READ C:\TUTOREPI\ACCIDEN.REC
FREQ SEXO
MEANS EDAD TURNO
SELECT SEXO="M"
PIE TURNO
QUIT
```

Como hemos indicado, este programa debe guardarse en un archivo, por ejemplo RUTINA.PGM. Una vez que el programa está guardado, para ejecutarlo sólo tendremos que escribir, estando en ANALYSIS, la orden RUN RUTINA.PGM (RUN es la orden y RUTINA.PGM es el objeto).

Los programas de ANALYSIS pueden ser tan complicados como queramos, por ejemplo, las tutorías que estás usando para conocer mejor este programa son archivos de programas de ANALYSIS. A veces, después de haber estado escribiendo varias órdenes en el teclado nos damos cuenta de que estaría bien guardarlas pa-

ra otro día, puedes hacer esto escribiendo SAVE <archivo.PGM>, donde <archivo.pgm> será el nombre de un archivo donde se guarden estas órdenes. Normalmente ANALYSIS es capaz de recordar las últimas 20 órdenes utilizadas. Este número se puede incrementar hasta 99 usando SET CMDSTACK=99. Si haces esto podrás guardar en el archivo hasta las últimas 99 órdenes.

Crear nuevas variables

En ocasiones, los datos disponibles en la tabla necesitan algunas modificaciones para ser presentados de una forma más agradable, o bien es mejor agrupar a los sujetos en una nueva categoría que incluya otras anteriores, o se necesita hacer algún cálculo de interés. ANALYSIS permite hacer estas cosas y para ello es útil crear primero nuevas variables donde almacenar los nuevos valores obtenidos. La orden que permite crear nuevas variables se llama DEFINE. La sintaxis general es:

DEFINE <nombre> <símbolos>

donde <nombre> es el nombre de la variable que quieres crear (no debe coincidir con otro ya existente) y <símbolos> es cualquiera de los siguientes símbolos:

____, el carácter de subrayado indica que la variable será de tipo alfanumérico, es decir podemos escribir en ella lo que queramos. El máximo espacio posible estará indicado por el número de caracteres de subrayado.

###, caracteres almohadillas, indican que la variable será de tipo numérico, es decir sólo podremos escribir en ella valores numéricos. El número más alto estará determinado por el número de caracteres.

<dd/mm/yyyy>, esta combinación, incluyendo los signos <> del principio y fin, indican que la variable será de tipo fecha en formato europeo.

<Y>, esta combinación, incluyendo los caracteres <> de principio y fin indican que el campo será de tipo lógico, es decir que sólo acepta Y para representar Sí y N para No.

Cuando se crea una variable nueva, dicha variable no contiene datos. Como práctica, abre la tabla C:\TUTORPEI\ACCIDEN.REC y define una variable a la que llamarás GENERO (en ella pondrás los valores MASCULINO, FEMENINO). Para hacer esto debes escribir:

DEFINE GENERO _____

Esto significa: "crea una variable que se llame GENERO, que sea de tipo alfanumérica y que tenga espacio para escribir hasta 15 letras" (debes poner 15 caracteres de subrayado). Ahora, para ver lo que contiene esta variable, escribe:

BROWSE SEXO GENERO

Con esta orden verás sólo las dos variables indicadas: SEXO que contendrá los valores M, H y algún otro y GENERO que debe estar vacía.

Acciones condicionales

Ahora deseamos rellenar la variable GENERO con los valores MASCULINO y FEMENINO que nos parecen más descriptivos que H y M. Para hacerlo de forma automática debemos indicarle a ANALYSIS que haga cosas distintas dependiendo de que se cumplan determinadas condiciones. Para ello debes usar una orden del tipo:

IF.....THEN.....ELSE..., es decir: SI (pasa algo) ENTONCES (haz una cosa) SI NO (haz otra cosa). ELSE y lo que va detrás es opcional.

Para hacer lo que deseamos, la orden que debes escribir es:

IF SEXO="H" THEN GENERO="MASCULINO" (y pulsar enter)

(Aparentemente nada ha ocurrido, sin embargo prueba a hacer BROWSE SEXO GENERO)

Con esto lo que has hecho ha sido rellenar la variable GENERO con el valor MASCULINO en aquellos registros en los que SEXO="H".

Ahora escribe:

IF SEXO="M" THEN GENERO="FEMENINO" (y pulsar enter). Vuelve a comprobar los datos con BROWSE SEXO GENERO. Ahora la mayoría de los registros deben tener rellena la variable GENERO con valores que dependen de los encontrados en SEXO. Observa que los registros en los que GENERO está vacío son aquellos en los que SEXO es un valor distinto de H y de M.

Consulta la ayuda de la orden IF para aprender más sobre ella.

Recodificar variables

Ahora deseamos agrupar a los sujetos de nuestra tabla en distintos grupos según su edad, por ejemplo queremos crear los grupos: Menores de 30 años, de 31 a 50 años y mayores de 50 años. Lo primero que necesitamos es crear una variable a la que llamarás GRUPEDAD, usando la orden:

DEFINE GRUPEDAD _____

Esto significa: "crea la variable GRUPEDAD que será de tipo alfanumérico y donde podré escribir hasta un máximo de 20 letras" (debes haber escrito 20 caracteres de subrayado). Si ahora haces BROWSE GRUPEDAD, verás que esta variable está vacía. Para recodificar GRUPEDAD con unos valores que depende de

los que pone en la variable edad podríamos utilizar el sistema IF... THEN... ELSE, pero ahora vamos a usar la orden RECODE.

La sintaxis es:

RECODE <varori> to <varden> <equivalencias>, donde

<varori> será el nombre de la variable que tiene los valores originales (en nuestro ejemplo la edad)

to es exactamente to (es decir escribe esas dos letras).;-)

<varden> será el nombre de la variable donde se pondrán los valores recodificados (en el ejemplo, GRUPEDAD).

<equivalencias> representa la forma en que se indican que valores de la variable original corresponden con los de la variable destino. En nuestro ejemplo será: LO-30="< 30", 31-50="Entre 31 y 50", 51-HI="> 50". En resumen debes escribir:

RECODE EDAD TO GRUPEDAD LO-30="<= 30" 31-50="Entre 31 y 50" 51-HI="> 50" (y pulsa enter). (ojo todo esto debes escribirlo en una sola línea). LO sirve para especificar el valor más bajo, mientras que HI se usa para especificar el más alto. Ahora escribe BROWSE EDAD GRUPEDAD para ver los resultados.

Cálculos con variables

ANALYSIS permite hacer cálculos utilizando una o más variables de las existentes en la tabla para obtener un resultado que se almacenará en otra variable. Un ejemplo de esto es calcular una edad a partir de dos fechas o calcular el Índice de Masa Corporal a partir de la talla y el peso. Supongamos que en la tabla no disponemos de la variable edad, podríamos calcularla a partir de dos fechas, la de nacimiento (FECHANAC) y la del accidente (FECHANOT). Primero debemos crear una variable donde guardar los cálculos. La llamaremos EDADCAL (edad calculada). Para hacerlo escribe:

DEFINE EDADCAL ##

Esto significa: "crea una variable llamada EDADCAL, que será de tipo numérico, con espacio suficiente para escribir como número más alto 99" (si has puesto dos caracteres de almohadilla). Si ahora escribes BROWSE EDADCAL verás que la variable está vacía.

A continuación deberás indicarle a ANALYSIS que las fecha con las que vas a trabajar están en formato europeo, para ello configura el parámetro EUROPEAN usando:

SET EUROPEAN=ON

Ahora haremos los cálculos, para lo cual no necesitamos ninguna orden, simplemente debemos escribir:

EDADCAL=(FECHANOT -FECHANAC)
DIV 365.25 (y pulsar enter)

Lo anterior significa: "Pon en la variable EDADCAL, para cada registro, el valor resultante en días de la diferencia de días entre las fecha que hay en FECHANOT y FECHANAC, dividiendo ese número por 365.25 para obtener el valor redondeado entero más cercano que será un número de años" (resulta más largo decirlo que hacerlo). Ahora escribe BROWSE EDAD EDADCAL y compara los valores obtenidos con los originales.

Se pueden realizar muchos tipos de cálculos y usar distintos operadores matemáticos, consulta la ayuda para obtener más información.

Guardar los resultados en otra tabla

Todas las variables definidas y los cálculos, recodificaciones etc. que has realizado hasta ahora, se perderán si no lo guardas en una tabla nueva. Para hacer esto debes usar dos órdenes seguidas:

ROUTE <archivo.rec> (y pulsar enter)

WRITE RECFILE (y pulsar enter de nuevo)

Al finalizar el proceso de esta segunda orden obtendrás una nueva tabla llamada como indicaste en el lugar <archivo.rec>. Esta tabla puedes leerla con ANALYSIS usando la orden READ. Por ejemplo para guardar todo lo realizado hasta ahora en una tabla llamada BORRAME.REC en la carpeta C:\TUTOREPI debes escribir:

ROUTE C:\TUTOREPI\BORRAME.REC (y pulsar enter)

WRITE RECFILE (y pulsar enter)

Cuando acabe el proceso escribe:

READ C:\TUTOREPI\BORRAME.REC

y comprueba que tiene las variables que antes has definido con sus datos.

EN EL PRÓXIMO NÚMERO

- Utilizarás el programa EPED para diseñar un cuestionario.
- Utilizarás el programa ENTER para grabar datos en una tabla.
- Aprenderás desplazarte por la tabla usando ENTER.
- Utilizarás el programa CHECK para controlar la calidad de la entrada de datos.

Recuerda que si deseas realizar alguna consulta puedes hacerlo usando el correo electrónico: p.arias@comadrid.es, o el de la revista: revista@seps.es.