

# Epidemiología

## Artículo original

# Cómo construir una base de datos

## How to building a database

Esther Sofía Angulo Martínez, Esp, MSc\* ; Leonardo Elías Ordóñez Ordóñez, MD\* \*

### RESUMEN

*La construcción de una base de datos es un aspecto importante durante la realización de una investigación. Aunque hay varios programas que permiten construir bases de datos, sugerimos utilizar paquetes estadísticos que provean el armamentario bioestadístico necesario para hacer los posteriores análisis. El programa SPSS es uno de los que tienen mayor acogida y es fácil de manejar, por ello lo utilizamos para los ejemplos.*

*Antes de comenzar a construir la base de datos es muy importante determinar el nivel de medición de la variables, ya que de ello depende en gran parte el tipo de pruebas estadísticas que se van a utilizar. Hay cuatro tipo de variables de acuerdo al nivel de medición: nominales, ordinales, interválicas y proporcionales. Para las variables nominales y ordinales se utilizan pruebas no paramétricas; mientras que para las interválicas y proporcionales se utilizan pruebas paramétricas.*

*Se describe paso a paso cómo ingresar una variable nominal y una variable proporcional, haciendo énfasis en aspectos de formato del paquete estadístico. Se hacen varias recomendaciones para sacar el máximo provecho a la base de datos.*

**Palabras clave:** base de datos, diseño de investigación, estadística no-paramétrica, metodología.

---

\* Epidemióloga. Magíster en Infecciones y Salud en el Trópico.

\*\* Otorrinolaringólogo, Fellow de en Otología. Servicio Integrado de Otorrinolaringología del Hospital Militar Central y Hospital Universitario Clínica San Rafael.

Correspondencia:

Leonardo Elías Ordóñez Ordóñez.  
Servicio de Otorrinolaringología. Hospital Militar Central.  
Transversal 3 # 49-00, tercer piso norte. Bogotá.  
otoleor@gmail.com, kokhlias@gmail.com.

Recibido: 09/VII/2008

Aceptado: 10/VIII/2008

## ABSTRACT

*A database construction is an important step in research development. Although several software packages permit you to building a database, we would suggest you to use a statistics software packages that supply you with biostatistics options to do any wished analysis. SPSS software package has wide reception and is easy to run, for that reason we used it for the following examples.*

*Before anything else, to determine measurement level of study variables it is a central point, which is the main determinant to select a statistic test. Four variables are described according to measurement level: nominal, ordinal, interval and ratio variables. For nominal and ordinal variables non-parametric tests are used; otherwise, interval and ratio variables are analyzed with parametric tests.*

*Step by step instructions to enter nominal and ratio variables are described; format topics of software package are highlighted. Some recommendations for maximum advantage of database are outlined.*

**Key words:** *database, research design, nonparametric statistics, methodology.*

## INTRODUCCIÓN

Durante la realización de una investigación clínica un aspecto importante para la recolección de la información es la construcción de una base de datos. Una base de datos es un archivo, físico o electrónico, en donde se recoge y almacena de manera sistemática la información necesaria, para luego analizarla. Aunque se puede realizar una base de datos en físico (un archivo de documentos), esto implica que al menos una parte del análisis estadístico se realizará de forma manual, con un importante consumo de tiempo y recurso humano. La realización de bases de datos electrónicas supone un importante ahorro en tiempo y recursos, es por ello que en adelante hablaremos de cómo construir una base de datos en un archivo electrónico. Dejamos las bases de datos en físico para utilizarlas sólo como copias de seguridad de la información, ya sea que la información se recoja en formularios que luego se digitan o que se trate de impresiones del archivo electrónico.

Para la realización de una base de datos electrónica necesitamos de un programa, software, que nos permita construir este tipo de archivos. Aunque algunos programas como Excel y Access son ampliamente disponibles y están diseñados para construir bases de datos, su perfil es más financiero y contable, con herramientas limitadas para el análisis bioestadístico. Por ello recomendamos usar programas especializados en bioestadística como SPSS, STATA, SAS y EpiInfo; que son fáciles de manejar y ofrecen todo el armamentario bioestadístico que usted puede requerir. EpiInfo es un programa con una gran ventaja: su descarga es gratuita (lo puede obtener en la página: <http://www.cdc.gov/>

eppiinfo/), pero el manejo de algunas de sus herramientas es un poco más dispendioso. Hay una herramienta muy ágil e invaluable de EpiInfo versión 6.04: su calculadora estadística (se encuentra en el menú programas como "STATCAL"), que le permitirá analizar rápidamente tablas de  $2 \times 2$  y  $2 \times n$  y hacer algunos cálculos del tamaño de la muestra, Figura 1. Entre los programas mencionados vamos a seguir la siguiente discusión haciendo los ejemplos con el programa SPSS, hay varias razones para su selección: es bastante fácil de usar (versión para el sistema operativo Windows), está disponible en varias universidades colombianas (es probable que lo esté en alguna a la cual pueda acceder en su ciudad) y es uno de los más utilizados en la literatura médica.

Como ejemplo vamos a suponer que usted desea hacer un estudio para analizar sus resultados en estapedotomía/estapedectomía. Revisa sus libros de programación de cirugía y encuentra que en los últimos 10 años ha operado 150 pacientes de cirugía del estribo. Decide que el objetivo de su estudio es analizar sus resultados y compararlos con los publicados en la literatura para determinar si el manejo que usted hace es acorde con el estándar de tratamiento internacional (El Servicio Nacional de Salud del Reino Unido acuñó un término, que ahora es término Mesh, para este tipo de estudios: "Clinical Audit", que se puede traducir como auditoría clínica de resultados. Estos estudios consisten en una revisión y evaluación sistemática de las historias clínicas con el fin de determinar si los resultados son concordantes con los resultados internacionales, el estándar de tratamiento, para mejorar la calidad de la atención y resultados de los tratamientos). Este es un estudio descriptivo y la información la va a recoger de manera retrospectiva.



**Figura 1.** La calculadora estadística de EpiInfo 6.04 es una herramienta fácil de usar y valiosa. Por ejemplo si deseas comparar el porcentaje de cierre de perforación timpánica de tu estudio con el de otro trabajo publicado, puedes hacer una tabla de 2 x 2 y determinar si hay una diferencia estadísticamente significativa entre estos dos trabajos, mediante una prueba de chi cuadrado.

Para construir una base de datos por convención la información de los casos o pacientes se coloca en las filas y la información de las variables se coloca en las columnas: es decir, que la información se lee horizontalmente. En la Figura 2 observará una imagen de la base de datos de SPSS para su estudio. Observe que para su primer paciente tiene algunas variables como: [instituc], [género], [etiolo] y [ladocomp] (Los nombres de las variables se escriben en minúsculas con caracteres alfanuméricos, sin signos ortográficos, con máximo ocho caracteres de longitud —la última versión del programa ya no tiene esta restricción de longitud— y no se puede usar la “ñ”. En adelante los nombres de las variables los escribiremos entre corchetes, [variable], y cuando el nombre no sea claro, por ejemplo [ladocomp], seguido por un guión medio colocaremos la etiqueta de la variable: [ladocomp]-Lado comprometido-). Pero, ¿cómo creo esas variables?, para explicarlo es necesario primero revisar algunos conceptos de estadística.

**Nociones estadísticas: variables, tipos de variables y nivel de medición de las variables**

Una *variable* no es más que una característica que está siendo medida u observada, que por definición tiene al menos

dos posibles valores (1). Las variables se clasifican en independientes y dependientes.

Una *variable independiente* es aquella que mide la intervención que está siendo aplicada (1). El tipo de cirugía realizada ([tipcx]: estapedotomía, estapedectomía, hemiplatinectomía anterior y hemiplatinectomía posterior) es un ejemplo de variable independiente.

Una *variable dependiente* es aquella que varía en respuesta a alguna intervención, es una medida que cambia en respuesta a la variable independiente (1). En nuestro ejemplo el valor de la vía aérea en la frecuencia 1000Hz ([f1khprea]: medición en decibeles) es un ejemplo de variable dependiente: esperamos que cambie, especialmente que mejore, luego de la cirugía.

Determinar cuáles variables son independientes y cuáles dependientes es un aspecto muy importante que se debe aclarar en el protocolo de investigación, ya que de esta distinción depende en parte el tipo de pruebas estadísticas que vamos a aplicar.

Un aspecto igual de importante que el anterior es determinar el tipo de variable de acuerdo al *nivel de medición*

que se utilice. Hay una diferencia fundamental en el tipo de medida que se utiliza para medir la variable [tipcx]-tipo de cirugía- (posibles valores: estapedotomía, estapedectomía, hemiplatinectomía anterior y hemiplatinectomía posterior) y la medida que se utiliza para la variable [f1khprea]-Vía aérea: umbral en 1000Hz audiometría PREOPERATORIA- (posibles valores: cualquiera de la escala audiométrica medida en decibeles). Como podemos ver la variable [tipcx] tiene solamente cuatro posibles valores, mientras que la variable [f1khprea] puede tomar cualquier valor de la escala de decibeles, de hecho el número de posibles valores es infinito.

Para concretar: de acuerdo al nivel de medición utilizado las variables pueden ser nominales (categóricas), ordinales, interválicas o proporcionales (1).

**VARIABLES NOMINALES:** *una variable nominal consiste en categorías a las que se asigna un nombre sin que exista un orden implícito entre ellas.*

Aparte de [tipcx], otros ejemplos de variables nominales en su base de datos son: [género] que tiene solamente dos posibles valores: masculino y femenino; [ladocomp]-lado comprometido- que tiene tres posibles valores: izquierdo, derecho o bilateral. Es importante resaltar que no existe un orden implícito entre los posibles valores de la variable: no podemos decir que ser mujer es mejor que ser hombre o viceversa.

**VARIABLES ORDINALES:** *una variable ordinal consiste en categorías a las que se les asigna un nombre existiendo un orden implícito entre ellas, pero las diferencias entre uno y otro valor no son iguales.*

Para explicar la definición utilicemos un ejemplo. La medición clínica de la función motora facial se hace a través de la escala de House-Brackmann (HB). Recordemos que sus posibles valores son: I, II, III, IV, V y VI; correspondiendo el grado I a una motricidad normal, el grado VI a una parálisis completa y los grados II a V a paresia de diferentes niveles de severidad. Es claro que un grado I significa una mejor función facial que un grado II, y que un grado II es mejor que un grado III, y así sucesivamente; sin embargo, también es cierto que no podemos decir que la diferencia entre un grado VI y un grado V sea "igual" a la diferencia entre un grado IV y un grado III. Como en la base de datos que está construyendo no hay ejemplos de variables ordinales, vamos a dar otros: los estadios del cáncer laríngeo (posibles valores: I, II, III, IV) y el grado de hipertrofia amigdalina (posibles valores: I, II, III, IV).

Hay algunos estadísticos que consideran que la medición de la intensidad del sonido en decibeles (dB) es una medición ordinal ya que no cumple con los requisitos para ser considerada como una medición interválica o proporcional

(la diferencia entre 80 dB y 70 dB no es la misma que entre 30 dB y 20 dB; y 0dB no significa ausencia de sonido. Véase el apartado siguiente). Esto es mucho más que una discusión Bizantina ya que si la consideramos una variable interválica o proporcional utilizaremos las poderosas *pruebas paramétricas*, mientras que si la consideramos una variable ordinal debemos utilizar *pruebas no paramétricas*. Podría darse el caso que si se utiliza una prueba paramétrica se logre significación estadística ( $p < 0.05$ ), mientras que si se usa una prueba no paramétrica, para el mismo conjunto de datos, no se logre la significación estadística.

Nuestro concepto, que es tomado de otra corriente de estadísticos, es que una medición en decibeles es una medición proporcional, ya que recordemos que los decibeles son el logaritmo de una relación de presiones de sonidos y por tanto, sigue siendo una proporción. Por consiguiente el análisis de una variable medida en decibeles se puede hacer a través de pruebas paramétricas.

**VARIABLES INTERVÁLICAS Y PROPORCIONALES:** vamos a considerar a estos dos tipos de variables en el mismo apartado por una situación práctica, el análisis estadístico se hace con el mismo tipo de pruebas: las pruebas paramétricas.

Una variable interválica toma valores de un conjunto de datos entre los que existen diferencias iguales entre ellos, sin embargo, el punto cero es arbitrario.

Una variable proporcional además de tener intervalos iguales entre los valores tiene un punto cero significativo.

Un ejemplo de una variable interválica es la medición de la temperatura en grados centígrados (grados Celsius  $^{\circ}\text{C}$ ). Es claro que  $39^{\circ}\text{C}$  es mayor que  $37^{\circ}\text{C}$  y que la diferencia entre ellos,  $2^{\circ}\text{C}$ , es igual a la diferencia entre  $37^{\circ}\text{C}$  y  $35^{\circ}\text{C}$  ( $2^{\circ}\text{C}$  a lo largo de toda la escala significan la misma cantidad de energía cinética). También es claro que el punto cero,  $0^{\circ}\text{C}$ , es arbitrario, ya que  $0^{\circ}\text{C}$  no significa ausencia de energía cinética. El punto cero de esta escala fue colocado, por una de arbitrariedad humana, en el punto en que el agua pasa de estado sólido a estado líquido. Bien pudo haberse colocado en el punto en que el alcohol pasa de líquido a gaseoso o en el punto en que el hierro pasa de sólido a líquido. Otros ejemplos de mediciones interválicas son: la medición del coeficiente intelectual (posibles valores: distribución normal con la media en 100 y una desviación estándar de 20), la evaluación del lenguaje mediante la escala IT-MAIS (posibles valores: de 0 a 40). La corriente de estadísticos que quieren rebajar los decibeles al nivel ordinal hacen reparos similares sobre estos últimos dos ejemplos.

Un ejemplo de una variable proporcional es la medición de la temperatura en grados Kelvin ( $^{\circ}\text{K}$ ). Esta escala térmica

aparte de tener diferencias iguales entre sus valores tiene un punto cero significativo: 0°K realmente significa ausencia de energía cinética. Ejemplos de su base de datos que son variables proporcionales son: [edad] (medida en años, meses, días, horas o...), [sdspre]-Porcentaje de discriminación del lenguaje en la logaudiometría preoperatoria- (que va de 0% a 100%), [longprot]-Longitud de la prótesis- (que puede ir desde unos 2 mm hasta unos 6.5 mm). Otros ejemplos de variables proporcionales son el peso corporal (medido en kg, g o...), la talla (medida en m, cm, mm o ...) o nivel de hemoglobina en sangre.

### **Importancia de clasificar correctamente el nivel de medición de las variables**

La importancia de clasificar adecuadamente el nivel de medición no puede ser mayor, ya que de ello depende el tipo de prueba estadística que va a realizar. Si la variable es nominal el análisis estadístico se hace a través de pruebas no paramétricas como el chi cuadrado, la prueba exacta de Fisher, el chi cuadrado de McNemar o el chi cuadrado de Mantel-Haenszel. Si la variable es ordinal va a utilizar pruebas no paramétricas como la suma de rangos de Wilcoxon, el ANOVA de un tratamiento de Kruskal-Wallis, la prueba de los rangos signados de Wilcoxon o el ANOVA de dos tratamientos de Friedman. Finalmente si su variable es interválica o proporcional usted puede utilizar las poderosas pruebas paramétricas (entiéndase por “poderosas” como la potencia o poder de una prueba que es la probabilidad de llegar a la conclusión de que hay una diferencia cuando efectivamente así es), entre las cuales tenemos la prueba de la t de Student, la prueba t por parejas (llamada también prueba t de medidas repetidas), del ANOVA, ANOVA factorial y ANOVA de medidas repetidas.

Dado que las pruebas paramétricas son de mayor poder estadístico que las no paramétricas se considera un *grave error estadístico* convertir para su análisis el nivel de medición de una variable de interválica/proporcional a ordinal o nominal, ya que esto supone una pérdida de información. Explicamos, si usted quiere determinar si las cirugías que realizó mejoraron el nivel auditivo de sus pacientes compare el GAP (diferencia entre los umbrales aéreos y óseos) preoperatorio con el GAP posoperatorio a través de una prueba t de por parejas y así podrá determinar

si hay diferencias estadísticamente significativas. Es probable que en este momento usted esté pensando que según las guías de la Academia Americana de Otorrinolaringología y Cirugía de Cabeza y Cuello (AAO-HNS), se considera que la cirugía es exitosa si se cierra el GAP a menos de 10dB (2). Bueno, pues esto es cierto, y es muy frecuente ver en la práctica clínica la reducción de variables proporcionales/interválicas a variables nominales con el fin de hacer más fácil la comprensión de los resultados, lo cual es válido bajo esta premisa (otros ejemplos: transformar la edad en mayores o menores de 60 años, transformar el ángulo nasolabial en subrotación, normal o sobrerrotación de la punta nasal). Sin embargo, la recomendación es hacer el análisis de los datos en bruto y luego sí clasificarlos, para la presentación de resultados, en categorías. Una de las pocas justificaciones para reducir datos interválicos/proporcionales a nominales, es utilizarlos para comparar sus resultados con los publicados en otros estudios que han hecho ese tipo de reducción de la información. Siguiendo el ejemplo, si usted clasifica la cirugía como exitosa o no exitosa dependiendo si el GAP posoperatorio se cerró o no a 10 dB (transformó una variable proporcional en una nominal) esta nueva variable le puede permitir comparar sus resultados con los de estudios que representen en estándar de tratamiento actual, a través de una prueba de chi cuadrado.

Bueno una vez hechas estas anotaciones estadísticas continuamos con la elaboración de la base de datos

### **¿Cómo construir una base de datos?**

La ventana del programa SPSS se puede alternar entre dos posibilidades (botones abajo y a la izquierda): la vista de datos (en la cual vemos la base de datos como tal, Figura 2) y la vista de variables (en la cual creamos y definimos las variables), Figura 3.

Una vez en la vista de variables observamos que en esta ventana las variables van en las filas, mientras en las columnas están las características de estas variables (Nombre, Tipo, Anchura, Decimales, Etiqueta, Valores, Perdidos, Columnas, Alineación y Medida). Como ejemplo vamos a crear dos variables una nominal y otra proporcional, comenzamos por la variable nominal.

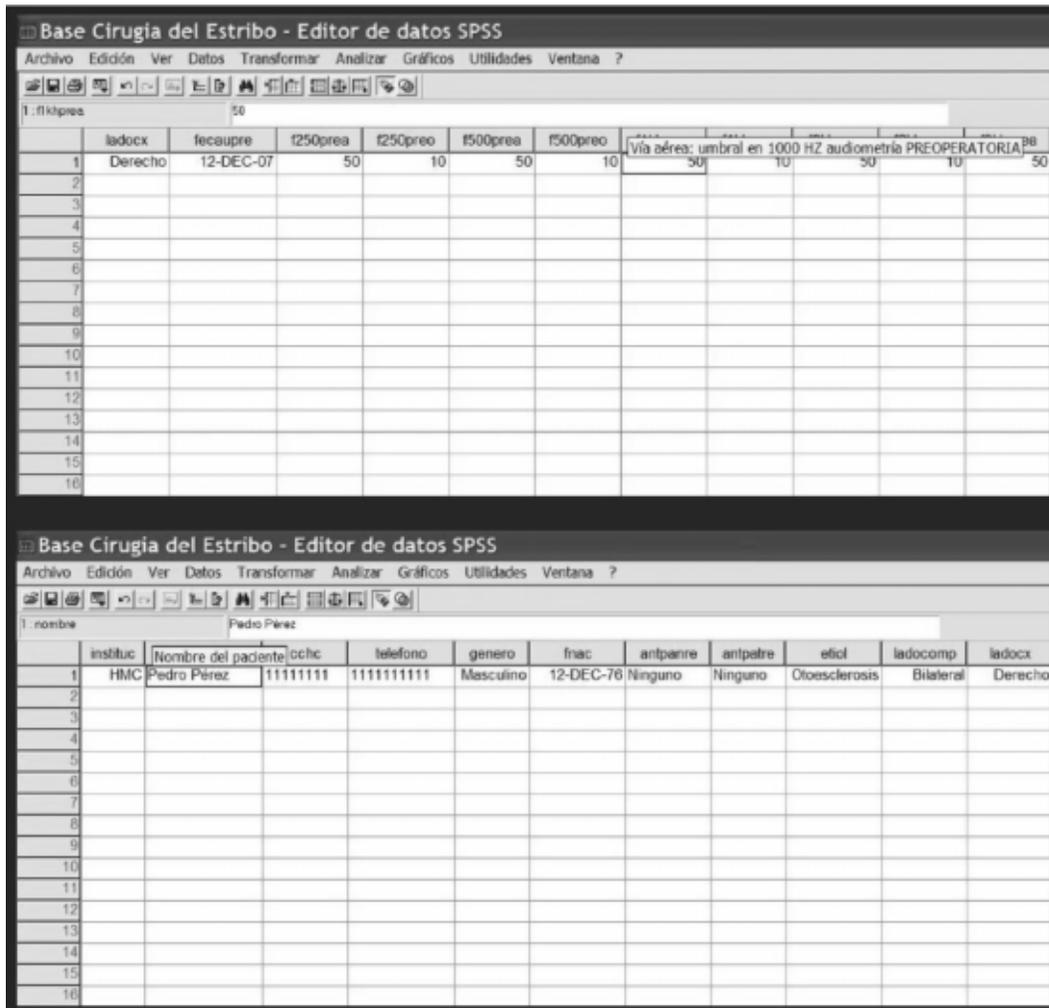


Figura 2. Imágenes de la pantalla de SPSS donde se encuentra una parte de su base de datos sobre cirugía del estribo. Observe como la información se dispone de manera horizontal, con los pacientes en las filas (numerados a partir del “1”) y las variables en las columnas.



Figura 3. En la vista de variables se crean y definen las características de las variables. Vemos en la fila 10 resaltada la variable [ladocomp]–Lado comprometido–, que utilizaremos como ejemplo de variable nominal. Observe en la parte inferior izquierda de la ventana los botones para alternar entre vista de datos y vista de variables.

### ¿Cómo crear una variable nominal?

En la fila número 10, resaltada en la Figura 3, encontramos la variable [ladocomp]-Lado comprometido-. Vemos que el nombre de la variable (columna 1) se ajusta a los requerimientos que impone el programa, aquí de todas formas es importante escoger un nombre que de por sí tenga significado ([ladocomp] no es el mejor ejemplo, pero en el contexto de esta base de datos se puede considerar un buen nombre). Para ingresar el nombre basta con dar doble-click e ingresar el nombre. En la columna 2 se define el tipo de variable, en la Figura 4 vemos que hay varias posibilidades; sin embargo, para efectos de manejo estadístico es mejor seleccionarla como una variable “Numérica” (sabemos que es una variable nominal cuyos valores son nombres: derecho, izquierdo o bilateral. Lo que pasa es que SPSS maneja mejor los datos cuando los valores son números). Los otros dos tipos de variables que usamos con frecuencia son el tipo “Fecha” y el tipo “Cadena”. Las variables del tipo fecha permiten ingresar ese tipo de datos y las de cadena ingresar valores alfanuméricos de hasta 255 caracteres de extensión.

Las variables de cadena son útiles para variables como el nombre del paciente, teléfonos, direcciones, anotaciones de la cirugía y otros. Al lado derecho de la ventana “Tipo de variable” nos da la posibilidad de seleccionar la “Anchura” y el número de “Cifras decimales”. La anchura hace referencia al número máximo de valores que puede tener la variable. “1” significa un dígito, es decir, un máximo de diez valores posibles para la variable (de 0 a 9); “2” significa dos dígitos, es decir, un máximo de cien valores posibles para la variable (de 0 a 99). Para nuestro ejemplo como la variable [ladocomp] solamente puede tener tres posibilidades, en la casilla de anchura tecleamos el número “1”. Para el caso de la variable [edad], que la vamos a medir en años podría presentarse el caso de un paciente mayor a cien años (bueno, quizá no para hacerse una estapedotomía, pero sí para tratarse con otra modalidad terapéutica como un audífono), por tanto la anchura debe ser de “3”. En cuanto a “Cifras decimales”, para el caso de variables nominales no tiene aplicación y se debe dejar como “0”, veremos con el ejemplo de variable proporcional la utilidad de esta casilla. Da usted aceptar y seguimos con las siguientes casillas.



Figura 4. Al dar clic sobre la segunda columna, Tipo, emerge la ventana “Tipo de variable”. En ella aparecen varias opciones de tipo de variable hacia el lado izquierdo de la ventana, vamos a seleccionar el tipo “Numérica”. Hacia la derecha nos da la posibilidad de seleccionar la anchura y el número de cifras decimales de la variable.

Las columnas tercera y cuarta, “Anchura” y “Decimales” son las mismas de la ventana “Tipo de variable”.

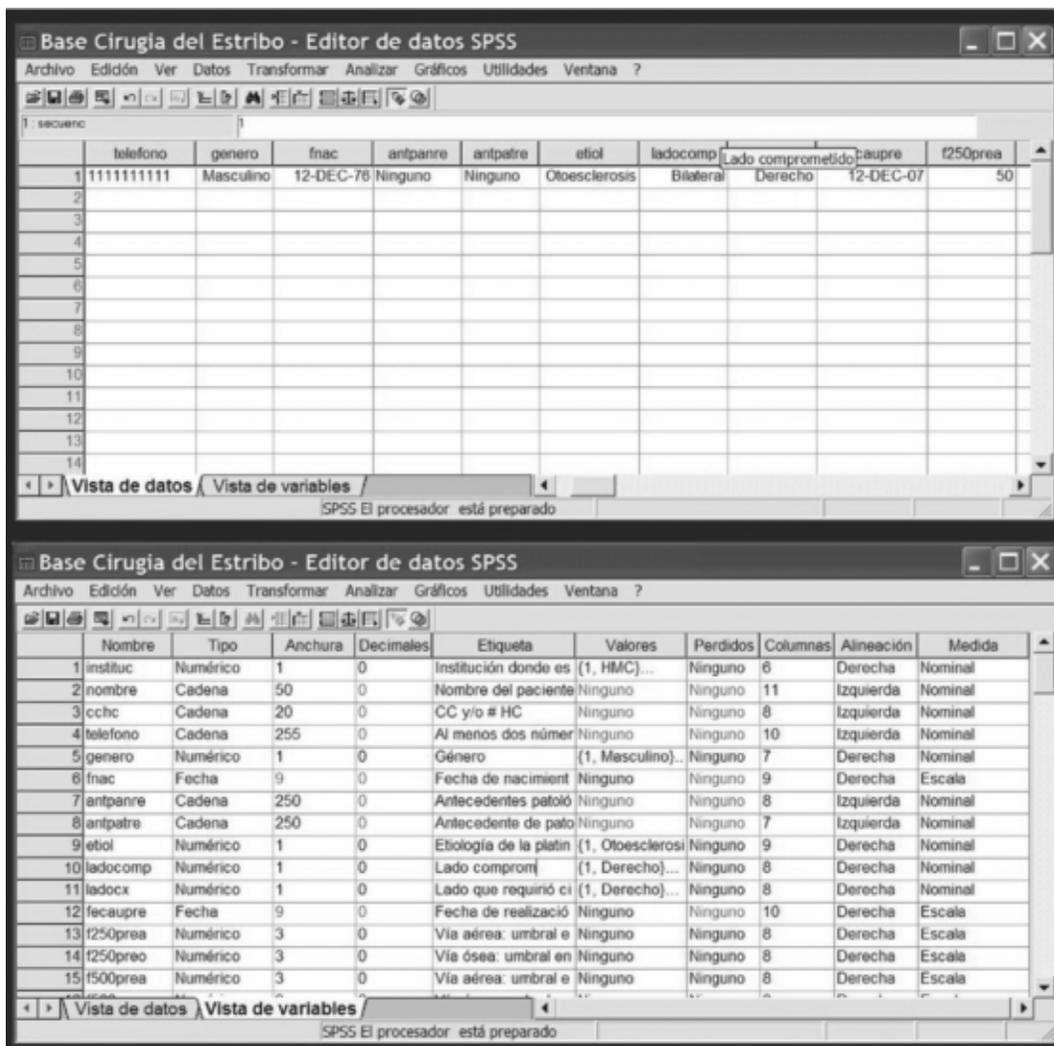
La quinta columna, “Etiqueta”, hace referencia al texto explicativo de la variable. Aquí usted puede escribir lo que crea conveniente para explicar el significado de la variable,

en nuestro caso dice: —Lado comprometido—. Es conveniente ser lo más explícito posible en la etiqueta ya que es probable que sean varias las personas encargadas de ingresar la información, y entre más información disponible tengan mejor. Por ejemplo para la variable [edad] Usted podría escribir en la etiqueta: —Edad medida en meses entre la

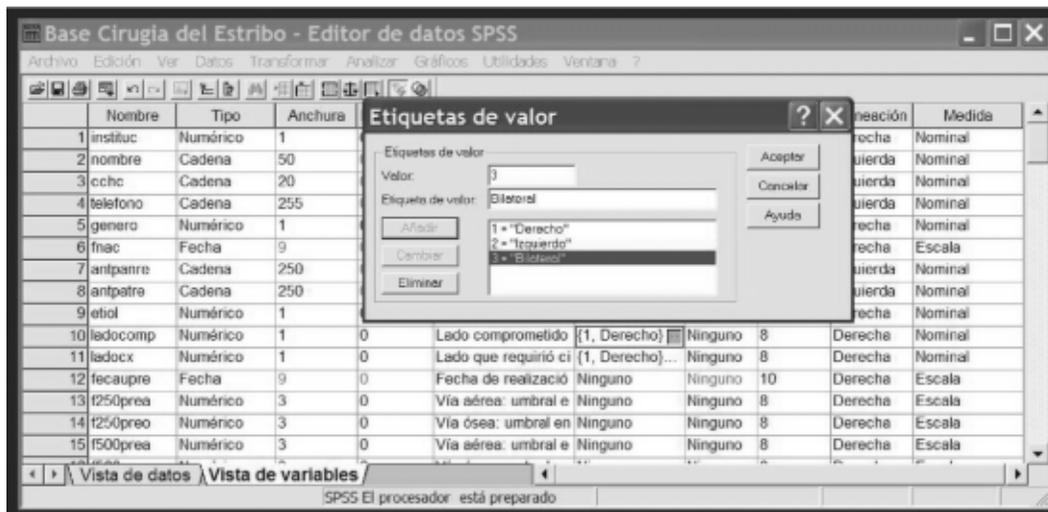
fecha de nacimiento y el día de la cirugía—. Para ingresar el texto de la etiqueta usted se coloca en la casilla, da doble click y escribe el texto que desee, Figura 5.

Cuando usted da click sobre la sexta columna, “Valores”, aparece otra ventana emergente que se llama “Etiquetas de valor”, Figura 6. Es aquí donde usted define y codifica los posibles valores para su variable y hace la correspondencia entre el número asignado a la variable (recuerde en el “Tipo de variable” seleccionamos: “Numérica”) y el valor real de la variable. Para nuestra variable [ladocomp] asignamos el valor “1” al lado derecho, “2” al lado izquierdo y “3” al compromiso bilateral. En la casilla “valor” usted ingresa el

número y en la casilla “Etiqueta de valor” el nombre del valor, luego da click en “Añadir”, pasa el siguiente valor y cuando termine de ingresar los posibles valores da clic en “Aceptar”. Aunque usted puede dejar de codificar los valores de sus variables nominales, es mejor que lo haga: quizás en cuatro meses cuando vaya a hacer el análisis estadístico se le puede haber olvidado si el “1” correspondía a “Izquierdo”, a “Derecho” o a “Bilateral”. Para las variables interválicas/proporcionales lo usual es no codificar las variables, de forma que esta columna no se llena: imagine lo dispendioso que sería codificar la escala de decibeles: tendría que ingresar valores entre -10dB a 140 dB, y aún peor, ¡que tal si se le ocurre ingresar una cifra decimal!



**Figura 5.** Imagen superior: la etiqueta de la variable es el texto explicativo de la misma. Sea lo más explícito posible de forma que si otra persona va a ingresar los datos entienda cuál es la variable, la unidad de medida utilizada y otros aspectos relevantes. Imagen inferior: para ver la etiqueta de la variable en la vista de datos colóquese con el cursor encima del nombre de la variable y aparecerá un recuadro con el texto de la variable.

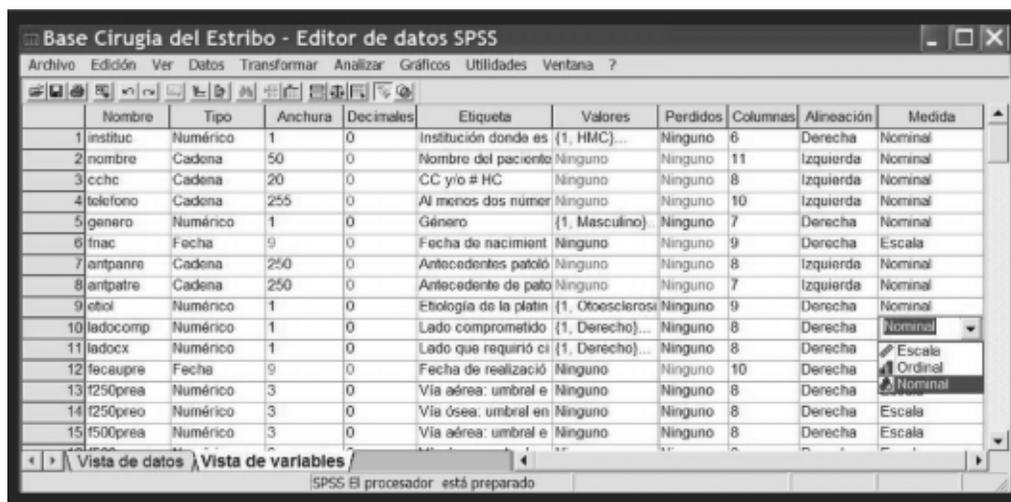


**Figura 6.** En la ventana emergente de “Etiquetas de valor” usted va a ingresar y codificar el valor de sus variables. Observe como el número “3” corresponde a “Bilateral” y el “1” y el “2” corresponden a “Derecho” e “Izquierdo”, respectivamente. En la “Etiqueta de valor” sea tan explícito como sea necesario.

La siguiente columna, “Perdidos”, hace referencia al número de datos no ingresados en la base de datos para la variable en cuestión, es una herramienta útil para encontrar casillas que no se han llenado (por ejemplo por un error de digitación).

“Columnas” y “Alineación” son parámetros netamente de formato con poca relevancia práctica. “Columnas” hace referencia al ancho de la columna en la vista de datos, una propiedad que puedes cambiar arrastrando el margen con el mouse. La columna “Alineación” se refiere a si usted quiere que los datos estén justificados a la derecha, izquierda o centrado.

Finalmente la última columna titulada “Medida”, hace referencia al nivel de medición de la variable, como habíamos dicho un aspecto de gran importancia. El programa SPSS considera tres posibilidades: “Escala”, “Ordinal” y “Nominal”, Figura 7. “Escala” incluye a las variables interválicas y las proporcionales. Una inadecuada selección del nivel de medición hace que cuando usted vaya a hacer las pruebas estadísticas no tenga disponible las variables que desea. Sea muy acucioso al seleccionar el nivel de medición para que sea el correcto.



**Figura 7.** La selección adecuada del nivel de medición es de gran importancia para el análisis estadístico. La variable [ladocomp]-Lado comprometido- es una variable nominal de forma que esa es nuestra selección. Recuerde que bajo el nivel de medición “Escala” se incluyen a las variables interválicas y proporcionales.

Bueno, usted ha creado satisfactoriamente una variable nominal, en el siguiente apartado va a crear una variable proporcional.

### Cómo crear una variable interválica o proporcional

Vamos a utilizar como ejemplo la variable [f1khprea]-Vía aérea: umbral en 1000 HZ audiometría preoperatoria-, ubicada en la fila 17 de la vista de variables. Muchos de los pasos son idénticos a los que siguió para crear la variable nominal de forma que sólo vamos a hablar de las diferencias.

Esta es una variable numérica que oscila entre -10 a unos 140 dB, de forma que la anchura es de “3” y no va a utilizar decimales. En columna “Etiqueta” usted es lo más explícito que puede y en la columna “Valores” no codifica ningún valor ya que se trata de una variable interválica (recuerde lo dicho en la sección precedente). Solamente en la última columna, “Medida”, debe estar atento a seleccionar “Escala”, ya que [f1khprea] es una variable proporcional, Figura 8.

El proceso para ingresar variables ordinales es muy similar con la excepción de que en la columna “Valores” suele ser conveniente codificar los valores y en la columna “Medida” debe escoger un nivel de medición “Ordinal”.

Bueno usted ingresó todas las variables que consideró relevantes para el estudio y así finaliza la construcción de su base de datos, ahora debe ingresar los datos. Para eso pasa a la ventana de “Vista de datos” e ingresa todos los casos/pacientes que desee. Observe como para las variables nominales, cuyos posibles valores codificó en el diseño de la base, aparece una pestaña que le permite seleccionar el valor de la variable, Figura 9. Para las variables con nivel de medida “Escala” no aparece esta pestaña ya que debe ingresar el valor numérico.

Una vez ingresados todos los datos su base de datos está lista para ser analizada. Más importante que poder hacer el análisis estadístico es que usted entienda cómo se crearon las variables y cómo se deben ingresar los datos. Siempre es recomendable buscar a una persona con experiencia en bioestadística para que le ayude con esta parte del proceso de investigación.

Una recomendación final: diseñe una base de datos lo más completa posible ya que en el futuro le puede ser muy útil. Por ejemplo, si usted ingresó las complicaciones u otros hallazgos quirúrgicos, podría hacer otro trabajo de investigación con esta misma base de datos cuyo objetivo puede ser por ejemplo buscar asociación entre algún tipo de prótesis específica (o la longitud de la prótesis) y complicaciones quirúrgicas.

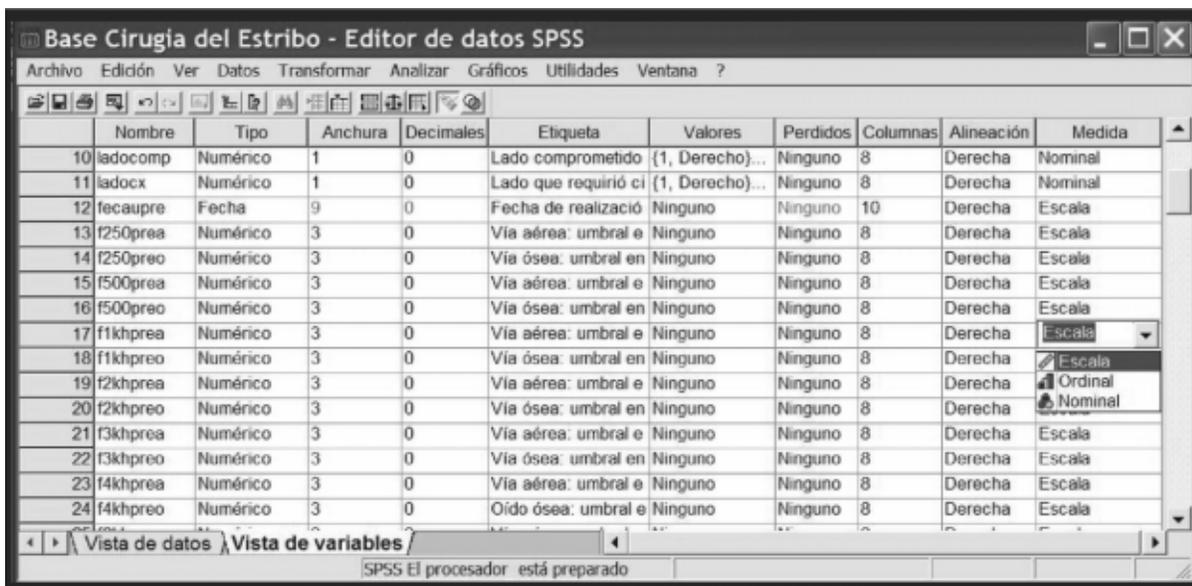
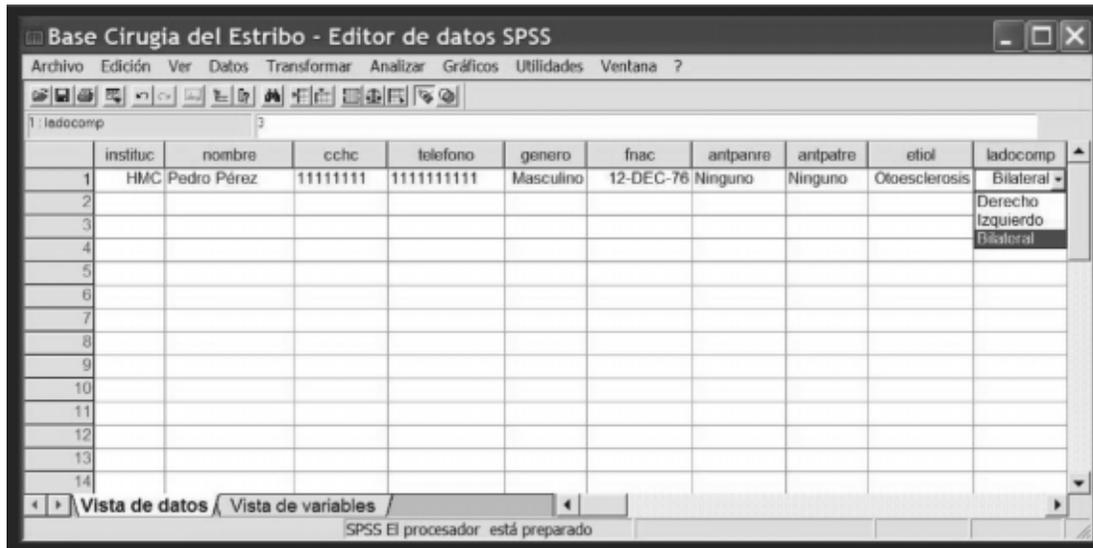


Figura 8. La creación de una variable interválica o proporcional es muy similar a la creación de una variable nominal. Dos diferencias importantes es que no se codifica ningún valor y que en la última columna, “Medida”, debe seleccionar un nivel de medición “Escala”.



**Figura 9.** Una vez creada la base de datos pasa a la vista de datos e ingresa los casos/pacientes que desee. Observe cómo para las variables nominales codificadas aparece una pestaña que permite seleccionar el valor de la variable, esto agiliza el ingreso y evita posibles errores de digitación. El que aparezca la etiqueta de valor o valor numérico de la variable (es decir, que aparezca “Bilateral” o “3”) depende de si tiene o no activo el penúltimo ícono de la fila de íconos que tiene forma de lápiz. Para las variables con nivel de medición “Escala”, debe ingresar el valor numérico correspondiente. Las variables ordinales suelen ingresarse de forma similar a las nominales.

## BIBLIOGRAFÍA

1. Norman GF, Streiner DL. *Bioestadística*. 1st ed. Madrid: Mosby/Doyma Libros. 1996.
2. Committee on Hearing and Equilibrium guidelines for the evaluation of results of treatment of conductive hearing loss. *Otolaryngol Head Neck Surg*. 1995; 113: 186-187.

