

### 3A. DISEÑOS DE ESTUDIOS DE INVESTIGACIÓN.

#### PROPIEDADES DE LAS VARIABLES.

#### DISEÑOS DE ESTUDIOS DE INVESTIGACIÓN

Los principales estudios de investigación biomédica pueden clasificarse según su diseño en dos grandes grupos: estudios de intervención y estudios de observación (Tabla 3.1).

▪ **Estudios de intervención.** El investigador introduce o manipula la variable exposición, modificando el curso natural del proceso. Pueden ser:

- **Estudios Cuasi-Experimentales.** Cuando no hay grupo control prospectivo (comparaciones históricas o pre-post) o cuando la variable exposición no se introduce de forma aleatoria.
- **Estudios Experimentales.** Cuando la intervención se introduce de manera aleatoria entre los grupos activo y control. El *ensayo clínico* es el prototipo de estudio experimental en el ámbito clínico.

**Ensayo clínico.** Tiene un *grupo experimental o índice* (el que recibe la intervención que se evalúa) y un *grupo control o de referencia* (el que recibe una intervención ya aceptada, placebo o nada). Se entiende por *Placebo* la administración de una intervención en el grupo control con las mismas características que la aplicada al grupo experimental pero sin ningún efecto. El ensayo clínico está basado en los principios de aleatorización y enmascaramiento.

- **Aleatorización.** Consiste en que todos los sujetos tienen la misma probabilidad de recibir la intervención y tiene como objeto minimizar los sesgos de confusión y de selección, no solo de las variables de confusión conocidas, sino también de aquellas otras desconocidas para el conocimiento científico. El *análisis por intención e tratar* consiste en comparar los resultados de los grupos formados inicialmente por la aleatorización, independientemente de la intervención recibida finalmente. Sujetos asignados inicialmente a la intervención que posteriormente reciben placebo por diversos motivos, se incluyen en el grupo de intervención en el análisis. Inversamente sujetos asignados por aleatorización al grupo placebo y que acaban recibiendo la intervención, se analizan en el grupo placebo. El motivo de este aparente contrasentido es respetar la aleatorización, que es la medida que aporta mayor inferencia causal.
- **Enmascaramiento.** Técnica que se aplica para reducir los errores en la recogida de información. Puede ser del observado, del observador y del analista, dando lugar al *simple ciego* (observado), *doble ciego* (observado y observador) y *triple ciego* (observado, observador y analista).

Según la fase de desarrollo de una nueva intervención, los ensayos clínicos pueden ser:

- **Fase I.** Estudio de farmacocinética y farmacodinamia en sanos o en ocasiones en pacientes.
- **Fase II.** Ensayo controlado y aleatorizado preliminar sobre la eficacia en la enfermedad diana.
- **Fase III.** Estudio controlado y aleatorizado para evaluar la eficacia y toxicidad de la intervención en las condiciones de uso habituales.
- **Fase IV.** Nuevos usos de una intervención ya autorizada y vigilancia de efectos indeseables.

El ensayo clínico admite varios formatos de diseño:

- **Factoriales.** Se combinan dos o más opciones terapéuticas en el mismo estudio. El diseño más simple es de 2 x 2 en el que se administra un tratamiento A y se combina con otro B, dando lugar a cuatro grupos: A+B+, A+B-, A-B+ y A-B-.
  - **Polietápicos.** Se aplican para tratamientos que se administran secuencialmente. Por ejemplo, se pueden comparar dos regímenes antineoplásicos para la inducción de remisión completa, y en una segunda fase otras dos pautas de antineoplásicos para el mantenimiento de la remisión en los mismos pacientes.
  - **Secuencia cruzada.** Cuando se administra una intervención a un grupo experimental con su grupo control y tras un período de tiempo de lavado, se administra la intervención al grupo control permaneciendo el grupo experimental como control. Se utiliza para valorar tratamientos sintomáticos en enfermedades crónicas recurrentes o recidivantes. Por ejemplo, valorar dos broncodilatadores en el tratamiento del asma.
- **Estudios de observación.** No hay variables manipuladas ni introducidas por el investigador, que no modifica el curso natural de los acontecimientos, actuando como mero notario de lo que sucede. Existen tres diseños epidemiológicos básicos: cohortes, caso-control (ambos son longitudinales) y transversal.

- **Estudios longitudinales.** Las variables exposición y respuesta se miden en momentos diferentes. Pueden ser de cohortes, caso-control y diseños híbridos.

**Estudios de cohortes.** La muestra se selecciona por la variable exposición y es seguida durante un período de tiempo hasta la detección de la respuesta. Según se haya producido la respuesta o no en el momento de realizar el estudio pueden ser *prospectivos* o *retrospectivos*, respectivamente. El ensayo clínico es un estudio experimental con diseño epidemiológico de cohortes.

El estudio del archivo *Factores.xls* sería de cohortes si los 112 casos se seleccionan por ser o no ser obesos y después se investiga el desarrollo de diabetes. Si en el momento de hacer el estudio ya ha transcurrido tiempo suficiente para desarrollar el efecto, y la información del desarrollo de diabetes se obtiene de registros previos existentes (analíticas, historia clínica, informes, etc.) sería un estudio de cohortes retrospectivo. Si por el contrario al iniciar el estudio, todavía no ha transcurrido tiempo suficiente para desarrollar el efecto, y monitorizamos el desarrollo de diabetes, estaríamos ante un estudio de cohortes prospectivo. Si además se registra el momento o tiempo en que se produce la respuesta se obtiene el ritmo en que se presenta el evento y se puede estudiar la dinámica del proceso. Son los estudios de supervivencia con seguimiento teniendo en cuenta el tiempo a un evento.

**Tabla 3A.1.** Diseños de estudios de investigación.

<b>Estudios de intervención</b>	Se introduce o “manipula” la variable exposición
<i>Experimentales</i>	La exposición se introduce de manera aleatoria entre los grupos activo y control
- Ensayo clínico	Estudio experimental en el ámbito clínico basado en la aleatorización y el enmascaramiento
<i>Cuasi-experimentales</i>	La exposición no se introduce de forma aleatoria o no hay un grupo control prospectivo
<b>Estudios de observación</b>	No hay variables introducidas o manipuladas
<i>Longitudinales</i>	Las variables exposición y respuesta se miden en momentos diferentes
Cohortes	La muestra se selecciona por la variable exposición
- Retrospectivos	La respuesta ya ha sucedido en el momento de realizar el estudio
- Prospectivos	La respuesta aún no ha sucedido en el momento de realizar el estudio
Caso-Control	La muestra se selecciona por la variable respuesta
Híbridos	Diseño mezcla de estudios de cohortes y caso-control
- Cohorte y casos	El muestreo de controles se produce al inicio del estudio
- Casos y controles anidados en una cohorte	El muestreo de controles se produce a medida que van apareciendo los casos
<i>Transversales, de corte o de Prevalencia</i>	Las variables exposición y respuesta se miden en el mismo momento

**Estudios caso-control.** A diferencia del estudio de cohortes, el muestreo inicial se realiza por la presencia (casos) o ausencia (controles) de la respuesta y después se evalúa retrospectivamente (hacia atrás) la presencia o ausencia de la exposición. La proporción de controles por caso es variable (1:1, 2:1, 3:1, etc) y es decidida por el investigador. Por encima de 4:1 no aumenta la eficiencia estadística.

El estudio del ejemplo sería caso-control si seleccionamos la muestra por la variable respuesta: un grupo de diabéticos (casos) y otro grupo de no diabéticos (controles). Posteriormente se investiga retrospectivamente la variable exposición: la presencia o ausencia de obesidad.

Pueden ser:

- *De base primaria.* Cuando los casos son los que se producen en una población definida en el tiempo y en el espacio y los controles representan a la colectividad no enferma.
- *De base secundaria.* Cuando los casos y controles proceden de una o varias instituciones, de modo que los casos definen una población hipotética, la que acudiría a dicha institución de ponerse enfermo por el proceso, y los controles representan a esa población ficticia.

Respecto a los estudios de cohortes, presentan las siguientes:

- *Ventajas*: requerir menor tiempo de realización, tener menor coste y mayor eficiencia estadística para el mismo tamaño muestral. Son de utilidad para estudiar respuestas raras o de latencia muy prolongada.
- *Inconvenientes*: tienen mayor posibilidad de sesgos derivados de la dificultad de la elección del grupo control adecuado (sesgo de selección) y de recoger correctamente la exposición conociendo la presencia o ausencia de la enfermedad (sesgo de información).

Los *diseños híbridos* de cohortes y casos y controles intentan eliminar algunos de los inconvenientes de los estudios de casos y controles. Se parte de una cohorte en la que se registran todos los casos que se producen. Pueden ser:

- *Estudio de cohorte y casos*. El muestreo de controles se realiza al principio, antes de producirse los casos.
- *Estudio de casos y controles anidado en una cohorte*. El muestreo de controles se realiza en el tiempo entre los que no tienen el proceso, a medida que se detectan los casos.
- *Estudios transversales*. Las variables exposición y respuesta se miden en el mismo momento. También se denominan estudios de corte o de prevalencia. El estudio del ejemplo sería transversal si la obesidad y la diabetes se evalúan en el mismo momento, por ejemplo en la misma visita de consulta.

El diseño es más simple pero tienen los inconvenientes de tener dificultades para cumplir el principio causal de precedencia temporal (la causa debe preceder al efecto) y mezclar efectos causales y pronósticos, ya que se seleccionan supervivientes.

### Diseños de estudios de grupos independientes o apareados.

Según la relación o dependencia que guarden los sujetos de los grupos a comparar, los diseños de los estudios pueden ser de dos tipos:

- **Grupos independientes**. También llamados diseños, muestras, medidas o datos *entresujetos*. Los valores de las variables de cada grupo proceden de elementos muestrales diferentes. Es decir, los individuos de uno de los grupos son diferentes o independientes de los individuos del otro grupo. Se comparan las respuestas de grupos de sujetos diferentes. Por ejemplo, evaluar la diferencia de glucemia entre el grupo de los obesos y el grupo de los no obesos.
- **Grupos apareados**. También llamados diseños, muestras, medidas o datos *intrasujetos, relacionados, agrupados o dependientes*. Los elementos de los dos grupos son los mismos, aunque en circunstancias distintas. La circunstancia diferenciadora más frecuente suele ser el tiempo, como analizar una variable (p. ej. *glucemia*) antes y después de administrar un tratamiento (p. ej. un nuevo antidiabético). Son los diseños pre-post. El análisis se realiza con pruebas estadísticas que permiten estudiar el cambio que se produce en la respuesta al pasar de un estado inicial a una respuesta final. Tienen la gran ventaja de eliminar la variabilidad entre sujetos, pero al no existir un grupo control disminuye la certeza de atribuir a la intervención el cambio observado en la respuesta, ya que entre ambos momentos pueden haber ocurrido otras circunstancias, además de la exposición, que sean las responsables del cambio observado.

### PROPIEDADES DE LAS VARIABLES

El *Nombre* en general viene determinado en el archivo origen de los datos y rara vez se modifica. La *Etiqueta de la variable*, la *Escala de medida* y el *Formato* son aplicables a todas las variables. La *Etiqueta de los valores* sólo es aplicable a las variables categóricas. *Valores perdidos* sólo es útil cuando se tengan valores perdidos definidos por el usuario. El *Roll o Papel* se ha introducido en la versión 18 y tiene escasa importancia.

#### Nombre

Debe comenzar por una letra (A-Z) o el símbolo @, puede tener hasta 64 caracteres (aunque es recomendable que sea corta, no más de 8-10 caracteres) entre letras, cifras y estos cuatro signos: punto (.) arroba (@) dólar (\$) guión bajo (\_). No permite espacios en blanco. No distingue mayúsculas de minúsculas, por lo que “urico”, “Urico” y “URICO” a pesar de sus diferencias son el mismo nombre. Si está instalado el teclado español admite la letra “ñ” y los acentos por lo que “úrico” y “urico” a pesar de su semejanza son nombres distintos, por lo que es aconsejable no utilizar acentos en el nombre. Están prohibidos los siguientes nombres de variables: los operadores lógicos (AND OR NOT), los operadores relacionales (EQ NE LT LE GT GE) y otras palabras clave (ALL BY TO WITH). En la Tabla 2A.2 se muestran ejemplos de nombres de variables. Corresponde a la columna Etiqueta (Label) de la Vista de variables del Editor de datos. El comando de sintaxis es [VARIABLE LABELS](#).

Tabla 2A.2. Ejemplos de Nombre de variables.	
Nombres Incorrectos	Nombres Correctos
IHTA	HTA1
Fecha Ingreso	FechaIngreso
OR	OR2

### Etiqueta de la variable

Es una breve descripción de la variable. Permite un texto de 260 caracteres aunque sólo se imprimen los 60 primeros y es deseable que sea lo más corto y explicativo posible y utilizando mayúsculas, minúscula y acentos ya que se verá en todas las tablas y gráficos que se realicen. Si la variable es cuantitativa se aconseja incluir al final las unidades de medida entre paréntesis. Corresponde a la columna Etiqueta de la Vista de variables del Editor de datos. El comando de sintaxis es `VARIABLE LABELS`.

### Etiquetas de los valores

Es una breve descripción de cada uno de los códigos de las categorías de las variables categóricas. No es aplicable a las variables cuantitativas. Admite 60 caracteres como máximo aunque se recomienda que sea lo más corto y explicativo posible. También deben ser redactadas correctamente porque aparecerán en las tablas y gráficos de resultados. Corresponde a columna Valores de la Vista de variables del Editor de datos. Comando de sintaxis: `VALUE LABELS`.

### Escala de medida

Corresponde a la columna Medida de la Vista de variables del Editor de datos. El comando de sintaxis es `VARIABLE LEVEL`. Se distinguen tres tipos de variables según la escala de medida:

- **Catagórica Nominal.** Utiliza una escala de medida Nominal. Son datos no métricos, categóricos nominales, sin categorías ordenadas, en los que sólo se pueden establecer relaciones de igualdad y desigualdad entre las categorías de la variable. Pueden tener dos categorías en cuyo caso se denominan variables Dicotómicas o Binarias, como la *Diabetes*, o bien tener más de dos categorías, llamadas Politómicas, como la variable *Hospital*. Una variable binaria siempre tiene escala de medida Nominal, nunca puede ser Ordinal. Se utiliza la palabra clave (`NOMINAL`) entre paréntesis para caracterizar esta escala de medida en la sintaxis.
- **Catagórica Ordinal.** Utilizan una escala de medida Ordinal. Son datos no métricos, categóricos ordinales, con categorías ordenadas, en los que además de relaciones de igualdad-desigualdad, se pueden establecer relaciones de orden entre sus categorías. Por ejemplo la variable categórica *ObesidadOr* con tres categorías según el índice de masa corporal (*Obesidad*, *Sobrepeso* e *Imc normal*) es ordinal porque entre sus tres categorías se puede establecer una gradación, un orden: *Imc normal* < *Sobrepeso* < *Obesidad*. Los intervalos entre los niveles de una variable ordinal no representan diferencias de igual magnitud. Por lo tanto sirve para situar la posición de cada sujeto y no es correcto calcular índices que comportan sumas y restas de sus valores como la media, la desviación estándar, etc. Se utiliza la palabra clave (`ORDINAL`) entre paréntesis para caracterizar esta escala de medida en la sintaxis. Es la escala de medida de una gran cantidad de variables utilizadas en ciencias de la salud como la puntuación de Apgar, la escala de disnea de la NYHA, la escala de la fuerza muscular, etc.
- **Cuantitativas.** Son datos métricos, cuantitativos, medibles o contables. Permiten establecer relaciones de igualdad-desigualdad, de orden y de suma y resta entre los valores de la variable. Se utiliza la palabra clave (`SCALE`) entre paréntesis para caracterizar esta escala de medida en la sintaxis. Las variables que se obtienen con puntuaciones, como los test de puntuación de calidad de vida, son cuantitativas. La variable de la escala visual analógica (VAS) para la puntuación del dolor que va desde 0 (ausencia de dolor) a 10 (máximo dolor posible) también es cuantitativa. El índice de actividad de la enfermedad de Crohn (Tabla 3A.3) es una puntuación de 0 a 600 obtenida con datos clínicos, exploratorios y analíticos, que tiene escala de medida cuantitativa. De esa puntuación se deriva la variable Actividad de la enfermedad (remisión, brote leve, moderado y grave) que es ordinal.

Tabla 3A.3. Crohn Disease Activity Index (CDAI).		
Puntuación	Actividad	
0	Remisión	0-150
.	Brote Leve	150-250
.	Brote Moderado	250-450
600	Brote Grave	450-600

Existen dos subtipos de escala de medida cuantitativa:

- **Interválicas.** Utilizan una escala de medida de intervalo, en la que el cero es una asignación arbitraria. Permite establecer relaciones de suma y resta entre los valores. El ejemplo típico de escala de intervalo es la temperatura: 20°C son 10°C más que 10°C (permite la suma y la resta), pero no es doble de 10°C (no permite

la multiplicación ni la división). El 0 no es una temperatura nula, es una más. El CI con el 100 como valor cero arbitrario también es cuantitativa interválica. También algunas escalas de ansiedad y depresión.

- **Proporcionales.** Utilizan una escala de medida de razón en la que existe un valor de cero objetivo (peso 0 gramos, longitud 0 m, concentración 0 mg/dL). Permite además establecer relaciones de multiplicación y división entre los valores, además de relaciones de suma y resta.

Según pueda ser el número de variaciones entre valores, las variables cuantitativas pueden ser de dos tipos:

- **Discretas.** El número de variaciones entre valores es finito. Son *recuentos*, como el número de factores de riesgo o número de hijos. El número de hijos/2 también, aunque tenga decimales, porque el número de variaciones es finito.
- **Continuas.** El número de variaciones entre valores es infinito. Son *medidas* como la edad, peso, presión arterial, etc. Pueden dar la apariencia de discretas debido a la falta de precisión del instrumento de medida apareciendo: “**Redondeadas**” al centro del intervalo como el peso, la talla, etc, o “**Truncadas**” al inicio del intervalo como el tiempo transcurrido o cumplido (p.ej. la edad).

La asignación de valores en una variable consiste en *clasificar* si se trata de una variable categórica nominal, *ordenar* si es una variable cuantitativa ordinal, *contar* si es una variable cuantitativa discreta o *medir* si es una cuantitativa continua. En la Tabla 3A.4 se muestran los tipos de variables, las escalas de medidas y los tipos de relaciones que permite cada una de ellas. Siempre que se pueda registrar una característica con diferentes escalas de medida es conveniente utilizar la que proporciona mayor información.

Relaciones permitidas	Escala de medida			
	Nominal	Ordinal	Cuantitativa	
			Intervalo	Razón
Igualdad-Desigualdad	X	X	X	X
Orden		X	X	X
Suma y resta			X	X
Multiplicación y división				X
Ejemplo	<i>Hospital</i>	<i>ObesidadOr</i>	<i>Temperatura</i>	<i>Tabaco</i>

### Formato

El Formato define el número de caracteres y la forma en que veremos la variable en la ventana de datos y en los listados. Las variables originales tienen el formato definido en la aplicación origen de los datos (Excel, Access, etc.). Corresponde a las columnas Tipo, Anchura y Decimales de la Vista de variables del Editor de datos. Se recomienda emplear sólo los formatos Numérico, Fecha y Cadena. El resto de formatos son variantes del formato numérico. El comando de sintaxis es [FORMATS](#).

- **Numérico.** Fa.d donde “a” es el número total de caracteres incluido el signo y el punto decimal y “d” es el número de cifras decimales. Este último se puede omitir si no las hay. Ejemplo: F4.1 mostrará los números entre -9.9 y 99.9. El Imc se le asigna formato F4.1, es decir un carácter decimal, otro el punto y otros dos para parte entera y signo que como no puede ser negativo mostrará la parte entera en decenas.
- **Fecha.** Existen varios formatos, pero se recomienda el formato de fecha europea dd-mm-yyyy (EDATE10), por ejemplo 19-11-2006. También están el formato de fecha americana o anglosajón, mm-dd-yyyy (ADATE10), ejemplo 11-19-2006. O con el mes en texto en inglés o formato estándar, dd-mmm-yyyy (DATE11), ejemplo 19-nov-2006. También el formato Fecha general dd-mmm-aaaa hh:mm:ss.ss (DATETIME). Todos los formatos se pueden encontrar con tres tipos de separadores: “-”, “.”, y “/” y con dos tipos de formato de año según se indique el año con 4 o con 2 cifras. Debido a la proximidad del cambio de siglo se deben descartar todos los formatos con el año en dos cifras para evitar problemas, errores o malentendidos.
- **Cadena (string).** An donde “n” es el número máximo de caracteres de la cadena. Se utiliza poco porque se prefiere tener las variables codificarlas como numéricas. Algunas variables como el NIF, las iniciales o el nombre necesariamente tienen que ser cadena.

### Valores perdidos

En ocasiones una variable tiene valores desconocidos. Pueden ser:

- **Valores no aplicables.** Son datos desconocidos que en realidad no existen. Solo pueden estar disponibles en una parte de los sujetos y es el propio diseño de la base de datos el que hace que se muestren como perdidos. Unos son evitables como el número de cigarrillos al día en no fumadores y otros no evitables como edad de la menopausia en hombres. No se dispone de herramientas que permitan diferenciarlos de los verdaderos valores perdidos.
- **Valores perdidos (missing).** Cuando desconocemos el verdadero valor que realmente existe. En la sintaxis se identifican con MISSING. Pueden ser:
  - **Valores perdidos del sistema (system missing).** Son los valores declarados perdidos de forma automática por el sistema por ser: a) datos sin valor o vacíos, bien porque no se ha preguntado o investigado, se han

olvidado o perdido o no se pueden conocer; b) valores incoherentes con el formato como 31-2-2006 en una fecha, o la letra O en un número; c) resultados imposibles de una transformación de datos (p. ej. 24/0). En la sintaxis se identifican con `SYSMIS`. Se debe hacer el máximo esfuerzo para reducirlos al mínimo, puesto que una proporción elevada de valores perdidos puede invalidar los resultados del estudio.

- **Valores perdidos declarados por el usuario o *user missing***. Son valores desconocidos por motivos concretos que nos interesa declarar como tales para que el sistema los identifique y poder manejarlos de forma especial. Es el caso en que se ha preguntado por el valor de una variable, por ejemplo si bebe alcohol, pero no lo sabe o no quiere contestar. Se permite diferenciarlos del resto de valores y de los *system missing*, declarándoles *user missing*, asignándole un código en las propiedades de variables, lo que permite tratarlos de forma particular por diversos procedimientos estadísticos. Además permite declarar varios tipos de valores *user missing*. En las variables numéricas se pueden declarar un máximo de tres valores o bien un valor más un intervalo de valores. En el ejemplo del número de cigarrillos podría ser “98” si no lo sabe y “99” si no lo quiere decir. Corresponde a la columna Perdidos de la Vista de variables del Editor de datos. En la sintaxis se identifican con el comando `MISSING VALUES`.

### Definición de las Propiedades de las variables

En la Tabla 3A.5 se muestran las propiedades de las variables del archivo *Recodificar2.sav*.

Las propiedades de las variables se pueden definir en la *Vista de variables* del *Editor de datos* o con el cuadro *Definir propiedades de variables*.

#### Videotutorial 3A1Propiedades1.avi

A partir del archivo *IntroCurso.sav* se muestra como se pueden definir las propiedades de las variables en la pestaña *Vista de variables* del *Editor de datos*.

- El **Nombre** de la variable se puede escribir o modificar en la columna Nombre.
- El **Formato** se define en el cuadro *Tipo de variable* que aparece al hacer clic sobre las celdas de la columna Tipo. Se recomienda utilizar sólo los formatos Numérico y Fecha para las variables numéricas y Cadena para las variables cadena. El resto de formatos (Coma, Puntos, Notación científica, Dólar y Moneda personalizada) son variantes del formato numérico. Se puede modificar su extensión en las columnas Anchura y Decimales.
- La **Etiqueta de la variable** se escribe en la columna Etiqueta.
- La **Etiqueta de los valores** de las variables categóricas se definen en el cuadro *Etiqueta de valor* que se abre haciendo ‘clic’ en las celdas de la columna Valores. Se debe escribir el valor y la etiqueta del valor, ambos sin comillas, y pulsar Añadir.
- Los **User missing** se definen en el cuadro *Valores perdidos* que aparece al hacer clic en las celdas de la columna Perdidos. Permite definir hasta tres valores discretos o bien (solo en variables numéricas) un rango de valores y un valor discreto.
- La **Escala de medida** se define en la columna Medida. Puede ser Escala, Ordinal o Nominal.

También se muestra la tabla de frecuencias de las variables *Tabaco* y *Tab* para demostrar la **importancia de la definición de las propiedades de las variables**. Ambas variables son idénticas salvo que *Tabaco* tiene definidas correctamente todas sus propiedades mientras que *Tab* no las tiene definidas. Si se comparan las tablas de frecuencias de ambas, se observan no sólo diferencias estéticas en la presentación de la tabla. También hay diferencias importantes en los porcentajes válidos de fumadores y no fumadores.

#### Videotutorial 3A2Propiedades2.avi

Se muestra la definición de las propiedades de las variables del archivo *Recodificar2.sav* con el cuadro *Definir propiedades de variables*.

- **Nombre**: aparece en Variable actual: al marcar la variable.
- **Etiqueta de la variable**: se escribe en Etiqueta.
- **Escala de Medida**: se selecciona Nominal, Ordinal o Escala en Nivel de medida:. El botón Sugerir ayuda en la elección.
- **Formato**: se especifica en:
  - Tipo: se recomienda utilizar sólo Numérico y Fecha para las variables numéricas y el formato Cadena para las variables cadena. El resto de formatos son variantes de número.
  - Anchura: escribir el número total de caracteres incluido decimales, separador decimal y signo para las variables numéricas, y el número de caracteres máximo de las variables cadena.
  - Decimales: especificar el número de decimales en las variables numéricas.
- **Etiqueta de valores**: se escriben en la columna Etiqueta de la Rejilla de etiquetas de valores.
- **Valores perdidos**: en la columna Perdido de la Rejilla de etiquetas de valores se marcan las casillas de los valores que se declaran *user missing*. Es el caso del valor 99 de *Tabaco* y del valor 9 de *TabacoBi* y *TabacoOr*.

**Tabla 3A.5.** Propiedades de las variables del estudio sobre factores de riesgo cardiovascular.

Nombre	Etiqueta de la variable	Formato			Escala de medida	Etiquetas de los valores
		Tipo	Anchura	Decimal		
Caso	Caso	Número	3	0	Nominal	
FNac	Fecha de nacimiento	Fecha europea	10	0	Escala	
FIntro	Fecha de introducción	Fecha europea	10	0	Escala	
Edad	Edad (años)	Número	3	0	Escala	
Sexo	Sexo	Número	1	0	Nominal	0 Mujer 1 Varón
Hospital	Hospital	Número	1	0	Nominal	1 Ramón y Cajal 2 Príncipe de Asturias 3 Gregorio Marañón 4 Clínico San Carlos 5 12 de Octubre
Peso	Peso (kg)	Número	5	1	Escala	
Talla	Talla (cm)	Número	3	0	Escala	
Imc	Índice de masa corporal (kg/m <sup>2</sup> )	Número	4	1	Escala	
Obesidad	Obesidad	Número	1	0	Nominal	0 No 1 Sí
ObesidadOr	Nivel de masa corporal	Número	1	0	Ordinal	0 Imc Normal 1 Sobrepeso 2 Obeso
Glucosa	Glucosa plasmática (mg/dl)	Número	3	0	Escala	
Diabetes	Diabetes mellitus	Número	1	0	Nominal	0 No 1 Sí
Tabaco	Tabaquismo (c/d)	Número	3	0	Escala	99 No contesta*
TabacoBi	Fumador	Número	1	0	Nominal	0 No 1 Sí 9 No contesta*
TabacoOr	Nivel de tabaquismo	Número	1	0	Ordinal	0 Nulo 1 Medio 2 Alto 9 No contesta*
Pas	Presión arterial sistólica (mmHg)	Número	3	0	Escala	
Pad	Presión arterial diastólica (mmHg)	Número	3	0	Escala	
Pam	Presión arterial media (mmHg)	Número	5	1	Escala	
Hta	Hipertensión arterial	Número	1	0	Nominal	0 No 1 Sí

\* Corresponde a valor perdido definido por el usuario (*user missing*)

#### Recomendaciones prácticas:

- Es recomendable desmarcar las casillas Limitar número de casos explorados a: y Limitar número de valores mostrados a: \_\_, para que la Rejilla de etiquetas de valores muestre en la columna Recuento, la distribución de frecuencias completa de la variable y permita detectar valores erróneos o fuera de rango.
- Se puede pasar de una variable a otra para definir las propiedades de todas las variables y al final se sale con Pegar. El programa va reteniendo todas las propiedades definidas. En la ventana de sintaxis se escriben todas las propiedades de todas las variables que se hayan definido en el cuadro.
- Se puede evitar la repetición de propiedades similares en varias variables. Si se quiere aplicar las mismas propiedades a otras variables, pulsando el botón A otras variables... se abre la ventana *Aplicar etiquetas y medida a* que permite seleccionar la o las variables a las que se transferirán las propiedades de la variable activa. Si se va a iniciar la definición de las propiedades de una variable y son idénticas a las de otra variable previamente

definida, pulsando el botón *De otra variable...* se abre la ventana *Aplicar etiquetas y medida de* que permite seleccionar la variable cuyas propiedades se van a transferir a la variable actualmente activa.

- Si la escala de medida ya esta definida como *Escala* y se mantiene como tal, no escribe su instrucción de sintaxis. La *Nominal*, la *Ordinal* y la *Escala* definida de nuevo (a partir de una *Nominal*, la *Ordinal* previa) sí las escribe.

En la ventana de sintaxis se escriben las propiedades de las variables, variable a variable y propiedad a propiedad. Es muy útil tenerlo así porque facilita añadir propiedades no definidas por olvido o cambiar las definidas erróneamente. En el videotutorial no se han definido la etiqueta de *Caso*, ni la escala de medida (Nominal) de *Caso* y *Hospital*, ni la etiqueta del valor 9 de *Tabaco* y *TabacoBi*, ni las de *TabacoOr* (0=Nulo, 1=Medio, 2=Alto y 9=No contesta), además el formato de *Edad*, *Talla* y *Glucosa* es F3 en vez de F1. Se ha corregido en la sintaxis y se ha vuelto a ejecutar. El resultado se guarda de nuevo en el archivo *Factores.sav* que será el que utilizaremos en el resto del curso. El archivo de sintaxis se guarda con el nombre *Propiedades2.sps*.

El cuadro *Variables* presenta a la izquierda el listado de las variables y a la derecha las propiedades de la variable seleccionada. Se utiliza para consultar las propiedades de las variables y para “Pegar” el nombre de la variable en el punto de la ventana de sintaxis donde esta el cursor, sin necesidad de escribirlos evitando posibles errores ortográficos.