

Capítulo X

Tablas de contingencia con mas de dos variables

1. Objetivos didácticos del capítulo

La explicación de las tablas de contingencia de dos variables –relaciones bivariantes– da paso a las relaciones *múltiples*, a las tablas de contingencia donde intervienen tres y más variables. Éstas pueden participar formando tablas de dos dimensiones donde cada fila o columna está formada por una o varias variables, o tablas con más de dos dimensiones.

En el primer caso, las tablas de dos dimensiones son elaboradas uniendo variables que recogen las *respuestas múltiples* de determinadas preguntas, concretamente aquellas con más de una respuesta (preguntas multirrespuesta). El cuestionario utilizado a lo largo del libro (presentado en la sección 6 del capítulo segundo) dispone de un gran número de preguntas de este tipo: ¿cuáles son las *dos* situaciones que mejor definen tu actividad en tu tiempo libre (pregunta 3); ¿qué asignaturas proponen libros de lectura obligatoria (pregunta 7); ¿qué periféricos o dispositivos tienen los ordenadores de tu hogar (pregunta 17a); etc. Tal y como señalamos en el capítulo siete se trata de unas preguntas muy utilizadas en la investigación con encuesta, y con un tratamiento diferente al resto de preguntas del cuestionario.

Las tablas de más de dos dimensiones relacionan más de dos variables eliminando la influencia de la última; la tercera en el caso de tres dimensiones, la cuarta cuando se trata de tablas cuatridimensionales, etc. Dedicaremos el segundo apartado a la interpretación de tablas de tres dimensiones, y en el siguiente se explicarán las diferentes situaciones que genera la *última variable* de la tabla de contingencia. Realizaremos, de este modo, un ligero acercamiento a las relaciones múltiples entre variables, al análisis multivariable, al que nos referimos someramente en el capítulo II (apartado 2.5).

Como en el resto de capítulos, la explicación se ha llevado a cabo utilizando ejemplos realizados con el archivo de datos obtenido del cuestionario presentado en el segundo capítulo, sección 7 (ENCUESTAS ESTUDIANTES 2003_04.SAV).

2. Tablas de contingencia de respuestas múltiples categóricas

Comenzaremos la explicación de las tablas de contingencia con varias variables analizando las preguntas que presentan categorías de respuestas no excluyentes, aquellas donde los entrevistados pueden seleccionar varias de las alternativas posibles. Recordemos que éstas fueron definidas en el capítulo siete (apartado 7.4 y 7.5) como preguntas *multirespuesta* o *respuesta múltiple*, diferenciando entre preguntas multirespuesta categóricas y dicotómicas. La propia temática del capítulo siete, dedicado al análisis univariable, limitó la exposición de estas preguntas a la *unión* de preguntas (definición de conjuntos) y a la obtención de frecuencias. En el presente apartado daremos un paso más y relacionar las preguntas múltiples con otras variables.

Antes de proceder con el cruce de tablas debe realizarse un análisis de las frecuencias, tal y como hemos mantenido a lo largo de todo el libro. Con el fin de simplificar la exposición volveremos de nuevo a la variable *situaciones que mejor definen la actividad durante el tiempo libre*, pregunta 3, y cuyas respuestas se recogen en las variables v03 y v04. Su distribución de frecuencias fue presentada en la tabla 7.1, y ahí señalamos que para el 66,3% de los entrevistados la situación que mejor define su tiempo libre es *estar con la gente*, un 36,8% señaló *dedicarme tranquilamente a mis cosas y aficiones*, y otro 22,1% mostró su preferencia por *hacer muchas cosas*.

Una vez interpretadas las frecuencias llega el momento de proceder con el cruce de tablas. Utilizaremos, para ello, el menú *Analizar⇒Respuestas múltiples⇒Tablas de contingencia...* que da lugar al cuadro de diálogo de la figura 10.1. En la ventana superior izquierda aparecen todas las variables del cuestionario, y en la inferior izquierda la agrupación de variables que forman las preguntas multirespuesta; concretamente las variables *situaciones que mejor definen la actividad durante el tiempo libre* (pregunta 3) y *número de dispositivos en el ordenador* (pregunta 17a); que fueron definidas en el séptimo capítulo¹³². En las ventanas centrales de la figura 10.1 se situarán las variables a *cruzar*, aquellas que formarán la tabla de contingencia.

En este caso nuestro objetivo es conocer las *situaciones que mejor definen la actividad durante el tiempo libre* diferenciando el sexo del entrevistado con el fin de descubrir si los hombres y las mujeres presentan diferencias en las situaciones que mejor definen su actividad durante el tiempo libre; continuando así con uno de los temas

132. Pudiera suceder que una vez analizadas las frecuencias de todo el cuestionario nos interesara volver a trabajar con estas variables en una sesión posterior. En este caso, antes de realizar la tabla de contingencia será necesario proceder de nuevo con la *Definición de conjuntos* mediante el menú *Analizar⇒Respuestas múltiples⇒Definir conjuntos...* Recuérdese que mientras no hay *conjuntos* (de variables) *definidos* la opción *Tablas de contingencia de respuestas múltiples* aparece inactiva.



Figura 10.1. Cuadro de diálogo para realizar una tabla de contingencia con variables de respuesta múltiple (categóricas).

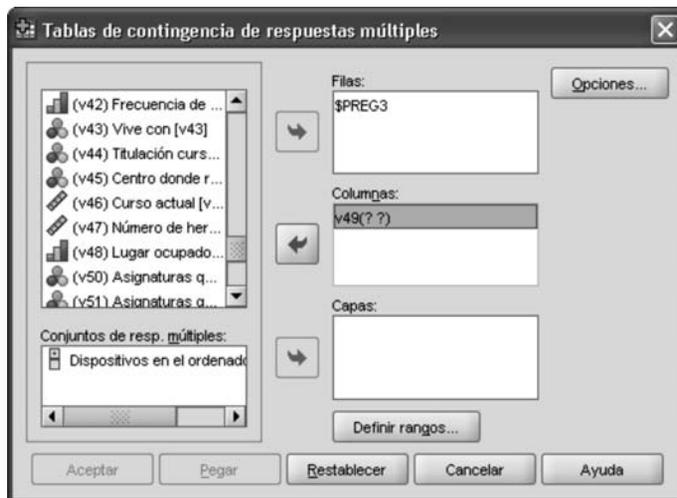


Figura 10.2. Tabla de contingencia de respuestas múltiples (categóricas) con dos variables.

planteados en el capítulo IX. Para ello será necesario desplazar *Situaciones definen...* a la ventana de las filas y el sexo del entrevistado (V49) a la ventana de las columnas (figura 10.2). En este momento el botón *Definir rangos...* cambia de color y, una vez pulsado, aparece un cuadro de diálogo para definir el valor mínimo y máximo de la variable situada en columnas (figura 10.3).

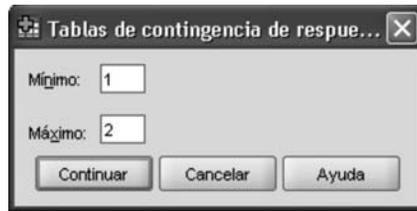


Figura 10.3. Tabla de contingencia con variables de respuesta múltiple (categóricas): definir rangos variable en columnas.

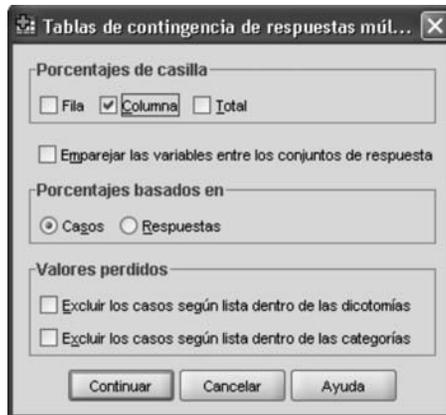


Figura 10.4. Tabla de contingencia con variables de respuesta múltiple: Opciones...

El botón *Continuar* del cuadro de diálogo de la figura 10.3 introduce estos valores en el cuadro de diálogo principal. Definida la tabla, recomendamos pulsar el botón *Opciones...* (figura 10.2) que muestra los elementos que aparecerán en el interior de las celdillas (figura 10.4). Nuestro objetivo, que recordemos consistía en conocer si los hombres y las mujeres presentan diferencias en las situaciones que mejor definen su actividad durante el tiempo libre, nos ha llevado a solicitar los porcentajes de columna. Del resto de elementos incluidos en este cuadro de diálogo la opción más importante es “*Porcentajes basados en*”, que lo dejaremos como está por la explicación realizada en el capítulo VII. Lo mismo hacemos con los valores perdidos¹³³. Pulsar el botón *Continuar* da paso al cuadro de diálogo principal que, tal y como se muestra en

133. Marcar la opción *Excluir los casos según lista dentro de las dicotomías/categorías* implica la eliminación de los casos que presentan valores perdidos en cualquiera de las variable que forman parte del conjunto de respuestas múltiples.

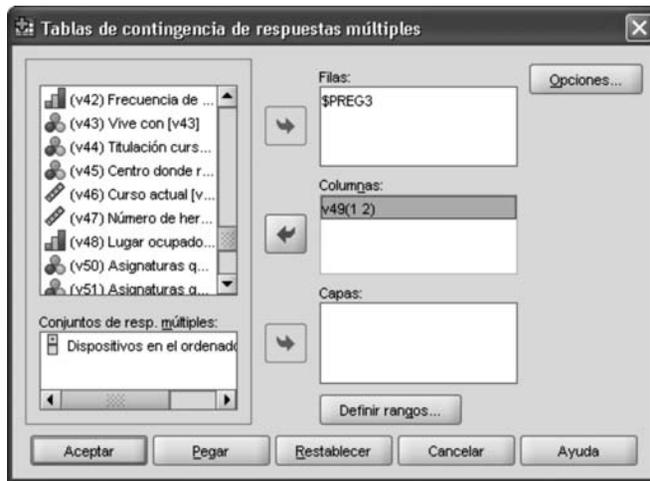


Figura 10.5. Tabla de contingencia con variables de respuesta múltiple.

la figura 10.5, aparece con el rango de la variable situada en columnas. Tras pulsar el botón *Aceptar* obtenemos los resultados que se muestran en la tabla 10.1.

En la parte superior de la tabla 10.1 se presenta, al igual que en las tablas de contingencia vistas en el capítulo anterior, el resumen de los casos; donde se indica el número de casos válidos y perdidos. 191 casos son válidos en este ejemplo. A continuación aparece la tabla de contingencia con la variable sexo en columnas y las situaciones que definen el tiempo libre en filas. Aunque en este caso la parte final de la etiqueta define con precisión el origen de esta variable (“v3+v4”), la nota (a) junto a ésta indica que se trata de una agrupación, de una variable formada por la unión de varias. Dentro de las celdillas aparece el recuento y los porcentajes verticales (“% dentro de v49”). Al pie de la tabla se indica que los porcentajes y los totales se basan en los encuestados; en los casos (tal y como se definió en la figura 10.4).

La columna de la derecha (rotulada con *total*) es la suma de las variables v03 y v04, es decir, la distribución univariante con las dos variables agrupadas que se mostró en la tabla 7.1. El análisis de las celdillas interiores debe hacerse tal y como señalamos en el capítulo anterior, en el apartado *análisis del interior de la tabla* (9.4). Centraremos la explicación en los porcentajes, en la medida que la interpretación de los números absolutos es complicado por la diferencia entre el número de hombres y mujeres entrevistadas (71 y 120 respectivamente). Se trata de porcentajes de columna, calculados dividiendo el número de casos de la celdilla entre el número de respuestas obtenido por cada una de las categorías de la variable columna (v049)¹³⁴. Ahora bien, a

134. $(6 / 71) * 100 = 8,5$; $(12 / 71) * 100 = 16,9$; $(10 / 71) * 100 = 14,1$;...

Resumen de los casos

	Casos					
	Válidos		Pérdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
\$PREG3*v49	191	100,0%	0	,0%	191	100,0%

Tabla de contingencia \$PREG3*v49

			(v49) Genero		Total
			Hombre	Mujer	
Situaciones definen tiempo libre (v3+v4)(a)	Pasarlo bien sin hacer nada	Recuento % dentro de v49	6 8,5%	8 6,7%	14 7,3%
	Hacer muchas cosas	Recuento % dentro de v49	12 ¹³⁵ <u>16,9%</u>	30 25,0%	42 22,0%
	Dedicarme a las personas más queridas	Recuento % dentro de v49	10 14,1%	26 21,7%	36 18,8%
	Hacer cosas de mi trabajo que tengo pendientes	Recuento % dentro de v49	6 <u>8,5%</u>	20 16,7%	26 13,6%
	Descansar, recuperar fuerzas	Recuento % dentro de v49	17 23,9%	24 20,0%	41 21,5%
	Estar con la gente, charlar, tratar a los amigos	Recuento % dentro de v49	42 <u>59,2%</u>	84 70,0%	126 66,0%
	Aburrirme	Recuento % dentro de v49	6 8,5%	0 ,0%	6 3,1%
	Pensar, meditar	Recuento % dentro de v49	4 5,6%	4 3,3%	8 4,2%
	Dedicarme tranquilamente a mis cosas, aficiones, deportes	Recuento % dentro de v49	34 47,9%	36 <u>30,0%</u>	70 36,6%
	Total	Recuento	71	120	191

Los porcentajes y los totales se basan en los encuestados.
a Agrupación

Tabla 10.1. Tabla de contingencia Pregunta 3 y sexo (v49); porcentajes basados en casos.

135. Se ha utilizado el subrayado para identificar aquellos porcentajes inferiores al promedio (columna total); mientras que la negrilla indica porcentajes notablemente superiores al promedio.

diferencia de las tablas de contingencia vistas en el capítulo anterior en este caso los porcentajes de columna no suman 100 en sentido vertical puesto que cada entrevistado ha proporcionado más de una respuesta, dos en este ejemplo¹³⁶.

Pese a este hecho, los porcentajes se interpretan exactamente igual como señalamos en el capítulo anterior: el 59,2% de los hombres señalan que la situación que mejor define su tiempo libre es *estar con la gente, charlar, tratar a los amigos*, un 47,9% de los hombres señala *dedicarme tranquilamente a mis cosas, aficiones y deportes*, el 23,9% muestra su preferencia por *descansar, recuperar fuerzas, ...*, por señalar las situaciones que mejor definen el tiempo de ocio (las categorías con más elecciones, con mayores porcentajes)¹³⁷.

En el capítulo anterior indicamos que la gran riqueza del análisis de tablas de contingencia es poder comparar *transversalmente* estos porcentajes, cotejando *linealmente* las celdas de las distintas columnas. En tal caso deberíamos resaltar únicamente los porcentajes donde existen diferencias importantes entre los hombres y las mujeres, comparando cada categoría con el total. Esto sucede, en el caso de los hombres, en *dedicarme tranquilamente a mis cosas*, diferencia de 11,3 puntos; *estar con la gente, charlar con los amigos*, diferencia de -6,8 puntos, *aburrirme*, diferencia 5,4; *hacer cosas del trabajo que tengo pendientes y hacer muchas cosas*, diferencias de -5,1 puntos y, por último, *dedicarme a las personas más queridas*, con una diferencia de -4,7 puntos. De este modo podríamos decir que las situaciones que mejor definen el tiempo libre de los hombres son *dedicarme tranquilamente a mis cosas* y *aburrirme*; mientras que las mujeres apuestan en mayor medida por *estar con la gente, charlar con los amigos* (diferencia 4 puntos), *hacer cosas del trabajo que tengo pendientes* (diferencia 3,1 puntos) y *hacer muchas cosas* (diferencia 3 puntos). Es importante precisar, en el caso de las mujeres, que las diferencias son muy pequeñas, menores de los valores mínimos recomendados señalados en el capítulo anterior (recordar nota a pie número 118). También existen, como no, situaciones de tiempo libre que no varían según el sexo del entrevistado; esto es, que son igualmente elegidas por hombres y por mujeres. Se trata, concretamente, de pasarlo bien sin hacer nada y descansar.

Buscando *fixar* los conocimientos aprendidos en esta sección, antes de considerar nuevos contenidos, proponemos un par de ejercicios utilizando la investigación sobre *Vida Cotidiana*.

136. Como vimos en el capítulo siete los porcentajes sumarían 100 si el cálculo se hubiera realizado según el *porcentaje de respuestas*.

137. En el capítulo VII recomendamos leer un artículo de prensa para observar cómo se ha presentado una pregunta similar en un medio de comunicación. Recomendamos, de nuevo, volverlo a leer.

- Considerando los objetivos más importantes a solucionar en España¹³⁸ (pregunta 6, variables a27, a29 y a31) y en el mundo (pregunta 8, variables a39, a41 y a43), ¿presentan variación en función del sexo? ¿Y en función de la edad?
- Describe los principales motivos de discusiones entre los españoles (pregunta 29, variables b43, b45, b47 y b49). No se trata especificar *con quién* se discute *sobre qué*, sino en señalar los motivos de discusión sin diferenciar la persona con la que se discute. Estos motivos de discusión, ¿presentan variación en función de la edad? ¿Y según la zona geográfica?

Los tipos de personas –al margen de los familiares– con los que los entrevistados se relacionan habitualmente (pregunta 34, variables b63, b64 y b65¹³⁹), ¿presentan variación en función de la edad de los entrevistados? Señala la *asociación* entre cada grupo de edad y los distintos *tipos de personas* que se presentan en la pregunta 34.

3. Tablas de contingencia de respuestas múltiples dicotómicas

El segundo tipo de preguntas multirespuesta son conocidas como dicotómicas y se diferencian de las anteriores en que consideran tan sólo una categoría de la variable. En el séptimo capítulo, apartado 7.5, se analizó la pregunta 17a que recordemos preguntaba a los entrevistados con ordenador por la presencia de una serie de dispositivos. Las respuestas eran codificadas con el valor 1 cuando se contaba con ese equipamiento, dejándose sin responder en caso contrario¹⁴⁰. En la tabla 7.2 vimos que el 94,4% de los entrevistados con ordenador disponían de también de impresora, un 92,2% de lectora de CD, un 91,1% de altavoces...

138. Recuérdese la recomendación realizada en el capítulo VII, concretamente en la nota a pie número 73 (página 177), donde se aconsejaba tener cuidado con la interpretación de esta pregunta puesto que las tres variables no pueden *agregarse* de esta forma al estar ordenadas por orden de importancia. En aquel momento se señaló que para la realización del ejercicio se considerará que se tratan de tres *objetivos* igualmente importantes, no ordenados según la mayor o menor trascendencia. Esto es, olvidando que la variable a39 recoge el objetivo más importante, la a41 el segundo más importante, y la a43 el tercero más importante.

139. Considerando que se trata de tres relaciones *con igual importancia*, esto es, sin considerar el *orden* según la mayor o menor frecuencia; al igual que en el ejercicio anterior.

140. Los que no disponen de cada equipamiento no responden la variable (ver cuestionario en el apartado 6 del capítulo II) pero –por la necesidad de introducir todos los valores de los entrevistados– en la codificación del cuestionarios (sección 6 del capítulo III) recomendamos codificar esta situación con el valor 0. Ver libro de códigos de la sección 9 del capítulo III.

En este momento nos interesa conocer si el equipamiento del ordenador varía según se tenga –o no– conexión a Internet desde el hogar (pregunta 17b, v49); planteando la hipótesis que *los ordenadores conectados precisarán de un mayor equipamiento*. Solicitando las frecuencias de la variable v30¹⁴¹ se aprecia que 140 estudiantes disponen de conexión a Internet desde su hogar. De los 191 entrevistados, el porcentaje de estudiantes con conexión a Internet en el hogar es del 77,8%.



Figura 10.6. Tabla de contingencia de respuestas múltiples (dicotómicas) con dos variables.

Para conocer si el equipamiento del ordenador cambia según la conexión a Internet desde el hogar será necesario realizar una tabla de contingencia entre ambas variables. Utilizaremos para ello el menú *Analizar*⇒*Respuestas múltiples*⇒*Tablas de contingencia...* colocando la variable *Número de dispositivos* en filas, y el *acceso a internet desde tu hogar* (v30) en columnas. A continuación definimos el rango de esta última, que identifica la categoría *Si* con el uno y el *No* con el dos¹⁴², obteniendo la figura 10.6. Pulsando el botón *Opciones* seleccionamos, al igual que en el ejemplo anterior, los porcentajes de columnas y dejamos *por defecto* los porcentajes basados en casos.

141. Menú *Analizar*⇒*Estadísticos Descriptivos*⇒*Frecuencias*⇒v30. En el apartado 2 no se han solicitado las frecuencias del sexo porque ya se conocen del capítulo nueve.

142. Esta variable tenía también los valores 90 y 98 (ver libro de códigos en el apartado 9 del capítulo III). Al seleccionar el 1 y el 2 los valores superiores no son considerados.

Resumen de los casos						
	Casos					
	Válidos		Péridos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
\$n°_dispo*v49	180	94,2%	11	5,8%	191	100,0%

Tabla de contingencia \$n°_dispo*v30

			(v30) Acceso a Internet desde el hogar		Total
			Si	No	
Dispositivos en el ordenador(a)	(v22) Dispositivos IMPRESORA	Recuento % dentro de v30	132 94,3%	38 95,0%	170 94,4%
	(v23) Dispositivos MODEM	Recuento % dentro de v30	128 91,4%	6 <u>15,0%</u>	134 74,4%
	(v24) Dispositivos ALTAVOCES	Recuento % dentro de v30	136 97,1%	28 <u>70,0%</u>	164 91,1%
	(v25) Dispositivos WEBCAM	Recuento % dentro de v30	36 25,7%	0 ,0%	36 20,0%
	(v26) Dispositivos LECTORA CD	Recuento % dentro de v30	134 95,7%	32 <u>80,0%</u>	166 92,2%
	(v27) Dispositivos GRABADORA CD	Recuento % dentro de v30	60 42,9%	8 <u>20,0%</u>	68 37,8%
	(v28) Dispositivos LECTORA DVD	Recuento % dentro de v30	90 64,3%	14 <u>35,0%</u>	104 57,8%
	(v29) Dispositivos GRABADORA DVD	Recuento % dentro de v30	24 17,1%	0 ,0%	24 13,3%
Total		Recuento	140	40	180

Los porcentajes y los totales se basan en los encuestados.
a Agrupación de dicotomías. Tabulado el valor 1.

Tabla 10.2. Tabla de contingencia "Número de Dispositivos" (Pregunta 17a) y conexión a internet en el hogar (v30); porcentajes basados en casos.

Los resultados se muestran en la tabla 10.2, donde se aprecia que de los 191 casos analizados el 94,2% han sido considerados como válidos, mientras que 11 –que suponen un 5,8% de los entrevistados– son definidos como perdidos. Se trata de 8 personas que no cuentan con ordenador (y que por lo tanto no se les ha preguntado si tienen o no conexión) y otras 3 que no la han respondido. La nota (a) junto al nombre de la variable en filas indica que se trata de una agrupación de dicotomías (pregunta multirespuesta dicotómica) realizada considerando el valor 1. Al pie de la tabla puede apreciarse también que los porcentajes se basan en los casos, en los encuestados.

La interpretación del interior de la tabla desvela, en primer lugar, que los entrevistados que disponen de conexión a internet en su hogar se caracterizan por tener ordenadores con más equipamientos, destacando fundamentalmente la presencia de modem (diferencia respecto al total de 17 puntos), altavoces (diferencia de 6 puntos), y lector de DVD (diferencia de 6,5 puntos). En segundo lugar puede apreciarse que la práctica totalidad de ordenadores *conectados* disponen de altavoces (únicamente el 2,9¹⁴³% no lo tienen), y 9 de cada diez de modem. Resulta sorprendente, por otro lado, que seis personas que no disponen de conexión a internet en el hogar cuenten con ordenador con modem; lo que puede llevarnos a dudar si se ha comprendido bien la pregunta. Un análisis exhaustivo de esta pregunta –que no detallaremos aquí por motivos de espacio y por alejarse de nuestros propósitos– nos ha llevado a la conclusión que se trata de ordenadores con modem incorporado. Es importante señalar también que ninguno de los entrevistados *no conectados* dispone de WebCam y grabadora de DVD; lo cual nos lleva a pensar que podría tratarse de ordenadores con más antigüedad que los conectados a Internet. La impresora, sin embargo, se distribuye por igual entre los que están conectados y no conectados a Internet.

Se trata, como hemos visto, de una interpretación muy similar a la realizada con las preguntas multirespuesta categóricas. Con el fin de añadir una mayor complejidad a la explicación –y así llegar a comprender mejor la diferencia entre los porcentajes basados en casos y en respuestas– repetiremos el análisis anterior solicitando *Porcentajes basados en respuestas*, modificando una opción del cuadro de diálogo *Opciones* (figura 10.7).

Los resultados se muestran en la tabla 10.3; si bien se ha eliminado el resumen de los casos puesto que es el mismo que el presentado en la tabla 10.2. Obsérvese que al pie de la tabla se indica que los porcentajes y totales se basan en las respuestas. Aparte del cambio en la magnitud de los coeficientes, que se han reducido notablemente y suman 100 en sentido vertical, se aprecia también algunas diferencias cuando se comparan estos resultados con la tabla 10.2. Resulta sorprendente, en primer lugar,

143. $100 - 97,1 = 2,9$

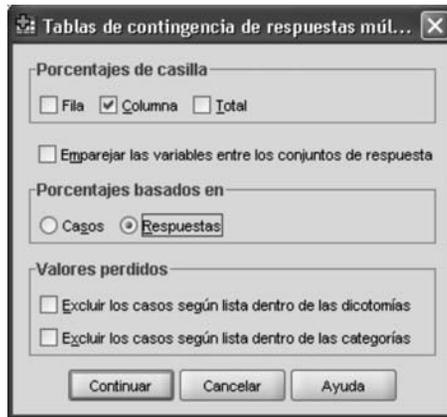


Figura 10.7. Tabla de contingencia con variables de respuesta múltiple, opciones: “Porcentajes basados en Respuestas”.

la magnitud de los que poseen impresora, un 12% más entre los que no tienen acceso a Internet desde el hogar (recordemos que en la tabla 10.2 la diferencia era –tan sólo– de un 0,7%). Otro aspecto interesante es la posesión de altavoces, que en esta tabla es ligeramente superior entre los que no están conectados a Internet; justo lo contrario que desvelaba la tabla anterior. Esta misma situación se produce con el lector de CD. A la hora de explicar el motivo de esta situación tendríamos que destacar el diferente cociente en ambas tablas: la tabla 10.2 divide las frecuencias observadas de cada celdilla entre el total de entrevistados (140 con Internet y 40 sin este equipamiento); mientras que en la tabla 10.3 las frecuencias observadas son divididas entre el total de respuestas, que ascienden a 740 en el caso de los conectados a Internet y a 126 en los no conectados¹⁴⁴.

Esta diferencia está originada por el *objeto* interpretado; por la información que contiene la tabla. Está claro que la tabla 10.2 (basada en casos) se refiere al equipamiento presente en los ordenadores de los entrevistados. Ahora bien, ¿qué información presta la tabla 10.3?, basada en respuestas. En realidad lo que está mostrando es *como se reparte el equipamiento* en la población analizada. Esta tabla indica que –en total– se cuenta de 866 dispositivos en el ordenador (respuestas), y de éstos 740 se encuentran en los ordenadores conectados a Internet desde el hogar y 126 en los no conectados. Del total de dispositivos en los ordenadores *conectados* los más extendidos son

144. La gran diferencia entre los cocientes utilizados en una y otra tabla *favorece* a los valores más pequeños, puesto que en la primera tabla la relación tiene/no tiene Internet es de 3 a 1 (exactamente 3,15; 140 / 40); mientras que la segunda presenta una relación de 5,9 a 1 (740 / 126 = 5,87).

los altavoces (18,4%), lector de CD (18,1%) y modem (17,3%). Entre los ordenadores no conectados destaca la presencia de impresoras (30,2%), lectoras de CD (25,4%) y altavoces (25,4%).

Tabla de contingencia \$n^{\circ}\$_dispo*v30

			(v30) Acceso a Internet desde el hogar		Total
			Si	No	
Dispositivos en el ordenador(a)	((v22) Dispositivos en el ordenador: IMPRESORA	Recuento % dentro de v30	132 17,8%	38 30,2%	170 19,6%
	(v23) Dispositivos en el ordenador: MODEM	Recuento % dentro de v30	128 17,3%	6 4,8%	134 15,5%
	(v24) Dispositivos en el ordenador: ALTAVOCES	Recuento % dentro de v30	136 18,4%	28 22,2%	164 18,9%
	(v25) Dispositivos en el ordenador: WEBCAM	Recuento % dentro de v30	36 4,9%	0 ,0%	36 4,2%
	(v26) Dispositivos en el ordenador: LECTORA CD	Recuento % dentro de v30	134 18,1%	32 25,4%	166 19,2%
	(v27) Dispositivos en el ordenador: GRABADORA CD	Recuento % dentro de v30	60 8,1%	8 6,3%	68 7,9%
	(v28) Dispositivos en el ordenador: LECTORA DVD	Recuento % dentro de v30	90 12,2%	14 11,1%	104 12,0%
	(v29) Dispositivos en el ordenador: GRABADORA DVD	Recuento % dentro de v30	24 3,2%	0 ,0%	24 2,8%
Total		Recuento	740	126	866

Los porcentajes y los totales se basan en las respuestas.
a Agrupación de dicotomías. Tabulado el valor 1.

Tabla 10.3. Tabla de contingencia "Número de Dispositivos" (Pregunta 17a) y conexión a Internet en el hogar (v30); porcentajes basados en respuestas.

Terminaremos la explicación con una reflexión sobre las limitaciones de las preguntas de respuesta múltiple. Además de las señaladas al final del apartado 7.5¹⁴⁵, estas variables presentan el inconveniente que no es posible recodificar la variable resultante, sino que debe hacerse en las variables originales. Es decir, hay que construir la respuesta múltiple, analizar las frecuencias y, en caso que sea preciso, recodificar las variables originales para elaborar de nuevo los *conjuntos de respuestas múltiples*.

El lector habrá apreciado otra de las diferencias existentes entre éstas y las tablas de contingencia mostradas en el capítulo anterior. En las tablas multirespuesta no hay *estadísticos* que indiquen si la relación entre variables es significativa¹⁴⁶, estadísticos que suponen un adecuado *instrumento* en la redacción del informe de investigación cuando se están manejando una gran cantidad de tablas. En las tablas de respuesta múltiple es necesario revisarlas una a una para *desechar* (o no considerar) aquellas con relación no significativa entre variables. Además, tampoco es posible utilizar los residuos para conocer las celdillas donde se producen una mayor relación entre variables. Estos motivos, junto con los indicados en el séptimo capítulo, nos llevan a recomendar una utilización *prudente* de este tipo de preguntas.

Fijar los conocimientos aprendidos es el fin de toda actividad docente. Con el fin de facilitar esta tarea proponemos cuatro ejercicios a realizar con el archivo de la investigación sobre Vida Cotidiana de la Fundación CIRES:

- El equipamiento disponible en el hogar, ¿varían según la comunidad autónoma de residencia? ¿Y con el tamaño del municipio?
- Considerando la pregunta sobre la frecuencia con la que se comen determinados productos alimenticios (pregunta 56, variables de la c58 a la c64), y teniendo en cuenta únicamente aquellos entrevistados que *comen todos o casi todos los días* (opción 1), ¿existe variación según la clase social de pertenencia? ¿Y considerando la relación con el cabeza de familia? ¿Y respecto al tamaño del hogar (número de miembros)?

145. Necesidad de definir los conjuntos de respuestas múltiples cada vez que comienza una nueva sesión de trabajo, y una interpretación más complicada de los resultados

146. No vendrá mal recordar uno de los requisitos citados por Reynolds (1984: 19) para utilizar el Chi-Cuadrado, empleado para variables nominales como las utilizadas en los ejemplos: las categorías de las variables deben ser exhaustivas y mutuamente excluyentes.

4. Relaciones múltiples con tablas de más de dos variables. Introducción al análisis multivariable

En el capítulo nueve, concretamente en el apartado 9.3, propusimos analizar si existe relación significativa entre el grado de felicidad y el estado civil. Los resultados obtenidos se encuentran en los *materiales complementarios* (web) del capítulo 9, y su análisis muestra una relación significativa: valor V de Cramer de 0,171; con una significación de 0,000. La interpretación del interior de la tabla considerando los porcentajes de columna desvela porcentajes superiores en la categoría *muy feliz* por parte de los solteros, los casados y los que conviven en pareja. Es importante señalar el bajo porcentaje en esta categoría de los que *han estado casados*¹⁴⁷. Los que conviven en pareja destacan también por sus porcentajes –más elevados que el promedio– en *bastante feliz*. Por último los que han estado casados (viudos, separados y divorciados) superan ampliamente el porcentaje promedio (total) en la categoría *Poco y nada feliz*.

Algunos lectores estarán dudando de la adecuación de la agrupación realizada en la categoría *han estado casados* en la medida que se han unido personas con perfiles muy diferentes¹⁴⁸. De acuerdo que los separados, divorciados y viudos *han estado casados* anteriormente, pero se trata de situaciones muy diferentes. De momento los últimos (viudos) no han decidido *dejar* de estar casados, sino que el fallecimiento de su cónyuge les ha llevado a esa situación. En relación a los separados y divorciados debemos considerar la fecha de realización del estudio (año 1993), y la legislación en la materia, que implica que para obtener el divorcio era necesario *certificar* una separación; por lo que el estar separado era una condición previa al divorcio. La fecha de realización del estudio explica también el bajo número de personas que se encuentran en esta situación, 30 entrevistados que suponen un 2,6% de los entrevistados.

Otra posibilidad de agrupación vendría de unir los separados y divorciados con los solteros, en la medida que todos son “no casados”. Ahora bien, debe tenerse en cuenta que se trata de grupos totalmente diferentes: el estilo de vida de los separados

147. Esta categoría se ha creado por la unión de los separados (22 entrevistados, un 1,9% de la muestra), divorciados (8 entrevistados, un 0,7%) y viudos (104 entrevistados, un 8,6%). En la explicación de las dos primeras magnitudes debemos tener en cuenta que se trata de un estudio cuyo trabajo de campo se realizó en el año 1993.

148. No vendrá mal considerar aquí la edad media de cada colectivo, obtenidos utilizando la segmentación del archivo explicada en el apartado 8.9:

Solteros: 28,54 años.

Casados: 47,28 años

Viviendo en pareja: 35,32 años.

Separados: 45,27 años

Divorciados: 48,20 años

Viudos: 67,81 años.

y divorciados (con hijos, pagando una hipoteca, etc.) poco tiene que ver con los solteros. Estos motivos nos llevan a considerar que lo más adecuado es eliminar a los separados y divorciados del análisis; en base a su reducido tamaño muestral (30 entrevistados, que alcanzan únicamente el 2,5% de la muestra). Debe tenerse en cuenta que lo treinta entrevistados separados y divorciados no pueden proporcionar una imagen representativa de la realidad de estos colectivos. Además, su escaso número no altera sustancialmente la muestra cuando se opta por eliminarlos.

En la tabla 10.4 se presenta la relación entre el grado de felicidad y estado civil sin los separados y divorciados, donde se aprecia su similitud con la tabla comentada anteriormente (mostrada en los *materiales complementarios* del capítulo 9). La V de Cramer presenta una magnitud de 0,170, con una significación de 0,000. Tan sólo la última columna presenta alguna diferencia, diferencias que no son importantes porque el tamaño muestral de los viudos (103 entrevistados) –frente al escaso número de separados y divorciados– ha permitido *mantener* los hallazgos detectados anteriormente. No obstante, estaremos de acuerdo que la tabla 10.4 representa la realidad mejor que la anterior.

			Estado civil			Total
			Solteros	Casados y conviven	Viudos	
Grado de Felicidad actual	Poco y nada feliz	Recuento % de est civil	49 16,3%	94 12,4%	45 43,7%	188 16,2%
	Bastante feliz	Recuento % de est civil	209 69,7%	565 74,2%	52 50,5%	826 71,0%
	Muy feliz	Recuento % de est civil	42 14,0%	102 13,4%	6 5,8%	150 12,9%
Total		Recuento	300	761	103	1164
		% de est civil	100,0%	100,0%	100,0%	100,0%

Tabla 10.4. Relación entre grado de felicidad y estado civil, porcentajes de columna.

El análisis de los residuos corregidos implicaría prestar atención a lo realmente significativo de esta tabla¹⁴⁹, como es la relación entre viudos y estar *poco y nada feliz*, con un residuo de 8,0 (tabla 10.5). Otras asociaciones importantes son estar viudo y NO¹⁵⁰

149. Aprovechamos para *dar un repaso* a la interpretación de los residuos estandarizados corregidos con el fin de mostrar la gran ayuda que proporcionan en la interpretación de tablas de contingencia.

150. Se añade el *no* por la relación negativa del porcentaje.

			Estado civil			Total
			Solteros	Casados y conviven	Viudos	
Grado de Felicidad actual	Poco y nada feliz	Recuento	49	94	45	188
		Res corregido	,1	-4,8	8,0	
	Bastante feliz	Recuento	209	565	52	826
		Res corregido	-,6	3,4	-4,8	
	Muy feliz	Recuento	42	102	6	150
		Res corregido	,7	,7	-2,2	
Total		Recuento	300	761	103	1164

Tabla 10.5. Relación entre grado de felicidad y estado civil, residuos estandarizados corregidos.

ser *bastante feliz*, con un residuo de -4,8; *convivir en pareja* y estar *bastante feliz*, con un residuo de 3,4; y estar viudo y NO ser *muy feliz*.

Resulta llamativo el bajo estado de felicidad de los viudos, como si la convivencia con una persona de otro sexo fuera la *situación determinante* en el estado de la felicidad. Revisando el resto de tablas de contingencia de la variable grado de felicidad (ver *materiales complementarios* del capítulo 9) resulta llamativo la relación de esta variable con la edad, con un valor Tau-c de -0,115. Observando el interior de la tabla 10.6, y considerando el porcentaje de personas *poco y nada feliz* destaca la baja felicidad de los *talludos*¹⁵¹ (20%) y más aún de los *mayores* (23,1%). Los *jóvenes* y *maduros* presentan –por otro lado– los porcentajes más bajos de *poco y nada feliz*. Dentro de la categoría *bastante feliz* destacan únicamente los maduros, con un 73,1%, si bien es necesario tener en cuenta que esta diferencia –de un par de puntos– no llega a ser significativa en los residuos. Respecto a los *muy felices* son reseñables los elevados porcentajes de los más jóvenes (18,4%) y algo más bajos en los *talludos* y *mayores* (interpretación contraria a la primera fila).

Reflexionando sobre los resultados presentados en ambas tablas resulta sorprendente el bajo grado de felicidad de los viudos y las personas de más edad. Otro aspecto llamativo es la coincidencia –en la categoría *bastante feliz*– de los que conviven en pareja y los que tienen entre 30 y 64 años; si bien la mayor relación se produce en el grupo entre 30 y 44 años. Ambas situaciones nos inducen a sospechar de la verdadera relación entre el estado civil y la felicidad. Ante la sospecha de contar con una relación no *directa*, una relación *intervenida* o *provocada* por una variable extraña a la relación considerada (variable *interveniente* que actúa como *perturbadora*) hemos llevado

151. Hemos adoptado aquí la definición de grupos de edad realizada por De Miguel (1997: 59) en sus estudios sobre la realidad española.

			Edad				Total
			De 18 a 29 años (jóvenes)	De 30 a 44 años (maduros)	De 45 a 64 años (talludos)	65 y más (mayores)	
Felicidad actual	Poco y nada feliz	Recuento	40	40	73	45	198
		% de Edad	12,9%	12,5%	19,7%	23,1%	16,6%
		Residuos corregidos	-2,0	-2,3	2,0	2,7	
Bastante feliz	Bastante feliz	Recuento	212	234	266	134	846
		% de Edad	68,6%	73,1%	71,9%	68,7%	70,9%
		Residuos corregidos	-1,0	1,0	,5	-,7	
Muy feliz	Muy feliz	Recuento	57	46	31	16	150
		% de Edad	18,4%	14,4%	8,4%	8,2%	12,6%
		Residuos corregidos	3,6	1,1	-2,9	-2,0	
Total		Recuento	309	320	370	195	1194
		% de Edad	100,0%	100,0%	100,0%	100,0%	100,0%

Tabla 10.6. Relación entre grado de felicidad y edad. Porcentajes de columna y residuos estandarizados corregidos.

a cabo un análisis entre el grado de felicidad y el estado civil eliminando la influencia de la edad. Para ello se ha utilizado las tablas de contingencia vistas en el capítulo anterior, mediante el menú *Analizar*⇒*Estadísticos descriptivos*⇒*Tablas de contingencia*. En el cuadro de diálogo resultante será introducido el grado de felicidad en filas, el estado civil en columnas, y la edad categorizada en la tercera ventana del cuadro de diálogo (figura 10.8). A continuación solicitaremos los estadísticos de asociación correspondientes, concretamente el Chi-cuadrado y la V de Cramer.

Esta orden crea una tabla que analiza –es importante insistir en ello– la relación entre el grado de felicidad y el estado civil eliminando la influencia de la edad. Es lo que se conoce como *relación parcial*; relación entre dos variables eliminando la influencia de una tercera. Al tratarse de un análisis exploratorio que genera una gran cantidad de resultados recomendamos marcar la opción suprimir tablas.

No presentaremos la totalidad de los resultados, tan sólo los niveles de significación de los estadísticos solicitados (tabla 10.7). El elevado número de celdillas con frecuencia esperada menor que cinco –que alcanza el 33,3% en los maduros y talludos– precisa un análisis detallado del interior de las tablas que permite verificar que esta situación se produce fundamentalmente por el escaso tamaño muestral de los viudos. Tras eliminar esta categoría se vuelven a efectuar los análisis, obteniendo los resulta-



Figura 10.8. Cuadro de diálogo Tablas de contingencia con tres variables.

dos de la tabla 10.8; que muestra la desaparición de la relación entre grado de felicidad y estado civil en TODOS los grupos de edad considerados. Obsérvese que –de nuevo– hay dos grupos de edad donde el 33,3% de las celdillas muestran una frecuencia esperada menor que cinco. En el capítulo IX, apartado 3.2, señalamos que este tamaño debe ser menor que el 20%, puesto que cuando es mayor el Chi-Cuadrado tiende a aumentar *ficticiamente*. Esto se produce porque el numerador es una suma cuadrática de diferencias entre las frecuencias observadas y la teóricas. Cuando las teóricas son muy pequeñas esta suma aumenta notablemente, lo que presentará grandes tamaños del Chi-Cuadrado y –lógicamente– relaciones significativas. Por este motivo se recomienda que las frecuencias teóricas inferiores a 5 no superen el 20% de las celdillas (Reynolds, 1984: 19): En definitiva, que cuando hay muchas celdillas con frecuencia esperada menor que cinco el valor del estadístico aumenta “artificialmente”.

Ahora bien, esta situación no afecta al caso que nos ocupa puesto que contamos con un valor de Chi-Cuadrado no significativo; un valor que se descendería aún más si se redujera el tamaño de las tablas. Con el fin de disminuir este gran porcentaje de celdillas con frecuencias menores a 5 procedemos a unir las categorías “Bastante feliz” y “muy feliz” en el colectivo mayor de 45 años¹⁵²; formando así una tabla de 2 x 2

152. Se trata del procedimiento *Recodificar en distintas variables... Si los casos*; mostrado en la figura 8.11, página 207.

Pruebas de chi-cuadrado

Edad (caracterizada en cuatro categorías)		Valor	gl	Sig. asintótica (bilateral)
De 18 a 29 años (jóvenes)	Chi-cuadrado de Pearson	2,483(a)	2	,289
	Razón de verosimilitudes	2,484	2	,289
	Asociación lineal por lineal	2,467	1	,116
	N de casos válidos	304		
De 30 a 44 años (maduros)	Chi-cuadrado de Pearson	6,414(b)	4	,170
	Razón de verosimilitudes	5,952	1	,203
	Asociación lineal por lineal	4,002	4	,045
	N de casos válidos	311		
De 45 a 64 años (talludos)	Chi-cuadrado de Pearson	24,868(c)	4	,000
	Razón de verosimilitudes	20,177	1	,000
	Asociación lineal por lineal	5,920	4	,015
	N de casos válidos	355		
65 y más (mayores)	Chi-cuadrado de Pearson	21,155(d)	4	,000
	Razón de verosimilitudes	21,976	4	,000
	Asociación lineal por lineal	7,287	1	,007
	N de casos válidos	192		

a 0 casillas (,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 11,88.

b 0 casillas (,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 6,45.

c 2 casillas (33,3%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 2,24.

d 2 casillas (33,3%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 1,14.

Medidas simétricas

Edad (caracterizada en cuatro categorías)		Valor	Sig. aproximada
De 18 a 29 años (jóvenes)	Nominal por nominal	Phi	,090
		V de Cramer	,090
	N de casos válidos		304
De 30 a 44 años (maduros)	Nominal por nominal	Phi	,144
		V de Cramer	,102
	N de casos válidos		311
De 45 a 64 años (talludos)	Nominal por nominal	Phi	,265
		V de Cramer	,187
	N de casos válidos		355
65 y más (mayores)	Nominal por nominal	Phi	,332
		V de Cramer	,235
	N de casos válidos		192

a Asumiendo la hipótesis alternativa.

b Empleando el error típico asintótico basado en la hipótesis nula.

Tabla 10.7. Estadísticos obtenidos de la relación grado de felicidad y estado civil, eliminado la influencia de la edad.

Pruebas de chi-cuadrado				
Edad (caracterizada en cuatro categorías)		Valor	gl	Sig. asintótica (bilateral)
De 18 a 29 años (jóvenes)	Chi-cuadrado de Pearson	2,483(a)	2	,289
	Razón de verosimilitudes	2,484	2	,289
	Asociación lineal por lineal	2,467	1	,116
	N de casos válidos	304		
De 30 a 44 años (maduros)	Chi-cuadrado de Pearson	5,279(b)	2	,071
	Razón de verosimilitudes	4,906	2	,086
	Asociación lineal por lineal	4,693	1	,030
	N de casos válidos	307		
De 45 a 64 años (talludos)	Chi-cuadrado de Pearson	1,056(c)	2	,590
	Razón de verosimilitudes	1,212	2	,546
	Asociación lineal por lineal	,945	1	,331
	N de casos válidos	324		
65 y más (mayores)	Chi-cuadrado de Pearson	5,509(d)	2	,064
	Razón de verosimilitudes	5,573	2	,062
	Asociación lineal por lineal	5,086	1	,024
	N de casos válidos	125		

a 0 casillas (,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 11,88.

b 0 casillas (,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 6,45.

c 2 casillas (33,3%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 2,24.

d 2 casillas (33,3%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 1,14.

Medidas simétricas				
Edad (caracterizada en cuatro categorías)		Valor	Sig. aproximada	
De 18 a 29 años (jóvenes)	Nominal por nominal	Phi	,090	,289
	V de Cramer	,090	,289	
	N de casos válidos	304		
De 30 a 44 años (maduros)	Nominal por nominal	Phi	,131	,071
	V de Cramer	,131	,071	
	N de casos válidos	307		
De 45 a 64 años (talludos)	Nominal por nominal	Phi	,057	,590
	V de Cramer	,057	,590	
	N de casos válidos	324		
65 y más (mayores)	Nominal por nominal	Phi	,210	,064
	V de Cramer	,210	,064	
	N de casos válidos	125		

a Asumiendo la hipótesis alternativa.

b Empleando el error típico asintótico basado en la hipótesis nula.

Tabla 10.8. Estadísticos obtenidos de la relación grado de felicidad y estado civil, eliminado la influencia de la edad. Análisis limitado a los solteros y convivientes en pareja (eliminados los que han estado casados).

donde las celdas con frecuencias esperadas (teóricas) inferiores a 5 se han reducido al 25% en ambas tablas, descendiendo los valores del Chi-Cuadrado al 0,38 en la tabla de 45 a 64 años (en la tabla 10.7 era de 1,056) y al 4,798 (anteriormente era de 5,509). Esta misma explicación sirve para interpretar las diferencias entre los valores Chi-Cuadrado de las tablas 10.7 y 10.8: las reducciones en esta última se producen por la disminución del número de celdillas con frecuencias teóricas menores que 5.

Para poder concluir si la variable neutralizada (la edad) influye más que la otra variable independiente (el estado civil) Reynolds (1984: 76-78) recomienda elaborar un *Índice de relación Parcial* mediante la suma ponderada de cada uno de los valores de la tabla. Si este índice está cercano a cero indicará que la relación entre las variables X e Y es muy débil cuando se elimina el influjo de la tercera. Si, por el contrario, el valor de este índice es muy alto interpretaremos que la tercera variable apenas influye. Como puede apreciarse en el cuadro 10.1 el índice de relación parcial obtenido es de 0,1059, notablemente menor que el obtenido en la relación directa entre grado de felicidad y estado civil (que recordemos era de 0,170); lo que implica que la edad influye en el grado de felicidad más que el estado civil.

$$\text{IRP} = \frac{(0,090 * 304) + (0,131 * 307) + (0,057 * 324) + (0,210 * 125)}{1.060} = 0,1059$$

Cuadro 10.1. Índice de relación parcial, con los estadísticos de la tabla 10.7.

Sintetizando, al eliminar la influencia de la edad desaparece la relación entre grado de felicidad y estado civil; lo que implica –necesariamente– una elevada relación entre estado civil y edad. Esta situación se aprecia con precisión en la tabla 10.9, puesto que el 70% de los solteros tienen menos de 29 años (en la nota a pie número 148 vimos que la media de edad de los solteros es de 28,54 años); uno de cada tres casados tienen entre 30 y 44 años, y un 39% entre 45 y 54 años. El 64,4% de los viudos tiene más de 64 años, y un 31,7% entre 45 y 64¹⁵³. De hecho, el valor del estadístico V de Cramer entre ambas variables asciende a 0,502; el mayor de todos los considerados a lo largo del libro. En numerosas ocasiones no aparecen relaciones tan claras, puesto que –normalmente– suele existir relación en algunos de los grupos de la variable *interveniente* (la edad en este caso) y no en otros.

153. La edad media de los casados y de los que viven en pareja es de 47,55 años, edad que aumenta hasta el 67,8 en el caso de los viudos (recordar nota a pie número 148, página 303).

Es en este momento –una vez que se ha comprobado la exactitud de todas las relaciones detectadas– cuando se procede con la elaboración de gráficos que ilustren el informe. Puede emplearse la opción *mostrar gráficos de barras agrupadas* del menú tablas de contingencia (parte inferior de la figura 10.8) o el menú *Gráficos* mostrado en la figura 4.9. Nuestra recomendación es emplear este último por el mayor número de gráficos disponibles (*Gráfico de Barras*⇒*Agrupado* ó *Apilado*; *Gráfico de Líneas*⇒*Múltiple*). Aunque cuando planificamos este trabajo pensábamos en dedicar un capítulo a los gráficos, las dimensiones actuales del libro –más largo de lo que planteamos en un primer momento–, unido al gran número de trabajos monográficos sobre el tema y a su excelente calidad, nos ha llevado a desistir de este empeño. Nos limitaremos a recomendar alguno, por ejemplo el elaborado por Sevilla Moróder (2006) y el –ya clásico– texto de Alaminos (1993).

			Estado civil			Total
			Soltero	Casado y convive	Viudo	
Edad	De 18 a 29 años (jóvenes)	Recuento	211	95	0	306
		% de Estado civil	70,1%	12,4%	,0%	26,2%
	De 30 a 44 años (maduros)	Recuento	55	253	4	312
		% de Estado civil	18,3%	33,1%	3,8%	26,7%
	De 45 a 64 años (talludos)	Recuento	25	300	33	358
		% de Estado civil	8,3%	39,3%	31,7%	30,6%
	65 y más (mayores)	Recuento	10	116	67	193
		% de Estado civil	3,3%	15,2%	64,4%	16,5%
Total		Recuento	301	764	104	1169
		% de Estado civil	100,0%	100,0%	100,0%	100,0%

Tabla 10.9. Relación entre estado civil y edad.

Señalar, por último, una limitación de esta técnica, como es la necesidad de contar con grandes tamaños muestrales con el fin de que los subgrupos creados por las distintas variables tengan tamaños adecuados para poder calcular el Chi-Cuadrado o los estadísticos correspondientes. Por este motivo se ha utilizado en la explicación el archivo de la investigación sobre vida cotidiana, basado en una muestra de 1.200 casos, en vez del archivo empleado a lo largo del libro, que recoge la información de 191 casos.

Buscando *fixar* los conocimientos aprendidos en esta sección, antes de considerar nuevos contenidos, proponemos un par de ejercicios utilizando la investigación

sobre *Vida Cotidiana*. El primero busca detectar si existe relación entre el nivel de estudios (e21) y el estado civil (e12) y, en caso de que así sea, comprobar si las personas con menos estudios *se casan más* que los tienen estudios elevados.

El segundo ejercicio se centra en la pregunta 17 (b3), referida a la propiedad de la vivienda: ¿La vivienda donde usted reside es de su propiedad, de su familia, o es alquilada? Se plantea como hipótesis a comprobar que las personas con bajos estudios (sin estudios y estudios primarios) presentan una gran propiedad en la vivienda (viviendas propias y pagadas); mientras que los colectivos con mayores estudios se caracterizan por no disponer de vivienda en propiedad.

5. Teoría sobre relaciones múltiples¹⁵⁴

Finalizaremos el capítulo con una reflexión teórica –que surge del ejemplo expuesto en el apartado anterior– con el fin de *explicitar* las distintas relaciones entre variables en tablas de más de dos dimensiones. Respecto a la influencia de otras variables en la relación entre grado de felicidad y estado civil, la experiencia de otras investigaciones recomienda adoptar una postura crítica hacia este hecho debido a las peculiaridades sociodemográficas de nuestro país, donde la mayor parte de los jóvenes están solteros y una gran parte de las personas de más edad están casadas, tal y como vimos en la tabla 10.9. Debemos dudar si la relación entre jóvenes-solteros y casados-mayores implica que la influencia del estado civil está condicionada por la edad, o son variables que influyen de forma independiente.

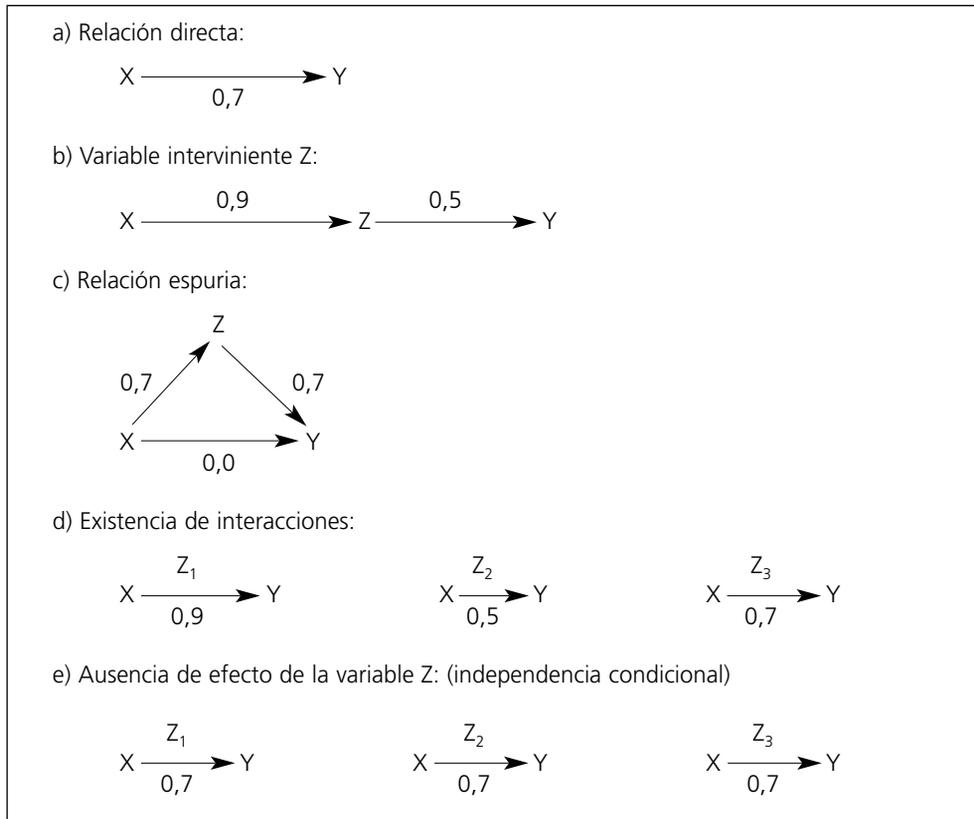
Esta *relación* entre el estado civil y la edad nos invita a elaborar una hipótesis de trabajo que plantea que *pese a la relación entre estado civil (X) y grado de felicidad (Y), ésta descenderá notablemente –llegando incluso a desaparecer– cuando se elimine el influjo de la edad (Z)*. Ahora bien, antes de probar la vigencia de esta hipótesis será preciso realizar una definición de conceptos donde se explique el rol que adopta la tercera variable (variable de control).

De modo que en este apartado contextualizaremos la relación entre variables realizada en el apartado anterior, explicando el interés de trabajar con tablas de contingencia formadas por tres o más variables. Se trata de análisis más complejos al introducir una tercera variable (llamémosle Z) cuya función es *controlar* la relación entre ambas variables X e Y. Lo normal es realizar varias *tablas parciales* para cada catego-

154. Hemos realizado un desarrollo más amplio de estos aspectos en Díaz de Rada, 1999: 199-221.

ría de Z: relación entre X e Y para Z_1 , relación entre X e Y para Z_2 , relación entre X e Y para Z_m . Las magnitudes de asociación entre X e Y en cada una de estas tablas parciales pueden cambiar al ser eliminado el influjo de la variable control Z ¹⁵⁵. Es decir, al construir una tabla (o tres tablas parciales) con las tres variables, la relación entre X e Y se convierte en una *relación parcial* al eliminar la influencia de Z, al controlar el efecto de esta última. La asociación entre variables de tablas parciales es conocida en la literatura especializada como *asociación* (o *relación*) *condicional*, mientras que la relación entre X e Y sin tener en cuenta ninguna otra variable es definida como *relación* (o *asociación*) *marginal*.

La variable que elegimos como *control*, la tercera variable de la tabla, puede adoptar diversos roles al detectar la relación entre X e Y; tal y como se muestra en el cuadro 10.2. En los siguientes párrafos describimos cada uno:

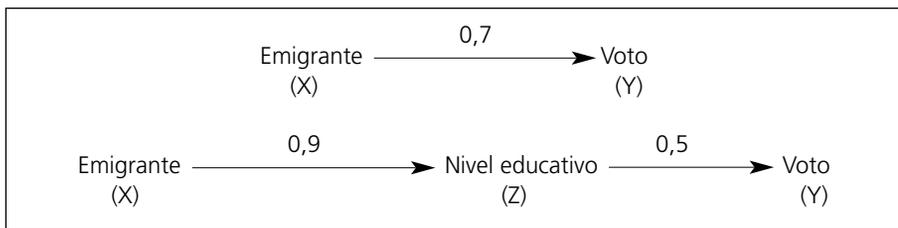


Cuadro 10.2. tipos de relaciones entre tres variables.

155. Debe quedar claro que siempre es una relación entre X e Y.

1. *Variable interviniente* para explicar fenómenos. La tercera variable (Z) adopta este nombre cuando es utilizada para explicar la relación entre la variable independiente y la dependiente.

Un ejemplo ayudará a comprender este concepto: supongamos una investigación sobre voto político donde un investigador ha encontrado una alta relación entre ser hijo de emigrantes (X) y el voto político (Y). Este investigador decide introducir una tercera variable (Z), el nivel educativo, a fin de comprobar si el *bajo nivel educativo* de los hijos de emigrantes influye en el voto político. Este investigador plantea la hipótesis que el ser hijo de emigrante está relacionado con el nivel educativo conseguido, pero no al revés; y por ello la educación es una variable interviniente. Si las personas con el mismo nivel educativo (sean hijos de emigrantes o no) muestran un comportamiento electoral similar, la variable *interviniente* estará desvelando que el ser hijo de emigrante afecta al voto político no tanto por sí misma, sino porque los hijos de emigrantes alcanzan un determinado nivel educativo.



Cuadro 10.3. ejemplo de variable interviniente.

Otro ejemplo muy ilustrativo es el análisis de la relación entre las pérdidas producidas por un incendio y el número de bomberos que han acudido a apagarlo. Si alguien recoge y analiza estos datos verá que existe una elevada relación entre ambas. La dificultad para interpretar esta relación podría llevarnos a creer que son los bomberos los que producen los destrozos. Es evidente que esta interpretación es muy poco consistente, de modo que nos encontramos con una relación estadística entre variables que –a priori– carece de interpretación sociológica. Las dificultades en la interpretación le llevan al investigador a introducir una tercera variable, la magnitud del incendio, pudiendo comprobar así que la relación entre número de bomberos y pérdidas producidas por el incendio cambia notablemente. Es decir, no es que el número de bomberos genere más pérdidas en cada incendio, sino que es la intensidad del fuego la que fija el número de bomberos necesarios, y esta intensidad es la que genera los destrozos.

2. *Detección de relaciones espurias*: Responde a la pregunta de si la naturaleza y la fuerza de la asociación cambia en las tablas parciales. La relación espuria se producirá cuando al tener en cuenta la tercera variable (Z) desaparezca la relación entre la variable independiente (X) y la dependiente (Y). Dicho de otra forma, cuando la *relación parcial* entre la variable independiente y la dependiente sea cero.

Supongamos que un investigador está analizando la relación entre la edad (X) y el voto político (Y), a fin de probar la hipótesis de que los más jóvenes cada vez presentan actitudes y comportamientos políticos más conservadores. Para ello se han recogido los datos mostrados en el tabla 10.10.a. En los jóvenes el partido conservador logra un 13,6% (56,4% – 42,8%) más de votos que en los adultos. El análisis del coeficiente Phi presenta un valor de 0,135 que es significativo al 0,000.

La *relación marginal* mostrada en la tabla “a” permite aceptar la hipótesis planteada, si bien se trata de una relación entre voto y edad no muy alta (aunque significativa). Esta situación crea dudas en el investigador y decide incluir una tercera variable en el modelo, el nivel de estudios (tabla b). Así, en primer lugar repite esta relación para el grupo de estudios bajos (valor *Phi* 0,003, no significativo) y para el colectivo de estudios medios (valor *Phi* 0,014, no significativo). Es decir, al eliminar la influencia del nivel de estudios la relación entre la edad y el voto político desaparece, detectándose así la presencia de una *relación espuria* entre ambas.

El nivel de estudios ha permitido detectar una relación espuria entre la edad y el voto político por lo que el investigador se plantea hasta que punto será el nivel educativo el factor que influye en el conservadurismo. Para ello elabora la tabla “c”, donde se analiza la relación marginal entre el nivel de estudios y el voto, obteniendo un valor *Phi* de 0,396; el mayor de los localizados hasta el momento.

A fin de detectar la posible influencia de la edad se vuelve a realizar este análisis por separado en los jóvenes y en los adultos (tabla “d”), encontrando valores de *Phi* altamente significativos que indican una relación de 0,341 en los jóvenes y de 0,405 en los adultos. Lo que desvela que existe relación entre los estudios y el voto político en ambos colectivos, si bien es mayor en los adultos que en los jóvenes.

Considerando los porcentajes de la tabla se observa una diferencia de 38,9 puntos en el caso de los jóvenes (66,2% – 27,3%), y de 40,6 puntos (66,5% – 25,9%) en el nivel de conservadurismo de los adultos (los que tienen bajos estudios muestran un mayor conservadurismo); manifestándose más conservadores las personas con menos estudios. Conviene tener en cuenta que esta

Tabla "a"

	Partido Conservador	Otros	Total	% Conserv.	Phi (**)	Signif
Jóvenes	270	209	479	56,4	0,135	0,000
Adultos	224	299	523	42,8		

Tabla "b"

	Partido Conservador	Otros	Total	% Conserv.	Phi (*)	Signif	
Estudios bajos	Jóvenes	237	121	358	66,2	0,003	0,939
	Adultos	145	73	218	66,5		
Estudios medios	Jóvenes	33	88	121	27,3	0,014	0,772
	Adultos	79	226	305	25,9		

Tabla "c"

	Partido Conservador	Otros	Total	% Conserv.	Phi (**)	Signif
Estudios bajos	382	194	572	66,3	0,396	0,000
Estudios medios	112	314	426	26,3		

Tabla "d"

	Partido Conservador	Otros	Total	% Conserv.	Phi (**)	Signif	
Jóvenes	Estudios bajos	237	121	358	66,2	0,341	0,000
	Estudios medios	33	88	121	27,3		
Adultos	Estudios bajos	145	73	218	66,5	0,405	0,000
	Estudios medios	79	226	305	25,9		

(*) Relación condicional.

(**) Relación marginal

Tabla 10.10. Relación entre edad, nivel de estudios y voto político.

diferencia es mayor que la detectada al principio: 13,6%, obtenido de restar 56,4% menos 42,8%. Obsérvese, por otro lado, las escasas diferencias entre los jóvenes y adultos con estudios bajos (66,5 – 66,2), y entre los jóvenes y adultos con estudios medios (25,9 – 27,3).

3. *Localización de interacciones*: Se produce cuando la relación entre la variable dependiente (X) y la independiente (Y) cambia según las distintas categorías de la tercera variable (Z): la relación entre X e Y es fuerte en la primera categoría de la variable Z, se debilita en la segunda, vuelve a aumentar en la tercera, etc. En esta situación estaríamos en presencia de una interacción entre la tercera variable (variable control) y el resto de variables de la tabla. Dicho con otras palabras la variable control determina el tipo de relación existente entre las variables.

Es posible también que la relación entre la variable dependiente y la independiente, tras introducir la tercera variable, genere una ligera reducción en los coeficientes de asociación de cada una de las tablas parciales, pero que esta relación continúe siendo significativa. Esta situación estaría provocada por la existencia de relaciones espurias en determinadas categorías de la variable control.

Podemos señalar la presencia de una pequeña interacción en el ejemplo anterior (tabla 10.10) cuando analizábamos la relación entre nivel de estudios y voto político. El coeficiente *Phi* en los jóvenes es de 0,341, mientras que en los adultos es de 0,405. Aunque la magnitud de la relación es diferente, consideramos que ésta es muy pequeña para poder hablar de presencia de una interacción entre variables.

Cuando no existen interacciones, es decir cuando la relación entre la variable X y la Y no cambian en ninguna de las categorías de Z estamos en presencia de una *asociación homogénea*.

4. *Ausencia de efecto*: Una última posibilidad es que la tercera variable no genere ninguna influencia en la relación entre la variable X y la Y, concluyendo que ésta no tiene ninguna influencia en la variable dependiente Y.

Aquí se han presentado la mayor parte de las relaciones entre variables Hemos prestado mayor importancia a la detección de relaciones espurias por su importancia en el ámbito de la investigación con encuesta.

6. Anexo 1: Lenguaje de sintaxis de los análisis realizados

Tal y como se ha procedido en los capítulos anteriores, finalizamos presentando el lenguaje de sintaxis de SPSS con los análisis realizados. Recuérdesse que en el apartado 7 del capítulo VII se explicó el origen de cada uno de estos mandatos y su proceso de *ejecución*.

Apartado 2: Tablas de contingencia con variables de respuesta múltiple categóricas

MULT RESPONSE

GROUPS=\$preg3 'Situaciones definen el tiempo libre (v3+v4)' (v03 v04 (1,11))
/FRECUENCIAS=\$preg3.

MULT RESPONSE

GROUPS=\$preg3 'Situaciones definen el tiempo libre (v3+v4)' (v03 v04 (1,11))
/VARIABLES=v49(1 2)
/TABLES=\$preg3 BY v49
/CELLS=COLUMN
/BASE=CASES.

Apartado 3: Tablas de contingencia con variables de respuesta múltiple dicotómicas

MULT RESPONSE

GROUPS=\$nº_dispo 'Dispositivos en el ordenador' (v22 v23 v24 v25 v26 v27
V28 v29 (1))
/FRECUENCIAS=\$nº_disp.

FRECUENCIAS

VARIABLES=v30
/ORDER ANALYSIS.

MULT RESPONSE

GROUPS=\$nº_dispo 'Dispositivos en el ordenador' (v22 v23 v24 v25 v26 v27
V28 v29 (1))
/VARIABLES=v30(1 2)
/TABLES=\$nº_dispo BY v30
/CELLS=COLUMN
/BASE=CASES.

MULT RESPONSE

GROUPS=\$nº_dispo 'Dispositivos en el ordenador' (v22 v23 v24 v25 v26 v27
V28 v29 (1))
/VARIABLES=v30(1 2)

```
/TABLES=$nº_dispo BY v30  
/CELLS=COLUMN  
/BASE=RESPONSES.
```

Apartado 4: Relaciones múltiples entre varias variables. Introducción al análisis multivariable

FREQUENCIES

```
VARIABLES=a54  
/ORDER ANALYSIS.
```

```
RECODE A54 (1=1)(2=1) (3=2) (4=3) (9=SYS) INTO A54RECO.  
VARIABLE LABELS A54RECO "GRADO DE FELICIDAD ACTUAL  
(RECODIFICADO EN 3 CATEGORÍAS)".  
VALUE LABELS A54RECO 1'POCO Y NADA FELIZ'  
2'BASTANTE FELIZ'  
3'MUY FELIZ'.
```

FREQUENCIES

```
VARIABLES=a54 a54reco  
/ORDER ANALYSIS.
```

FREQUENCIES

```
VARIABLES=e12  
/ORDER ANALYSIS.
```

MEANS

```
TABLES=e10 BY e12  
/CELLS MEAN COUNT STDDEV  
/STATISTICS ANOVA LINEARITY.
```

```
RECODE E12 (1=1)(2=2)(3=2)(6=3) INTO E12RECO.  
VARIABLE LABELS E12RECO "Estado civil (recodificado en 3 categorías)".  
VALUE LABELS E12RECO 1"Solteros" 2"Casados y conviven" 3"Viudos".  
FREQUENCIES
```

```
VARIABLES=e12 e12reco  
/ORDER ANALYSIS.
```

CROSSTABS

```
/TABLES=a54reco BY e12reco  
/STATISTIC=CHISQ PHI  
/FORMAT= AVALUE TABLES  
/CELLS= COUNT COLUMN ASRESID  
/COUNT ROUND CELL.
```

FREQUENCIES

```
VARIABLES=e10  
/ORDER ANALYSIS.
```

Recode e10 (18 thru 29=1) (30 thru 44=2) (45 thru 64=3) (65 thru 99=4) into EDAD.

Variable label EDAD "Edad (categorizada en cuatro categorías)".

Value labels EDAD 1"De 18 a 29 años (jóvenes)"

2"De 30 a 44 años (maduros)"

3"De 45 a 64 años (talludos)"

4"65 y más (mayores)".

FREQUENCIES

```
VARIABLES=e10 EDAD  
/ORDER ANALYSIS.
```

CROSSTABS

```
/TABLES=a54reco BY e12reco BY edad  
/STATISTIC=CHISQ PHI  
/FORMAT= AVALUE TABLES  
/CELLS= COUNT COLUMN ASRESID  
/COUNT ROUND CELL.
```

RECODE E12reco (3=sysmis).

CROSSTABS

```
/TABLES=a54reco BY e12reco BY edad  
/STATISTIC=CHISQ PHI  
/FORMAT= AVALUE TABLES  
/CELLS= COUNT COLUMN ASRESID  
/COUNT ROUND CELL.
```

```
DO IF (edad > 2).  
RECODE A54Reco (1=1)(2 3=2) INTO A54REC_2.  
END IF.  
EXECUTE.  
VARIABLE LABELS A54REC_2 "GRADO DE FELICIDAD ACTUAL  
  (RECODIFICADO EN 2 CATEGORÍAS)".  
VALUE LABELS A54REC_2 1'POCO Y NADA FELIZ'  
                2'BASTANTE Y MUY FELIZ'.
```

CROSSTABS

```
/TABLES=a54reco BY e12reco BY edad  
/STATISTIC=CHISQ PHI  
/FORMAT= AVALUE TABLES  
/CELLS= COUNT COLUMN ASRESID  
/COUNT ROUND CELL.
```

MEANS

```
TABLES=e10 BY e12  
/CELLS MEAN COUNT STDDEV.
```

CROSSTABS

```
/TABLES=edad BY E12reco  
/STATISTIC=CHISQ PHI  
/FORMAT= AVALUE TABLES  
/CELLS= COUNT COLUMN ASRESID  
/COUNT ROUND CELL.
```