

Análisis de datos en psicología I

Breves orientaciones sobre gráficas

Vicente Manzano - 2006

Las posibilidades gráficas son impresionantes en diversidad y resultados. Este documento pretende orientar con brevedad, por lo que vamos a ignorar la inmensa mayoría y nos vamos a quedar con unas pocas estrategias para la representación gráfica. El objetivo es tener recursos para decidir cómo escoger unas u otras de entre la gran variedad disponibles. Se indica, además, cómo realizar estas representaciones con SPSS.

Para tomar las decisiones, vamos a considerar dos criterios: el tipo de objetivo (unitario o relacional) y el tipo de variable (nominal, ordinal o cuantitativa). En el caso de variables nominales, distinguiremos también cuando nos encontramos con pocos o muchos valores.

Para mostrar un ejemplo completo, partimos del archivo de datos que se encuentra en el anexo a este documento. Se trata de un estudio con 50 casos y 6 variables, de las que 2 son nominales, 2 ordinales y 2 cuantitativas (una de ellas, una ordinal *fin*).

Objetivos unitarios

A. Variables nominales con pocos valores.

Una representación muy ilustrativa es el diagrama de sectores o ciclograma. Pedimos a SPSS un ciclograma mediante la sucesión

```
Gráficos / Sectores... / Resúmenes para grupos de casos /  
Definir / N° de casos / Definir sectores por: Estado civil  
/ Aceptar
```

O mediante el comando

```
GRAPH  
  /PIE=COUNT BY civil .
```

Y obtenemos el resultado de la representación 1, que ha sido editada en SPSS antes de darla por definitiva. El ciclograma es más orientativo que el diagrama de barras, puesto que éste hace suponer un orden que en las variables nominales no tiene sentido. Y el diagrama de barras, además, al expresarse en ejes cartesianos, es

sensible a las escalas que se utilicen para los ejes vertical (abcisas) y horizontal (ordenadas). Se puede observar que hay aproximadamente la misma cantidad de gente soltera que casada, y menos de la categoría *otros*.

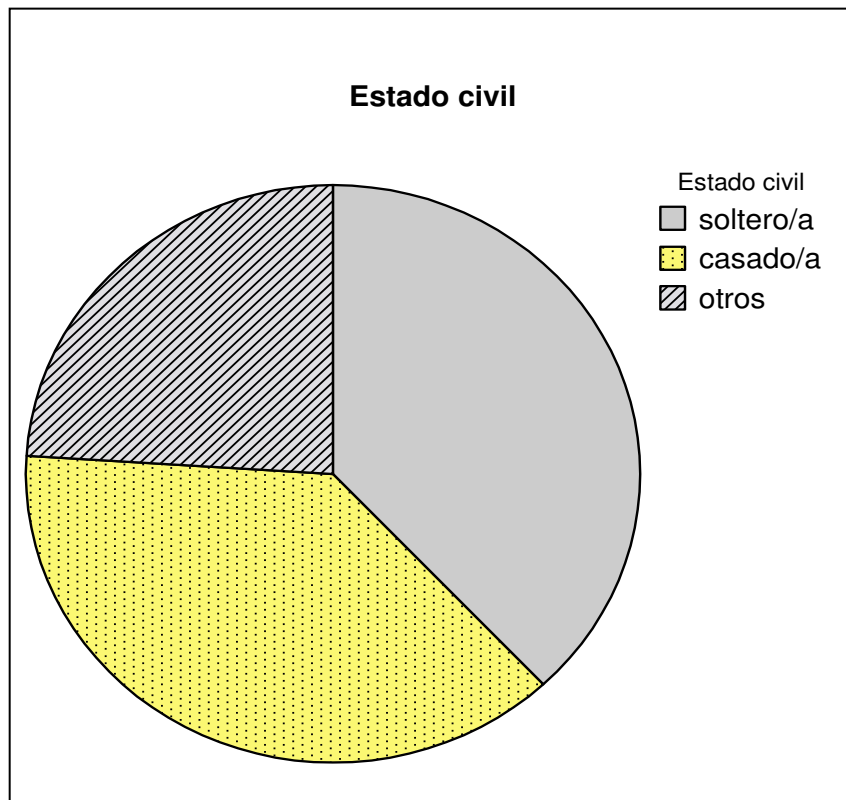


Gráfico 1. Ciclograma o diagrama de sectores.

B. Variables nominales con muchos valores.

Cuando se cuenta con muchos valores, el ciclograma se carga de sectores y pierde claridad. En estos casos, es preferible acudir a un diagrama de barras. Así hemos hecho con la variable *vota* y que muestra el partido a quien la persona entrevistada dice haber votado en las últimas elecciones.

Gráficos / Barras... / Simple / Definir / N° de casos / Eje de categorías: a qué partido votó en las últimas elecciones / Aceptar

O mediante el comando

```
GRAPH  
/BAR(SIMPLE)=COUNT BY vota .
```

Cuyo resultado se muestra en la representación gráfica 2. Vemos que el partido más votado ha sido DPH, seguido muy de cerca por URT y

PhP. Los menos votados fueron (siempre según las confesiones de quienes respondieron a la encuesta) los partidos PGG y PPP.

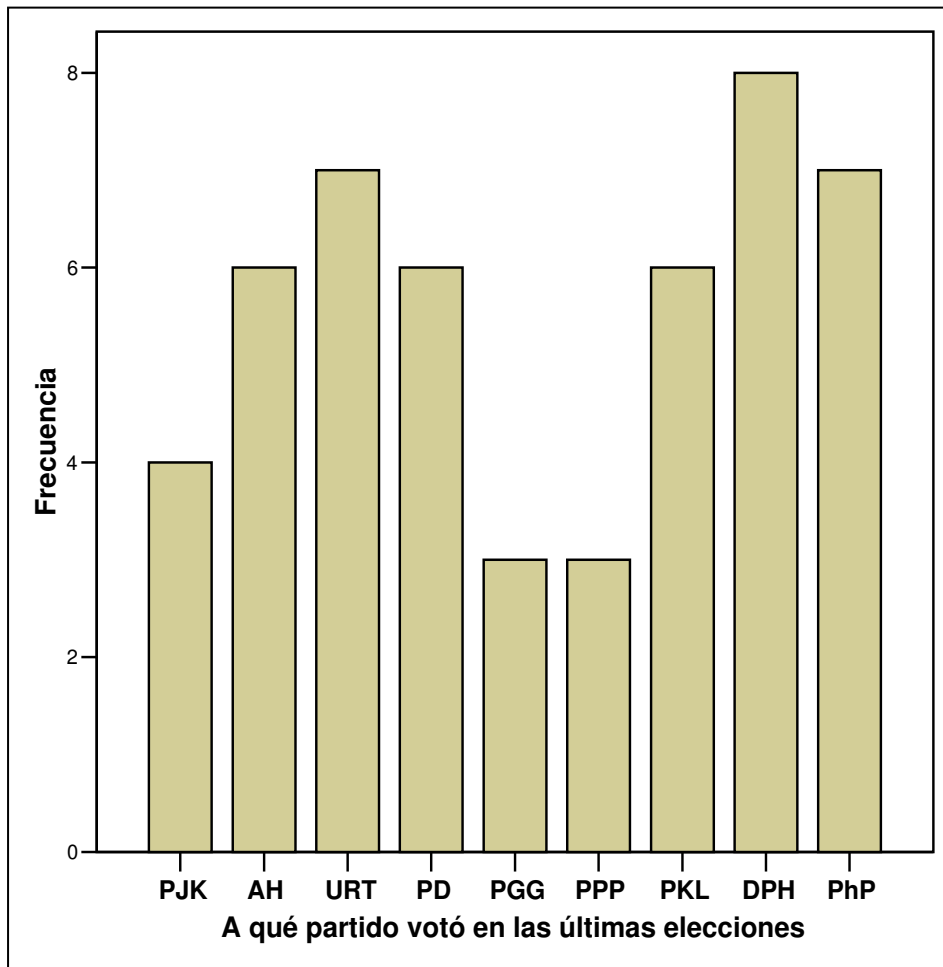


Gráfico 2. Diagrama de barras.

C. Variables ordinales (burdas).

Para las variables ordinales burdas, es decir, cuya distancia entre valores sea claramente dispar, es recomendable acudir a un diagrama de barras, por lo que lo indicado en el caso anterior es válido igualmente para éste.

D. Variables cuantitativas (y ordinales finas).

Dado que éstas permiten operar con distancias, contamos con una gran cantidad de posibilidades. Nos interesan dos de ellas: los histogramas y los box-plot. Los primeros superan a los diagramas de barras al corregir dos deficiencias de éstos en cuantitativas: respetan las distancias (lo que provoca mantener barras de altura 0 si no hay valores, cuando los diagramas de barras ignorarían el hueco) y acumulan los valores en intervalos cuando la cantidad de aquéllos

hace desaconsejable contemplarlos todos en el eje de abcisas. La gráfica 3 muestra un histograma para la variable edad. La gráfica 4 representa la misma variable pero mediante un diagrama de caja y patillas (box-plot).

Para generar el histograma se ha acudido a la siguiente sucesión:

Gráficos / Histograma... / Variable: estado civil / Aceptar

O mediante el comando

```
GRAPH  
  /HISTOGRAM=civil .
```

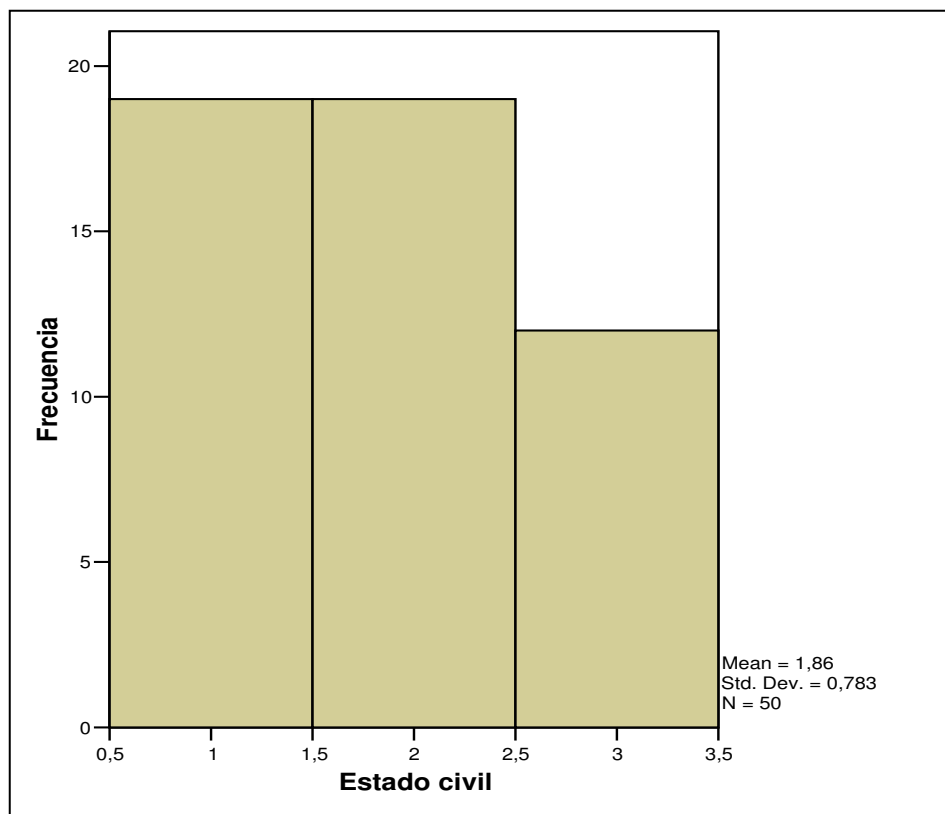


Gráfico 3. Histograma.

El blox-plot permite superar el inconveniente del histograma cuando éste acumula los valores. Las acumulaciones pueden dar resultados divergentes y con apariencia arbitraria. Observamos, por ejemplo, que el resultado que muestra la gráfica 3 es difícil de interpretar, puesto que SPSS ha acumulado en exceso los valores y perdemos información en los intervalos finales. El diagrama de caja y patillas, además, representa un índice central (la mediana) y permite obtener más información sobre la forma de la distribución. Su fuerte es la relación, por lo que lo veremos más adelante. Recuerdo, además, que

este documento no tiene por motivo explicar las representaciones, sino justificar por qué unas y no otras en qué situaciones.

Para generar el diagrama de caja y patillas se ha acudido a la siguiente sucesión de acciones:

```
Analizar / Estadísticos descriptivos / Explorar... /  
Dependientes: edad / Mostrar gráficos / (botón) Gráficos /  
(Desactivar Descriptivos) / Continuar / Aceptar
```

O mediante el comando

```
EXAMINE  
VARIABLES=edad  
/PLOT BOXPLOT  
/COMPARE GROUP  
/STATISTICS NONE  
/CINTERVAL 95  
/MISSING LISTWISE  
/NOTOTAL.
```

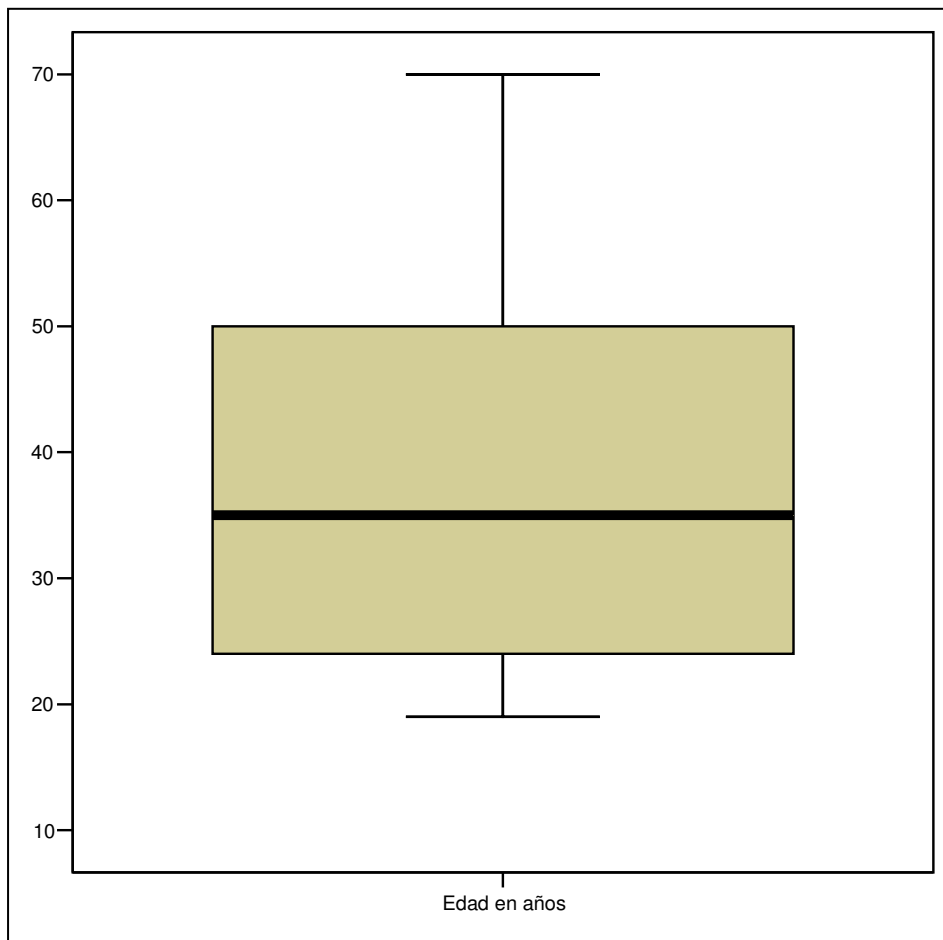


Gráfico 4. Diagrama de caja y patillas o box-plot.

Objetivos relacionales

E. Entre nominales.

Cuando el número de valores es pequeño, el mejor recurso es una sencilla tabla de contingencia. Si se requiere una representación gráfica, un recurso aceptable es un diagrama de barras con datos agrupados. La gráfica 5 muestra un ejemplo donde el eje de abscisas muestra el partido al que se votó y las barras se reproducen para cada estado civil. Se observa que lo más frecuente es los solteros que votaron a URT. Los solteros se reparten entre los partidos PJK, AH, URT y PD, mientras que los que se encuentran en otro estado civil se decantan más por DPH, PhP y PKL, seguido más lejos por PPP. Los casados se encuentran muy repartidos en las diferentes opciones, principalmente en DPH, seguido de PGG, PKL y PhP.

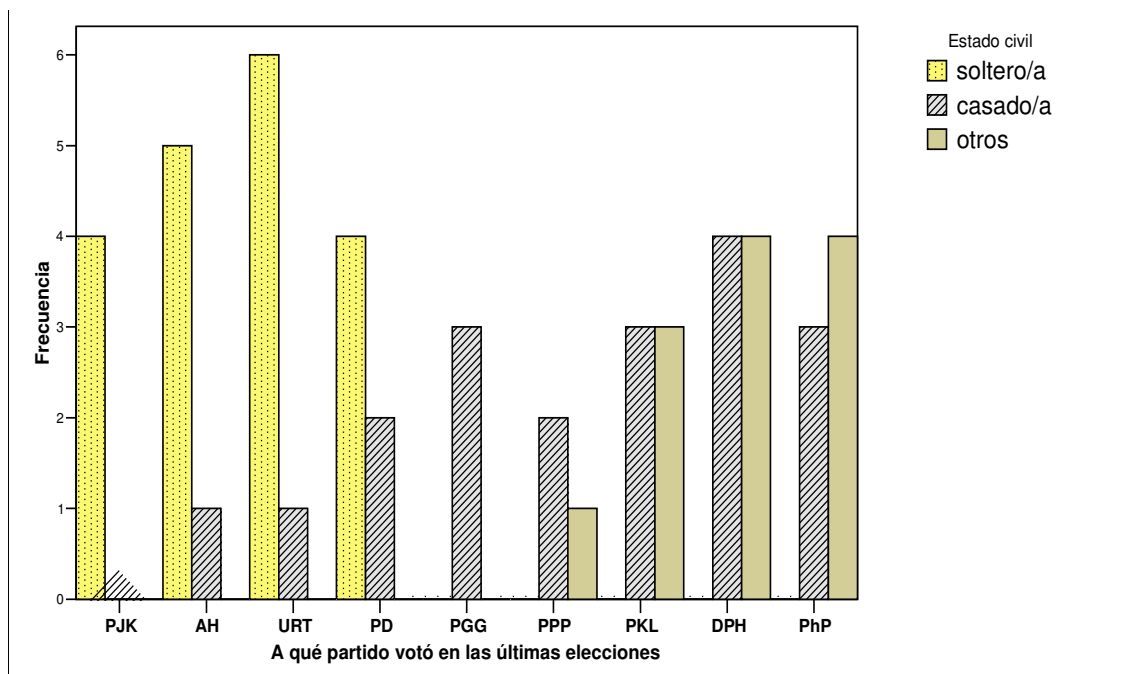


Gráfico 5. Diagrama de barras agrupadas.

Para conseguir este resultado, se atraviesa la secuencia:

Gráficos / Barras... / Agrupado / Definir / Eje de categorías: a qué partido votó / Definir grupos por: estado civil / Aceptar

O bien el comando de sintaxis:

```
GRAPH  
/BAR(GROUPED)=PCT BY vota BY civil .
```

Pero la representación es algo limitada puesto que no muestra bien la comparación de las categorías combinadas. Es mejor el recurso de una tabla de contingencia, como ya se ha indicado. Lo que sigue es la misma información, dispuesta en dos tablas donde se ha recurrido al porcentaje por filas y al porcentaje por columnas. Se observa mucho mejor que ambas variables (preferencia de voto y estado civil) tienen una marcada relación.

% columnas	PJK	AH	URT	PD	PGG	PPP	PKL	PDH	PhP	Total
soltero/a	100,0	83,3	85,7	66,7						38,0
casado/a		16,7	14,3	33,3	100,0	66,7	50,0	50,0	42,9	38,0
otros						33,3	50,0	50,0	57,1	24,0

% filas	PJK	AH	URT	PD	PGG	PPP	PKL	PDH	PhP
soltero/a	21,1	26,3	31,6	21,1					
casado/a		5,3	5,3	10,5	15,8	10,5	15,8	21,1	15,8
otros						8,3	25,0	33,3	33,3
Total	8,0	12,0	14,0	12,0	6,0	6,0	12,0	16,0	14,0

F. Entre ordinales.

Entre variables ordinales se puede acudir a la misma representación gráfica de las barras apiladas, con más razón que en el caso de las nominales, porque el orden de las barras tiene sentido. No obstante, cuando el número de valores es elevado (como ocurre en el ejemplo anterior, con las preferencias por los partidos políticos), hay que recurrir a otras estrategias. Un consejo aceptable es fundir categorías, es decir, crear categorías nuevas que engloben a otras. Dentro de esta medida, la opción más habitual es crear una categoría "otros" que englobe las respuestas minoritarias.

Si la ordinal es del tipo *fin*, puede ser objeto del mismo tratamiento que las variables cuantitativas.

G. Entre cuantitativas.

La mejor representación inicial para dos variables cuantitativas es el diagrama de dispersión o nube de puntos. En el siguiente ejemplo, contenido en el gráfico 6, se muestra la posible relación entre la ansiedad y la edad.

Para generar la representación se ha acudido a la secuencia:

Gráficos / Dispersión... / Simple / Definir / Eje Y: nivel de ansiedad / Eje X: edad / Aceptar

O a la sintaxis:

```
GRAPH  
/SCATTERPLOT(BIVAR)=edad WITH ansiedad  
/MISSING=LISTWISE .
```

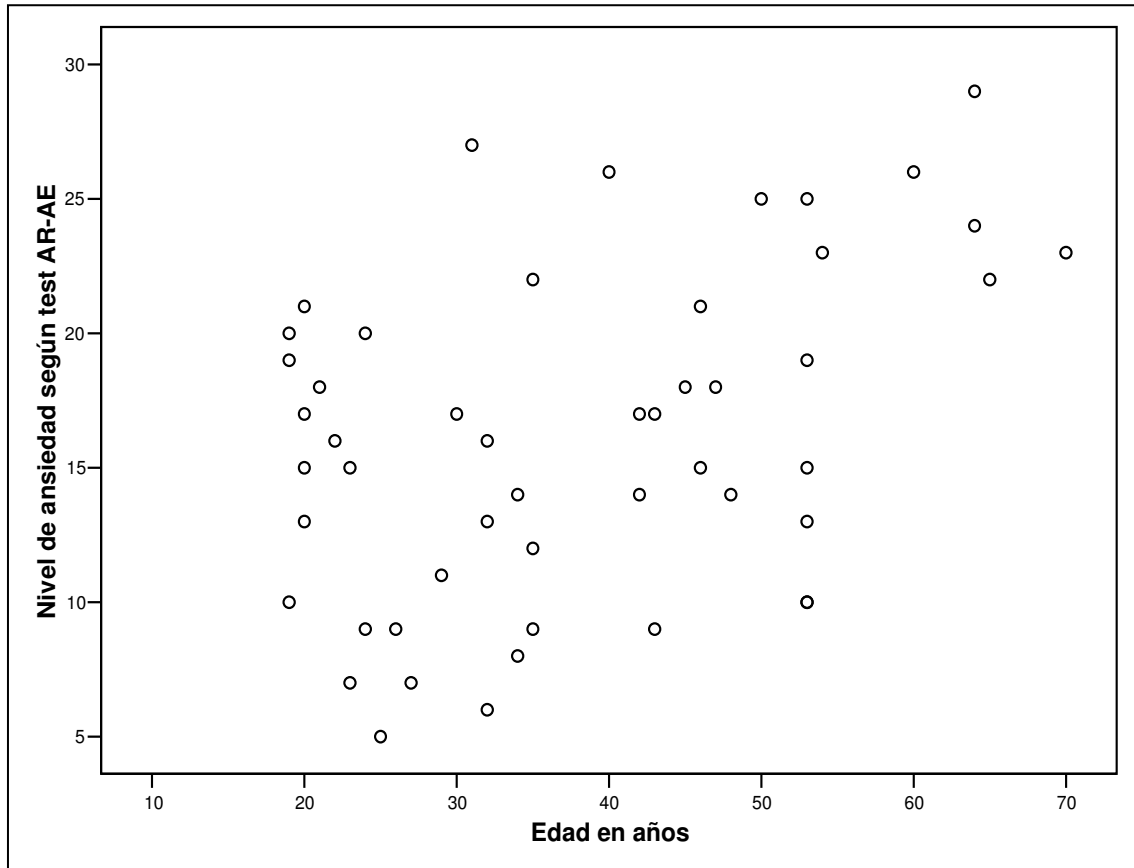


Gráfico 6. Diagrama de dispersión o nube de puntos.

En la representación se observa una leve relación: se insinúa que al aumentar la edad también lo hace la ansiedad. Hay coincidencia en los valores altos para ambas variables, pero no así en los medios o bajos, por lo que la relación no parece tener peso.

H. Nominal con ordinal.

Como viene siendo la tónica, depende de la calidad de la variable ordinal, será tratada como una nominal (ver opción E) o una cuantitativa (ver opción I).

I. Nominal con cuantitativa.

Hay muchos recursos disponibles. Las barras agrupadas o apiladas donde la medida no es la frecuencia sino un índice de la cuantitativa

(como la media aritmética) es una opción disponible e interesante. Pero aquí vamos a recurrir únicamente al diagrama de caja y patillas que ofrece muchas ventajas. Con sensible claridad (siempre y cuando se sepa interpretar la información, como en todo), se puede interpretar la posible relación entre las dos variables. Para ellos, se genera un diagrama de caja y patillas de la variable cuantitativa para cada una de las categorías de la variable nominal. En el ejemplo de este apartado, se ha recurrido a observar la relación entre la ansiedad y el estado civil.

La ejecución se realiza mediante:

```
Gráficos / Diagramas de caja... / Simple / Definir /  
Variable: edad / Eje de categorías: estado civil / Aceptar
```

O con la sintaxis:

```
EXAMINE  
  VARIABLES=edad BY civil  
  /PLOT=BOXPLOT/STATISTICS=NONE/NOTOTAL.
```

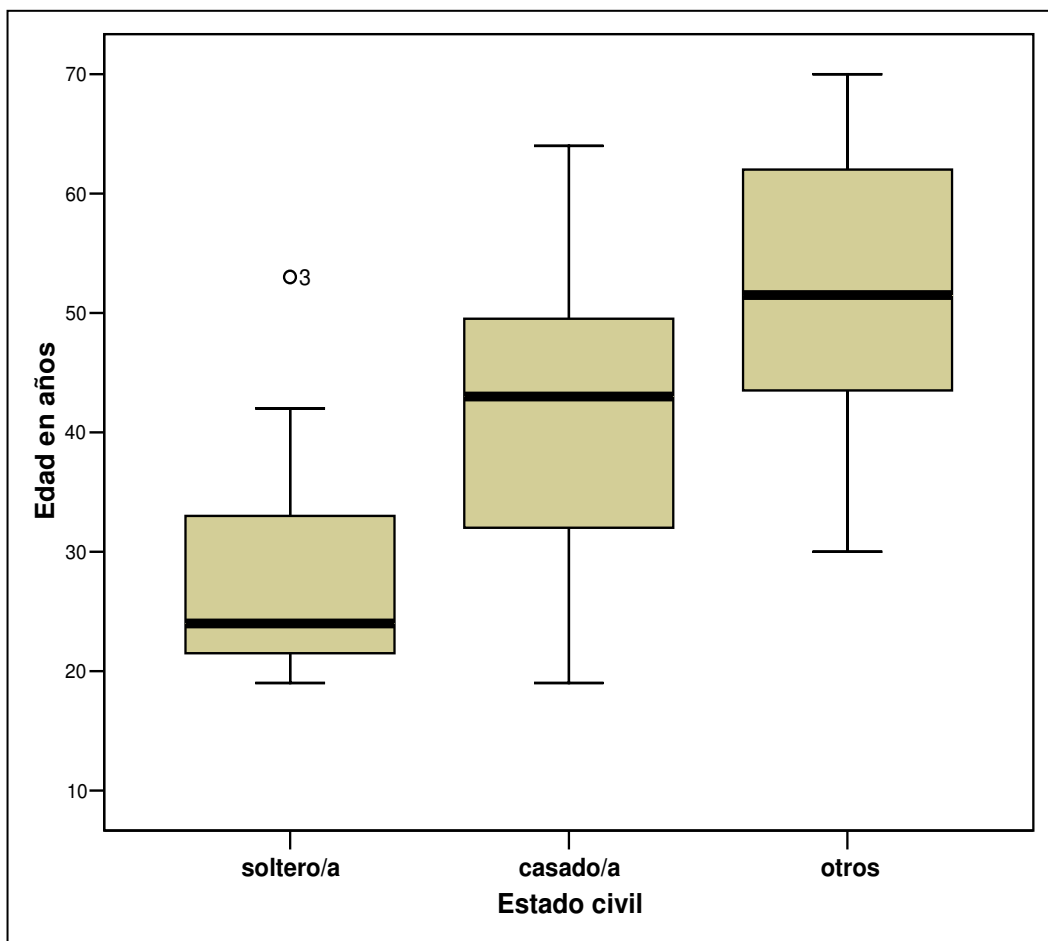


Gráfico 7. Diagramas de caja combinados.

En el resultado puede observarse que existe una relación muy clara. La mediana de edad ya de por sí es sensiblemente diferente en los tres grupos de estado civil, especialmente entre solteros/as y casados/as. Pero las cajas muestran también información útil. Así, entre solteros/as y otros no hay solapamiento del 50% central de sus respectivas distribuciones. La gente soltera, además, es la menos dispersa en años, es decir, la que posee más homogeneidad en cuanto a su edad (excepción de un caso raro, el 3).

J. Ordinal con cuantitativa.

Por último, reflexiónese también aquí sobre la calidad de la variable ordinal, para seguir los pasos del punto I (ordinal burda) o G (ordinal fina).

Anexo

Variables:

Civil: Estado civil:
soltero/a (1), casado/a (2) y otros (3).
Vota: A qué partido votó en las últimas elecciones:
PJK (1), AH (2), URT (3), PD (4), PGG (5), PPP (6),
PKL (7), DPH (8) y PhP (9)
Estudios: Nivel de estudios:
analfabeto/a (1), sin estudios (2), primarios (3),
medios (4) y superiores (5)
Ansiedad: Nivel de ansiedad según AR-AE.
Satisfa: Grado de satisfacción con la vida:
pésimo (1), bajo (2), suficiente (3), bastante (4),
elevado (5) y óptimo (6)
Edad: Edad en años.

Los datos, siguiendo el mismo orden de las variables especificadas, son:

1	1	5	12	6	35	1	4	4	7	5	27	2	9	2	23	1	54
1	1	4	14	6	34	1	4	4	10	2	19	2	9	3	10	3	53
1	1	3	15	4	53	2	4	5	19	4	19	2	8	4	14	2	42
1	1	2	17	5	42	2	4	5	20	3	19	2	8	4	16	3	32
1	2	4	8	3	34	2	5	3	21	1	20	3	7	4	17	2	30
1	2	3	9	4	24	2	5	4	22	3	35	3	7	1	27	4	31
1	2	2	6	2	32	2	5	2	24	2	64	3	6	2	26	3	40
1	2	5	17	3	20	2	6	3	25	4	53	3	9	3	25	5	50
1	2	2	15	2	20	2	6	1	18	3	45	3	9	2	26	4	60
1	3	3	13	1	20	2	2	2	15	5	46	3	9	3	23	3	70
1	3	1	15	4	23	2	7	3	10	4	53	3	9	1	22	5	65
1	3	5	16	2	22	2	7	4	11	5	29	3	8	2	29	4	64
1	3	5	18	3	21	2	7	3	13	6	32	3	8	5	19	2	53
1	3	4	20	5	24	2	3	4	17	5	43	3	8	3	18	3	47
1	3	3	5	4	25	2	8	4	9	4	35	3	8	2	14	4	48
1	4	2	7	3	23	2	8	3	9	3	43	3	7	3	13	5	53
1	4	3	9	6	26	2	9	4	21	2	46						