

6. DESCRIPCIÓN BIVARIANTE

El contexto sustantivo correspondiente a los datos de la tabla 6.1 se encuentra disponible en el capítulo 5.

Tabla 6.1. Matriz de datos correspondientes al estudio ficticio.

Id	Género	Nivel socio €	Terapia	Calidad	STAI-E	BDI	BDI.Cat
1	H	Bajo	PsicoD	3	45	29	grave
2	H	Bajo	PsicoD	5	40	27	grave
3	H	Medio	PsicoD	4	42	22	grave
4	H	Medio	PsicoD	2	38	26	grave
5	H	Alto	PsicoD	1	47	32	grave
6	M	Bajo	PsicoD	2	50	52	severa
7	M	Medio	PsicoD	4	25	20	interm
8	M	Medio	PsicoD	3	35	25	grave
9	M	Alto	PsicoD	2	48	35	grave
10	M	Alto	PsicoD	2	44	28	grave
11	H	Bajo	TCognitiva	5	20	18	interm
12	H	Medio	TCognitiva	3	18	13	interm
13	H	Medio	TCognitiva	4	14	10	interm
14	H	Medio	TCognitiva	2	19	21	grave
15	H	Alto	TCognitiva	2	39	27	grave
16	M	Medio	TCognitiva	5	11	9	interm
17	M	Medio	TCognitiva	3	32	25	grave
18	M	Medio	TCognitiva	3	17	12	interm
19	M	Alto	TCognitiva	4	14	16	interm
20	M	Alto	TCognitiva	3	20	19	interm

Ejercicio 1. Variables medidas en escala nominal

- a) Estudiar la fuerza de la relación entre el grado de depresión (variable recodificada con categorías «interm», «grave» y «severa») y el género.

Solución. Puesto que estamos interesados en estudiar la relación o asociación entre dos variables, se trataría de realizar un análisis estadístico propio de la descripción bivalente. Debe notarse que la pregunta hace referencia a una variable ordinal (grado de depresión) y una variable categórica (género). Así, sería correcto utilizar índices apropiados para dos variables categóricas (i.e., la escala de medida inferior de las

dos): *ji*-cuadrado para estudiar si las variables son independientes o no y *V* de Cramér para cuantificar la fuerza o intensidad de la relación cuando esta exista.

En primer lugar, es necesario construir la tabla de contingencia para contar el número de veces que se da cada cruce de categorías. Habiendo tres categorías de grados de depresión y dos categorías de género, hay un total de seis cruces posibles: son las casillas enmarcadas mediante una línea gruesa.

Tabla 6.2. Tabla de contingencia: frecuencias observadas o empíricas en la distribución conjunta para las variables depresión y género.

FRECUENCIAS OBSERVADAS	GÉNERO		Suma
	Hombre	Mujer	
Depresión			
Interm.	3	5	8
Grave	7	4	11
Severa	0	1	1
Suma	10	10	20

Empezamos el recuento con el número de casos en los que el participante es hombre y su grado de depresión después de la terapia es «interm»: hay tres personas que cumplen estas características: identificadores 11, 12 y 13. Contando el número de hombres con depresión «grave» obtenemos 7 casos: identificadores 1, 2, 3, 4, 5, 14 y 15. Saber que hay 10 hombres en la muestra (la suma de la columna) implica que no hay hombres con depresión «severa»: a la correspondiente casilla se le asigna el valor 0. Centrándonos en las mujeres, 5 de ellas presentan depresión «interm» (identificadores 7, 16, 18, 19 y 20), mientras que hay 4 con depresión «grave» (identificadores 8, 9, 10 y 17), presentando solo la persona con identificador 15 una puntuación categorizada como depresión «severa».

Una vez hecho el recuento de las frecuencias observadas, parece haber más hombres con depresión grave y más mujeres con depresión moderada. Por tanto, aparentemente hay relación entre las dos variables, puesto que el hecho de ser hombre o mujer cambia la probabilidad, en esta muestra, de presentar uno u otro grado de depresión. Sin embargo, para poder valorar mejor si hay relación entre las variables, es necesario obtener las frecuencias esperadas en el caso de que fueran independientes, y comparar ambos conjuntos de frecuencias: observadas (o empíricas) y esperadas (o teóricas).

Las frecuencias esperadas se obtienen a partir de los marginales de la tabla, es decir, basándonos en las sumas de filas y columnas. Por ejemplo, para reconstruir las frecuencias esperadas al azar de la primera casilla, se parte de la probabilidad de ser hombre (igual a 0,5, puesto que 10 de los 20 participantes en el estudio son hombres) y la probabilidad de tener depresión más leve (igual a 0,4, puesto que 8 de los 20 participantes en el estudio están categorizados con depresión «interm»). Para obtener la frecuencia esperada en caso de independencia entre género y grado de depresión se han de multiplicar las probabilidades de ambos sucesos ($0,5 \cdot 0,4 = 0,2$) y tener en cuenta el número de participantes en la muestra; esto es, si la probabilidad de que se dé la intersección de ambos sucesos (ser hombre y presentar depresión «interm») es de 0,2 y hay 20 personas, se espera que 4 de ellos ($0,2 \cdot 20 = 4$) sean hombres con depresión «interm».

A continuación se obtienen todas las frecuencias esperadas, teniendo en cuenta que, al haber la misma cantidad de hombres y mujeres en la muestra, las frecuencias esperadas para las tres categorías de depresión serán las mismas para ambos géneros:

$$f_{e.H.interm} = \frac{f_{o.H}}{n} \frac{f_{o.interm}}{n} n = \frac{10}{20} \frac{8}{20} 20 = 0,5 \cdot 0,4 \cdot 20 = 0,2 \cdot 20 = 4.$$

$$f_{e.M.interm} = \frac{f_{o.M}}{n} \frac{f_{o.interm}}{n} n = \frac{10}{20} \frac{8}{20} 20 = 0,5 \cdot 0,4 \cdot 20 = 0,2 \cdot 20 = 4.$$

$$f_{e.H.grave} = \frac{f_{o.H}}{n} \frac{f_{o.grave}}{n} n = \frac{10}{20} \frac{11}{20} 20 = 0,5 \cdot 0,55 \cdot 20 = 0,275 \cdot 20 = 5,5.$$

$$f_{e.M.grave} = \frac{f_{o.M}}{n} \frac{f_{o.grave}}{n} n = \frac{10}{20} \frac{11}{20} 20 = 0,5 \cdot 0,55 \cdot 20 = 0,275 \cdot 20 = 5,5.$$

$$f_{e.H.severa} = \frac{f_{o.H}}{n} \frac{f_{o.severa}}{n} n = \frac{10}{20} \frac{1}{20} 20 = 0,5 \cdot 0,05 \cdot 20 = 0,025 \cdot 20 = 0,5.$$

$$f_{e.M.severa} = \frac{f_{o.M}}{n} \frac{f_{o.severa}}{n} n = \frac{10}{20} \frac{1}{20} 20 = 0,5 \cdot 0,05 \cdot 20 = 0,025 \cdot 20 = 0,5.$$

Tabla 6.3. Tabla de contingencia: frecuencias esperadas o teóricas en la distribución conjunta para las variables depresión y género.

FRECUENCIAS ESPERADAS	GÉNERO		
	Depresión	Hombre	Mujer
Interm		4	4
Grave		5,5	5,5
Severa		0,5	0,5
Suma		10	10

El lector puede advertir que hay más hombres con depresión grave y menos con depresión más leve o intermedia que lo esperable en el caso de que las variables fueran independientes. No obstante, la comparación no puede hacerse solo de manera aproximada, sino que se formaliza mediante el índice *ji*-cuadrado, que precisamente calcula el cuadrado de la diferencia entre frecuencias observadas y esperadas para cada uno de los cruces entre las $I = 2$ categorías de género y $J = 3$ grados de depresión.

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \left[\frac{(f_{o.ij} - f_{e.ij})^2}{f_{e.ij}} \right]$$

$$\chi^2 = \frac{(3-4)^2}{4} + \frac{(5-4)^2}{4} + \frac{(7-5,5)^2}{5,5} + \frac{(4-5,5)^2}{5,5} + \frac{(0-0,5)^2}{0,5} + \frac{(1-0,5)^2}{0,5}$$

$$\chi^2 = \frac{1}{4} + \frac{1}{4} + \frac{2,25}{5,5} + \frac{2,25}{5,5} + \frac{0,25}{0,5} + \frac{0,25}{0,5} \approx 0,25 + 0,25 + 0,41 + 0,41 + 0,5 + 0,5.$$

$$\chi^2 = 2,32.$$

Observamos que el valor de *ji*-cuadrado es superior a 0, lo que nos indica que las dos variables, género y grado de depresión, no son independientes. Sin embargo, este

índice no está superiormente acotado y su valor se ve influido, no solo por la existencia de relación, sino también por el tamaño muestral y el número de categorías. Por tanto, para cuantificar la fuerza de la relación se utilizará un índice con idéntica cota superior, independiente de las dimensiones de la tabla de contingencia y del número de participantes. Este índice se conoce como V de Cramér, cuya fórmula incorpora la dimensión de la tabla ($q = I$ o $q = J$, en función de cuál de las dos variables dispone de menos categorías) y el tamaño de la muestra, calculándose como aparece a continuación:

$$V = \sqrt{\frac{\chi^2}{n(q-1)}} = \sqrt{\frac{2,32}{20(2-1)}} = \sqrt{0,116} \approx 0,34.$$

Este valor nos indica que se trata de una relación de intensidad moderada (i. e., entre 0,3 y 0,5, siendo 0 el mínimo de V y 1 el máximo). La relación entre las variables también puede ilustrarse mediante representaciones gráficas:

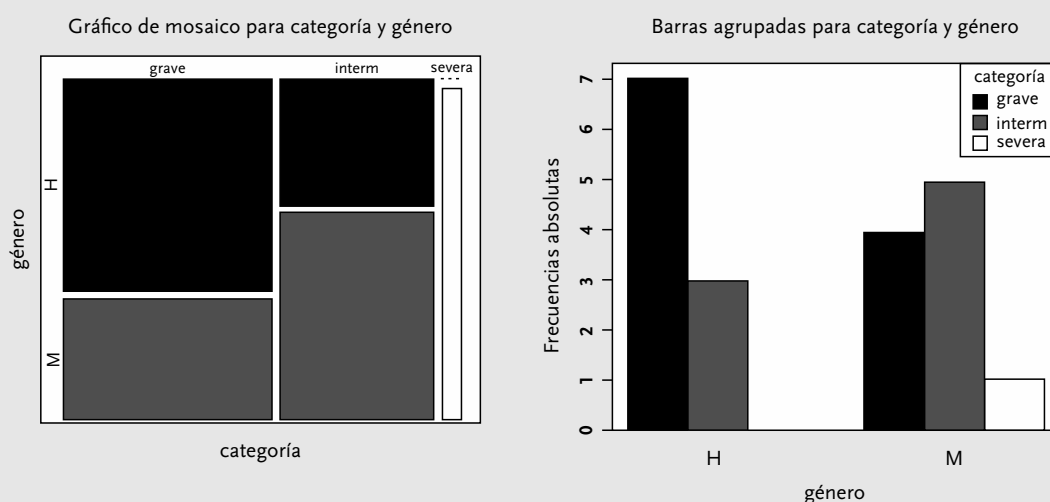


Figura 6.1. Gráfico de mosaico (a la izquierda) y diagrama de barras agrupadas (a la derecha). Representan las frecuencias relativas y absolutas del cruce de categorías de las variables género y grado de depresión.

- b) Explorar si existe relación entre la terapia administrada y la intensidad de la depresión padecida según los puntos de corte del *BDI*.

Solución. Puesto que estamos interesados en estudiar la relación o asociación entre dos variables, se trataría de realizar un análisis estadístico propio de la descripción bivalente. La pregunta hace referencia a una variable nominal (tipo de terapia) y una variable ordinal (la gravedad de la depresión). En este caso, sería correcto utilizar índices apropiados para dos variables nominales (i. e., la escala de medida inferior de las dos). Concretamente, puesto que la pregunta está redactada en términos de riesgo de padecer una enfermedad, el índice apropiado es la *odds ratio*.

Para poder responder a la pregunta es necesario recategorizar la variable *BDI_Cat*, que representa un resumen de las puntuaciones en el *BDI*. Categorizaremos como «depresión» a aquellos valores superiores a 20 (i. e., las categorías grave y

severa), mientras que etiquetaremos mediante «no depresión» al resto de casos (i.e., la categoría interm). Con esta recategorización, la tabla de contingencia tendrá dos niveles para «Terapia» y dos niveles para «Estado». Para construir la tabla de contingencia hay que contar la cantidad de personas que han seguido cada una de las dos terapias y que han obtenido puntuaciones *BDI* inferiores o iguales a 20 («no depresión») o superiores a 20 («depresión»). La única persona sin depresión en el grupo que siguió terapia psicodinámica es el Id.7, mientras que entre las personas que siguieron terapia cognitiva hay solo tres que presentaron puntuaciones propias de una depresión: identificadores 14, 15 y 17.

Tabla 6.4. Tabla de contingencia: frecuencias observadas o empíricas en la distribución conjunta para las variables terapia y depresión, esta última definida como la presencia o ausencia de psicopatología en función de los puntos de corte.

FRECUENCIAS OBSERVADAS	ESTADO		Suma
	Depresión	No depresión	
Terapia			
Psicodinámica	9	9	10
Cognitiva	3	7	10
Suma	12	8	20

Para conocer la *odds ratio*, en primer lugar es preciso obtener las *odds* para padecer depresión en cada uno de los grupos por separado:

$$Odds_{PsicoD} = \frac{f_{depresión}}{f_{no}} = \frac{9}{1} = 9.$$

$$Odds_{TCognitiva} = \frac{f_{depresión}}{f_{no}} = \frac{3}{7} = 0,43.$$

Se observa como en el grupo que siguió terapia psicodinámica es nueve veces más probable padecer depresión que no padecerla: es decir hay 9 personas que la padecen por cada persona que no. En cambio, entre las personas que siguieron terapia cognitiva, la probabilidad de padecer depresión es menos de la mitad de la probabilidad de no padecerla. Si expresáramos la *odds* de una manera alternativa $Odds_{TCognitiva} = f_{no}/f_{depresión} = 7/3 = 2,33$, se podría decir que hay 2,33 personas sin depresión por cada persona que la padece tras la terapia cognitiva. De esta manera, ya se intuye que la terapia psicodinámica se asocia con una mayor probabilidad de padecer depresión, pero ¿cuánto? La *odds ratio* nos ayudará a responder a esta pregunta.

$$OR = \frac{Odds_{PsicoD}}{Odds_{TCognitiva}} = \frac{\frac{f_{PsicoD.depresión}}{f_{PsicoD.no}}}{\frac{f_{TCognitiva.depresión}}{f_{TCognitiva.no}}} = \frac{f_{PsicoD.depresión}}{f_{PsicoD.no}} \cdot \frac{f_{TCognitiva.no}}{f_{TCognitiva.depresión}}$$

$$OR = \frac{Odds_{PsicoD}}{Odds_{TCognitiva}} = \frac{9}{0,43} \approx 20,93$$

$$OR = \frac{f_{PsicoD.depresión}}{f_{PsicoD.no}} \cdot \frac{f_{TCognitiva.no}}{f_{TCognitiva.depresión}} = \frac{9}{1} \cdot \frac{7}{3} \approx 20,93.$$

Observamos que el hecho de haber seguido terapia psicodinámica se asocia con un riesgo veinte veces mayor de padecer depresión.

Ejercicio 2. Variables medidas en escala ordinal

- a) Estudiar la fuerza o intensidad de la relación entre el grado de depresión (variable recodificada «categoría») y el nivel socioeconómico en las mujeres que han seguido terapia psicodinámica.

Solución. Dado que estamos interesados en estudiar la relación o asociación entre dos variables, se trataría de realizar un análisis estadístico propio de la descripción bivariable. La pregunta hace referencia a dos variables ordinales. En este caso, sería correcto utilizar el índice *tau-b* de Kendall para tener en cuenta el número de empates y considerando que ambas variables tienen tres categorías diferentes (i. e., se obtiene una tabla de contingencia cuadrada).

Para calcular el índice *tau-b* de Kendall, es necesario obtener el número de concordancias, discordancias y empates entre el total de comparaciones posibles entre pares de individuos. De hecho, como se puede comprobar en la fórmula que se presenta a continuación, el número de concordancias (C) y discordancias (D) es la base del cálculo del numerador, mientras que el denominador incluye el número de comparaciones $n(n-1)/2$, restándose la cantidad de empates (E_X y E_Y) en cada una de las dos variables:

$$\tau_b = \frac{C - D}{\sqrt{\left(\frac{n(n-1)}{2} - E_Y\right)\left(\frac{n(n-1)}{2} - E_X\right)}}$$

Para obtener estos valores, recuperaremos parte de la matriz de datos original —la que contiene la información referente a las mujeres que han seguido terapia psicodinámica—. Para facilitar la comparación asignaremos rangos que reflejen el orden de los valores en las dos variables de interés: nivel socioeconómico y grado de depresión.

Tabla 6.5. Datos correspondientes al presente ejercicio.

Id	Género	Nivel socio €	Nivel socio € rango	BDI_Cat	BDI_Cat-rango
6	M	Bajo	1	severa	3
7	M	Medio	2	interm	1
8	M	Medio	2	grave	2
9	M	Alto	3	grave	2
10	M	Alto	3	grave	2

- *Comparación 1:* Id.6 con Id.7: Id.6 tiene un nivel socioeconómico más bajo y depresión más grave. Por tanto, contabilizamos la primera discordancia.
- *Comparación 2:* Id.6 con Id.8: Id.6 tiene un nivel socioeconómico más bajo y depresión más grave. Por tanto, contabilizamos la segunda discordancia.

- *Comparación 3:* Id.6 con Id.9: Id.6 tiene un nivel socioeconómico más bajo y depresión más grave. Por tanto, contabilizamos la tercera discordancia.
- *Comparación 4:* Id.6 con Id.10: Id.6 tiene un nivel socioeconómico más bajo y depresión más grave. Por tanto, contabilizamos la cuarta discordancia.
- *Comparación 5:* Id.7 con Id.8: Los dos individuos tienen el mismo nivel socioeconómico. Por tanto, contabilizamos el primer empate en E_X .
- *Comparación 6:* Id.7 con Id.9: Id.7 tiene un nivel socioeconómico más bajo y depresión menos grave que Id.9. Por tanto, contabilizamos la primera concordancia.
- *Comparación 7:* Id.7 con Id.10: Id.7 tiene un nivel socioeconómico más bajo y depresión menos grave que Id.10. Por tanto, contabilizamos la segunda concordancia.
- *Comparación 8:* Id.8 con Id.9: Los dos individuos tienen el mismo grado de depresión. Por tanto, contabilizamos el primer empate en E_Y .
- *Comparación 9:* Id.8 con Id.10: Los dos individuos tienen el mismo grado de depresión. Por tanto, contabilizamos el segundo empate en E_Y .
- *Comparación 10:* Id.9 con Id.10: Los dos individuos tienen el mismo grado de depresión y el mismo nivel socioeconómico. Contabilizamos este resultado como el tercer empate en E_Y , pero no lo contabilizaremos también como empate en E_X para no contarlo dos veces.

Se puede observar que el número de comparaciones posibles es 10, tal como indica la expresión

$$\frac{n(n-1)}{2} = \frac{5(5-1)}{2} = \frac{20}{2} = 10.$$

En resumen: a) número de concordancias, $C = 2$; b) número de discordancias, $D = 4$; c) número de empates en el nivel socioeconómico, $E_X = 1$; d) número de empates en el grado de depresión, $E_Y = 3$. Utilizando estos valores calcularemos el índice *tau-b* de Kendall para valorar la existencia, el sentido y la fuerza de la relación entre el nivel socioeconómico y el grado de depresión en las 5 mujeres que han seguido terapia psicodinámica.

$$\tau_b = \frac{C - D}{\sqrt{\left(\frac{n(n-1)}{2} - E_Y\right)\left(\frac{n(n-1)}{2} - E_X\right)}} = \frac{2 - 4}{\sqrt{(10-3)(10-1)}}$$

$$\tau_b = \frac{-2}{\sqrt{7 \cdot 9}} = \frac{-2}{\sqrt{63}} \approx \frac{-2}{7,94} \approx -0,25.$$

El resultado obtenido sugiere que: a) hay relación puesto que $\tau_b \neq 0$; b) la relación es inversa, ya que $\tau_b < 0$, es decir, a mayor nivel socioeconómico menor grado de depresión; c) la fuerza de la relación es más bien baja, puesto que el valor obtenido es más cercano al 0 que a la cota inferior, -1 . Este mismo resultado se puede representar gráficamente. En el gráfico también se observa por qué decimos que la tabla de contingencia sería cuadrada en este caso (en ambas variables los rangos posibles son 1, 2 y 3 según el orden de las categorías).

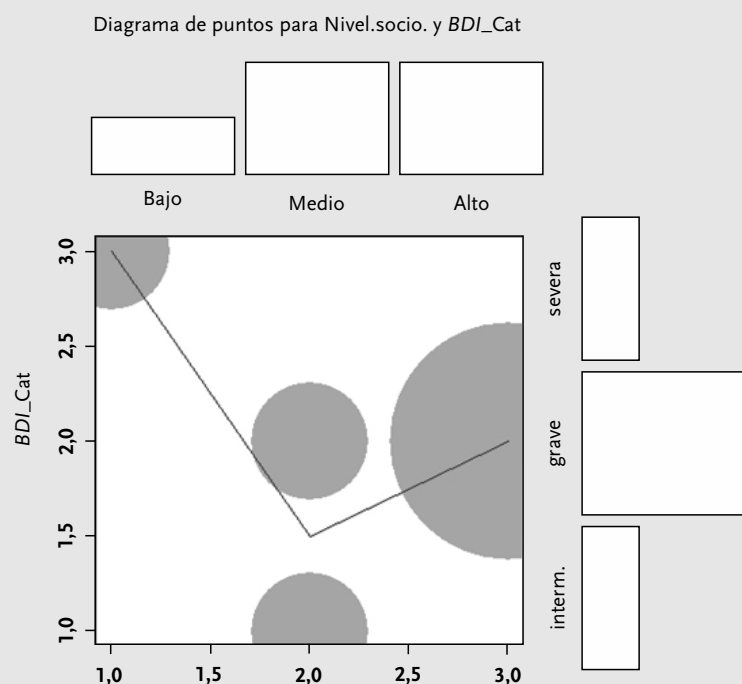


Figura 6.2. Diagrama de puntos que representa la relación entre el nivel socioeconómico y el grado de depresión.

La tabla de contingencia en la que se basa este gráfico es la siguiente:

Tabla 6.6. Tabla de contingencia para representar los datos de la Figura 6.2.

FRECUENCIAS OBSERVADAS	NIVEL SOCIOECONÓMICO		
	Bajo	Medio	Alto
BDI_Cat			
Severa	1	0	0
Grave	0	1	2
Interm	0	1	0

- b) ¿Están relacionados los valores de la prueba estandarizada *BDI* con las respuestas al ítem sobre la calidad de vida para las personas de nivel socioeconómico medio que han seguido terapia psicodinámica?

Solución. Puesto que estamos interesados en estudiar la relación o asociación entre dos variables, se trataría de realizar un análisis estadístico propio de la descripción bivalente. La pregunta hace referencia a una variable ordinal (la puntuación en el ítem de calidad de vida) y una variable de escala de intervalo (puntuación en la prueba estandarizada *BDI*), es decir, se trata de variables de diferentes escalas de medida. Así pues, trataremos la variable de intervalo como si fuera ordinal y utilizaremos índices para estudiar la relación entre variables ordinales. Siempre se puede «bajar» de escala a una variable, pero no es posible «subir» de escala (i.e., no es posible estudiar la relación entre calidad de vida y puntuación en el *BDI* como si ambas fueran variables medidas en escala de intervalo).

Empezaremos el estudio de la relación mediante la *tau-c* de Kendall y Stuart, apropiada para casos en los cuales las dos variables no tienen la misma cantidad de valores. En otros términos, *tau-c* es adecuada para tablas de contingencia rectangulares, pero nótese que no tiene en cuenta el número de empates, aunque estos últimos son menos probables si una de las dos variables (puntuación en el *BDI*) está medida en escala de intervalo con valores posibles entre 0 y 63. La expresión para el cálculo es:

$$\tau_c = \frac{2q(C-D)}{n^2(q-1)}$$

Para obtener el valor de *tau-c* hay que contar el número de concordancias y discordancias para la parte de la matriz de datos a la que la pregunta se refiere: las $n = 4$ personas con nivel socioeconómico medio y que siguen terapia psicodinámica.

Tabla 6.7. Datos correspondientes al presente ejercicio.

Id	Nivel socio €	Terapia	Calidad	BDI
3	Medio	PsicoD	4	22
4	Medio	PsicoD	2	26
7	Medio	PsicoD	4	20
8	Medio	PsicoD	3	25

Se tienen que llevar a cabo las siguientes comparaciones para determinar las concordancias y discordancias:

- *Comparación 1:* Id.3 con Id.4: Id.3 evalúa su calidad de vida con un valor más alto y tiene una puntuación más baja en depresión. Por tanto, contabilizamos la primera discordancia.
- *Comparación 2:* Id.3 con Id.7: Id.3 evalúa su calidad de vida con el mismo valor y tiene una puntuación más baja en depresión. Por tanto, contabilizamos el primer empate, aunque estos no forman parte del cálculo de *tau-c*.
- *Comparación 3:* Id.3 con Id.8: Id.3 evalúa su calidad de vida con un valor más alto y tiene una puntuación más baja en depresión. Por tanto, contabilizamos la segunda discordancia.
- *Comparación 4:* Id.4 con Id.7: Id.4 evalúa su calidad de vida con un valor más bajo y tiene una puntuación más alta en depresión. Por tanto, contabilizamos la tercera discordancia.
- *Comparación 5:* Id.4 con Id.8: Id.4 evalúa su calidad de vida con un valor más bajo y tiene una puntuación más alta en depresión. Por tanto, contabilizamos la cuarta discordancia.
- *Comparación 6:* Id.7 con Id.8: Id.7 evalúa su calidad de vida con un valor más alto y tiene una puntuación más baja en depresión. Por tanto, contabilizamos la quinta discordancia.

Esta información todavía no es suficiente para obtener *tau-c*, puesto que necesitamos conocer q , que representa el número de valores diferentes de la variable que

dispone de menos valores. En este caso, la variable calidad de vida presenta solo 3 valores diferentes y, por consiguiente, $q = 3$. Así,

$$\tau_c = \frac{2q(C-D)}{n^2(q-1)} = \frac{2 \cdot 3(0-5)}{4^2(3-1)} = \frac{6(-5)}{16 \cdot 2} = \frac{-30}{32} \approx -0,94.$$

Como se desprende de la gran proporción de discordancias observadas, existe una asociación entre ambas variables, siendo esta relación inversa (a menor depresión, mayor calidad de vida). Además, observamos que dicha relación es muy intensa, puesto que el valor de *tau-c* se encuentra próximo a su cota inferior de -1 . Esta relación se puede apreciar también en el diagrama de dispersión.

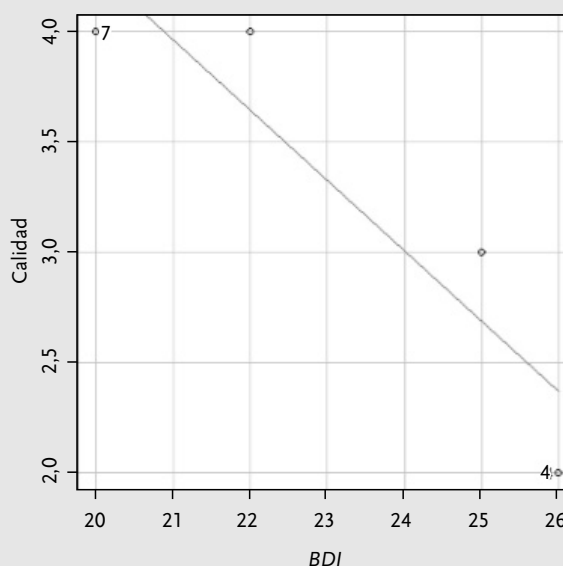


Figura 6.3. Diagrama de dispersión que representa la relación entre la puntuación en la escala *BDI* y la valoración de la calidad de vida.

Finalmente, construiremos la tabla de contingencia teniendo en cuenta los valores de cada variable que, de hecho, se han dado (en vez de considerar todos los valores que podrían haberse dado) para las personas de nivel socioeconómico medio que siguieron terapia psicodinámica. Como puede observarse, se trata de una tabla rectangular; también lo sería si se tuviera en cuenta que la variable de calidad de vida puede tomar 5 valores diferentes, mientras que la puntuación en el *BDI* puede tomar 64 valores, de 0 a 63, ambos inclusive.

Tabla 6.8. Tabla de contingencia para representar los datos de la Figura 6.3.

FRECUENCIAS OBSERVADAS	PUNTUACIÓN EN EL <i>BDI</i>			
	20	22	25	26
Calidad				
4	1	1	0	0
3	0	0	1	0
2	0	0	0	1

Aparte de calcular el índice *tau-c* de Kendall, en el caso de una variable ordinal y una variable medida en escala de intervalo, es posible utilizar el coeficiente de corre-

lación por rangos de Spearman: este índice, a diferencia del coeficiente de correlación de Pearson, es apropiado tanto para escalas ordinales, como para escalas superiores. Para calcular el coeficiente de correlación de Spearman, o coeficiente de correlación por rangos, se hallan diversas expresiones en la literatura, pero solo proporcionan el mismo valor cuando no hay empates. Es muy común encontrar la siguiente expresión para obtener el coeficiente de correlación por rangos, y es la que, en principio, se utilizará:

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

Una vez más, los valores originales (x_i e y_i) se transforman en rangos (r_i y s_i), asignando el rango 1 al valor más pequeño y el rango n al valor más grande. Posteriormente se calcula la diferencia d_i entre estos rangos:

Tabla 6.9. Asignación de rangos y cálculo de la diferencia entre estos.

Id	Nivel socio €	Terapia	Calidad x_i	Calidad Rango r_i	BDI y_i	BDI Rango s_i	Diferencia de rangos $d_i = r_i - s_i$
3	Medio	PsicoD	4	3,5	22	2	1,5
4	Medio	PsicoD	2	1,0	26	4	-3,0
7	Medio	PsicoD	4	3,5	20	1	2,5
8	Medio	PsicoD	3	2,0	25	3	-1,0

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} = 1 - \frac{6[(1,5)^2 + (-3)^2 + (2,5)^2 + (-1)^2]}{4(4^2 - 1)}$$

$$r_s = 1 - \frac{6[2,25 + 9 + 6,25 + 1]}{4(16 - 1)} = 1 - \frac{6 \cdot 18,5}{4 \cdot 15} = 1 - \frac{111}{60}$$

$$r_s = 1 - 1,85 = -0,85.$$

El resultado obtenido indica una fuerte relación negativa entre las dos variables, tal como se desprendía del índice *tau-c* de Kendall y Stuart. Ahora bien, se ha indicado que, cuando existen empates, las distintas expresiones de cálculo conducen a resultados que discrepan. Al respecto, en los datos tenemos un empate; concretamente, los identificadores 3 y 7 tienen el mismo rango para la variable calidad de vida. Utilizaremos a continuación, para calcular el coeficiente de correlación por rangos, la expresión que se fundamenta en el coeficiente de correlación de Pearson, que es como se define el coeficiente de correlación de Spearman:

$$r_s = \frac{\sum_{i=1}^n (r_i - \bar{r})(s_i - \bar{s})}{\sqrt{\sum_{i=1}^n (r_i - \bar{r})^2 \sum_{i=1}^n (s_i - \bar{s})^2}}$$

Ya disponemos de los datos convertidos en rangos y las medias de estos rangos también se conocen, pero su cálculo se ilustrará igualmente:

$$\overline{\text{Calidad(rango)}} \equiv \bar{r} = \frac{\sum_{i=1}^n r_i}{n} = \frac{3,5 + 1 + 3,5 + 2}{4} = \frac{10}{4} = 2,5$$

$$\overline{\text{BDI(rango)}} \equiv \bar{s} = \frac{\sum_{i=1}^n s_i}{n} = \frac{2 + 4 + 1 + 3}{4} = \frac{10}{4} = 2,5.$$

Es posible calcular la correlación de Spearman obteniendo en primer lugar el numerador (la suma de productos cruzados), para posteriormente obtener el denominador (las desviaciones elevadas al cuadrado).

Numerador:

$$\sum_{i=1}^n (r_i - \bar{r})(s_i - \bar{s}) = (3,5 - 2,5)(2 - 2,5) + \dots + (2 - 2,5)(3 - 2,5)$$

$$\sum_{i=1}^n (r_i - \bar{r})(s_i - \bar{s}) = (1)(-0,5) + \dots + (-0,5)(0,5) = -0,5 + \dots + (-0,25) = -4,5.$$

Parte izquierda de la raíz cuadrada del denominador:

$$\sum_{i=1}^n (r_i - \bar{r})^2 = (3,5 - 2,5)^2 + (1 - 2,5)^2 + (3,5 - 2,5)^2 + (2 - 2,5)^2$$

$$\sum_{i=1}^n (r_i - \bar{r})^2 = (1)^2 + (-1,5)^2 + (1)^2 + (-0,5)^2 = 1 + 2,25 + 1 + 0,25 = 4,5.$$

Parte derecha de la raíz cuadrada del denominador:

$$\sum_{i=1}^n (s_i - \bar{s})^2 = (2 - 2,5)^2 + (4 - 2,5)^2 + (1 - 2,5)^2 + (3 - 2,5)^2$$

$$\sum_{i=1}^n (s_i - \bar{s})^2 = (-0,5)^2 + (1,5)^2 + (-1,5)^2 + (0,5)^2 = 0,25 + 2,25 + 2,25 + 0,25$$

$$\sum_{i=1}^n (s_i - \bar{s})^2 = 5.$$

Estos mismos cálculos se pueden realizar utilizando una tabla con los rangos en la que se incorporan como columnas las variables que representan diferencias respecto a las medias, tal como se muestra a continuación:

Tabla 6.10. Cálculos necesarios para obtener el valor del coeficiente de correlación por rangos.

Id	BDI s_i	BDI.d $(s_i - \bar{s})$	BDI.d2 $(s_i - \bar{s})^2$	Calidad r_i	Calidad.d $(r_i - \bar{r})$	Calidad.d2 $(r_i - \bar{r})^2$	BDI.d*Calidad.d $(s_i - \bar{s})(r_i - \bar{r})$
3	2	-0,50	0,25	4	1,00	1,00	-0,50
4	4	1,50	2,25	2	-1,50	2,25	-2,25
7	1	-1,50	2,25	4	1,00	1,00	-1,50
8	3	0,50	0,25	3	-0,50	0,25	-0,25
$\bar{s} = 2,5$		$\sum_{i=1}^n = 5$		$\bar{r} = 2,5$		$\sum_{i=1}^n = 4,5$	$\sum_{i=1}^n = -4,5$

Finalmente,

$$r_s = \frac{\sum_{i=1}^n (r_i - \bar{r})(s_i - \bar{s})}{\sqrt{\sum_{i=1}^n (r_i - \bar{r})^2 \sum_{i=1}^n (s_i - \bar{s})^2}} = \frac{-4,5}{\sqrt{4,5 \cdot 5}} = \frac{-4,5}{\sqrt{22,5}} \approx \frac{-4,5}{4,74} \approx -0,949.$$

Se puede comprobar que el valor obtenido mediante esta última fórmula es diferente del que se obtuvo inicialmente. De hecho, cuando existen empates, se debe utilizar una expresión de cálculo que aquí no se trata.

Ejercicio 3. Variables medidas en escala de intervalo o razón

- a) Estudiar si la ansiedad estado podría ser una de las explicaciones de la variabilidad en las puntuaciones *BDI*.

Solución. Por una parte, puesto que estamos interesados en estudiar la relación o asociación entre dos variables, se debería realizar un análisis estadístico propio de la descripción bivalente. Por otra, la pregunta hace referencia a dos variables cuantitativas, suponiendo, como es común, que la suma de ítems en la escala Likert (ordinal) pudiera dar lugar a una variable medida en escala de intervalo. En este caso, sería correcto utilizar el coeficiente de correlación producto-momento de Pearson para cuantificar una relación lineal, o el coeficiente de correlación por rangos de Spearman si se detecta un patrón no lineal. Por tanto, antes de proceder con el cálculo, es necesario explorar visualmente el diagrama de dispersión.

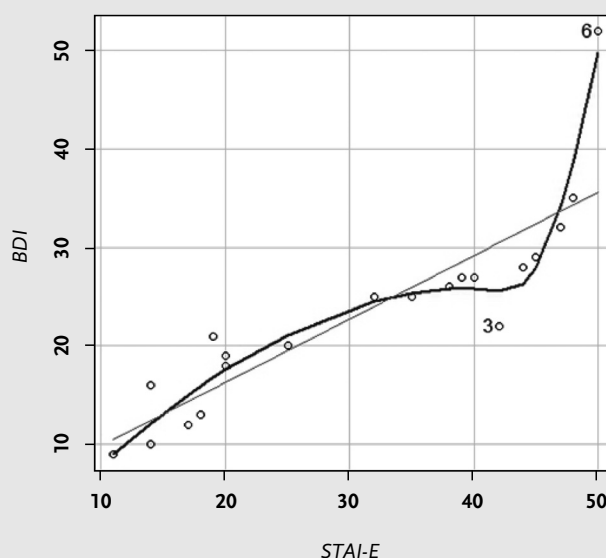


Figura 6.4. Diagrama de dispersión que representa la relación entre la puntuación en la escala STAI que mide ansiedad como estado y la puntuación en el inventario de depresión de Beck.

Si analizamos el gráfico, parece claro que hay un componente lineal en los datos y, concretamente, en el incremento que se da en la puntuación en el *BDI* al mismo

tiempo que incrementa la ansiedad estado. En el gráfico se identifican dos valores alejados del conjunto de medidas, ilustrando uno de ellos (Id.6) la tendencia de que valores altos en una variable se corresponden con valores altos en la otra, pero, al mismo tiempo, no concuerda con la relación lineal. Por consiguiente, aparte del coeficiente de correlación de Pearson para cuantificar el componente lineal de la relación entre las variables, puede también utilizarse el coeficiente de correlación por rangos para verificar si se da una relación monótona creciente que no sea necesariamente lineal. En primer lugar, procederemos a calcular la correlación de Pearson.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Como se puede apreciar en la fórmula, es necesario disponer de la media de ambas variables (puntuación en el *BDI*, que aquí llamaremos *X*, y en el *STAI-E*, que llamaremos *Y*) para posteriormente calcular las desviaciones de cada dato de su media.

$$\overline{BDI} \equiv \bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{29 + 27 + \dots + 28 + 18 + 13 + \dots + 19}{20} = \frac{464}{20} = 23,3$$

$$\overline{STAI-E} \equiv \bar{y} = \frac{\sum_{i=1}^n y_i}{n} = \frac{45 + 40 + \dots + 44 + 20 + 18 + \dots + 20}{20} = \frac{618}{20} = 30,9.$$

Es posible determinar la correlación de Pearson calculando en primer lugar el numerador (la suma de productos cruzados) y luego el denominador (desviaciones al cuadrado).

Numerador (que de hecho es el numerador de la covariancia):

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = (29 - 23,3)(45 - 30,9) + \dots + (19 - 23,3)(20 - 30,9)$$

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = (5,7)(14,1) + \dots + (-4,3)(-10,9) = 80,37 + \dots + 46,87 = 2.181,6.$$

Parte izquierda de la raíz cuadrada del denominador (numerador de la variancia de la puntuación en el *BDI*):

$$\sum_{i=1}^n (x_i - \bar{x})^2 = (29 - 23,3)^2 + (27 - 23,3)^2 + \dots + (19 - 23,3)^2$$

$$\sum_{i=1}^n (x_i - \bar{x})^2 = (5,7)^2 + (3,7)^2 + \dots + (-4,3)^2 = 32,49 + 13,69 + \dots + 18,49$$

$$\sum_{i=1}^n (x_i - \bar{x})^2 = 1.864,2.$$

Parte derecha de la raíz cuadrada del denominador (que, de hecho, es el numerador de la variancia de la puntuación en el *STAI-E*):

$$\sum_{i=1}^n (y_i - \bar{y})^2 = (45 - 30,9)^2 + (40 - 30,9)^2 + \dots + (20 - 30,9)^2$$

$$\sum_{i=1}^n (y_i - \bar{y})^2 = (14,1)^2 + (9,9)^2 + \dots + (-10,9)^2 = 198,81 + 82,81 + \dots + 118,81$$

$$\sum_{i=1}^n (y_i - \bar{y})^2 = 3.367,8.$$

Estos mismos cálculos se pueden realizar utilizando la Tabla 6.11, en la que se incorporan como columnas las variables que representan diferencias respecto a las medias.

Tabla 6.11. Cálculos necesarios para determinar el valor del coeficiente de correlación lineal producto-momento.

Id	<i>STAI</i> y_i	<i>STAI.d</i> $(y_i - \bar{y})$	<i>STAI.d2</i> $(y_i - \bar{y})^2$	<i>BDI</i> x_i	<i>BDI.d</i> $(x_i - \bar{x})$	<i>BDI.d2</i> $(x_i - \bar{x})^2$	<i>STAI.d*BDI.d</i> $(y_i - \bar{y})(x_i - \bar{x})$
1	45	14,1	198,81	29	5,7	32,49	80,37
2	40	9,1	82,81	27	3,7	13,69	33,67
3	42	11,1	123,21	22	-1,3	1,69	-14,43
4	38	7,1	50,41	26	2,7	7,29	19,17
5	47	16,1	259,21	32	8,7	75,69	140,07
6	50	19,1	364,81	52	28,7	823,69	548,17
7	25	-5,9	34,81	20	-3,3	10,89	19,47
8	35	4,1	16,81	25	1,7	2,89	6,97
9	48	17,1	292,41	35	11,7	136,89	200,07
10	44	13,1	171,61	28	4,7	22,09	61,57
11	20	-10,9	118,81	18	-5,3	28,09	57,77
12	18	-12,9	166,41	13	-10,3	106,09	132,87
13	14	-16,9	285,61	10	-13,3	176,89	224,77
14	19	-11,9	141,61	21	-2,3	5,29	27,37
15	39	8,1	65,61	27	3,7	13,69	29,97
16	11	-19,9	396,01	9	-14,3	204,49	284,57
17	32	1,1	1,21	25	1,7	2,89	1,87
18	17	-13,9	193,21	12	-11,3	127,69	157,07
19	14	-16,9	285,61	16	-7,3	53,29	123,37
20	20	-10,9	118,81	19	-4,3	18,49	46,87

$$\bar{y} = 30,9$$

$$\sum_{i=1}^n = 3.367,8 \quad \bar{x} = 23,3$$

$$\sum_{i=1}^n = 1864,2 \quad \sum_{i=1}^n = 2.181,6$$

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{2.181,6}{\sqrt{1.864,2 \cdot 3.367,8}} = \frac{2.181,6}{\sqrt{6.278.252,76}}$$

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{2.181,6}{2.505,64} = 0,87.$$

El resultado obtenido muestra que el componente lineal de la relación es muy pronunciado, puesto que el valor obtenido es superior a 0,5 y muy cercano a la cota máxima, 1.

Una vez cuantificado el componente lineal, calcularemos el coeficiente de correlación de Spearman para indagar sobre la existencia de una relación que no sea lineal pero sí monótona creciente (creciente, puesto que el gráfico sugiere relación directa o positiva).

$$r_s = \frac{\sum_{i=1}^n (r_i - \bar{r})(s_i - \bar{s})}{\sqrt{\sum_{i=1}^n (r_i - \bar{r})^2 \sum_{i=1}^n (s_i - \bar{s})^2}}.$$

Tabla 6.12. Asignación de rangos.

Id	STAI-E y_i	STAI rango s_i	BDI x_i	BDI rango r_i
1	45	17	29	17
2	40	14	27	14,5
3	42	15	22	10
4	38	12	26	13
5	47	18	32	18
6	50	20	52	20
7	25	9	20	8
8	35	11	25	11,5
9	48	19	35	19
10	44	16	28	16
11	20	7,5	18	6
12	18	5	13	4
13	14	2,5	10	2
14	19	6	21	9
15	39	13	27	14,5
16	11	1	9	1
17	32	10	25	11,5
18	17	4	12	3
19	14	2,5	16	5
20	20	7,5	19	7
	$\bar{y} = 30,9$	$\bar{s} = 10,5$	$\bar{x} = 23,3$	$\bar{r} = 10,5$

Para obtener el valor del coeficiente de correlación por rangos, hay que llevar a cabo los mismos cálculos que el coeficiente de correlación producto-momento, pero

sobre los datos transformados en rangos, en vez de sobre los datos originales. Por tanto, procederemos a llevar a cabo dicha transformación. Asignaremos el rango 1 al valor más pequeño, 2 al segundo más pequeño, y así sucesivamente hasta el rango 20 para el valor más elevado. También se tendrán en cuenta los empates.

Como se puede apreciar en la fórmula, es necesario disponer de la media de ambas variables transformadas (rango en el *BDI*, que aquí llamaremos R , y en el *STAI-E*, que llamaremos S) para poder calcular luego las desviaciones de cada dato de su media. De hecho, la media de los rangos ya es conocida, puesto que se trata del promedio de los veinte números enteros entre 1 y 20. Aun así, la calcularemos con fines ilustrativos:

$$\overline{BDI(rango)} \equiv \bar{r} = \frac{\sum_{i=1}^n r_i}{n} = \frac{17 + 14,5 + \dots + 16 + 6 + 4 + \dots + 7}{20} = \frac{210}{20} = 10,5$$

$$\overline{STAI-E(rango)} \equiv \bar{s} = \frac{\sum_{i=1}^n s_i}{n} = \frac{17 + 14 + \dots + 16 + 7,5 + 5 + \dots + 7,5}{20} = \frac{210}{20} = 10,5.$$

Es posible determinar la correlación de Spearman calculando en primer lugar el numerador (la suma de productos cruzados) y luego el denominador (las desviaciones elevadas al cuadrado).

Numerador:

$$\sum_{i=1}^n (r_i - \bar{r})(s_i - \bar{s}) = (17 - 10,5)(17 - 10,5) + \dots + (7 - 10,5)(7,5 - 10,5)$$

$$\sum_{i=1}^n (r_i - \bar{r})(s_i - \bar{s}) = (6,5)(6,5) + \dots + (-3,5)(-3) = 42,25 + \dots + 10,50 = 638.$$

Parte izquierda de la raíz cuadrada del denominador:

$$\sum_{i=1}^n (r_i - \bar{r})^2 = (17 - 10,5)^2 + (14,5 - 10,5)^2 + \dots + (7 - 10,5)^2$$

$$\sum_{i=1}^n (r_i - \bar{r})^2 = (6,5)^2 + (4)^2 + \dots + (-3,5)^2 = 41,25 + 16 + \dots + 12,25 = 664.$$

Parte derecha de la raíz cuadrada del denominador:

$$\sum_{i=1}^n (s_i - \bar{s})^2 = (17 - 10,5)^2 + (14 - 10,5)^2 + \dots + (7,5 - 10,5)^2$$

$$\sum_{i=1}^n (s_i - \bar{s})^2 = (6,5)^2 + (3,5)^2 + \dots + (-3)^2 = 41,15 + 12,25 + \dots + 9 = 664.$$

Estos mismos cálculos se pueden realizar utilizando la tabla siguiente, en la que se incorporan columnas con las variables que representan las diferencias respecto a las medias.

Tabla 6.13. Cálculos necesarios para determinar el valor del coeficiente de correlación por rangos.

Id	$STAI$ s_i	$STAI.d$ $(s_i - \bar{s})$	$STAI.d2$ $(s_i - \bar{s})^2$	BDI r_i	$BDI.d$ $(r_i - \bar{r})$	$BDI.d2$ $(r_i - \bar{r})^2$	$STAI.d * BDI.d$ $(s_i - \bar{s})(r_i - \bar{r})$
1	17,0	6,5	42,25	17,0	6,5	42,25	42,25
2	14,0	3,5	12,25	14,5	4,0	16,00	14,00
3	15,0	4,5	20,25	10,0	-0,5	0,25	-2,25
4	12,0	1,5	2,25	13,0	2,5	6,25	3,75
5	18,0	7,5	56,25	18,0	7,5	56,25	56,25
6	20,0	9,5	90,25	20,0	9,5	90,25	90,25
7	9,0	-1,5	2,25	8,0	-2,5	6,25	3,75
8	11,0	0,5	0,25	11,5	1,0	1,00	0,50
9	19,0	8,5	72,25	19,0	8,5	72,25	72,25
10	16,0	5,5	30,25	16,0	5,5	30,25	30,25
11	7,5	-3,0	9,00	6,0	-4,5	20,25	13,50
12	5,0	-5,5	30,25	4,0	-6,5	42,25	35,75
13	2,5	-8,0	64,00	2,0	-8,5	72,25	68,00
14	6,0	-4,5	20,25	9,0	-1,5	2,25	6,75
15	13,0	2,5	6,25	14,5	4,0	16,00	10,00
16	1,0	-9,5	90,25	1,0	-9,5	90,25	90,25
17	10,0	-0,5	0,25	11,5	1,0	1,00	-0,50
18	4,0	-6,5	42,25	3,0	-7,5	56,25	48,75
19	2,5	-8,0	64,00	5,0	-5,5	30,25	44,00
20	7,5	-3,0	9,00	7,0	-3,5	12,25	10,50
$\bar{s} = 10,5$		$\sum_{i=1}^n = 664$		$\bar{r} = 10,5$		$\sum_{i=1}^n = 664$	$\sum_{i=1}^n 638$

$$r_s = \frac{\sum_{i=1}^n (r_i - \bar{r})(s_i - \bar{s})}{\sqrt{\sum_{i=1}^n (r_i - \bar{r})^2 \sum_{i=1}^n (s_i - \bar{s})^2}} = \frac{638}{\sqrt{664 \cdot 664}} = \frac{638}{664} \approx 0,96.$$

El resultado obtenido muestra que la relación es monótona creciente de forma casi perfecta, puesto que el valor obtenido es prácticamente igual a la cota máxima, 1. Por consiguiente, se puede concluir que las puntuaciones de depresión y ansiedad estado están fuertemente relacionadas.

Finalmente, más allá de la cuantificación, es necesario hacer un apunte metodológico. Nótese que, a pesar de que la pregunta está redactada en términos de la ansiedad explicando la depresión, de hecho los índices estadísticos ofrecen información sobre covariación o asociación y no sobre causalidad, puesto que esta vendría determinada por el diseño, por las teorías existentes o por el momento de aparición de cada una de las condiciones mentales.

- b) Si se utiliza la ansiedad estado como variable de control, ¿cuánta variabilidad en las puntuaciones *BDI* quedaría por explicar después de haberla considerado?

Solución. Puesto que esta pregunta se centra en las mismas variables que la anterior, vuelve a tratarse de una descripción bivariante de la relación entre ellas, siendo el índice utilizado propio de las variables cuantitativas. Para dar respuesta a la pregunta, es necesario utilizar un indicador que devuelva un resultado en términos de variabilidad compartida. El indicador adecuado sería el coeficiente de determinación, *R*-cuadrado:

$$R^2 = (r)^2 = 0,87^2 = 0,7569.$$

Este resultado indica que la ansiedad estado y la depresión comparten el 75,69 % de la variabilidad y, por tanto, solo queda un 24,31 % de variabilidad en las puntuaciones de la depresión que no está relacionado con la ansiedad y que se podría intentar explicar mediante las otras variables medidas en el estudio. En realidad, el resultado obtenido solo hace referencia a la variabilidad que comparten en términos de relación lineal. En este sentido, es posible cuantificar la variabilidad compartida si no se restringiera la relación a la linealidad, mediante el coeficiente de determinación que se obtendría a partir del coeficiente de correlación por rangos, aunque no se trate de una práctica habitual:

$$R^2 = (r_s)^2 = 0,96^2 = 0,9216.$$

Si modeláramos la relación permitiendo la no linealidad, sería posible dar cuenta incluso de una mayor proporción de la variabilidad en las puntuaciones en el *BDI*, concretamente el 92,16 %, quedando solo el 7,84 % por explicar.

Tal como se puntualizó en la pregunta anterior, hablaríamos de variabilidad «explicada» en vez de «compartida» solo en aquellos casos en los cuales se puede asegurar una relación causal, teniendo en cuenta la precedencia temporal y la ausencia de explicaciones alternativas de esta variabilidad en la variable que se considera «efecto» o dependiente.

Ejercicio 4. Variables en diferentes escalas de medida

- a) Si se utiliza el nivel socioeconómico como variable de control, ¿cuánta variabilidad en las puntuaciones *BDI* quedaría por explicar después de haberla considerado?

Solución. Puesto que estamos interesados en estudiar la relación o asociación entre dos variables, se trataría de realizar un análisis estadístico propio de la descripción bivariante. Teniendo en cuenta que la pregunta hace referencia a una variable cuantitativa (la puntuación en el *BDI*) y una variable ordinal (nivel socioeconómico), existe más de una opción para la cuantificación, por ejemplo: a) tratar las dos variables como ordinales y calcular algunos de los índices tau de Kendall o la lambda de Goodman-Krusal; o b) tratar la variable ordinal como nominal y, por consiguiente, utilizar índices para estudiar la relación entre variables de diferentes escalas de

medida. En esta pregunta nos centraremos en la segunda opción y en el índice *eta*-cuadrado (η^2). Mediante este índice es posible cuantificar la variabilidad compartida entre la variable ordinal (nivel socioeconómico) y la variable cuantitativa (puntuación en el *BDI*) y, por tanto, se trata de una cuantificación análoga a la que proporciona el coeficiente de determinación.

Es importante destacar que el índice *eta*-cuadrado se utiliza principalmente para cuantificar la intensidad de una variable nominal y una variable cuantitativa, como podría haber sido el caso si se estudiara la asociación entre el tipo de terapia y la puntuación en el *BDI*. No obstante, para los casos en que la variable nominal tiene solo dos categorías (terapia psicodinámica y terapia cognitiva) es más habitual utilizar el índice *d* de Cohen, que también es objeto de otro de los problemas incluidos en el presente ejercicio. Así, hemos optado por centrarnos en una variable como el nivel socioeconómico que tiene más de dos valores o niveles. Dicho esto, es necesario resaltar que el uso que se le da aquí al índice *eta*-cuadrado es correcto, puesto que sirve para comparar los valores de la variable cuantitativa (puntuación en el *BDI*) en diferentes grupos definidos por otra variable de la cual únicamente se utiliza el hecho de que distingue categorías (nivel bajo, medio y alto) y no la posibilidad de ordenar dichas categorías.

En este contexto original de aplicación de η^2 , la variable ordinal se trata como independiente o causa, pero esta consideración, como siempre, depende del corpus teórico y del diseño, y no del tipo de análisis utilizado. La expresión para obtener *eta*-cuadrado es:

$$\eta^2 = \frac{SC_{factor}}{SC_{total}} = \frac{SC_{factor}}{SC_{factor} + SC_{error}}$$

$$SC_{total} = \sum_{j=1}^a \sum_{i=1}^{n_j} (y_{ij} - \bar{y})^2$$

$$SC_{factor} = \sum_{j=1}^a n_j (\bar{y}_j - \bar{y})^2$$

$$SC_{error} = \sum_{j=1}^a \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2$$

Como se puede observar, para calcular *eta*-cuadrado es necesario obtener la media global de todas las puntuaciones y la media en cada uno de los grupos definidos por los valores de la variable ordinal (i.e., la media en la puntuación en el *BDI* para personas con nivel socioeconómico bajo, medio y alto). A partir de estas medias se obtiene: a) la variabilidad total: la desviación de las puntuaciones individuales (y_{ij}) de la gran media de todas las puntuaciones (\bar{y}); b) la variabilidad explicada por el factor (variable ordinal): la diferencia entre la media de la variable cuantitativa en cada valor de la variable ordinal (\bar{y}_j) respecto a la gran media, y c) la variabilidad que queda sin explicar por el factor, conceptualizada como «error»: la diferencia entre las puntuaciones de un valor de la variable ordinal y su propia media. De hecho, es necesario calcular solo dos de estas variabilidades, puesto que a partir de ellas la tercera variabilidad está determinada:

$$SC_{total} = SC_{factor} + SC_{error}$$

$$\sum_{j=1}^a \sum_{i=1}^{n_j} (y_{ij} - \bar{y})^2 = \sum_{j=1}^a n_j (\bar{y}_j - \bar{y})^2 + \sum_{j=1}^a \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2$$

En primer lugar, calcularemos la media de la puntuación en el *BDI* para cada valor de la variable ordinal:

$$\overline{BDI}_{bajo} = \frac{\sum_{i=1}^{n_j} BDI_i}{n_j} = \frac{29 + 27 + 52 + 18}{4} = \frac{126}{4} = 31,5$$

$$\overline{BDI}_{medio} = \frac{\sum_{i=1}^{n_j} BDI_i}{n_j} = \frac{22 + 26 + 20 + 25 + 13 + 10 + 21 + 9 + 25 + 12}{10}$$

$$\overline{BDI}_{medio} = \frac{183}{10} = 18,3$$

$$\overline{BDI}_{alto} = \frac{\sum_{i=1}^{n_j} BDI_i}{n_j} = \frac{32 + 35 + 28 + 27 + 16 + 19}{6} = \frac{157}{6} \approx 26,2.$$

A partir de estos valores ya se puede observar que hay diferencias entre los grupos, siendo las personas con nivel socioeconómico medio las que presentan niveles más bajos de depresión, mientras que las personas con nivel socioeconómico bajo obtienen puntuaciones más elevadas en el *BDI*. Ahora procederemos a cuantificar la parte de la variabilidad total que se explica por el hecho de tener diferentes niveles socioeconómicos. Para obtener la variabilidad asignable a la variable ordinal, es necesario obtener la gran media:

$$\overline{BDI} = \frac{\sum_{i=1}^n BDI_i}{n} = \frac{29 + 27 + \dots + 19}{20} = \frac{466}{20} = 23,3$$

$$SC_{factor} = \sum_{j=1}^a n_j (\bar{y}_j - \bar{y})^2 = 4(31,5 - 23,3)^2 + 10(18,3 - 23,3)^2 + 6(26,2 - 23,3)^2$$

$$SC_{factor} = 4(8,2)^2 + 10(-5)^2 + 6(2,9)^2 = 4 \cdot 67,24 + 10 \cdot 25 + 6 \cdot 8,41$$

$$SC_{factor} = 268,96 + 250 + 50,46 = 569,42.$$

Esta cantidad no es fácil de valorar si no se considera la variabilidad total o la variabilidad del error. Por tanto, procederemos a calcular la variabilidad del error (las variaciones intragrupo) que no puede ser explicada por el nivel socioeconómico, sino que se debe a otra(s) variable(s):

$$SC_{error} = \sum_{j=1}^a \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2 = SC_{error.bajo} + SC_{error.medio} + SC_{error.alto}$$

Calcularemos la variabilidad no explicada en cada grupo, i. e., para cada valor de la variable nivel socioeconómico y, posteriormente, sumaremos estos tres valores:

$$SC_{error.bajo} = \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2 = (29 - 31,5)^2 + (27 - 31,5)^2 + (52 - 31,5)^2 + (18 - 31,5)^2$$

$$SC_{error.bajo} = (-2,5)^2 + (-4,5)^2 + 20,5^2 + (-13,5)^2 = 6,25 + 20,25 + 420,25 + 182,25$$

$$SC_{error.bajo} = 629$$

$$SC_{error.medio} = \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2 = (22 - 18,3)^2 + (26 - 18,3)^2 + \dots + (12 - 18,3)^2$$

$$SC_{error.medio} = (3,7)^2 + (7,7)^2 + \dots + (-6,3)^2 = 13,69 + 59,29 + \dots + 39,69$$

$$SC_{error.medio} = 396,1$$

$$SC_{error.alto} = \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2$$

$$= (32 - 26,2)^2 + (35 - 26,2)^2 + (28 - 26,2)^2 + (27 - 26,2)^2 + (16 - 26,2)^2 + (19 - 26,2)^2$$

$$SC_{error.alto} = (5,8)^2 + (8,8)^2 + (1,8)^2 + (0,8)^2 + (-10,2)^2 + (-7,2)^2$$

$$SC_{error.alto} = 33,64 + 77,44 + 3,24 + 0,64 + 104,04 + 51,84 = 270,84.$$

A partir de estos valores, se podría pensar que, aparentemente, entre las personas con nivel socioeconómico bajo hay mayor variabilidad en su puntuación en el *BDI* que en los otros dos valores de la variable ordinal. Sin embargo, hay que tener presente que la suma de cuadrados es una cuantificación que se ha de expresar en términos relativos (i.e., dividirla por la cantidad de valores n cada grupo) para que pueda ser comparable entre grupos. Por tanto, para obtener un indicador comparable (tal como se procede al calcular la variancia), es necesario dividir por el tamaño muestral n . De hecho, la división de la suma de cuadrados por los grados de libertad (que están estrechamente relacionados con la cantidad de medidas disponibles) llevaría a la obtención de los «cuadrados medios» que se utilizan en el análisis de la variancia (se trata en el último capítulo del bloque de inferencia estadística incluido en el presente documento).

La variabilidad del error es:

$$SC_{error} = \sum_{j=1}^a \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2 = SC_{error.bajo} + SC_{error.medio} + SC_{error.alto}$$

$$SC_{error} = \sum_{j=1}^a \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2 = 629 + 396,1 + 270,84 = 1.295,94.$$

Ya disponemos de todos los valores necesarios para calcular la *eta*-cuadrado:

$$\eta^2 = \frac{SC_{factor}}{SC_{factor} + SC_{error}} = \frac{569,42}{569,42 + 1.295,94} = \frac{569,42}{1.865,36} \approx 0,3053.$$

Este valor nos indica que la variabilidad compartida es del 30,53 % y, por consiguiente, la variabilidad en la puntuación en el *BDI* que quedaría por explicar, después de dar cuenta de las diferencias en el nivel socioeconómico, sería del 69,47 %. Las diferencias entre los valores de la variable ordinal también se pueden representar visual-

mente, mediante un gráfico de medias con las respectivas desviaciones estándares en cada categoría (Figura 6.5 izquierda) o mediante diagramas de caja (derecha).

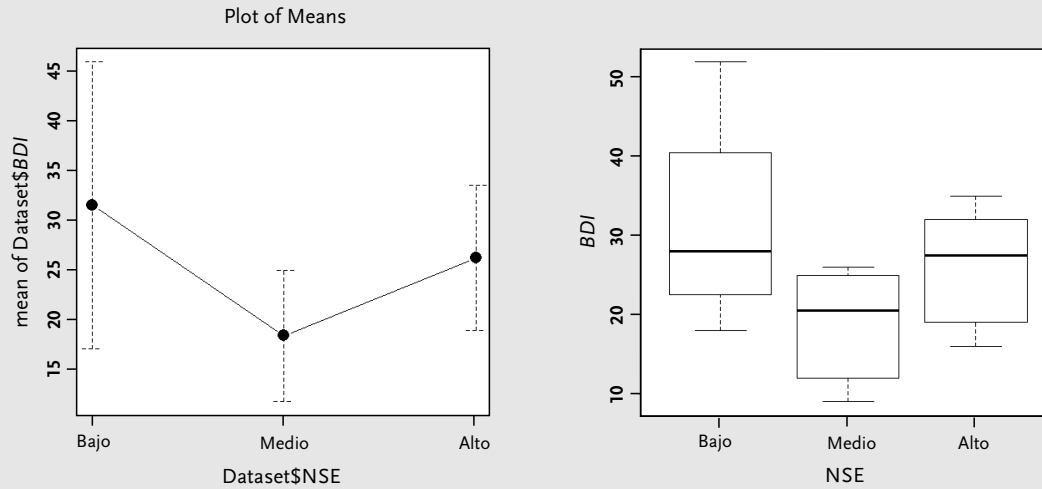


Figura 6.5. Gráfico de medias con barras que representan una desviación estándar (izquierda), y diagramas de caja (derecha) que muestran las diferencias entre los tres niveles socioeconómicos en cuanto a la puntuación en el *BDI*.

- b) Calcular el índice d de Cohen para comparar los dos tipos de terapia en cuanto a las puntuaciones de *BDI*.

Solución. Puesto que estamos interesados en estudiar la relación o asociación entre dos variables, se trataría de realizar un análisis estadístico propio de la descripción bivalente. Por otra parte, la pregunta hace referencia a una variable cuantitativa (la puntuación en el *BDI*) y una variable nominal (tipo de terapia) y, por tanto, hay que utilizar índices para estudiar la relación entre variables de diferentes escalas de medida, concretamente la d de Cohen, puesto que es un indicador apropiado cuando la variable nominal es dicotómica:

$$d = \frac{\bar{x}_{PsicoD} - \bar{x}_{TCognitiva}}{\sqrt{\frac{s_{PsicoD}^2 + s_{TCognitiva}^2}{2}}}$$

De la expresión para la d de Cohen se desprende que es necesario obtener las medias de ambos grupos y las desviaciones estándares:

$$\begin{aligned} \bar{BDI}_{PsicoD} &= \frac{\sum_{i=1}^n BDI_i}{n} = \frac{29 + 27 + 22 + 26 + 32 + 52 + 20 + 25 + 35 + 28}{10} \\ \bar{BDI}_{PsicoD} &= \frac{296}{10} = 29,6 \\ \bar{BDI}_{TCognitiva} &= \frac{\sum_{i=1}^n BDI_i}{n} = \frac{18 + 13 + 10 + 21 + 27 + 9 + 25 + 12 + 16 + 19}{10} \end{aligned}$$

$$\bar{BDI}_{TCognitiva} = \frac{170}{10} = 17.$$

Si comparáramos las medias de ambos grupos, obtendríamos la diferencia de medias no estandarizada: $29,6 - 17 = 12,6$. Este resultado indica que las personas que han seguido terapia psicodinámica obtienen, en promedio, 12,6 puntos más en la escala del *BDI*, presentando, por tanto, niveles más elevados de depresión. Para calcular la diferencia de medias estandarizada (d), tendremos en cuenta también la variabilidad en ambos grupos:

$$s_{PsicoD}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{(29-29,6)^2 + (27-29,6)^2 + \dots + (28-29,6)^2}{10}$$

$$s_{PsicoD}^2 = \frac{0,36 + 6,76 + \dots + 501,76 + \dots + 2,56}{10} = \frac{730,4}{10} = 73,04$$

$$s_{TCognitiva}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{(18-17)^2 + (13-17)^2 + \dots + (19-17)^2}{10}$$

$$s_{TCognitiva}^2 = \frac{1^2 + (-4)^2 + \dots + 2^2}{10} = \frac{1 + 16 + \dots + 4}{10} = \frac{340}{10} = 34.$$

A partir de los valores disponibles, determinamos la d de Cohen:

$$d = \frac{\bar{x}_{PsicoD} - \bar{x}_{TCognitiva}}{\sqrt{\frac{s_{PsicoD}^2 + s_{TCognitiva}^2}{2}}} = \frac{29,6 - 17}{\sqrt{\frac{73,04 + 34}{2}}} = \frac{12,6}{\sqrt{53,52}} \approx \frac{12,6}{7,32} \approx 1,72.$$

Obtenemos que la diferencia entre ambos grupos es de 1,72 desviaciones estándares. Para interpretar este valor podríamos utilizar los criterios de Cohen, que implican etiquetar los valores superiores a 0,8 como relación «fuerte» o tamaño del efecto «grande». Esta práctica no es incorrecta, pero hay que tener en cuenta que unos criterios generales para todas las ciencias sociales podrían no ser apropiados, y que es importante disponer de puntos de corte empíricos según la disciplina.

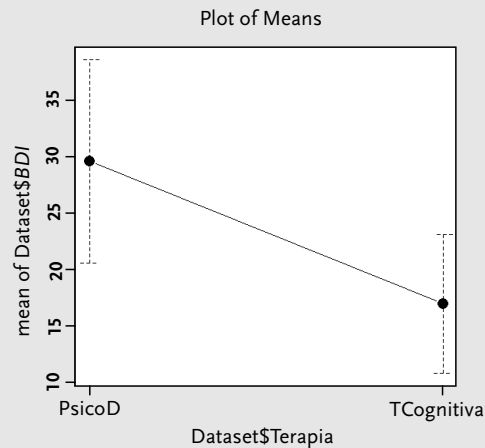


Figura 6.6. Gráfico de medias con barras que representan una desviación estándar y muestran la diferencia entre los dos tipos de terapia en cuanto a la puntuación en el *BDI*.

La diferencia entre los grupos se puede observar visualmente mediante el gráfico de medias con barras que representan una desviación estándar de distancia respecto a la media. Como se puede observar, en la Figura 6.6, la media del grupo que siguió terapia psicodinámica (29,6) está a una distancia superior a una desviación estándar de la media del grupo de terapia cognitiva (la barra superior está por debajo de 25). Análogamente, la media del grupo que siguió terapia cognitiva (17) está a más de una desviación estándar de distancia de la media del otro grupo (la barra inferior está por encima de 20). Esta es una manera gráfica de entender el resultado de la d de Cohen, que es superior a 1 desviación, aunque para este indicador se ha utilizado una estimación conjunta de la variancia.

- c) Evaluar el significado de la d de Cohen como indicador del grado de solapamiento entre las distribuciones de puntuaciones en el *BDI* en los dos grupos de terapia. Utilice el indicador no paramétrico denominado «probabilidad de superioridad» para cuantificar el grado de solapamiento observado.

Solución. Puesto que en esta pregunta seguimos tratando con la d de Cohen, se está utilizando otra vez la estadística descriptiva bivalente y un indicador apropiado para estudiar la relación entre variables de diferentes escalas de medida, concretamente entre una variable nominal dicotómica (tipo de terapia) y una variable medida en escala de intervalo (puntuación en el *BDI*). Adicionalmente, se menciona la versión no paramétrica de la «probabilidad de superioridad». Este indicador permite comparar, por pares, dos distribuciones de valores, siendo el resultado de la comparación: igual, mayor o menor. Por consiguiente, se trata de un indicador apropiado para una variable nominal dicotómica (que permite separar los valores en dos conjuntos) y una variable medida en escala ordinal (o superior), puesto que la única operación que se realiza tiene que ver con el orden y no con la diferencia de la magnitud.

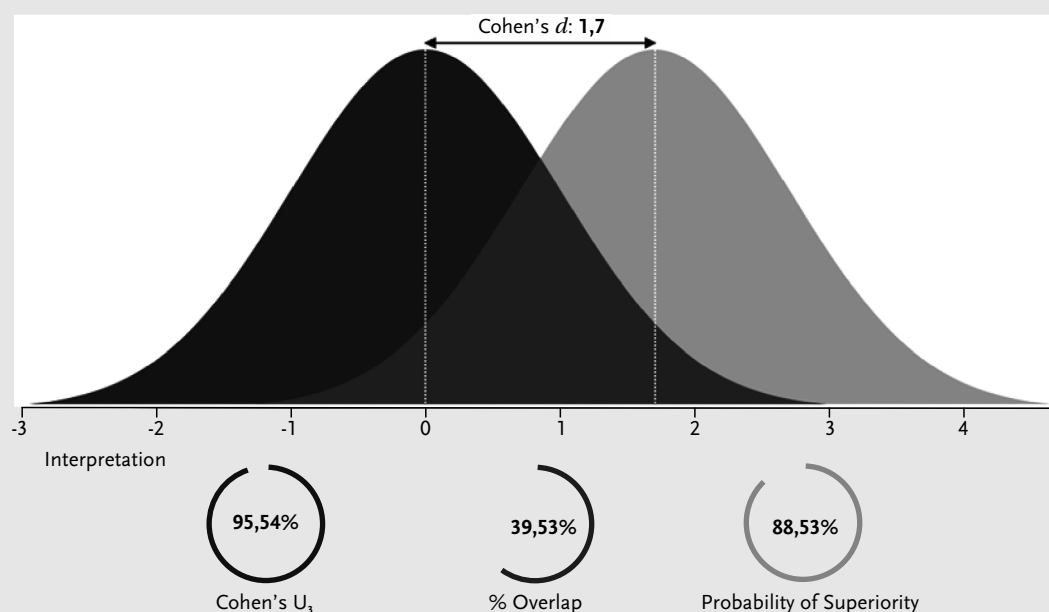


Figura 6.7. Grado de solapamiento que se observaría con una $d = 1,7$ si la variable cuantitativa se distribuyera normalmente en ambos grupos.

Aparte de ser una cuantificación de la diferencia de medias en términos de desviación estándar, la d de Cohen puede traducirse, en ciertas condiciones, en la proporción de solapamiento entre los datos y en la probabilidad de que un valor extraído al azar de una distribución (e.g., puntuación en el *BDI* después de la terapia psicodinámica) sea superior a un valor extraído al azar de la otra (e.g., puntuación en el *BDI* después de la terapia cognitiva). El valor de U_3 que le correspondería la d que hemos hallado puede obtenerse y visualizarse mediante la página web de Kristoffer Magnusson: <http://rpsychologist.com/d3/cohend/>. Debido a que esta página permite calcular la d con solo un decimal, hemos tomado como referencia el valor de 1,7, redondeando el resultado obtenido en una pregunta anterior ($d = 1,72$).

En la última figura se ofrece información referente a la probabilidad de superioridad: en el 88,53 % de los casos se espera que un valor en el *BDI* seleccionado al azar del grupo de terapia psicodinámica sea superior a un valor en el *BDI* seleccionado al azar del grupo de terapia cognitiva. Sin embargo, esta cuantificación parte del supuesto de que las dos distribuciones son normales. Para comprobar, al menos visualmente, la plausibilidad de este supuesto, podemos utilizar el histograma de la Figura 6.8. Se observa que ninguna de las distribuciones parece ser normal. No obstante, para contrastar si la desviación de la normalidad presente es suficientemente sustancial, sería mejor complementar la exploración visual mediante pruebas de conformidad propias de la estadística inferencial.

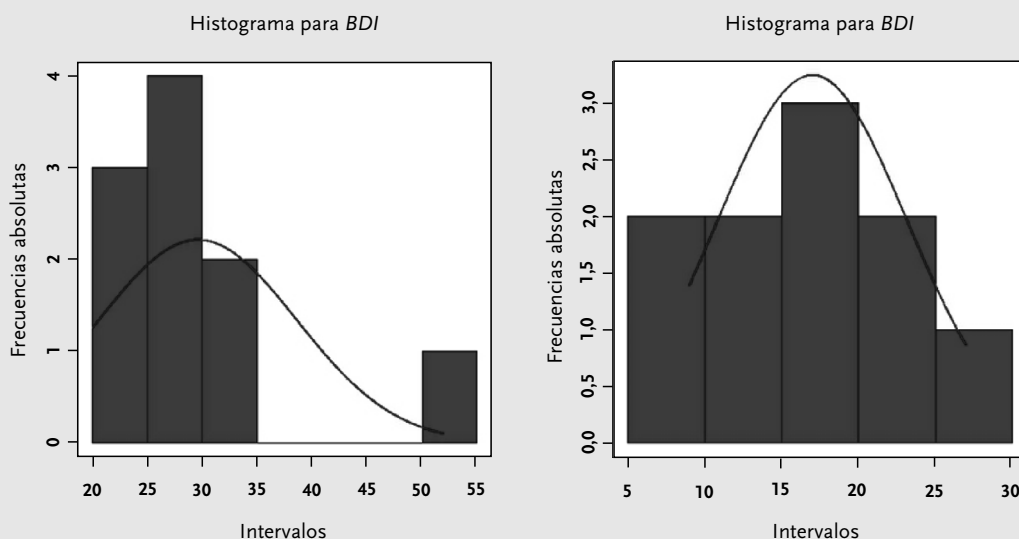


Figura 6.8. Histogramas para la puntuación en el *BDI* en los dos grupos de terapia.

Para calcular la probabilidad de superioridad en los datos observados, es necesario comparar cada valor de *BDI* en un grupo con cada valor en el otro grupo. Por tanto, es necesario realizar un total de $n_{PsicoD} \cdot n_{TCognitiva} = 10 \cdot 10 = 100$ comparaciones. Se cuenta la cantidad de veces que un valor *BDI* del grupo de terapia psicodinámica es superior a un valor *BDI* del grupo de terapia cognitiva, además de tener en cuenta (con la mitad de peso), la cantidad de empates:

$$PS = \frac{[\#(X_{PsicoD} > X_{TCognitiva}) + 0,5\#(X_{PsicoD} = X_{TCognitiva})]}{n_{PsicoD} \cdot n_{TCognitiva}}$$

Podemos realizar estas comparaciones a partir de la matriz de datos o con la ayuda del siguiente gráfico:

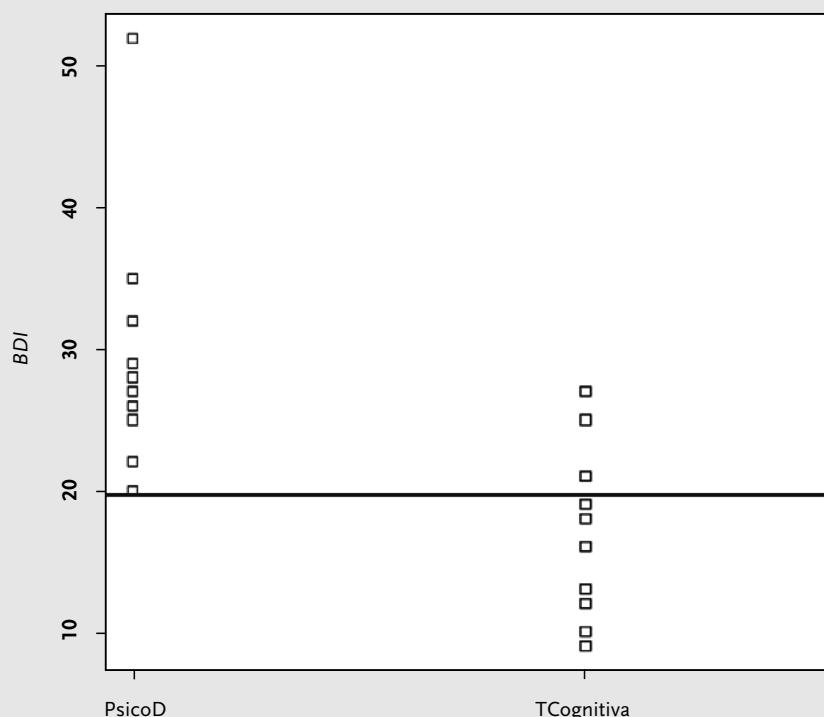


Figura 6.9. Stripchart para la puntuación en el BDI en los dos grupos de terapia.

Puesto que hablar de superioridad en la puntuación en el BDI del grupo de terapia psicodinámica es equivalente a hablar de inferioridad del grupo de terapia cognitiva, podemos escoger una de las dos perspectivas a la hora de hacer el recuento. En la Figura 6.9 podemos comprobar que los 7 valores BDI más bajos para el grupo de terapia cognitiva son todos ellos inferiores a cualquier valor BDI del grupo de terapia psicodinámica. Por tanto, ya disponemos de $7 \cdot 10 = 70$ casos de $(X_{PsicoD} > X_{TCognitiva})$. Compararemos el resto de los valores BDI del grupo de terapia cognitiva que sí se solapan con valores del otro grupo siguiendo un orden ascendente.

El valor 21, correspondiente a Id.14 (terapia cognitiva), es de hecho superior al valor 20 del Id.7 (terapia psicodinámica), pero inferior al resto de los valores de este último grupo. Así, teniendo en cuenta que cada valor en un grupo se compara con 10 valores del otro grupo, disponemos de otros $10 - 1(X_{PsicoD} < X_{TCognitiva}) = 9$ casos de $(X_{PsicoD} > X_{TCognitiva})$.

El valor 25, correspondiente a Id.17 (terapia cognitiva), es de hecho superior al valor 20 del Id.7, al valor 22 de Id.3 e igual al valor 25 de Id.8 (todos ellos de terapia psicodinámica), pero inferior al resto de los valores de este último grupo. Por consiguiente, disponemos de otros $10 - 2(X_{PsicoD} < X_{TCognitiva}) - 0,5(X_{PsicoD} = X_{TCognitiva}) = 7,5$ casos de $(X_{PsicoD} > X_{TCognitiva})$.

Por último, el valor 27, correspondiente a Id.17 (terapia cognitiva), es de hecho superior al valor 20 del Id.7, al valor 22 de Id.3, al valor 25 de Id.8, al valor 26 de Id.4 e igual al valor 27 de Id.2 (todos ellos de terapia psicodinámica), pero inferior al resto de los valores de este último grupo. Por tanto, disponemos de otros $10 - 4(X_{PsicoD} < X_{TCognitiva}) - 0,5(X_{PsicoD} = X_{TCognitiva}) = 5,5$ casos de $(X_{PsicoD} > X_{TCognitiva})$.

Así,

$$PS = \frac{[\#(X_{PsicoD} > X_{TCognitiva}) + 0,5\#(X_{PsicoD} = X_{TCognitiva})]}{n_{PsicoD} \cdot n_{TCognitiva}}$$

$$PS = \frac{70 + 9 + 7,5 + 5,5}{100} = \frac{92}{100} = 0,92.$$

En la anterior expresión, debe leerse # como «número de veces que ocurre un suceso». La comparación también puede presentarse en formato tabular, desde la perspectiva del grupo de terapia cognitiva. Mostramos abajo un ejemplo, centrándonos en el valor 27 de Id.15 (perteneciente al grupo de terapia cognitiva).

Tabla 6.14. Ilustración de la comparación por pares para obtener el valor de la probabilidad de superioridad centrándonos en el valor 27 del grupo de terapia cognitiva.

TCognitiva	Inferior a ($X_{PsicoD} > X_{TCognitiva}$)	Igual a ($X_{PsicoD} = X_{TCognitiva}$)	PsicoD
18	10	0	29
13	10	0	27
10	10	0	22
21	9	0	26
27	5	1	32
9	10	0	52
25	7	1	20
12	10	0	25
16	10	0	35
19	10	0	28
# ($X_{PsicoD} > X_{TCognitiva}$) = 91		# 0,5 ($X_{PsicoD} = X_{TCognitiva}$) = 1	

La misma tabla se puede elaborar también desde la perspectiva del grupo de terapia psicodinámica. Mostramos abajo un ejemplo, centrándonos en el valor 20 de Id.7 (perteneciente al grupo de terapia psicodinámica).

Tabla 6.15. Ilustración de la comparación por pares para obtener el valor de la probabilidad de superioridad centrándonos en el valor 7 del grupo de terapia psicodinámica.

PsicoD	Inferior a ($X_{PsicoD} > X_{TCognitiva}$)	Igual a ($X_{PsicoD} = X_{TCognitiva}$)	TCognitiva
29	10	0	18
27	9	1	13
22	8	0	10
26	9	0	21
32	10	1	27
52	10	0	9
20	7	0	25
25	8	1	12
35	10	0	16
28	10	0	19
# ($X_{PsicoD} > X_{TCognitiva}$) = 91		# 0,5 ($X_{PsicoD} = X_{TCognitiva}$) = 1	

El valor de probabilidad de superioridad obtenido aquí es más elevado que el que se indicaba en la página web, basándose este último en el supuesto de que las dos distribuciones de puntuaciones en el *BDI* son normales. Por tanto, parece razonable no confiar en los valores derivados bajo supuestos que no se han podido comprobar. En cuanto a la probabilidad obtenida, esta indica que en el 92 % de los casos el valor seleccionado al azar del grupo de terapia psicodinámica sería superior a la puntuación en el *BDI* del grupo de terapia cognitiva; se trata de una expresión más de que el nivel de depresión después de la terapia cognitiva es inferior al nivel de depresión después de la terapia psicodinámica.

Finalmente, relacionando la probabilidad de superioridad en su versión no paramétrica con el grado de solapamiento, se ha sugerido que la probabilidad de 0,92 anteriormente obtenida puede también interpretarse como la proporción de no solapamiento entre los datos. Por consiguiente, el porcentaje de solapamiento sería $100 - 92 = 8\%$, resultado que contrasta con el que se obtiene suponiendo la normalidad de ambas distribuciones de *BDI*.

- d) Evalúe el significado de la d de Cohen en su relación con el indicador U_3 : ¿a qué percentil en la distribución del grupo de personas que siguió terapia cognitiva corresponde la media en la puntuación en el *BDI* del grupo que siguió terapia psicodinámica?

Solución. Puesto que en esta pregunta seguimos tratando con la d de Cohen, se está utilizando otra vez la estadística descriptiva bivalente y un indicador apropiado para estudiar la relación entre variables de diferentes escalas de medida, concretamente entre una variable nominal dicotómica (tipo de terapia) y una variable medida en escala de intervalo (puntuación en el *BDI*).

Aparte de ser una cuantificación de la diferencia de medias en términos de desviación estándar, la d de Cohen puede traducirse, en ciertas condiciones, en un indicador denominado U_3 que cuantifica la proporción o el porcentaje de valores de una distribución (e.g., puntuación en el *BDI* después de la terapia cognitiva) que son inferiores a la media de la otra distribución (e.g., puntuación en el *BDI* después de la terapia psicodinámica). Alternativamente, U_3 puede entenderse como el percentil que le correspondería a la media de *BDI* para las personas que siguieron terapia psicodinámica en la distribución de valores del grupo de terapia cognitiva.

El valor de U_3 que le correspondería la d que hemos hallado puede obtenerse y visualizarse en la página web de Kristoffer Magnusson: <http://rpsychologist.com/d3/cohend/>. Debido a que esta página permite especificar la d con un solo decimal, hemos tomado como referencia el valor de 1,7, redondeando el resultado obtenido en una pregunta anterior ($d = 1,72$).

En la figura se representa la separación de las distribuciones que se espera con un valor de $d = 1,7$. También podemos observar que el valor de $U_3 = 95,54\%$. Es decir, se espera que la media de la puntuación en el *BDI* en el grupo de terapia psicodinámica sea superior al 95,54 % de los valores en el grupo de terapia cognitiva. Sin embargo, nótese que estamos utilizando el término «se espera», puesto que esta cuantificación corresponde al caso (representado gráficamente) en el que ambas distribuciones son normales. Los datos de *BDI* en los dos grupos, no obstante, no son exacta, ni aproximadamente normales, como se ha visto en la pregunta anterior.

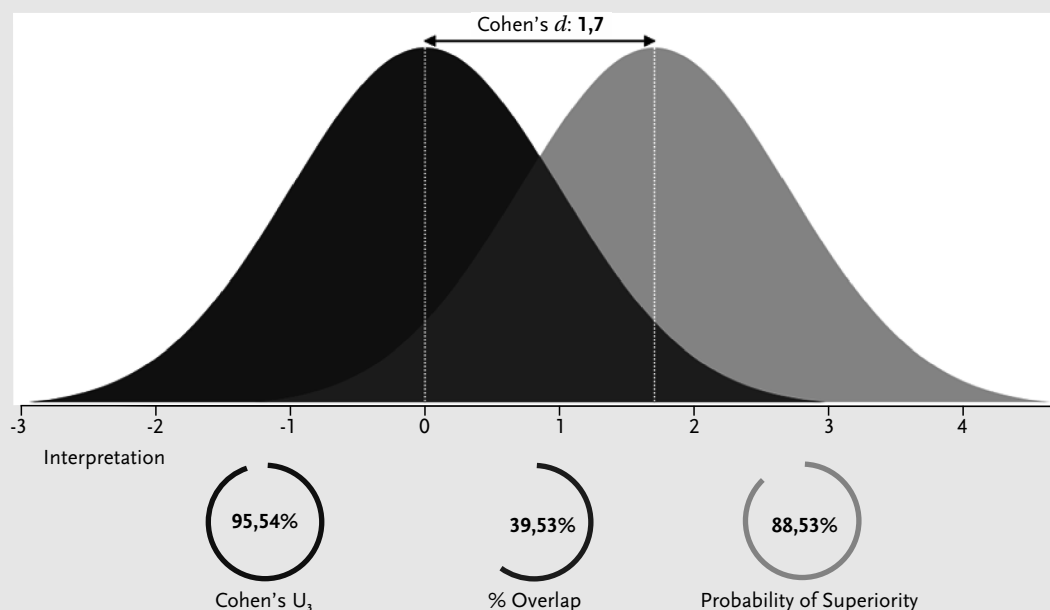


Figura 6.10. Valor del índice U_3 que se obtendría con una $d = 1,7$ si la variable cuantitativa se distribuyera normalmente en ambos grupos.

A continuación se presentan los diagramas de caja, que muestran cómo la mediana en el grupo de psicodinámica es superior que el valor más elevado en el grupo de terapia cognitiva. ¿Y la media? Como se vio en la pregunta 1, y como es esperable en presencia del valor anómalo en la parte alta de la distribución, la media en el grupo de terapia psicodinámica es más elevada que la mediana. De hecho, de las respuestas a preguntas anteriores sabemos que la media en el grupo de terapia psicodinámica es 29,6, mientras que, si nos fijamos en la matriz de datos, podemos comprobar cómo el valor más alto obtenido en el grupo de terapia cognitiva es 27. Por tanto, sigue siendo superior al 100 % de los valores de *BDI* en el grupo de terapia cognitiva. La falta de correspondencia entre este valor y el U_3 que nos ofrece la página web se relaciona con el hecho de que las dos distribuciones de valores *BDI* no son normales.

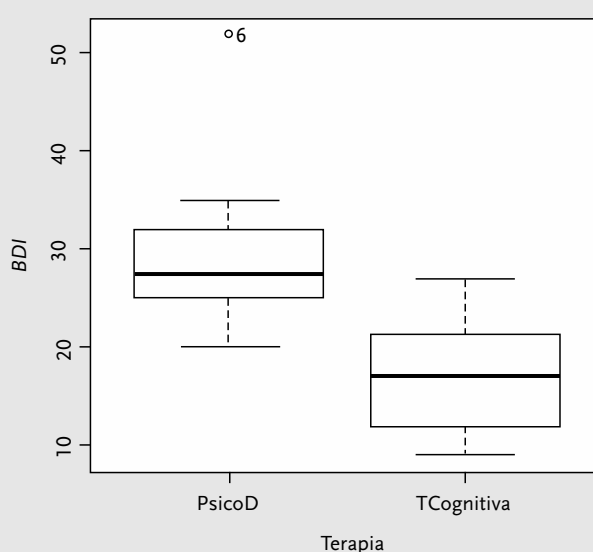


Figura 6.11. Diagramas de caja para la puntuación en el *BDI* en los dos grupos de terapia.