

Transformación e interpretación de las puntuaciones

Sergi Valero

PID_00198630

Índice

Introducción	5
1. Interpretación de una puntuación	7
2. Transformación de las puntuaciones	9
2.1. Percentiles	9
2.2. Puntuaciones estandarizadas	13
2.3. Puntuaciones estandarizadas derivadas	17
2.4. Puntuaciones estandarizadas normalizadas	18
2.5. Normas cronológicas	20
3. Baremación	22
4. Equiparación de puntuaciones	26
Resumen	30
Bibliografía	33

Introducción

Interpretar el valor obtenido por una persona en una medida psicológica, y hacerlo de un modo razonable y razonado, exige no solo conocer con cierto detalle las características técnicas del instrumento que es empleado, sino también conocer el constructo al que se hace referencia y entender cómo este se estructura y se explica en un marco teórico que tiene sentido, a la vez, en un marco disciplinario concreto, ya sea educacional, clínico o de cualquier otro tipo.

Es evidente que un módulo como este no puede alcanzar satisfactoriamente un objetivo de esta magnitud. Por esta razón, cuando en este texto se hable de **interpretar una puntuación**, se deberá siempre considerar desde una perspectiva eminentemente metodológica y matemática y, lo más importante, inferida de acuerdo con la respuesta de un conjunto de individuos que, evaluados con una misma medida, proporcionarán el modelo de comportamiento que será dispuesto a modo de norma, de marco de referencia respecto al cual la puntuación de una persona concreta será interpretada. Bajo esta perspectiva tendrán su sentido las puntuaciones transformadas. Los percentiles, las puntuaciones estandarizadas y las normas cronológicas se recogerán aquí, exponiendo sus características más importantes, la manera de calcularlas y, donde se considere oportuno, sus limitaciones más relevantes.

Un segundo término de gran relevancia, y que será presentado en el tercer apartado, es el concepto de baremo. Se trata de una herramienta de gran importancia en la que se establece, de manera estructurada, una conexión entre la puntuación de un individuo y el comportamiento normativo proporcionado por una muestra relevante de sujetos, que, en algunos casos, aparecerán estratificados según conveniencia. Se dedicará un especial interés a la determinación de cuál debe ser esta muestra normativa. En este sentido, se expondrán los criterios de relevancia, representatividad y homogeneidad como factores fundamentales en la determinación de la calidad de esta muestra de referencia.

El módulo finalizará exponiendo las diferentes estrategias disponibles para establecer equiparaciones entre puntuaciones de instrumentos diferentes que evalúan un mismo rasgo. Un conjunto de estrategias, también metodológicas y matemáticas, que son especialmente interesantes en aquellos ámbitos de la evaluación en los que los procesos de memoria o aprendizaje de las personas participantes pueden convertirse en una limitación relevante de las inferencias que se derivan de ellos.

1. Interpretación de una puntuación

Consideremos una circunstancia nada excepcional en la que para aprobar una asignatura, por ejemplo *Lengua inglesa*, es necesario obtener una puntuación de 5, asumiendo que estamos ante un examen clásico en el que la puntuación mínima posible es 0 y la máxima 10. ¿Alguna vez os habéis preguntado qué propiedad intrínseca tiene el número 5 como para considerar que con él y por encima de él uno ya domina esta lengua, y que por el contrario, cuando se está debajo, nuestras aptitudes no satisfacen el nivel exigido? Excepto que nos sintiéramos cómodos con algún postulado más propio de la numerología que de la psicometría, habría que considerar que, *a priori*, el número 5 carece de cualquier atributo intrínseco que lo haga especial. Y lo mismo se podría afirmar de una puntuación de 4 o de cualquier otra. Si no se nos informa de nada más, deberemos asumir que aprobar o suspender en función de superar o no una puntuación de 5 responde a una simple convención. Hay que establecer un punto de corte porque muchas veces es necesario tomar una decisión respecto a quién supera un curso, y el valor de 5, cifra que ocupa la posición central, suele ser el valor de referencia utilizado más frecuentemente.

En otras circunstancias, sin embargo, el criterio empleado para discernir entre personas con o sin un determinado rasgo no responde a un juicio arbitrario o convencional. El criterio puede responder a consideraciones empíricamente contrastadas, surgido generalmente de la comparación de poblaciones de sujetos muy diferenciados en el rasgo objeto de interés y que permiten la obtención de un punto de corte, de una puntuación de referencia dotada de capacidad discriminativa (Ramos-Quiroga et al., 2009).

En muchas situaciones, no obstante, no se dispone de criterios externos que proporcionen un marco de interpretación respecto al que otorgar relevancia o significación a una puntuación obtenida en una medida. De hecho, se puede afirmar que la inmensa mayoría de las medidas psicológicas carecen de criterios externos e independientes con los que inferir el estatus actitudinal, aptitudinal o clínico de un individuo. ¿Qué se puede inferir entonces de la puntuación de una persona en una medida cuando no están disponibles estos referentes?

Mirar **dentro** del instrumento, es decir, remitirse a alguna de las propiedades estructurales de la medida es una alternativa posible. Una aproximación sencilla, por ejemplo, consiste en atender las puntuaciones mínima y máxima que es posible alcanzar. Esta aproximación absoluta permite contextualizar una puntuación indicando su excepcionalidad. Cualquier manual de uso o publicación de referencia de una medida debería proporcionar información clara respecto a estos valores mínimo y máximo. La sencillez evidente de esta aproximación, no obstante, suele ser a la vez su limitación más importante:

probablemente tendremos menos problemas para emitir un juicio sobre una persona que obtiene en una media una puntuación de 17 que para una de 14, sabiendo que la puntuación máxima alcanzable es 18. Ahora bien, ¿cuánto debe estar un valor alejado del valor máximo (o mínimo) posible para considerar que es mucho o poco? No es posible dar una respuesta operativa y universal a esta pregunta.

Una manera alternativa de proceder consiste en comparar la puntuación obtenida por una persona con la puntuación que obtienen otras personas que fueron evaluadas con la misma medida. La media (y también la mediana) suele ser una manera frecuente de resumir el comportamiento de este grupo de personas. De este modo, si una persona obtiene en una medida de aptitudes mnésicas una puntuación de 7 y la media de sus colegas fue de 9, evaluados siguiendo la misma pauta, podemos afirmar que esta persona presenta una capacidad de memoria inferior a la presentada por sus compañeros. Y aquí descansa una de las claves de esta aproximación relativa: la inferencia que se hace de una observación concreta es una función de la tendencia de comportamiento de otras personas en la misma medida. Retomaremos esta observación más adelante.

2. Transformación de las puntuaciones

Según lo que se ha apuntado hasta el momento, es posible disponer de estrategias que permiten la interpretación de una medida y que, además, no suponen ninguna manipulación de los datos obtenidos. No obstante, no son las únicas estrategias al alcance para dar sentido a una determinada puntuación. Otros procedimientos, menos simples que los expuestos hasta el momento, ofrecen al usuario una aproximación más precisa e informativa. Estos procedimientos implican, no obstante, la transformación de las puntuaciones observadas.

Transformar las puntuaciones de una medida consiste en aplicar una estrategia de codificación en la que las puntuaciones obtenidas en la medida, sus puntuaciones directas, son recodificadas (transformadas) en un nuevo sistema de valores que facilitan al usuario su interpretación.

Una propiedad necesaria de las puntuaciones transformadas es que no alteran el escalamiento de las puntuaciones directas, es decir, respetan la diferente disposición de los distintos sujetos según la magnitud de sus puntuaciones. Y este mantenimiento de la ordenación original se lleva a cabo, como se ha mencionado, otorgando a las nuevas puntuaciones un sentido eminentemente práctico. Los percentiles, las puntuaciones estandarizadas, las puntuaciones estandarizadas derivadas y normalizadas y las normas cronológicas serán las distintas estrategias de transformación que serán tratadas a continuación.

2.1. Percentiles

Para describir el concepto de percentil, también conocidos como centiles, se expondrá en primer lugar el concepto de mediana. Una vez ordenados los valores de un conjunto de observaciones, la mediana será aquel valor que ocupa la posición u orden central. Un ejemplo sencillo ayudará en esta primera aproximación.

Ejemplo

Se dispone del siguiente conjunto de valores: 4, 5, 4, 3, 7, 8, 3, 1 y 6. En primer lugar habrá que ordenar la secuencia: 1, 3, 3, 4, 4, 5, 6, 7, 8. La mediana de este conjunto de valores es el 4 (el segundo 4), dado que es el valor que ocupa la posición central de la distribución. La mitad de las personas están dispuestas a un lado y otro de este valor.

En este ejemplo ha sido fácil identificar el valor central de la distribución, ya que el número de observaciones es impar. Toda distribución impar tiene siempre un valor que ocupa la posición central. ¿Y si la distribución de valores fuese par? Añadimos a la ya conocida y ordenada distribución de valores un nuevo valor, cualquiera: 1, 3, 3, 4, 4, 5, 6, 7, 8 y 9. En una distribución con número par de observaciones no es uno, sino dos, los números que se ubican

en posición central. En el supuesto que nos ocupa son los valores 4 y 5. Aplicando una media aritmética entre los dos valores se concluirá que la mediana de esta nueva distribución de valores es 4,5. Todo es correcto: una mediana que proviene de valores enteros puede ser decimal y, además, no tiene por qué coincidir con ninguna de las observaciones originales.

¿Qué tienen que ver, no obstante, los percentiles con la mediana? Los percentiles son una generalización de la mediana. La transformación basada en los percentiles consiste en asignar a cada puntuación directa una puntuación porcentual, según la posición de las observaciones dentro del conjunto de observaciones. Si el valor porcentual es del 25%, se estará hablando del percentil 25. Y bajo esta denominación se estará identificando aquella puntuación directa que deja por debajo una cuarta parte de todas las observaciones. O expresado de otra manera, aquel valor que con él y por encima de él se encuentra el 75% de todas las observaciones. Si el porcentaje asciende a un 50%, se hablará del percentil 50, que se corresponderá con el valor que ocupará la posición central de la distribución. El percentil 50 es, efectivamente, la mediana.

La primera formulación práctica de los percentiles se remonta a finales del siglo XIX, y proviene del explorador y científico Francis Galton, primo de Charles Darwin, que entre otras aficiones estaba bastante interesado en la medida de multitud de rasgos antropométricos. Hoy en día, y sin abandonar el ámbito de la medida de las propiedades del cuerpo, los percentiles continúan siendo una estrategia de interés, por ejemplo, para aquellos padres y madres que desean conocer en qué grado sus hijos se ajustan a los otros niños de la misma edad en términos de altura y peso.

Los percentiles han resultado una medida muy empleada no solo en el contexto de la medida del peso o de la altura, sino también en el contexto de las medidas psicológicas. Su simplicidad y universalidad, dado que no hay que tener amplios conocimientos en matemáticas ni de estadística en particular, ha estimulado sin duda su difusión como estrategia interpretativa. Expongamos un caso práctico que servirá para ejemplificar el sentido y generalizar su cálculo.

En una muestra de 1.000 personas de nivel académico, edad y género diferentes (Gomà-i-Freixament et al., 2008) fue administrado el Zuckerman-Kuhlman Personality Questionnaire (ZKPQ, Zuckerman, Kuhlman, Joireman, Teta y Kraft, 1993), un instrumento que mide cinco dimensiones básicas de personalidad normal.

De las cinco dimensiones que el modelo permite medir, se ha centrado la atención, por motivos de simplificación, en una de las variables de personalidad, el neuroticismo¹. En este instrumento, esta dimensión de personalidad es operativizada mediante un total de 19 ítems de respuesta tipo verdadero-falso, donde el sumatorio de todos ellos (teniendo en cuenta que algunos ítems son directos y otros inversos) resulta ser una estimación del grado en el que un individuo tiende a presentar este rasgo de personalidad. Hay que estar atento a las peculiaridades de cada instrumento para saber cómo debe ser obtenida una puntuación en la variable de interés. No siempre se sigue una misma pauta para calcular las puntuaciones totales.

En el supuesto que nos ocupa, la puntuación máxima alcanzable es 19 y la mínima 0. Una mayor puntuación equivale a una mayor intensidad del rasgo. La distribución de valores absolutos y relativos de este grupo de 1.000 personas en esta variable aparece recogida en la tabla 1.

⁽¹⁾El neuroticismo expresa la tendencia de toda persona a mostrarse alterada, con tensión, preocupada, con indecisión y con labilidad emocional (estado de humor frecuentemente cambiante).

Tabla 1. Puntuaciones directas, frecuencias, porcentajes y percentiles

Puntuación en la escala	Frecuencia absoluta (fi)	Frecuencia acumulada (fa)	Porcentaje (%)	Porcentaje acumulado (%)	Percentil (Pc)
0	21	21	2,1	2,1	0
1	55	76	5,5	7,6	3
2	49	125	4,9	12,5	8
3	77	202	7,7	20,2	13
4	63	265	6,3	26,5	21
5	70	335	7,0	33,5	27
6	76	411	7,6	41,1	34
7	67	478	6,7	47,8	42
8	68	546	6,8	54,6	48
9	72	618	7,2	61,8	55
10	59	677	5,9	67,7	62
11	60	737	6,0	73,7	68
12	57	794	5,7	79,4	74
13	44	838	4,4	83,8	80
14	43	881	4,3	88,1	84
15	41	922	4,1	92,2	89
16	38	960	3,8	96,0	93
17	23	983	2,3	98,3	97
18	11	994	1,1	99,4	99

Puntuación en la escala	Frecuencia absoluta (fi)	Frecuencia acumulada (fa)	Porcentaje (%)	Porcentaje acumulado (%)	Percentil (Pc)
19	6	1000	0,6	100,0	100

Si se toma por ejemplo la puntuación 8, se puede observar que han sido 68 las personas que han presentado esta puntuación (columna de frecuencia absoluta). Dado que son un total de 1.000 personas en esta muestra, se puede afirmar que un 6,8% de los participantes tienen una puntuación directa de 8 (columna de porcentaje). La penúltima columna, identificada como porcentaje acumulado, ha exigido que los valores directos fueran dispuestos de manera ordenada. Los porcentajes correspondientes a cada puntuación se van sumando y acumulando. En el caso de la puntuación directa 8, el porcentaje acumulado es 54,6. Esto significa que el 54,6% de las personas de la muestra han obtenido puntuaciones de como máximo 8 unidades (8 o menos). ¿Cuál es el percentil asociado a este valor directo? Dado que debe ser el porcentaje de personas que quedan por debajo de este valor, habrá que observar que hasta llegar a una puntuación directa de 7 se observa un porcentaje acumulado del 47,8%. Este valor, una vez redondeado al entero inmediatamente superior, pasa a ser de un 48%. La puntuación directa 8 se asocia a un percentil de 48.

Si la correspondencia entre puntuaciones directas y percentiles se enfocara de manera inversa, en cambio, se podría formular una pregunta del tipo “¿A qué valor directo correspondería un percentil de 55?”. Según los porcentajes acumulados, hasta un valor directo de 8 le corresponde un porcentaje acumulado de 54,6%. Un porcentaje de 55, a pesar de que por poco, no se ha producido todavía. El siguiente valor directo es el 9, y hasta él se ha acumulado el 61,8% de las observaciones. En este caso sí, el 55% objeto de interés ya se ha producido, dado que está incluido en el 61,8%. La puntuación directa 9, por lo tanto, se asocia a un percentil de 55.

De acuerdo con este razonamiento habrá que entender por qué una puntuación de 9 es también percentil 56. Pero también 57, 58, 59, 60 y 61. Solo cuando se llega a la puntuación directa 10 el percentil pasa a ser 62. Esta circunstancia, en la que un mismo valor directo se vincula a más de un percentil, es del todo plausible cuando la variable objeto de interés no dispone de valores suficientes para cubrir las cien unidades para las que el percentil sí está preparado.

Hay que tener en cuenta que esta estrategia para calcular percentiles, basada en los porcentajes acumulados, no es la única posible. Existen otras aproximaciones que son también aceptadas. Y decimos aproximaciones porque el estudiante deberá tolerar que no todas ellas lleguen necesariamente a idéntico resultado. Una estrategia alternativa a la ya expuesta para calcular un percentil, cualquiera, es la que se propone a continuación:

$$P = \frac{fa + 0,5fi}{N} \times 100$$

Así pues, y de acuerdo con la tabla 1, el percentil que se asocia a una puntuación directa de 9 es 58:

$$P9 = \frac{546 + (0,5 \times 72)}{1000} \times 100 \approx 58$$

Lectura de la fórmula

f_a : Frecuencia acumulada previa a la puntuación directa de la que se quiere calcular el percentil.
 f_i : Frecuencia absoluta en la que se encuentra la puntuación directa.
 N : Número de personas que constituyen la muestra.

Dado que los percentiles no tienen decimales, se ha ignorado el decimal del resultado exacto obtenido (58,2). Hay que recordar que siguiendo la primera estrategia expuesta para calcular percentiles, la puntuación directa 9 era, a la vez, los percentiles 55, 56, 57, 58, 59, 60 y 61. El valor obtenido en la fórmula anterior es solo uno de los valores, el central.

Enlace recomendado

En Wikipedia (en inglés) y en otros recursos de Internet, encontraréis otras estrategias alternativas, también válidas, para calcular los percentiles.

La primera de las ventajas del uso de los percentiles, como ya se ha indicado, es su simplicidad conceptual. Una segunda ventaja es que en el cálculo y la correcta interpretación de este, dado que centra su atención solo en la orden en el que las observaciones son dispuestas, resulta irrelevante el modo como se distribuyen los valores objeto de análisis. Esto implica que, a pesar de que hay estrategias de transformación que requieren una distribución de valores esencialmente ajustada a una curva normal, más adelante se verá, en el caso de los percentiles es posible abordar cualquier tipo de distribución. Esta característica es especialmente interesante en psicología, donde muchas de las variables de interés raramente se ajustan a una distribución normal de valores.

Como se ha visto, los percentiles consisten en dividir una distribución de valores en cien posiciones ordinales. No obstante, existen otros abordajes que simplifican la cantidad de órdenes posibles y que pueden ser suficientes para resumir el comportamiento de un conjunto de valores. Los más comunes son:

- Los **cuartiles**. Dividen una distribución en cuatro partes. Se suelen identificar como cuartil 1, cuartil 2 y cuartil 3. Coinciden con los percentiles 25, 50 y 75, respectivamente.
- Los **quintiles**, que al dividir la distribución en cinco partes se equiparan a los percentiles 20, 40, 60 y 80.
- Los **deciles**. Suponen una división en diez partes y se corresponden con los percentiles 10, 20, 30, 40, etc.

2.2. Puntuaciones estandarizadas

Si bien la media es un estadístico de tendencia central que permite expresar el comportamiento medio de un conjunto de observaciones, la desviación típica (o estándar) permite cuantificar lo diferentes que son entre sí estas observaciones. Este índice de dispersión expresa la discrepancia promedio entre una observación cualquiera y la media del conjunto de observaciones. La media y la desviación típica son la materia prima imprescindible para calcular una puntuación estandarizada. Si bien los percentiles dirigían la atención hacia la posición que ocupa un individuo respecto a un grupo de referencia, en el caso de las puntuaciones estandarizadas el interés se focaliza en la discrepancia que presenta este individuo respecto al modo de comportamiento promedio del grupo, su media.

La puntuación estandarizada o puntuación típica es aquella que permite expresar cuántas desviaciones típicas por encima o por debajo de una media se sitúa una observación. Este proceso de estandarización tiene su sentido cuando la distribución de valores objeto de interés se ajusta esencialmente a una curva normal. El valor resultante de esta transformación se constituye en una nueva puntuación, que recibe el nombre de puntuación estandarizada z . Su cálculo responde a la fórmula siguiente:

$$Z_x = \frac{X - \bar{X}}{S_x}$$

Hay que tener presente que esta puntuación z tendrá siempre una media de 0 y una desviación estándar de 1. En la tabla 2, y retomando el estudio sobre neuroticismo en una muestra de 1.000 personas:

Lectura de la fórmula

X : Puntuación directa.
 \bar{X} : Media de la muestra.
 S_x : Desviación típica de la muestra.

Tabla 2. Puntuaciones directas, frecuencias y puntuaciones estandarizadas Z

Puntuaciones directas	Frecuencia absoluta	Puntuaciones estandarizadas z
0	21	-1,72
1	55	-1,51
2	49	-1,30
3	77	-1,09
4	63	-0,88
5	70	-0,66
6	76	-0,45
7	67	-0,24
8	68	-0,03
9	72	0,18
10	59	0,39
11	60	0,61
12	57	0,81
13	44	1,03
14	43	1,24
15	41	1,45
16	38	1,67
17	23	1,88
18	11	2,09
19	6	2,30

Para poder calcular los valores z , ubicados en la última columna de la tabla, es necesario conocer la media y la desviación típica de la variable en esta muestra: la media era 8,14 y la desviación típica 4,72. Así pues, en el caso concreto de una persona que ha obtenido una puntuación directa de 8, su puntuación z será:

$$-0,03 = \frac{8 - 8,14}{4,72}$$

¿Qué significa este valor de $-0,03$? Dado que la media de z siempre es 0 y que su desviación típica también es siempre 1, observar una puntuación típica de $-0,03$ significa que los sujetos que tienen una puntuación directa de 8, y que son un total de 68 personas, se ubican en 0,03 desviaciones típicas de la media de la muestra. Sabemos, además, que este valor directo se ubica por debajo de la media porque su puntuación típica es negativa. Necesariamente, todos los sujetos que presentan puntuaciones directas que se disponen por debajo de la media obtendrán puntuaciones z negativas, mientras que aquellos que se ubican por encima de la media presentan puntuaciones típicas positivas. Si nos fijamos en la parte final de la tabla 2, veremos que las 6 personas que obtienen una puntuación de 19 tienen una puntuación típica de 2,30, lo que implica que estas se ubican a más de 2 desviaciones típicas por encima de la media. En cambio, las 21 personas que obtuvieron una puntuación directa de 0 presentan una puntuación típica de $-1,72$. Son personas que quedan, por lo tanto, a 1,72 desviaciones por debajo de la media.

Una pregunta importante que nos podemos formular es si distanciarse en 1,72 o 2,30 desviaciones típicas de una media es mucho o poco. Está claro que alejarse de una media en 2,30 desviaciones típicas es hacerlo más que si lo hacemos 1,72 veces. Podéis ver en la tabla 2 que en esta muestra las personas con la máxima puntuación se sitúan más alejadas de la media que las personas que obtienen la puntuación más baja posible. Ahora bien, ¿es mucho hacerlo 2,3 veces?

En el estudio de Valero et al. (2012), se comparan dos muestras de sujetos adultos: una constituida por 217 personas adultas afectadas por un trastorno por déficit de atención e hiperactividad (TDAH) y la otra formada por 434 sujetos control emparejados por edad y género. Distintas variables de personalidad fueron contrastadas, pero para simplificar la exposición se centrará de nuevo la atención sobre la dimensión de neuroticismo. La tabla 3 expone las puntuaciones típicas en esta variable para cada uno de los dos grupos de estudio.

Tabla 3. Puntuaciones directas y estandarizadas para una muestra de personas con TDAH y controles

Puntuaciones directas	Puntuación estandarizada z	
	TDAH	Controles
0	-2,54	-1,67

Puntuaciones directas	Puntuación estandarizada z	
1	-2,38	-1,46
2	-2,16	-1,25
3	-1,94	-1,04
4	-1,73	-0,83
5	-1,51	-0,62
6	-1,30	-0,41
7	-1,08	-0,19
8	-0,87	0,02
9	-0,65	0,23
10	-0,43	0,44
11	-0,22	0,65
12	0	0,86
13	0,21	1,08
14	0,43	1,29
15	0,65	1,50
16	0,86	1,71
17	1,08	1,92
18	1,29	2,14
19	1,51	2,35
	Media = 12,01	Media = 7,91
	Desviación típica = 4,63	Desviación típica = 4,72

Como se puede ver a pie de tabla, los sujetos afectados por TDAH presentan una media en la medida de neuroticismo superior a la presentada por los sujetos control (12,01 frente a 7,91). Se trata de un resultado esperable, teniendo en cuenta que una es muestra psiquiátrica y la otra no. Las respectivas desviaciones típicas, en cambio, y en este estudio, son muy comparables. Si para cada puntuación directa se comparan las respectivas puntuaciones típicas de las dos muestras, se observará que no coinciden en ningún caso. De hecho, se muestran claramente diferenciadas.

En el caso de las personas afectadas por TDAH, las puntuaciones tipificadas tienden a estar sistemáticamente desplazadas hacia la parte baja de la distribución, respecto a las presentadas por los sujetos control. La discrepancia entre las dos medias condiciona este comportamiento. Si bien una puntuación directa de 0 en el caso de los sujetos control implica ubicarse en 1,67 desviaciones

típicas por debajo de la media, en el caso de los pacientes afectados de TDAH, la misma puntuación directa los aleja más de dos desviaciones típicas y media de la respectiva media. Todo depende del grupo en el que se esté ubicado.

Y aquí descansa una de las consideraciones más relevantes que hay que tener presente cuando se interpretan puntuaciones estandarizadas: las inferencias a las que podemos llegar mediante el uso de estas puntuaciones dependen de la muestra que es empleada, dado que es de ella de donde se obtienen la media y desviación típica que permiten el cálculo de las puntuaciones transformadas. Elegir cuidadosamente la muestra de referencia resulta imprescindible. Más adelante, cuando se hable del concepto de baremos, se retomará esta idea.

2.3. Puntuaciones estandarizadas derivadas

Aunque intrínsecamente no es un problema, las puntuaciones z se expresan con valores negativos y decimales. El sentido básico de las puntuaciones estandarizadas derivadas es transformar las puntuaciones típicas según un sistema de codificación que elimina estos valores negativos y decimales. Una de las puntuaciones típicas derivadas más comúnmente empleadas es la puntuación T de McCall. Su formulación es sencilla:

$$T = 50 + 10z$$

Consiste en partir de una constante, en este caso 50, y sumarle diez veces el valor de la puntuación típica z . Para interpretar debidamente esta puntuación solo hay que tener presente que la media de esta puntuación es 50 y su desviación típica 10. De este modo, por ejemplo, una puntuación z de 2 tendrá un valor de 70 una vez transformada en puntuación T . En ambos casos (z y T), y necesariamente, sería indicador de una puntuación directa que se ubica por encima de la correspondiente media en dos desviaciones típicas.

La tabla 4 recoge los valores de la tabla 2 añadiendo las puntuaciones T correspondientes:

Tabla 4. Puntuaciones directas, frecuencias y puntuaciones estandarizadas z y T

Puntuaciones en la escala	Frecuencia absoluta	Puntuación estandarizada z	Puntuación estandarizada derivada T
0	21	-1,72	33
1	55	-1,51	35
2	49	-1,30	37
3	77	-1,09	39
4	63	-0,88	41
5	70	-0,66	43
6	76	-0,45	46

Puntuaciones en la escala	Frecuencia absoluta	Puntuación estandarizada z	Puntuación estandarizada derivada T
7	67	-0,24	48
8	68	-0,03	50
9	72	0,18	52
10	59	0,39	54
11	60	0,61	56
12	57	0,81	58
13	44	1,03	60
14	43	1,24	62
15	41	1,45	65
16	38	1,67	67
17	23	1,88	69
18	11	2,09	71
19	6	2,30	73

Es posible encontrar puntuaciones típicas derivadas de distinta naturaleza, y es que estrictamente cualquiera puede crearse la suya, dado que una puntuación derivada se formula según criterios arbitrarios. Ejemplos diferentes los encontraremos en el test de inteligencia Stanford-Binet, que emplea una media de 100 y una desviación típica de 16, la Wechsler Adult Intelligence Scale (WAIS), con una media también de 100, pero con una desviación típica de 15, o el Minnesota Multiphasic Personality Inventory (MMPI), que, coincidiendo con la escala de McCall, se interpreta en torno a una media de 50 y una desviación típica de 10.

2.4. Puntuaciones estandarizadas normalizadas

Si se asume que la distribución de valores que es objeto de análisis se ajusta a una curva normal, es posible transformar los percentiles en aquellas puntuaciones estandarizadas que se corresponderían si la distribución fuera, efectivamente, normal. Estas puntuaciones recibirían el nombre de *puntuaciones estandarizadas normalizadas*.

Tabla 5. Puntuaciones directas, frecuencias, puntuaciones estandarizadas z y T y normalizada

Puntuación en la escala	Frecuencia absoluta (f_i)	Frecuencia acumulada (f_n)	Porcentaje acumulado	Percentil (P_c)	Puntuación estandarizada z	Puntuación estandarizada normalizada
0	21	21	2,1	2	-1,72	-2,05
1	55	76	7,6	3	-1,51	-1,88
2	49	125	12,5	8	-1,30	-1,41

Puntuación en la escala	Frecuencia absoluta (f_i)	Frecuencia acumulada (f_a)	Porcentaje acumulado	Percentil (P_c)	Puntuación estandarizada z	Puntuación estandarizada normalizada
3	77	202	20,2	13	-1,09	-1,13
4	63	265	26,5	21	-0,88	-0,81
5	70	335	33,5	27	-0,66	-0,61
6	76	411	41,1	34	-0,45	-0,41
7	67	478	47,8	42	-0,24	-0,20
8	68	546	54,6	48	-0,03	-0,05
9	72	618	61,8	55	0,18	0,13
10	59	677	67,7	62	0,39	0,31
11	60	737	73,7	68	0,61	0,47
12	57	794	79,4	74	0,81	0,64
13	44	838	83,8	80	1,03	0,84
14	43	881	88,1	84	1,24	0,99
15	41	922	92,2	89	1,45	1,23
16	38	960	96,0	93	1,67	1,48
17	23	983	98,3	97	1,88	1,88
18	11	994	99,4	99	2,09	2,33
19	6	1000	100,0	100	2,30	-

De acuerdo con la tabla 5, si se toma por ejemplo la puntuación directa 9, se observará que, como ya se sabía, su puntuación estandarizada z es 0,18 y su percentil 55. El paso siguiente es buscar una tabla estandarizada de puntuaciones normales, también conocida como tabla de puntuaciones z . En Internet encontraréis muchas². La tabla 6 es el extracto de una de estas tablas. Si en ella se busca una proporción de 0,55 se localizará el valor z correspondiente.

⁽²⁾En un buscador teclead, por ejemplo, z tablas.

En el supuesto que nos ocupa la proporción más cercana es 0,5517, y en consecuencia el valor z que se vincula a ella es 0,13. Cualquier percentil, a modo de proporción, tiene asociado un valor z y viceversa.

Tabla 6. Extracto tabla puntuaciones z

z	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517

z	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224

Reiteramos que el valor obtenido en una tabla de valores z asume que la distribución de valores que es objeto de interés se ajusta básicamente a una curva normal. La discrepancia observada entre las puntuaciones 0,18 y 0,13 (tabla 5), o genéricamente entre cualquier puntuación estandarizada calculada y la correspondiente puntuación estandarizada normalizada, responde a la discrepancia existente entre la distribución de valores empíricos y una distribución teórica normal. Hay que tener presente que la presencia de discrepancia entre las dos distribuciones de valores tipificados no será la excepción, sino la norma. Como ya hemos mencionado con anterioridad, son muchas las variables que no se ajustarán a una (perfecta) distribución normal. Para aquellas que lo hagan esencialmente, he aquí una manera posible de transformar una puntuación.

Del mismo modo que en el caso de las puntuaciones estandarizadas derivadas, es posible disponer también de puntuaciones estandarizadas normalizadas derivadas. En psicología las más habituales son las estaninas o eneatis y los decatipos. Las respectivas formulaciones son:

$$\text{Eneatis} = 5 + 2z_n$$

$$\text{Decatipos} = 5,5 + 2z_n$$

Donde la z es la puntuación estandarizada normalizada.

Asumiendo una distribución normal de valores, si clicáis en un buscador términos como *z distribution percentile*, podréis ver gráficamente la correspondencia que se establece entre algunos de los procedimientos de transformación presentados en este módulo.

2.5. Normas cronológicas

Otra posible transformación de las puntuaciones directas consiste en hacerlo como normas cronológicas o de edades. La estrategia a seguir consiste en administrar a varios grupos de edad, representativos de una determinada población, una misma medida, recogiendo para cada grupo de edad la respectiva media. Posteriormente, cuando un individuo es evaluado, su puntuación es transformada en una edad, aquella edad que en la población de referencia se aproximaría a la puntuación obtenida. La idea básica que justifica esta fórmula es simple: asumir que existen formas de comportamiento que tienden a ser típicos en una edad. Así pues, si después de seleccionar una muestra representativa de una determinada población se administra una determinada medida

a los niños de 6 años y la media obtenida es de 12, y la misma medida administrada a los niños de 8 años es de 17, si un niño concreto obtiene una puntuación de 17, se afirmará que en esta medida el niño tiene una edad de 8 años (aunque cronológicamente tenga 7).

Uno de los contextos donde ha sido más popular esta estrategia es el del cociente de inteligencia. Su formulación es:

$$CI = \frac{EM}{EC} \times 100$$

La transformación en edades es una estrategia antigua, probablemente tanto como lo ha sido la medida de los constructos psicológicos. No obstante, no se puede dejar de mencionar que esta transformación se acompaña de limitaciones nada desdeñables, lo que han ido reduciendo a lo largo del tiempo su presencia en la bibliografía. Una de sus limitaciones más relevantes es que el modelo asume que los cambios que se van produciendo a lo largo del tiempo, según se analizan varios grupos de edad, son cambios esencialmente lineales, es decir, que el cambio producido entre las personas de 6 a 7 años es comparable a los cambios que se producen entre las personas de 8 respecto a las de 9 años. Desgraciadamente, asumir este principio es excesivo para muchos procesos psicológicos. Cambios producidos entre determinadas edades cronológicas no tienen por qué tener la misma relevancia o significado a lo largo de todas las edades mentales ni, por supuesto, al revés.

Lectura de la fórmula

EM: Edad mental.
EC: Edad cronológica.

Lectura recomendada

En Fancher (1987) encontraréis un texto interesante, en el que no solo se aborda el concepto de edad mental sino que también se hace un repaso histórico del concepto de inteligencia, así como de algunas controversias en torno a su medida.

3. Baremación

La baremación, también conocida como escalamiento de una medida, es un elemento de gran relevancia en el proceso de elaboración y uso de un instrumento de medida. Como se ha visto en el presente módulo, la transformación de una puntuación ayuda al usuario a inyectar significado a un determinado valor de acuerdo con el comportamiento de un conjunto de observaciones. Si el instrumento dispone de un manual, este debería reportar los baremos de la medida, con las transformaciones necesarias si es pertinente, además de las indicaciones básicas para interpretar debidamente cualquier puntuación. En otros casos no habrá manual de referencia, pero sí probablemente una o más publicaciones periódicas a donde podremos remitirnos y hallar la información necesaria.

Independientemente del tipo de publicación con la que el baremo de una medida se reporte, para que toda interpretación que se derive de este se haga según un elevado estándar de calidad, es necesario que la muestra que proporciona los datos que permiten satisfacer una transformación (por ejemplo, medias y desviaciones típicas), y que recibirá el nombre de *muestra normativa*, sea una muestra relevante, representativa y homogénea.

Relevante porque hay que garantizar que la muestra que conforma los baremos de un instrumento está haciendo referencia a una población que es significativa de acuerdo con la satisfacción de unos objetivos específicos. Tal y como se ha visto durante la exposición de las puntuaciones estandarizadas, el sentido que puede tener una determinada puntuación tipificada debe ser entendido siempre dentro del marco de la población que proporciona los parámetros de referencia. No hay una población *a priori* ideal sobre la que desarrollar el baremo de un instrumento de medida. La población ideal depende del tipo de persona que se desea describir y, sobre todo, de los términos en los que se proponga hacerlo.

Probablemente todo el mundo estaría de acuerdo en que si se quiere inferir algo relevante sobre el grado de conocimiento de los jóvenes de un instituto sobre la Revolución Industrial sería conveniente poder disponer de una amplia muestra de otros jóvenes provenientes de otros institutos con los que contrastar el rendimiento. No obstante, no siempre es posible identificar una única población respecto a la que inferir satisfactoriamente el estado de un individuo.

De acuerdo con la misma línea argumental expuesta al hablar del neuroticismo y TDAH, si se desea interpretar la intensidad de la sintomatología depresiva de una persona mediante un determinado instrumento clínico, una puntuación concreta no tiene la misma consideración si los baremos empleados provienen de la población en general, de varios centros de atención primaria o de unidades de hospitalización psiquiátrica. *A priori*, las tres poblaciones son igualmente válidas, en cuanto que en todas ellas puede ser pertinente obtener datos normativos e interpretar la gravedad de la sintomatología depresiva de esta persona. Ahora bien, no es lo mismo ubicarse en un percentil de 70 tomando como muestra normativa población general que población clínica. La excepcionalidad (o no) de la sintomatología de este individuo dependerá necesariamente de cuál sea la población de referencia expresada mediante los baremos, y esto condicionará la interpretación que se pueda derivar, así como las decisiones que habría que tomar al respecto (McIntire y Miller, 2007). Hay que delimitar debidamente, a modo de objetivos de trabajo claros, qué debe ser evaluado y, sobre todo, con qué finalidad.

Se ha afirmado también que la muestra constitutiva de los baremos ha de ser **representativa**. Una vez definida cuál debe ser la población diana, la estrategia empleada para seleccionar a las personas de la población que formarán parte de la muestra normativa se transforma en el foco de atención principal. De acuerdo con los estándares que proporciona la American Psychological Association (APA, 1999), hay que exponer con cuidado y claridad cuál ha sido la población objetivo y, a la vez, la muestra que se ha seleccionado. Solo así es posible delimitar el tipo de inferencia a la que se puede llegar cuando unos determinados baremos son empleados.

Un muestreo probabilístico es la mejor estrategia metodológica disponible para asegurar la obtención de una muestra representativa, dado que solo es en este contexto aleatorio donde cualquier individuo de la población dispone de idéntica probabilidad de formar parte de la muestra. Y tal y como expone Aiken (1996), el muestreo aleatorio estratificado debería ser la estrategia de elección. Este tipo de muestreo debería proporcionar la debida representatividad para todos los parámetros que han sido considerados relevantes. En este sentido, la edad y el género suelen ser algunos de los factores que más frecuentemente articulan un baremo, pero cualquier otra variable puede ser objeto de interés y por lo tanto de estratificación. El ámbito de estudio y los objetivos que deben ser resueltos condicionarán el tipo de variables que se someten a escalamiento y el grado de estratificación con el que estas variables deberían expresarse.

Desgraciadamente no siempre se explicita en los textos especializados, pero no se puede dejar de mencionar, que no suele ser fácil poder satisfacer un muestreo estrictamente probabilístico. Son muchas las condiciones que, en contexto aplicado, reducen el éxito de un muestreo aleatorio. Considérese por ejemplo la presencia de personas institucionalizadas a las que difícilmente se tendrá acceso (prisioneros, personas hospitalizadas, etc.), aquellas que no tienen

las aptitudes necesarias para dar respuesta a una determinada medida (analfabetas, con problemas de comprensión, etc.), las evaluadas en contextos en los que puede haber una alta deseabilidad social (selección de personal, presencia de preguntas comprometidas, etc.), personas que simplemente no quieren colaborar, etc. Los baremos de la mayoría de los instrumentos de evaluación, si no todos, están sujetos a circunstancias que limitan la generalización de sus valores. El esfuerzo de todo investigador debería dirigirse a minimizar o atenuar dentro de lo posible estos sesgos y, en todo caso, informar siempre detenidamente a los usuarios del instrumento respecto a qué se ha hecho y cómo se ha resuelto.

Cuando se habla de la **homogeneidad** de la muestra se está haciendo referencia a la necesidad de controlar la presencia de efectos de confusión o extraños que puedan sesgar o incluso invalidar las interpretaciones que se deriven de esa muestra. Las personas que serán evaluadas mediante una determinada medida deberían ser comparables a las personas que formaron parte de la muestra de participantes que constituyeron el baremo de esta. Discrepancias consistentes entre las dos muestras pueden derivar en inferencias que sobreestimarían o infraestimarían el estado real de las personas evaluadas.

Estas mismas consideraciones justifican la necesidad de que haya, también, una continua actualización de los datos que estructuran los baremos. Se debería poder disponer de baremos que se sometan a revisiones con una determinada regularidad. Y es que habría que asumir por defecto que los datos contenidos en un baremo caducan. Las variables que son objeto de interés son dinámicas y, por lo tanto, cambiantes, moduladas por el decurso de los tiempos. Son sensibles a las transformaciones científicas, sociales y culturales que se van produciendo y, por esta razón, todo baremo debería poder reflejar fielmente el estado actual del concepto que es objeto de consideración. A pesar de que las consideraciones teóricas y empíricas son diferentes, emplear en la actualidad baremos de calidad de vida en muestras y bajo criterios de hace treinta años, por ejemplo, sería comparable a emplear baremos de altura incluyendo a personas evaluadas durante los años sesenta. En cualquiera de las dos opciones se estaría cometiendo un error o, en el mejor de los casos, un sesgo difícilmente asumible.

La adaptación del instrumento a las características lingüísticas y culturales de las personas que han de ser evaluadas debe ser también motivo de una necesaria atención. En la mayoría de los casos los instrumentos de evaluación que son empleados en psicología en nuestro entorno cultural tienen su origen en el mundo anglosajón. La ventaja que presenta esta circunstancia es que permite que, una vez traducido un instrumento, resulte más fácil la homologación del instrumento desde un punto de vista internacional, a la vez que facilita la satisfacción de estudios de tipo transcultural en los que un mismo rasgo contrasta entre varios grupos culturalmente diferentes. No obstante, hay que tener presente que traducir un instrumento es condición necesaria pero no siempre suficiente. Siguiendo las indicaciones de los estándares de la American

Psychological Association (APA, 1999), es necesario que haya una adecuada adaptación cultural que sea sensible a las peculiaridades de los individuos que serán evaluados con el objetivo. Una correcta adaptación de la medida constituye un objetivo dirigido, de nuevo, a proporcionar estimaciones carentes de sesgo y que, por lo tanto, permitan inferir con las suficientes garantías de validez y fiabilidad el estado de los individuos.

En la configuración del baremo de un instrumento es posible emplear todas las estrategias expuestas en el presente módulo. Los datos descriptivos, dispuestos en tablas, se pueden expresar de un modo distinto, por ejemplo mediante percentiles, y agrupados según estratificaciones por edad (Rodríguez et al., 2012). En otros casos, los percentiles se dispondrán sin estratificación (Alegret et al., 2012; Romero et al., 2000). En algunas publicaciones se ofrecerán medianas y desviaciones típicas que permitirán, si se deseara, una transformación en puntuaciones estandarizadas. Y lo pueden hacer según dos factores, por ejemplo, por edad y género (Gomà-i-Freixenet y Valero, 2008), o bien sin estratificación alguna (Aluja et al., 2008). Otros proporcionarán puntuaciones estandarizadas derivadas, a modo de puntuaciones T, y de manera estratificada (Yueh-Hsien et al., 2012). Las combinaciones son evidentemente múltiples, no son más que algunos de los muchos ejemplos que pueden ser presentados. Queda en manos del investigador proporcionar la información necesaria para justificar una u otra opción.

4. Equiparación de puntuaciones

Equiparar las puntuaciones de dos o más medidas consiste en establecer una correspondencia entre las puntuaciones de estas. Esta estrategia debe permitir que, recogida la puntuación en una de las medidas, sea posible estimar la puntuación de las otras medidas.

Uno de los contextos en el que puede ser especialmente interesante resolver equiparaciones entre tests es aquel en el que es necesario medir las aptitudes de personas repetidamente y en el que, por lo tanto, se pueden dar procesos de aprendizaje o memoria, o simplemente cansancio, que podrían conducir hacia una más que probable estimación sesgada de las capacidades de las personas. En estas condiciones sería interesante poder disponer de dos o más formas alternativas de la medida que permitiera su uso indistintamente. No obstante, las condiciones que deberán ser satisfechas para una equiparación adecuada son las siguientes:

- a) Que se mida la misma característica.
- b) Que las dos medidas sean igualmente fiables.
- c) Que la transformación sea invertible, es decir, la transformación debe ser posible de A a B, pero también de B a A.

Dos tipos de transformaciones son posibles: la equiparación horizontal y la vertical. El primer caso se daría cuando el rasgo de interés debe mantener su dificultad entre las diferentes medidas alternativas. Un solo grupo de sujetos suele ser el que, generalmente, proporciona las dos o más medidas. En el caso vertical, en cambio, se asume que la dificultad entre formas alternativas es diferente. En esta última circunstancia, son evaluados dos o más grupos de personas que presentan capacidades diferenciadas.

Los distintos diseños que permiten crear equiparaciones entre medidas son los siguientes:

- los de grupo único,
- los de grupos equivalentes,
- los tests de anclaje.

En el caso de los diseños de **grupo único**, un mismo grupo de personas extraído aleatoriamente de la población destinataria de las medidas equivalentes es evaluado mediante las diferentes medidas de interés. Como estrategia de con-

trol, para evitar procesos de aprendizaje y memoria, y por lo tanto un posible sesgo como consecuencia de haber administrado primero una forma alternativa y posteriormente la otra, se recomienda que las muestras respondan a las medidas de manera contrabalanceada. Así, se puede dividir la muestra aleatoriamente en dos grupos y administrar a cada subgrupo las distintas formas alternativas en un orden diferente, por ejemplo, en un grupo la secuencia A-B y en el otro la B-A.

En los diseños que implican **grupos equivalentes** son dos los grupos que se seleccionan aleatoriamente, pero en este caso es administrada una medida a la primera muestra y otra a la segunda. Poder asumir con garantías que las dos muestras de sujetos son comparables, por ejemplo, mediante una asignación aleatoria de los sujetos, es condición necesaria para asegurar la validez del procedimiento.

La ventaja de emplear dos grupos de personas es que, en comparación con la estrategia anterior, reduce el efecto cansancio y no expone a los participantes a procesos de aprendizaje o memoria que podrían ser arrastrados entre las diversas medidas, dado que los participantes solo se exponen en una de ellas. La limitación más importante de esta aproximación es que, en general, exige incluir una cantidad de participantes superior a la necesaria en el caso del primer diseño.

Los **tests de anclaje** parten, también, de disponer de dos grupos de participantes que reciben cada uno de ellos una de las medidas dirigidas a ser equiparadas. Sin embargo, en este caso, a cada muestra se le administran algunos ítems que forman parte de la medida que es administrada en la otra muestra. Estos ítems comunes pasan a ser así ítems de anclaje. La estrategia es especialmente interesante cuando no es posible garantizar que los dos grupos de personas sean comparables, situación frecuente en un contexto aplicado en el que no ha sido posible generar los dos grupos de personas de manera aleatoria. Bajo esta circunstancia, los ítems de anclaje proporcionarán una estimación del grado en el que las dos muestras resultan comparables.

Independientemente de cuál de las tres estrategias haya sido empleada para satisfacer la equivalencia entre medidas, dos son los métodos más comunes que pueden ser ejecutados para resolver la equiparación entre medidas: equipercantil y transformación lineal.

La **equiparación equipercantil**, que suele ser la estrategia más habitual, consiste en considerar como equivalentes aquellas puntuaciones directas que se vinculan a un mismo percentil. Así pues, si las puntuaciones directas 10 y 14 obtenidas con la medida A y B, respectivamente, se asocian ambas a un percentil 30, se considerará que las dos puntuaciones son equivalentes. El mismo razonamiento sería aplicable al resto de los pares de puntuaciones directas que convergieran en un mismo percentil.

Lectura recomendada

En el texto de Muñiz (2003) encontraréis más información respecto a esta estrategia.

Las consideraciones que deben ser tenidas en cuenta a la hora de aplicar esta estrategia son las mismas que ya han sido expuestas al hablar de los percentiles. Se trata, pues, de una estrategia de fácil comunicación y ejecución.

La **transformación lineal** consiste en equiparar puntuaciones estandarizadas. Concretamente consiste en equiparar aquellas puntuaciones directas que convergen en una misma puntuación típica. Expresado de otro modo, y de acuerdo con la fórmula que permite calcular una puntuación típica z , se puede decir que:

$$\frac{X - \bar{X}}{S_x} = \frac{Y - \bar{Y}}{S_y}$$

De donde si se aísla Y :

$$Y = \frac{S_y}{S_x}(X - \bar{X}) + \bar{Y}$$

Ejemplo

Consideremos el ejemplo de dos muestras equivalentes de personas extraídas aleatoriamente de una determinada población (diseño de grupos equivalentes). La primera muestra obtiene, en una medida de aptitudes aritméticas (X), una media de 10 y una desviación típica de 4. La segunda muestra, en la segunda medida también de aptitudes aritméticas (Y), obtiene una media de 8 y una desviación típica de 2. Según esta información, una puntuación de 9 en la medida A ¿a qué puntuación directa correspondería en la medida B ?

$$Y = \frac{4}{2}(9 - 10) + 8 = 6$$

Obtener un 9 en la medida A es equivalente a obtener una puntuación de 6 en la medida B .

En el caso del diseño de un solo grupo en el que ha sido empleada una estrategia de contrabalanceo, hay que tener en consideración el orden en el que fueron administradas las dos medidas (orden $A-B$ y orden $B-A$). Bajo esta circunstancia se calcula la desviación típica y media en cada uno de los grupos de balanceo. Así pues, empezando por la desviación típica:

$$S_x = \sqrt{S_{x1}^2 + S_{x2}^2}$$

$$S_y = \sqrt{S_{y1}^2 + S_{y2}^2}$$

Donde los estadísticos incluidos en los radicales son las varianzas correspondientes al primer y segundo órdenes en la administración de las dos medidas.

En el caso de las medias:

$$\bar{X} = \frac{\bar{X}_{x1} + \bar{X}_{x2}}{2}$$

$$\bar{Y} = \frac{\bar{X}_{y1} + \bar{X}_{y2}}{2}$$

Siendo los estadísticos de los numeradores las medias correspondientes al primer y segundo orden de las dos medidas. Calculadas las medias y desviaciones típicas, solo falta aplicar la fórmula ya conocida que permite establecer la correspondencia entre las puntuaciones directas equivalentes de las dos medidas.

Ejemplo

Las medidas X e Y han sido administradas a dos grupos de personas de manera contrabalanceada. La medida X , cuando fue administrada en primer lugar, presentó una media de 10 y una desviación típica de 2. Cuando fue administrada en segundo lugar, la media fue de 8 y la desviación típica de 1. En el caso de la medida Y , administrada en primer lugar, presentó una media de 12 y una desviación típica de 3, mientras que administrada en un segundo orden su media fue 14 y su desviación típica 4. ¿Qué puntuación se correspondería en la medida Y , si en la medida X se obtuviera una puntuación de 7?

Calculando las desviaciones típicas de las medidas X e Y según el orden en el que fueron administradas, respectivamente:

$$2,24 = \sqrt{4+1}$$

$$5 = \sqrt{9+16}$$

En el caso de las medias, respectivamente:

$$9 = \frac{10+8}{2}$$

$$13 = \frac{12+14}{2}$$

Estableciendo la correspondencia entre las dos medidas, se llegaría a la conclusión de que la puntuación en la medida Y debería ser 9:

$$\frac{5}{2,24}(7-9)+13 = 8,54 \approx 9$$

La aplicación de esta forma de equiparación entre medidas debe ajustarse a las mismas consideraciones que fueron recogidas al hablar de las puntuaciones estandarizadas. Esto implica, por lo tanto, depositar el interés sobre la tendencia de respuesta de los individuos en las dos medidas y asumir que las dos distribuciones de valores objeto de equiparación se ajustan básicamente a una distribución normal.

Resumen

En la primera parte del presente módulo se ha hablado de los percentiles, las puntuaciones estandarizadas y las normas cronológicas, puntuaciones transformadas comúnmente empleadas, dirigidas a otorgar interpretabilidad a las puntuaciones obtenidas en una medida. Los percentiles, calculados de acuerdo con la posición de una persona en un continuo y expresados como porcentajes, se convierten en una aproximación fácil de entender y por lo tanto de difundir, incluso a personas poco acostumbradas al uso de este tipo de estrategias matemáticas. Además, el hecho de ser esencialmente insensibles a los grupos de valores no ajustados a una distribución normal permite que se puedan considerar una transformación de amplio espectro, una estrategia útil y aplicable para la mayoría de las variables cuantitativas que son de interés en psicología.

Las puntuaciones estandarizadas presentan más limitaciones en términos de aplicabilidad. Requieren que las observaciones que son objeto de análisis se ajusten básicamente a una distribución normal. Y tal y como se ha recogido, son muchas las variables psicológicas que, evaluadas en una muestra concreta de personas, acaban sin satisfacer este modo de distribución. No obstante, cuando esta circunstancia puede ser asumida, la estandarización supone una aproximación diferenciada a la de los percentiles, que puede resultar especialmente interesante. A diferencia de los percentiles, se focalizan en la tendencia de respuesta, la media de comportamiento que proporciona el grupo normativo. La desviación típica, la diferencia entre los individuos del grupo normativo, se convierte en la manera de cuantificar la discrepancia entre la puntuación de una persona concreta y ese comportamiento promedio.

Concluyendo el conjunto de estrategias de transformación, se ha expuesto la puntuación a modo de normas cronológicas o edad. En este caso, y de acuerdo con la tendencia de respuestas observadas de nuevo en un grupo normativo, constituido por edades diferentes, la puntuación de un individuo pasa a ser transformada e interpretada de acuerdo con esta variable cronológica. Su razonamiento, como en el caso de los percentiles, es también simple, fácilmente comunicable. No obstante, y tal y como se ha hecho notar, presenta inconvenientes que no se pueden menospreciar. Pese a estas limitaciones, ha sido una estrategia muy empleada en el ámbito de la medida de la inteligencia, lo que ha justificado el espacio dedicado a ella en el presente módulo.

El baremo ha sido uno de los conceptos fundamentales del módulo. Gracias a él el usuario de un instrumento puede realizar inferencias respecto a la puntuación de un individuo, de acuerdo con el perfil de comportamiento de una muestra normativa, y empleando tablas que establecen esta correspondencia, muchas veces mediante el uso de puntuaciones transformadas. No obstante,

más allá de las consideraciones que se derivan del uso de una u otra estrategia de transformación, el elemento más relevante que hay que tener en consideración es la necesaria y adecuada justificación de cuál debe ser esta muestra normativa que estructura este baremo. Esta muestra determinará fuertemente las inferencias que se puedan derivar de ella y podrá condicionar, cuando se esté trabajando en un contexto aplicado, el sentido de las decisiones que deban ser tomadas.

El módulo concluye con algunas de las estrategias más frecuentes empleadas dirigidas a equiparar puntuaciones entre medidas diferentes que miden un mismo objeto psicológico. La equiparación equipercantil o la transformación lineal devienen estrategias que pueden ser especialmente útiles, por ejemplo, en aquellos contextos de trabajo en los que una evaluación repetida de una misma variable en un mismo individuo puede conducir a una estimación sesgada. Las ventajas y limitaciones que el usuario debe tener presentes cuando se desee emplear alguna de estas dos estrategias son homólogas a las descritas en el caso de los percentiles o de las puntuaciones estandarizadas.

Bibliografía

- Aiken, L. R. (1996). *Tests psicológicos y evaluación* (8.ª ed. México: Prentice Hall.
- Alegret, M., Espinosa, A., Vinyes-Junqué, G., Valero, S., Hernández, I., Tárraga, L. et al. (2012). Normative data of a brief neuropsychological battery for Spanish individuals older than 49. *Journal of Clinical and Experimental Neuropsychology*, 34, 209-219.
- Aluja, A., Blanch, A., Solé, D., Dolcet J.M., y Gallart, S. (2008). Validez convergente y estructural del NEO-PI-R. Baremos orientativos. *Boletín de Psicología*, 92, 7-25.
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washintong, DC: American Psychological Association.
- Gomà-i-Freixanet, M. y Valero, S. (2008). Spanish normative data of the Zuckerman-Kuhlman Personality Questionnaire in a general population sample. *Psicothema*, 20, 324-330.
- Muñiz, J. (2003). *Teoría clásica de los tests*. Madrid: Pirámide.
- Ramos-Quiroga, J. A., Draigue, C., Valero, S., Bosch, R., Gómez-Barros, N., Nogueira, M., Palomar, G., Roncero, C., y Casas, M. (2009). Validación al español de la escala de cribado del trastorno por déficit de Atención/hiperactividad en adultos (ASRS v. 1.1): una nueva estrategia de puntuación. *Revista de Neurología*, 48, 449-452.
- Rodríguez, C., Jiménez, J. E., Díaz, A., García, E., Martín, R., y Fernández, S. (2012). Datos normativos para el Test de los Cinco Dígitos: desarrollo evolutivo de la flexibilidad en Educación Primaria. *European Journal of Education and Psychology*, 5, 27-38.
- Valero, S., Ramos-Quiroga, J. A., Gomà-i-Freixanet, M., Bosch, R., Gómez-Barros, N., Nogueira, M., Palomar, G. et al. (2012). Personality profile of adult ADHD: The alternative five factor model. *Personality Research*, 198, 130-134.
- Yueh-Hsien, L., Chwen-Yng, S., Wei-Yuan, G., y Yee-Pay, W. (2012). Psychometric validation and normative data of the second Chinese version of the Hooper Visual Organization test in Children. *Research in Developmental Disabilities*, 33, 1919-1927.
- Zuckerman, M., Kuhlman, D. M., Joireman, J., Teta, P., y Kraft, M. (1993). A comparison of the three structural models for personality: the big three, the big five, and the alternative five. *Journal of Personality and Social Psychology*, 65, 757-768.

