

Datos perdidos y propiedades psicométricas en los tests de personalidad

Marcelino Cuesta^{1*}, Eduardo Fonseca-Pedrero², Guillermo Vallejo¹ y José Muñiz¹

¹ Universidad de Oviedo, España

² Universidad de la Rioja, España

Resumen: Se investigó la influencia que tienen los valores perdidos sobre la estimación de las propiedades psicométricas de los tests de personalidad. Se ha utilizado un enfoque aplicado, simulando situaciones que puedan tener relevancia para la práctica profesional. Se diseñaron dos estudios de simulación a partir de datos reales obtenidos de la aplicación de la prueba ESQUIZO-Q que evalúa esquizotipia. En el primero de ellos se utilizó una muestra de 3056 personas y en el segundo una de 200; en ambos casos se emplearon cuatro niveles de pérdida de respuestas y ocho procedimientos de imputación de los valores perdidos. Se estudió su influencia sobre las estimaciones del coeficiente α de Cronbach, la estructura factorial de la prueba y la ordenación de las puntuaciones en el cuestionario. Los resultados apuntan a que en presencia de niveles bajos de valores perdidos, incluso los métodos más simples, ofrecen soluciones muy razonables desde el punto de vista práctico. Desde la perspectiva más estadística el procedimiento de Expectación-Maximización (EM) es el que presenta un mejor comportamiento global en los diferentes criterios manejados. Destaca también el pobre comportamiento de los métodos de sustitución por el valor anterior o posterior de cara a mantener la estructura factorial de los datos.

Palabras clave: Valores perdidos; fiabilidad; personalidad; imputación.

Title: Missing data and psychometric properties of personality tests.

Abstract: The aim of this study was to analyze the influence of missing values on the psychometric properties of personality tests. An applied approach was used, trying to simulate conditions similar to those found in the professional practice. Two simulation studies were designed, based on actual data from the administration of ESQUIZO-Q test, which assesses schizotypy. In the first study a large sample was used (N=3056), and in the second one a smaller sample (N=200) was analyzed. In both cases four levels of missing values, and eight procedures for handling missing values were simulated. The influence of these conditions on the estimates of Cronbach's α , the factor structure of the test, and the arrangement of test scores were analyzed. The results suggest that in the presence of low levels of missing values, even the simplest imputation methods offer appropriate solutions from a applied point of view. From a statistical perspective the Expectation-Maximization (EM) method is the one with a better overall performance in the different criteria handled. Also noteworthy is the poor performance of replacement procedures when using the value of the previous or posterior item in order to maintain the factor structure of the data.

Key words: Missing values; reliability; personality; imputation.

Introducción

Aunque la mejor solución al problema de los datos perdidos sería no tenerlos (Allison, 2002), lo cierto es que tarde o temprano cualquier investigador que maneje datos empíricos acaba encontrándose con el problema. Tal y como indican van der Ark y Vermunt (2010), ignorar el problema puede llevar a resultados estadísticamente sesgados y conclusiones erróneas. La preocupación por el problema de los valores perdidos no es reciente aunque es a finales del siglo pasado cuando comienza su estudio serio (Graham, 2009; Little y Rubin, 1987; Rubin, 1987). A estas alturas parecen asentadas las definiciones de los diferentes patrones de pérdida de datos (Pérdidas completamente al azar –MCAR–; Pérdidas al azar –MAR–; Pérdidas no aleatorias –NMAR–) y como indica Howell (2008), el tratamiento de estos valores perdidos no resulta un área especialmente controvertido a nivel estadístico, existiendo un buen número de procedimientos alternativos sobre los que se da un acuerdo bastante amplio respecto a sus fortalezas y debilidades. Sin embargo, no parece que se haya dado una adecuada transferencia de esos conocimientos desde los entornos estadístico-metodológicos a la investigación en contextos aplicados (Baraldi y Enders, 2010; Graham, 2009; Roth, 1994; Schafer y Graham, 2002).

Siguiendo a Acock (2005) y Baraldi y Enders (2010), los diferentes métodos de manejo de los valores perdidos pueden agruparse en métodos tradicionales y métodos modernos. Los métodos tradicionales a su vez se pueden diferen-

ciar en: a) Métodos de borrado, en este apartado se encuadran algunos procedimientos muy populares como *Listwise* (o análisis de casos completos) o *Pairwise* (o análisis de casos disponibles); b) Métodos de imputación simple, tales como emplear algún tipo de media (de la escala, del ítem, del sujeto...), regresión determinista o estocástica, o los procedimientos *Hot Deck*. En los métodos modernos se situarían los procedimientos de Máxima Verosimilitud y los de imputación múltiple.

Si bien son los métodos modernos los recomendados por los estadísticos debido a sus cualidades respecto a las propiedades de las estimaciones obtenidas, son los procedimientos tradicionales los más empleados por los investigadores aplicados. En el campo de las ciencias de la conducta y de la salud es muy habitual que la recogida de datos se realice mediante algún test o cuestionario formado por múltiples ítems que conforman una escala sumativa que trata de medir uno o más constructos. Como queda bien establecido en la literatura psicométrica (AERA, APA y NCME, 1999; Brennan, 2006) estos instrumentos deben cumplir ciertas características que garanticen la calidad de los datos obtenidos. Es aquí donde radica la importancia de los instrumentos de medida, pues están en la base de toda la cadena de análisis estadísticos, si medimos con error no podemos diagnosticar con precisión, y por lo tanto no se puede intervenir con eficacia. Como bien indica Lewis-Beck (1995) si entra basura, sale basura. Por esta razón resulta de nuestro interés conocer como la presencia de valores perdidos afecta a las características psicométricas de las pruebas, en tanto que garantes de la calidad de nuestras medidas. Roth, Switzer y Switzer (1999) señalan que hasta ese momento no existía apenas literatura respecto al problema de los valores perdidos a nivel de

* Dirección para correspondencia [Correspondence address]:
Marcelino Cuesta. Facultad de Psicología. Universidad de Oviedo. Plaza
Feijoo s/n. 33003 Oviedo (España). E-mail: mcuesta@uniovi.es

ítem, desde entonces han ido surgiendo trabajos en este ámbito si bien no siempre ocupándose directamente de aspectos psicométricos (Carpita y Manisera, 2011; Enders, 2004; Fernández-Alonso, Suárez-Alvarez y Muñiz, 2012; Gmel, 2001; McDonald, Thurston y Nelson, 2000; Shrive, Stuart, Quan y Ghali, 2006; Sijtsma y Van de Ark, 2003; Van Ginkel, van der Ark y Sijtsma, 2007).

En este contexto el objetivo de nuestro trabajo es abordar el problema de los datos perdidos y su influencia sobre la estimación de las propiedades psicométricas de los tests de personalidad. Se trata, en suma, de saber cómo afectan los datos perdidos, y las distintas formas de recuperarlos, a las estimaciones de las propiedades psicométricas de las pruebas. La pérdida de datos cuando se utilizan tests de personalidad es extremadamente frecuente, siendo el procedimiento *Listwise* de manejo de los valores perdidos mayoritariamente utilizado como muestran Van Ginkel, Sijtsma, van der Ark y Vermunt (2010). Este método tiene muchos inconvenientes y se trata de explorar otras opciones más adecuadas.

Para ello se han diseñado dos estudios de simulación a partir de datos reales obtenidos de la aplicación de una prueba de personalidad. En el primero de ellos se asume una situación en la que se dispone de una muestra amplia, mientras que en el segundo se trata de simular la situación de un usuario que se ve obligado a trabajar con muestras modestas. En ambos casos la atención se ha centrado en un estimador clásico de la fiabilidad de la prueba como es el coeficiente α de Cronbach, la capacidad para mantener estable la posición relativa de los sujetos y, en el primer estudio, en la estructura factorial de la prueba. Respecto a la selección de los procedimientos de manejo de los valores perdidos el criterio ha sido la sencillez, bien conceptual, bien por hallarse implementado en un paquete estadístico de amplio uso. Todo ello con la vista puesta en recrear una situación típica de trabajo de un usuario de datos generados por pruebas de personalidad en un contexto aplicado.

Estudio 1

Método

Participantes

Los individuos empleados fueron los utilizados en la construcción del Cuestionario Oviedo para la Evaluación de la Esquizotipia (ESQUIZO-Q) (Fonseca-Pedrero, Muñiz, Lemos, Paino y Villazón, 2010) obtenidos mediante un muestreo aleatorio estratificado, por conglomerados, a nivel de aula, en la Comunidad Autónoma del Principado de Asturias a lo largo de dos cursos académicos. Los alumnos pertenecían a diferentes centros escolares, públicos, concertados y privados, de Educación Secundaria Obligatoria (ESO) y Ciclos de Formación Profesional. Los estratos se crearon en función de la zona geográfica del Principado de Asturias (Oriente, Occidente y Central) y la etapa escolar (obligatoria y post-obligatoria), en donde la probabilidad de extracción

del centro escolar venía dada en función del número de alumnos. De esta manera se obtuvo una muestra final de 3056 sujetos de los cuales el 48.1% eran varones. Las edades variaron entre 14 y 18 años con una media de 15.9 años y una desviación típica de 1.17.

Instrumento

El ESQUIZO-Q (Fonseca-Pedrero et al. 2010) es un autoinforme creado para evaluar los rasgos esquizotípicos de la personalidad en población adolescente (Fonseca-Pedrero, Paino, Lemos y Muñiz, 2011), el cual está fundamentado en los criterios diagnósticos propuestos para este trastorno en el DSM-IV-TR (American Psychiatric Association, 2000) y en el modelo de esquizotaxia de Meehl (1962) sobre predisposición genética a la esquizofrenia. Los ítems del ESQUIZO-Q fueron seleccionados a partir de una revisión exhaustiva de la literatura previa (Fonseca-Pedrero et al., 2008), considerando un total de 10 facetas, a saber: Ideación Referencial, Pensamiento Mágico, Experiencias Perceptivas Extrañas, Pensamiento y Lenguaje Extraños, Ideación Paranoide, Anhedonia Física, Anhedonia Social, Comportamiento Raro, Falta de Amigos Íntimos y Ansiedad Social Excesiva. El formato de respuesta utilizado fue tipo Likert de 5 categorías (desde 1 “*Completamente en desacuerdo*” a 5 “*Completamente de acuerdo*”).

Diseño

Se manipularon dos variables a fin de comprobar su efecto sobre las estimaciones de las propiedades psicométricas del cuestionario: los métodos de tratamiento de los valores perdidos y el porcentaje de respuestas perdidas en la base de datos.

Los procedimientos de manejo de los valores perdidos empleados fueron:

Listwise: eliminar de los análisis a los sujetos con algún valor perdido en cualquiera de las variables a analizar.

Imputar a todos los valores perdidos la *media global* de todos los ítems que componen el test.

Imputar a los valores perdidos la *media del sujeto* en el resto de los ítems de la escala con valores válidos.

Imputar los valores perdidos mediante la *media del ítem* en los sujetos con respuesta válida.

Imputar mediante el *valor del ítem posterior* al que presenta el valor perdido.

Imputar mediante el *valor del ítem anterior* al que presenta el valor perdido.

Imputar por el procedimiento *EM* (expectation-maximization), tal y como es implementado en el programa SAS – proc MI. Algoritmo que permite obtener estimaciones de Máxima Verosimilitud mediante un procedimiento en dos pasos. En el primero (E) se imputan los valores generalmente empleando ecuaciones de regresión y en el segundo paso (M) se calculan de nuevo los valores para las medias y la matriz de covarianzas empleando los valores imputados y los no perdidos. Una vez que se tienen las nuevas estimaciones de

medias y covarianzas se empieza de nuevo con el paso E y el proceso continua hasta que las estimaciones convergen.

Imputar mediante *regresión lineal múltiple* (tal y como se implementa en el programa SPSS versión 18). La puntuación del ítem se imputa mediante un modelo de regresión múltiple empleando las puntuaciones de los sujetos con todas las respuestas, con el ítem con valor perdido como variable dependiente y el resto de ítems como variables independientes. A la puntuación pronosticada con el modelo se le añade un error aleatorio extraído de una distribución normal (con media 0 y desviación típica igual a la raíz cuadrada del término error cuadrático medio de la regresión), obteniéndose así la puntuación a imputar.

En todos los casos excepto en el procedimiento de regresión la implementación del procedimiento de manejo de valores perdidos y el cálculo del coeficiente alfa se realizó empleando el programa SAS.

En cuanto al porcentaje de valores perdidos se manejaron cuatro valores: 5%, 10%, 20% y 30%.

Procedimiento de simulación

El punto de partida en la generación de datos para este estudio fue la matriz completa de respuestas de los 3056 sujetos a los 51 ítems del cuestionario ESQUIZO-Q. A partir

de esta matriz se generaron patrones de pérdidas de datos completamente aleatorios (MCAR) para cada ítem de forma que se obtuviera el porcentaje de valores perdidos previamente establecido (5%, 10%, 20% y 30%), con la excepción de los ítems 1 y 51 que se mantuvieron sin valores perdidos para asegurar la aplicación de los procedimientos de imputación que implican la sustitución por el valor del ítem anterior o posterior. Para cada uno de los procedimientos de manejo de los valores perdidos se generaron 10 bases, lo que resulta en un total de 320 bases (4 porcentajes de valores perdidos x 8 métodos de manejo de los valores perdidos x 10) sometidas a análisis.

Para conseguir el patrón de pérdida deseado para cada ítem se generó, excepto para el 1 y el 51, una variable aleatoria uniformemente distribuida para cada uno de ellos. Cuando la variable tomaba valores inferiores a la proporción de valores perdidos que se deseaba lograr (.05; .1; .2; .3) se sustituía la respuesta dada al ítem por un valor perdido, sin establecer límite para el número de ítems con valor perdido que podía presentar un sujeto. Este proceso se llevó a cabo mediante la función RANUNI del programa SAS permitiendo que la semilla variara aleatoriamente para cada variable uniforme generada.

Tabla 1. Estadísticos Descriptivos de las bases generadas para el estudio 1.

% de valores perdidos pretendido	% de valores perdidos en el total de los ítems		% de casos afectado		Número de ítems perdidos en los casos afectados	
5%	mín = 3.7 media = 4.98	máx = 6.5	mín = 91.4 media = 91.88	máx = 92.6	mín = 1 media = 2.66	máx = 11
10%	mín = 8.5 media = 10	máx = 11.8	mín = 99.1 media = 99.39	máx = 99.5	mín = 1 media = 4.93	máx = 14
20%	mín = 17.5 media = 20.02	máx = 22.4	mín = 100 media = 100	máx = 100	mín = 1 media = 9.81	máx = 22
30%	mín = 27.7 media = 30	máx = 32.5	mín = 100 media = 100	máx = 100	mín = 3 media = 14.7	máx = 27

Análisis de datos

Sobre la matriz original se calculó el coeficiente α de Cronbach que fue tomado como referencia para los valores obtenidos en las matrices generadas.

Para cada matriz se calculó el valor de α , y a partir de sus valores dos indicadores de las diferencias entre el valor de la matriz original y las estimaciones en las bases imputadas.

La raíz del error cuadrático medio (*Root mean square error – RMSE*) que es el promedio de la diferencia entre (la fiabilidad estimada en las bases imputadas) y α (la fiabilidad de la matriz completa original) y que se emplea como un indicador de la variabilidad de las estimaciones

$$RMSE = \sqrt{\frac{\sum(\hat{\alpha} - \alpha)^2}{10}}$$

También se calculó el sesgo promedio siguiendo la expresión

$$SM = \frac{\sum \hat{\alpha} / \alpha}{10}$$

que tomará valor inferior a 1 cuando se produzca una infraestimación y superior a uno en los casos de sobreestimación.

Además para cada matriz se calculó la correlación entre las puntuaciones de los sujetos obtenidas empleando la base original y las obtenidas empleando las bases imputadas (excepto para el procedimiento *listwise* donde la correlación siempre es 1), a partir de las cuales se calculó la correlación media para cada procedimiento de manejo de los valores perdidos.

Finalmente, para cada matriz de datos imputada se comprobó si se reproducía la estructura factorial postulada por los autores del ESQUIZO-Q, con 11 factores de primer orden. Para ello cada matriz fue sometida a un análisis de componentes principales, forzando la extracción de 11 compo-

nerentes, con rotación *oblimin*. Sobre estas soluciones factoriales se determinaron tres posibles resultados: No se reproduce la estructura factorial pues los ítems se agrupan de forma distinta, Sí se reproduce la estructura en cuanto al número y composición de los factores pero hay cambios en el orden de los mismos y Sí se reproduce la estructura en número, composición y orden de los factores.

Resultados

Como puede observarse en la Tabla 2, en lo que se refiere a la precisión de la estimación del valor de α se observa un deterioro (mayor dispersión) a medida que aumenta la proporción de valores perdidos en la base de datos. Moviéndonos siempre dentro de unos valores muy pequeños del estadístico *RMSE*, con un 5% de valores perdidos todos los métodos se comportan de forma razonablemente adecuada con la excepción del procedimiento *listwise*, siendo a este nivel de datos perdidos el método de la regresión, EM, valor posterior y valor anterior los que manifiestan un mejor comportamiento. Para los niveles del 10%, 20% y 30% se repite el patrón expuesto pero ampliándose las diferencias entre los procedimientos con un mejor y peor desempeño. Señalar que en estos últimos niveles no es posible calcular α si se aplica el método *listwise* pues todos los sujetos presentan al menos un valor perdido.

Respecto al sesgo los procedimientos de la media global y media del ítem tienden a la infraestimación, más cuanto

mayor sea el porcentaje de valores perdidos. El procedimiento de la media del sujeto presenta el patrón contrario, tiende a sobreestimar más cuantos más valores perdidos. Los métodos EM, valor posterior, valor anterior, regresión y *listwise* (en los niveles 5% y 10%) presentan tendencias despreciables a la sobre o infraestimación, manteniéndose estables con independencia del porcentaje de valores perdidos.

Otra forma de abordar el comportamiento de los diferentes métodos es planteándose lo siguiente. Si nuestro interés como usuarios del test no se centra en una estimación exacta del nivel de los sujetos en el constructo psicológico de interés (en este caso la esquizotipia) sino en una adecuada ordenación de los sujetos (posición relativa) respecto al constructo, ¿en qué medida los diferentes procedimientos sometidos a prueba afectan a ese ordenamiento? Para dar respuesta a esta pregunta se ha calculado la correlación entre las puntuaciones de los sujetos en la base de datos completa y las puntuaciones tras aplicar algún procedimiento de imputación.

A este respecto la Tabla 2 nos permite comprobar que el valor de la correlación es siempre muy alto (ningún valor por debajo de 0.94), que se produce un ligero descenso en el valor a medida que aumenta la proporción de valores perdidos y que se mantiene una ordenación constante de los diferentes procedimientos (de mayor a menor): media del ítem, media del sujeto, media global, regresión, EM, valor posterior y valor anterior.

Tabla 2. Valores medios de RMSE, sesgo y correlaciones para los diferentes procedimientos de manejo de los valores perdidos y porcentajes de valores perdidos. (Estudio 1).

% de valores perdidos		Procedimiento de manejo de los valores perdidos							
		Media Global	Media del sujeto	Media del ítem	EM	Valor posterior	Valor anterior	Listwise	Regresión
α media		0.879	0.891	0.880	0.886	0.885	0.885	0.885	0.886
5%	RMSE	0.007	0.005	0.006	0.0005	0.0006	0.0005	0.013	0.0004
	Sesgo medio	0.992	1.006	0.993	1.000	1.000	1.000	0.999	1.000
	r media	0.996	0.996	0.997	0.995	0.993	0.993		0.996
α media		0.871	0.896	0.874	0.886	0.886	0.885	0.883	0.885
10%	RMSE	0.014	0.010	0.012	0.0006	0.0003	0.001	0.040	0.001
	Sesgo medio	0.984	1.012	0.986	1.000	1.000	0.999	0.997	1.000
	r media	0.992	0.993	0.993	0.990	0.986	0.986		0.990
α media		0.853	0.907	0.858	0.885	0.886	0.886		0.885
20%	RMSE	0.033	0.021	0.028	0.001	0.002	0.001		0.002
	Sesgo medio	0.963	1.024	0.968	0.999	1.001	1.000		0.999
	r media	0.981	0.983	0.985	0.978	0.969	0.969		0.980
α media		0.830	0.917	0.839	0.885	0.889	0.888		0.883
30%	RMSE	0.056	0.032	0.047	0.001	0.004	0.002		0.003
	Sesgo medio	0.937	1.036	0.947	0.999	1.004	1.002		0.996
	r media	0.969	0.971	0.974	0.963	0.948	0.947		0.967

También resulta de interés comprobar el efecto de los valores perdidos respecto a la validez de la prueba, en este caso respecto a las evidencias de estructura interna. Revisando el comportamiento de los diferentes métodos aplicados (Tabla 3) queda claro que tanto el método *listwise* como los del valor posterior y anterior resultan completamente inadecuados dado que, independientemente del porcentaje de valores perdidos que se esté manejando, son incapaces de man-

tener la estructura factorial de la prueba. En el resto de los procedimientos se da un comportamiento adecuado con un nivel del 5% de valores perdidos, que se va haciendo menos exitoso a medida que aumenta la proporción de valores perdidos. Destaca el buen funcionamiento, respecto a este criterio, del procedimiento de la regresión incluso con un 20% de valores perdidos. Cuando la cantidad de valores perdidos es del 30 % ningún método presenta un desempeño adecuado.

Tabla 3. Comportamiento de la estructura factorial para los diferentes procedimientos de manejo de los valores perdidos y porcentajes de valores perdidos en las bases (Estudio 1).

% de valores perdidos	Procedimiento de manejo de los valores perdidos	Procedimiento de manejo de los valores perdidos							
		Media Global	Media del sujeto	Media del ítem	EM	Valor posterior	Valor anterior	Listwise	Regresión
5%	No					90%	80%	100%	
	Cambia orden	80%	100%	60%	50%	10%	10%		60%
	Sí	20%		40%	50%		10%		40%
10%	No		10%			100%	100%	*	10%
	Cambia orden	70%	90%	40%	50%				20%
	Sí	30%		60%	50%				70%
20%	No	30%	30%	40%	40%	100%	100%	*	10%
	Cambia orden	30%	70%	40%	30%				50%
	Sí	40%		20%	30%				40%
30%	No	60%	80%	60%	70%	100%	100%	*	70%
	Cambia orden	40%	20%	40%	10%				30%
	Sí				20%				

* No hay datos suficientes para realizar el análisis

Estudio 2

Método

Participantes, instrumento y diseño

Se han seguido las mismas pautas que las indicadas para el primer estudio.

Procedimiento de simulación

En este segundo estudio se partió de la misma base completa obtenida de la aplicación del ESQUIZO-Q, toma-

da ahora como población. A partir de esta “población de referencia” se extrajeron muestras aleatorias sin reposición de tamaño 200 mediante un procedimiento creado a tal fin en SAS. Aplicando el mismo procedimiento descrito para el primer estudio se obtuvieron bloques de 10 bases completas de tamaño 200, sobre las cuales se generaron patrones de pérdida completamente aleatorios (MCAR) y se les aplicó el correspondiente procedimiento de manejo de valores perdidos. Se trabajó, por tanto, con 360 bases (40 bases completas + 320 con valores perdidos).

Tabla 4. Descriptivos de las bases generadas para el Estudio 2.

% de valores perdidos pretendido	% de valores perdidos en el total de los ítems	% de casos afectado	número de ítems perdidos en los casos afectados
5%	mín = 1.5 media = 5.09 máx = 10	mín = 88.5 máx = 96.5 media = 92.45	mín = 1 máx = 11 media = 2.70
10%	mín = 5 media = 9.90 máx = 16.5	mín = 98.5 máx = 100 media = 99.40	mín = 1 máx = 12 media = 4.88
20%	mín = 12 media = 19.84 máx = 28.5	mín = 100 máx = 100 media = 100	mín = 2 máx = 20 media = 9.73
30%	mín = 21 media = 29.91 máx = 40.5	mín = 100 máx = 100 media = 100	mín = 4 máx = 28 media = 14.66

Análisis de datos

Además del valor del α de Cronbach sobre la base de datos global, que para este estudio se tomará como valor paramétrico, se calculó dicho estadístico para cada una de las muestras completas de tamaño 200 y para cada una de las bases en las que se aplicó algún procedimiento de manejo de los valores perdidos. A partir de estos valores se calculó, al igual que en el primer estudio, los valores del RMSE y el sesgo medio.

Por otro lado, se halló la diferencia entre el valor del estadístico de Cronbach calculado sobre una base completa (sin valores perdidos) de tamaño 200 (α_c) y el valor calculado en una base en la que se ha aplicado algún procedimiento de manejo de los valores perdidos generada a partir de la dicha base completa (α_p), lo cual hemos definido como *discre-*

pancia ($\alpha_c - \alpha_p$). (Van Ginkel, van der Ark y Sijtsma, 2007).

También fueron calculadas las correlaciones entre las puntuaciones de los sujetos obtenidas en las bases completas y en las bases sobre las que se aplicó algún procedimiento de imputación.

Resultados

La Tabla 5 recoge los valores de los estadísticos calculados en el estudio segundo. Como puede observarse en comparación con los resultados del primer estudio los valores de RMSE son más altos, como era esperable al reducir el tamaño de la muestra. Al igual que en el caso anterior la variabilidad va aumentando a medida que se incrementa el porcentaje de valores perdidos en los datos, si bien cuando la imputa-

ción se hace mediante EM, valor posterior o valor inferior mantienen niveles similares con independencia de la cantidad de valores perdidos. Si tomamos como referencia el valor del estadístico alcanzado en las bases completas podemos ver que con un 5% de valores perdidos los diferentes procedimientos dan valores correctos, excepto *listwise* y regresión que presentan un comportamiento deficiente desde este primer nivel. A medida que incrementamos el porcentaje de valores perdidos se muestra una mayor diferencia en el comportamiento de la estimación realizada en las bases completas y en las bases imputadas, siendo menor esta diferencia en los procedimientos anteriormente mencionados.

En lo referente al sesgo al igual que en el tamaño de muestra grande los procedimientos de media global, media del ítem y regresión tienden a la infraestimación, leve en los niveles de 5% y 10% y más acusada en los niveles del 20% y 30%, especialmente el método de la regresión con una infraestimación muy acusada con el 30% de valores perdidos. El procedimiento EM apenas muestra sesgo en los dos primeros niveles y tiende ligeramente a la infraestimación en los dos niveles más altos de presencia de valores perdidos. La media del sujeto presenta una leve tendencia a la sobrestima-

ción, en tanto que el valor posterior y el valor anterior no presentan prácticamente sesgo. En relación a lo hallado en el primer estudio no hay diferencias reseñables en cuanto a los valores de sesgo encontrados a excepción del ya comentado mal comportamiento del procedimiento de regresión en el nivel más alto de presencia de valores perdidos. En la comparación con los valores calculados sobre las bases completas se observa un patrón similar al comentado para el *RMSE*.

Respecto a la discrepancia entre las estimaciones de α realizadas en las bases completas y en las imputadas toma en todos los casos valores poco significativos a nivel práctico, destacando los procedimientos del valor posterior y valor anterior por su buen funcionamiento aun con niveles altos de valores perdidos y, en el otro extremo, el muy mal resultado ofrecido por el método de regresión en las bases con un 30% de valores perdidos.

Por último, las correlaciones entre las puntuaciones de las bases completas y de las bases imputadas muestran el mismo patrón que en el primer estudio con valores muy altos y con los métodos basados en las medias alcanzando los valores más altos con independencia del porcentaje de valores perdidos.

Tabla 5. Valores medios de *RMSE*, sesgo, discrepancias y correlaciones para los diferentes procedimientos de manejo de los valores perdidos y porcentajes de valores perdidos (Estudio 2).

$\alpha = 0,886$		Procedimiento de manejo de los valores perdidos								
% de valores perdidos		Completa	Media Global	Media del sujeto	Media del ítem	EM	Valor posterior	Valor anterior	Listwise	Regresión
5%	α media	0.880	0.872	0.884	0.873	0.877	0.879	0.878	0.884	0.869
	<i>RMSE</i>	0.013	0.018	0.010	0.017	0.015	0.013	0.013	0.035	0.034
	Sesgo medio	0.993	0.984	0.999	0.986	0.991	0.993	0.992	0.998	0.981
	Discrepancia media		0.008	0.005	0.007	0.003	0.002	0.002	0.035	0.012
	r media		0.996	0.996	0.996	0.994	0.993	0.993		0.995
10%	α media	0.889	0.875	0.899	0.878	0.885	0.887	0.888	0.912*	0.885
	<i>RMSE</i>	0.011	0.016	0.016	0.014	0.011	0.011	0.012	0.033*	0.012
	Sesgo medio	1.004	0.988	1.015	0.991	1.000	1.002	1.002	1.030*	0.999
	Discrepancia media		0.014	0.010	0.012	0.004	0.003	0.003	0.055*	0.005
	r media		0.992	0.993	0.993	0.989	0.987	0.986		0.991
20%	α media	0.883	0.850	0.904	0.855	0.867	0.885	0.883		0.855
	<i>RMSE</i>	0.012	0.039	0.020	0.034	0.024	0.010	0.011		0.034
	Sesgo medio	0.997	0.959	1.021	0.965	0.979	0.999	0.996		0.965
	Discrepancia media		0.033	0.021	0.028	0.016	0.003	0.003		0.028
	r media		0.982	0.983	0.985	0.968	0.969	0.970		0.969
30%	α media	0.888	0.833	0.920	0.840	0.872	0.893	0.890		0.741
	<i>RMSE</i>	0.011	0.0548	0.035	0.048	0.018	0.015	0.013		0.150
	Sesgo medio	1.002	0.940	1.039	0.949	0.985	1.007	1.005		0.836
	Discrepancia media		0.055	0.032	0.048	0.016	0.005	0.004		0.147
	r media		0.969	0.973	0.974	0.954	0.947	0.950		0.879

* sólo dos bases válidas

Discusión y conclusiones

A la hora de establecer conclusiones a partir de los resultados hallados en nuestros dos estudios quizás sea conveniente diferenciarlas en virtud de los objetivos con los que el investigador se acerca a la prueba psicométrica: a) como usuario de la prueba interesado en emplearla para medir una variable relevante en el contexto de un área sustanti-

va; b) como investigador psicométrico interesado en la calidad de la estimación de las propiedades de la prueba.

A nivel general, y situándonos desde la perspectiva de un usuario de pruebas psicométricas de personalidad cuyo interés se centra en poder calcular de forma razonable la fiabilidad que presenta una determinada prueba en su muestra, podríamos decir que en tanto en cuanto el nivel de valores perdidos se mantengan bajos (5%-10%) y las

pérdidas sean aleatorias los resultados que obtenga no distorsionarán de forma relevante su visión sobre la adecuación de la prueba, sea cual sea el método elegido para el manejo de los valores perdidos. Si bien no debe perder de vista que ha quedado bien establecido en la literatura que el procedimiento *listwise* aunque bajo determinadas circunstancias puede dar lugar a estimaciones insesgadas siempre produce una disminución del tamaño de muestra, con su consiguiente influencia en la potencia de prueba de los análisis en los que interviene la variable afectada por los valores perdidos (Botella, 2002; Howell, 2008; Enders, 2010).

Siguiendo con este enfoque pragmático también se observa que la ordenación relativa que de los sujetos se pueda realizar con el test no se verá afectada por el método de tratamiento de los valores perdidos, lo que se refleja en las altas correlaciones obtenidas entre las puntuaciones calculadas en las bases completas y en las imputadas. Estos resultados están en la línea de los hallados en otros trabajos con cierta similitud al que aquí presentamos (Downey y King, 1998; Gmel, 2001; Shrive, Stuart, Quan y Ghali, 2006).

Deteniéndonos en la comparación entre métodos, y acercándonos más a la perspectiva del investigador psicométrico, en lo que respecta a la variabilidad de las estimaciones obtenidas tras las imputaciones son los procedimientos EM, valor posterior y valor inferior los que ofrecen un mejor comportamiento. De nuevo, nuestros resultados numéricos son coherentes con los presentados en trabajos anteriores tales como McDonald, Thurston y Nelson (2000) (emplean los métodos *listwise*, media del sujeto, regresión y media del ítem) y Enders (2004) (emplea *listwise*, EM y media del ítem) aunque en estos casos el número de ítems era mucho menor (entre 3 y 7) y el porcentaje de valores perdidos se restringe al 20% y 40%.

Respecto al sesgo el procedimiento que sale mejor parado es EM, confirmando lo aportado en el trabajo de Enders (2004) y la literatura genérica al respecto (véase por ejemplo Allison, 2002). Se confirma también en nuestro trabajo la tendencia apuntada en la literatura de que la media del ítem tiende a la infraestimación y la media del sujeto a la sobreestimación (Downey y King, 1998; Enders, 2004; Huisman, 2000; McDonald, Thurston y Nelson, 2000).

Para finalizar no podemos dejar de comentar el mal comportamiento que los procedimientos del valor posterior y valor anterior tienen a la hora de reproducir la estructura factorial del test, lo que contrasta con su buen comportamiento en los otros criterios considerados. La explicación a estos resultados quizás debamos buscarla en la propia estructura física del test, en la que los ítems correspondientes a una dimensión no van necesariamente

colocados de manera consecutiva en la prueba. Hemos de tener en cuenta que los datos con los que hemos trabajado proceden de un estudio empírico y que la estructura factorial de la prueba viene dada por la propia estructura del constructo evaluado y no por artefactos estadísticos como pudiera ocurrir en simulaciones puras. De esta forma al sustituir por el valor anterior o posterior estamos empleando una puntuación que puede estar generada por una dimensión conceptualmente distinta a la que trata de medir el ítem imputado, y parece que tiene cierta lógica que la estructura factorial de la base imputada se resienta.

En resumen, la buena noticia es que desde el punto de vista del usuario de pruebas psicométricas con similares características a la aquí empleada, cualquiera de los métodos de manejo de valores perdidos presentados, a pesar de su simplicidad, le permitirán tener una idea cabal de la calidad del instrumento siempre y cuando no se dispare el porcentaje de valores perdidos. No debe olvidar el usuario, sin embargo, las limitaciones propias del popular procedimiento *listwise* al emplear la variable medida en el contexto de análisis más amplios. Si el interés está centrado, más que en los asuntos “prácticos”, en las virtudes estadísticas de las estimaciones de las propiedades psicométricas de las pruebas parece que el método que, en general, ofrece más garantías es EM. Siendo, desde este segundo punto de vista, claramente descartables los procedimientos de sustituir por el valor anterior o posterior dada su negativa influencia sobre la estructura factorial de la prueba.

A pesar de que los resultados encontrados en este trabajo parecen coherentes con la literatura, no debe perderse de vista que han sido alcanzados en unas condiciones muy concretas. Parece necesario que se siga indagando en el comportamiento de diferentes procedimientos de manejo de datos perdidos en estas y otras condiciones (por ejemplo, quizás los patrones de pérdida MAR o NMAR sean más realistas que los completamente aleatorios utilizados aquí, véase por ejemplo Fernández-Alonso, Suárez-Alvarez y Muñoz, 2012 ó Vallejo, Fernández, Livacic-Rojas y Tuero-Herrero, 2011) a fin de poder seguir dando pautas a los usuarios de pruebas psicométricas en diferentes campos. Finalmente señalar que aquí se han estudiado únicamente tres propiedades psicométricas de la prueba en relación con los datos perdidos, a sabiendas de que otras muchas son posibles (Botella y Ponte, 2011; Guillén-Riquelme y Buela-Casal, 2011), por lo que sería aconsejable extender la investigación a otras características psicométricas que podrían venir afectadas.

Nota. Este trabajo ha sido financiado por los proyectos de investigación PSI2011-28638 y PSI2011-23095 del Ministerio de Ciencia e Innovación de España.

Referencias

- Acock, A.C. (2005). Working with missing values. *Journal of Marriage and Family*, 67, 1012-1028.
- Allison, P.D. (2002). *Missing Data*. Thousand Oaks, CA: SAGE Publications.
- American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed. revised) Washington, DC: American Psychiatric Association.
- American Educational Research Association, American Psychological Association y National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Botella, J. (2002). Potencia de pruebas alternativas para dos muestras relacionadas con datos perdidos. *Psicothema*, 14(1), 174-180.
- Botella, J. y Ponte, G. (2011). Effects of the heterogeneity of the variances on reliability generalization: An example with the Beck Depression Inventory. *Psicothema*, 23, 516-522.
- Brennan, R.L. (2006). *Educational measurement* (4ª edición). Westport, C.T.: American Council on Education/Praeger Publishers.
- Carpita, M. y Manisera, M. (2011). On the imputation of missing data in surveys with Likert-type scales. *Journal of Classification*, 28(1), 93-112.
- Downey, R.G. y King, C.V. (1998). Missing data in Likert ratings: A comparison of replacement methods. *The Journal of General Psychology*, 125(2), 175-191.
- Enders, C. (2004). The impact of missing data on sample reliability estimates: Implications for reliability reporting practices. *Educational and Psychological Measurement*, 64(3), 419-436.
- Enders, C. (2010). *Applied missing data analysis*. New York: The Guilford Press.
- Fernández-Alonso, R., Suárez-Alvarez, J. y Muñoz, J. (2012). Imputación de datos perdidos en las evaluaciones diagnósticas educativas. *Psicothema*, 24(1), 167-175.
- Fonseca-Pedrero, E., Paino, M., Lemos-Giráldez, S., García-Cueto, E., Campillo-Álvarez, A., Villazón-García, U. y Muñoz, J. (2008). Schizotypy assessment: State of the art and future prospects. *International Journal of Clinical and Health Psychology*, 8, 577-593.
- Fonseca-Pedrero, E., Muñoz, J., Lemos, S., Paino, M. y Villazón, U. (2010). *Esquizo-Q. Cuestionario Oviedo para la evaluación de la esquizotipia*. Madrid: TEA Ediciones.
- Fonseca-Pedrero, E., Paino, M., Lemos, S. y Muñoz, J. (2011). Nuevos desarrollos en la detección de jóvenes con riesgo de trastornos del espectro esquizofrénico. *Anales de Psicología*, 27(2), 333-341.
- Gmel, G. (2001). Imputation of missing values in the case of a multiple item instrument measuring alcohol consumption. *Statistics in Medicine*, 20, 2369-2381.
- Graham, J.W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, 60, 549-576.
- Guillén-Riquelme, A. y Buela-Casal, G. (2011). Actualización psicométrica y funcionamiento diferencial de los ítems en el State Trait Anxiety Inventory (STAI). *Psicothema*, 23, 510-515.
- Howell, D.G. (2008). The analysis of missing data. En Outhwaite, W. y Turner, S. (Eds), *Handbook of Social Science Methodology*. London: Sage.
- Huisman, M. (2000). Imputation of missing item responses: Some simple techniques. *Quality & Quantity*, 34, 331-351.
- Lewis-Beck, M.F. (1995). *Data analysis: An introduction*. London: Sage.
- Little, R.J.A. y Rubin, D.B. (1987). *Statistical analysis with missing data*. New York: Wiley.
- Mcdonald, R.A., Thurston, P.W. y Nelson, M.R. (2000). A Monte Carlo study of missing item methods. *Organizational Research Methods*, 3(1), 71-92.
- Meehl, P. E. (1962). Schizotaxia, schizotypy, schizophrenia. *American Psychologist*, 17, 827-838.
- Roth, P.L. (1994). Missing data: A conceptual review for applied psychologist. *Personnel Psychology*, 47, 537-560.
- Roth, P.L., Switzer, F.S. y Switzer, D.M. (1999). Missing data in multiple item scales: A Monte Carlo analysis of missing data techniques. *Organizational Research Methods*, 2(3), 211-232.
- Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.
- Schafer, J.J. y Graham, J.W. (2002). Missing data: Our view of the state of art. *Psychological Methods*, 7(2), 147-177.
- Shrive, F., Stuart, H., Quan, H. y Ghali, W.A. (2006). Dealing with missing data in a multi-question depression scale: a comparison of imputation methods. *BMC Medical Research Methodology*, 6:57 (<http://www.biomedcentral.com/1471-2288-6-57>) (descargado el 25/01/2011).
- Sijtsma, K. y van de Ark, L.A. (2003). Investigation and treatment of missing item scores in test and questionnaire data. *Multivariate Behavioral Research*, 38(4), 505-528.
- Vallejo, G., Fernández, P., Livacic-Rojas, P. y Tuero-Herrero, E. (2011). Comparison of modern methods for analyzing unbalanced repeated measures data. *Multivariate Behavioral Research*, 46, 900-937.
- Van der Ark, L.A. y Vermunt, J.K. (2010). New developments in missing data analysis (editorial). *Methodology*, 6(1), 1-2.
- Van Ginkel, J.R., Sijtsma, K., van der Ark, L.A. y Vermunt, J.K. (2010). Incidence of missing item scores in personality measurement, and simple item-score imputation. *Methodology*, 6(1), 17-30.
- Van Ginkel, J.R., van der Ark, L.A. y Sijtsma, K. (2007). Multiple imputation of item scores in test and questionnaire data, and influence on psychometric results. *Multivariate Behavioral Research*, 42(2), 387-414.

(Artículo recibido: 13-10-2011, revisado: 18-01-2012, aceptado: 19-02-2012)