



REVISIONES

Análisis estadístico de escalas ordinales. Aplicaciones en el Área de Salud infantil y Pediatría

Análise estatística de escalas ordinais. Aplicações na Área da Saúde Infantil e Pediatria
Statistical analysis of ordinal scales. Field applications of Pediatrics and Child Health

***Santos Curado, M^a Alice **Vitorino Teles, Júlia Maria ***Marôco, João**

*Profesora Coordinadora. Departamento de Salud del Niño y del Joven. Escuela Superior de Enfermería de Lisboa E-mail: malicecurado@gmail.com **Profesora Auxiliar. Departamento de Métodos Matemáticos. Facultad de Motricidade Humana, CIPER. Professor Auxiliar. Departamento de Estatística. Instituto Superior de Psicologia Aplicada. Portugal.

Palabras clave: escalas; muestra; análisis estadístico; salud infantil; pediatría.

Palavras-chave: escalas; amostra; análise estatística; saúde infantil; pediatria.

Keywords: scales; sample; statistical analysis; child health; pediatrics

RESUMEN

La investigación en diferentes áreas del conocimiento implica un gran número de variables que pueden ser evaluadas con técnicas univariantes y multivariantes y test paramétricos o no paramétricos. La decisión depende de la validación de los supuestos inherentes a cada uno, el tipo de variables y tamaño de la muestra. Comenzamos con una revisión sistemática de la literatura, que incluía artículos científicos cuyo instrumento de recolección de datos era una escala de evaluación/psicometría en el área de salud de niños y jóvenes. Definidos los criterios de inclusión, se seleccionaron 189 trabajos y 127 fueron sometidos a análisis exploratorio de las metodologías de investigación utilizadas. El análisis exploratorio de datos ha permitido identificar el tipo de técnicas estadísticas utilizadas por los investigadores cuando los instrumentos utilizados están compuestos de variables ordinales y analizar si se cumplían los supuestos inherentes al uso de las técnicas estadísticas con este tipo de variables, y con muestra de pequeñas o de gran tamaño.

RESUMO

A investigação nas diferentes áreas do conhecimento envolve grande número de variáveis que podem ser avaliadas com técnicas univariadas ou multivariadas e com testes paramétricos ou não paramétricos. A decisão depende da validação dos pressupostos inerentes a cada um deles, do tipo de variáveis e das dimensões das amostras. Começámos com uma revisão sistemática da literatura, que incluía artigos científicos cujo instrumento de recolha de dados era uma escala de avaliação/psicométrica na área da saúde da criança e jovem. Definidos os critérios de inclusão, foram seleccionados 189 artigos científicos, passando à análise exploratória das metodologias de investigação utilizadas 127 artigos. A análise exploratória de dados permitiu-nos identificar o tipo de

técnicas estadísticas utilizadas pelos investigadores quando os instrumentos utilizados são compostos por variáveis ordinais e analisar se são cumpridos os pressupostos inerentes à utilização das técnicas estadísticas com este tipo de variáveis e com amostra de pequena e grande dimensão.

ABSTRACT

Research in different areas of knowledge involves a large number of variables that can be evaluated with mono-variable and multi-variable techniques and parametric or nonparametric tests. The researcher decision depends on the validation of the assumptions inherent in each, the type of variables and sample sizes. We began with a systematic literature review, which included scientific articles that had as an instrument of data collection, a range of assessment/psychometrics in the health of children and youth. After the inclusion criteria were defined, 189 papers were selected and 127 articles were used for the exploratory analysis of research methodologies. The exploratory data analysis allowed us to identify the type of statistical techniques used by investigators when the instruments are composed of ordinal variables and analyze whether they met the assumptions inherent in the use of statistical techniques with this type of variables with small and large samples.

INTRODUCCIÓN

La investigación en las diferentes áreas del conocimiento abarca el análisis de variables a través de diferentes técnicas estadísticas dependiendo de la validación de los presupuestos inherentes a cada una de ellas, del tipo de variables y del tamaño de la muestra. El proceso de investigación incluye el concepto de variable y el tipo de medida que el investigador utiliza. En Ciencias sociales y humanas y en las Ciencias de la Salud, tenemos muchas veces que utilizar estrategias que nos permitan cuantificar lo no observable. Es en este contexto que emergen las «escalas de evaluación»⁽¹⁾ compuestas por variables cualitativas con una escala de medida ordinal que permiten estimar un constructo y que el investigador, muchas veces, trata como una variable continua⁽²⁾. Los items cualitativos que constituyen estas escalas de medida pueden presentarse con diverso número de clases de puntuación (3, 5, 7), habiendo una gran variabilidad en la utilización de los test estadísticos. Como refiere Marôco³ una de las necesidades más frecuentes en análisis estadístico pasa por la comparación de parámetros poblacionales basados en muestras aleatorias, en que se pretende testar si el tratamiento, intervención o manipulación tuvo un efecto significativo en la variable respuesta, en el caso univariante, o en la matriz de variables respuesta, en el caso multivariante. Las metodologías estadísticas inferenciales usadas para este tipo de análisis pueden ser de cariz paramétrico o no paramétrico.

Si se cumplen sus requisitos el investigador puede usar la estadística paramétrica (más potente que la no paramétrica)^(3,4). En caso contrario, la tradición de análisis recomienda el uso de la estadística no paramétrica, que no impone la especificación de una familia de distribuciones para la distribución muestral.

Tradicionalmente, en las ciencias sociales y humanas, la estadística no paramétrica ha sido utilizada como alternativa a la estadística paramétrica cuando sus presupuestos de aplicación no se verifican. Con todo, estudios de simulación⁽⁵⁾ han sugerido que algunos test no paramétricos (e.g. Mann-Whitney e Kruskal-Wallis) son tan sensibles como los paramétricos al incumplimiento de determinadas condiciones, sobre todo la homocedasticidad, y que la no verificación de estas condiciones conduce generalmente a la inflación de los errores tipo I y tipo II⁽³⁾. Así, la utilización de la estadística no paramétrica como alternativa a la paramétrica puede no ser la mejor estrategia de análisis. Por otro lado, es frecuente, en las ciencias sociales, ver

variables cualitativas ordinales analizadas con recurso a la estadística paramétrica. Urdan⁽⁶⁾ cuestiona la validez de los métodos paramétricos cuando aplicados a variables ordinales y refiere que las variables ordinales son “variables measured with numerical values where the numbers are meaningful (e.g., 2 is larger than 1) but the distance between the number is not constant” (p.6)⁽⁶⁾. La ausencia de recomendaciones claras relativas al uso de la estadística paramétrica versus no paramétrica es aún más pronunciada en el caso multivariante. Con frecuencia, en las ciencias sociales y humanas la naturaleza latente de un gran número de variables de interés obliga a la utilización de escalas sociométricas y psicométricas. Estos instrumentos de medida están generalmente constituídos por un conjunto de items ordinales, reconocidamente manifestaciones de las variables latentes que, en el dominio de las ciencias sociales y humanas no es posible observar directamente. De una forma general, este conjunto de items está sujeto a transformaciones lineales para operacionalizar la medida, pero esta práctica ha sido cuestionada por unos y defendida por otros^(7,8).

El análisis multivariante de datos engloba varias técnicas que pasan por el análisis de varianza multivariante (MANOVA), de componentes principales, de agrupamientos, de correlación canónica, discriminante y factorial. Varios autores consideran que la MANOVA tiene ventajas relativamente a la utilización de varias ANOVAs, de las cuales destacan la habilidad de medir los distintos ángulos del problema, pudiendo en algunos casos mejorar la potencia del test y presentar un valor de error tipo I más reducido⁽⁹⁾.

En muchos estudios, el uso de datos multivariantes es relevante si la entidad medida tiene una variedad de componentes conceptuales, como la competencia en una habilidad cognitiva como la lectura o la matemática, y si el investigador pretende comparar simultáneamente grupos en esas componentes. Destacar que en algunos casos donde se recogen respuestas múltiples para dos o más grupos de sujetos, las hipótesis de interés son univariantes por naturaleza y, en ese caso, cada variable de respuesta debe contemplarse aisladamente⁽³⁾.

Para que la MANOVA produzca conclusiones válidas, existen postulados que deben ser cumplidos y que son muy similares a aquellos que son requeridos para el uso de la ANOVA: (1) las observaciones deben ser independientes, (2) el vector de variables respuesta debe tener distribución normal multivariante para las diferentes poblaciones, y (3) las matrices de covarianzas de las poblaciones deben ser homogéneas^(10,3). Percibir el rendimiento de la MANOVA cuando los presupuestos no son cumplidos ha sido el objetivo de muchos investigadores. Finch⁽¹¹⁾ encontró algunos casos en que el incumplimiento del presupuesto de normalidad tiene un impacto limitado en la tasa de error de tipo I, refiriendo también algunos estudios que han mostrado que el poder de la MANOVA decrece sustancialmente cuando las variables del vector de respuesta son platicúrticas. Estos resultados son similares con el que se encontró en investigaciones sobre la potencia de la ANOVA.

Algunos estudios indican que cuando los grupos no varían en tamaño, la heterogeneidad de las matrices de covarianza tienen un efecto muy pequeño en el error de tipo I y en la potencia del test, pero cuando las muestras no presentan las mismas dimensiones, la heterogeneidad de las matrices de covarianza puede causar un aumento del error de tipo I y el poder del test estadístico puede disminuir. Estos resultados son más evidentes en la medida en que las «desigualdades» en las matrices de covarianzas y en las dimensiones de las muestras se van acentuando⁽¹¹⁾.

Ante los aspectos presentados y porque en ciencias sociales y humanas y en las ciencias de la salud tenemos muchas veces que utilizar estrategias que nos permitan cuantificar lo no observable, emergen las escalas, compuestas por variables cualitativas con una escala de medida ordinal y que el investigador, muchas veces, trata como si el constructo fuese de una variable continua (escalas, test psicológicos, inventarios). El problema surge porque en ausencia de un factor de calibración consensual (*gold standard*) nunca conseguiremos tener la certeza de que el instrumento medirá lo que se quiere medir, de forma válida, fiable y sensible⁽²⁾.

Al analizar las variables cualitativas, cuya escala de medida es ordinal, no podemos dejar de hablar de Rensis Likert por todo el trabajo desarrollado por este investigador en torno a las variables ordinales. Las escalas de tipo Likert son las más usadas en diversos campos de la investigación y particularmente en el área de la salud (enfermería y educación médica), psicología y otras ciencias sociales. En cuanto a la clasificación del tipo de estas variables no verificamos la existencia de dudas, son variables cualitativas ordinales, con escalas de medida que pueden presentar diferentes formas (ej. de las más comunes 3, 4, 5, 6, 7, 10 puntos), no podemos decir lo mismo cuando a través de operaciones algebraicas (suma o media) con los conjuntos de items, transformamos la variable ordinal a un nivel de medida intervalar (*Summated scales*) obteniendo *scores* referentes a constructo teóricos. En este punto existe una gran controversia entre los que son favorables a esta transformación y los que son desfavorables^{(7,8) (12-17)}.

No todos los autores están de acuerdo en que se adicionen los items, pues no consiguen comprender que, por ejemplo, al sumar « completamente de acuerdo» con «totalmente en desacuerdo», lo que el investigador suma es el grado y no el concepto y consideran que las variables son ordinales y su tratamiento estadístico tiene que privilegiar esta medida y no otra⁶. Hay también otros autores que consideran que esta operacionalización puede hacerse si los items de las escalas tuvieran por lo menos 5 puntos, en la escala de medida, y se consideraran como interválicas o casi interválicas, estando subyacente un constructo continuo^(18,2). Los investigadores que usan estos tipos de escala citan los estudios que valoran la utilización sin perjuicio de la investigación y recomiendan que debemos estar atentos: a la escala de medida, al número de items, si la escala está constituida por subescalas (si estas tienen un número de items iguales o diferentes) y al tamaño de la muestra.

En lo que concierne a los test estadísticos, las opiniones divergen de nuevo en lo relativo a su aplicación, pues algunos autores consideran que con variables ordinales solamente pueden usarse test no paramétricos, otros consideran que tras la transformación de las variables (*summated scales*) deben ser usados los paramétricos. Pell⁽¹⁹⁾ es uno de los autores que está de acuerdo con este presupuesto afirmando que los grupos de items pueden ser analizados con estadística paramétrica desde que se verifican los presupuestos de los test con todos los beneficios de su fortaleza. Carifo e Perla⁽⁸⁾ refuerzan esta idea, expresando que es perfectamente adecuado usar técnicas paramétricas con este tipo de variables.

En la literatura científica verificamos que muchos investigadores, en el caso univariante, estudian la normalidad de la distribución de las variables (usando por ejemplo el test de *Kolmogorov-Smirnov*) y verifican la homogeneidad de las varianzas (usando por ejemplo el test de *Levene*) con el objetivo de decidir si optan por los test paramétricos (e.g., teste t o ANOVA) o por los no paramétricos (e.g., U-Mann Whitney). También verificamos que algunos investigadores tienen esta práctica, pero

ante una situación en que las variables no tienen una distribución normal, consideran que los test paramétricos “(...) ANOVA, MANOVA, t-Student son firmes incluso cuando las distribuciones en estudio presentan sesgo o aplanamiento considerable (...)” (p.170)⁽²⁰⁾ y su decisión pasa por los test paramétricos en detrimento de los no paramétricos. Si por un lado los tests paramétricos exigen que la forma de distribución muestral sea conocida (ej: Normal), los tests no paramétricos no tienen este tipo de exigencia. Entonces, ¿por qué no se utilizan siempre los tests no paramétricos? Esta es una pregunta que puede surgir. Algunos autores justifican este aspecto diciendo que la potencia del test y la probabilidad de despreciar correctamente la hipótesis nula es mayor en los tests paramétricos que en los tests no paramétricos. Para Marôco⁽³⁾ “(...) los tests no paramétricos solo deben usarse cuando no existe alternativa, i. e., cuando no es posible validar las condiciones de aplicación de los tests paramétricos o cuando las variables no son cuantitativas” (p.199)⁽³⁾.

De la reflexión sobre las diferentes cuestiones presentadas y analizadas con este tipo de variables y porque en nuestro cotidiano nos enfrentamos diariamente con estos problemas, resolvimos analizar de forma más sistematizada lo que emerge de la literatura científica. Este análisis exploratorio de las metodologías de investigación va en el sentido de comprender cuál es el plan de análisis de los datos, una vez que en el área de las ciencias sociales y humanas, de la salud y motricidad humana, áreas que utilizan este tipo de variables, los investigadores muchas veces se interrogan si están siguiendo el camino adecuado. Así, este estudio tiene como objetivos: identificar el tipo de análisis estadístico (test paramétricos o no paramétricos) utilizado por los investigadores cuando los instrumentos utilizados están compuestos por variables ordinales (con 3, 4, 5, 7 y 10 puntos en la escala de medida) y analizar si se cumplen los presupuestos inherentes a la utilización de las diferentes técnicas estadísticas.

METODOLOGÍA

Este estudio comenzó con una revisión sistemática de la literatura (publicaciones periódicas en tres idiomas, portugués, español e inglés), donde se incluyeron los artículos científicos que tenían como instrumento de recogida de datos una escala de evaluación/psicométrica. En el contexto de este estudio definimos como criterios de inclusión, que estas escalas: 1) estuviesen compuestas por variables ordinales, 2) tuviesen una escala de medida de 3, 4, 5, 7, 10 puntos; 3) fuesen escalas usadas en el área de salud infantil y pediatría, pudiendo abarcar desde el recién nacido al joven adulto (con y sin deficiencia motora), y 4) uso en el análisis de datos test paramétricos o no paramétricos.

Estrategias de investigación

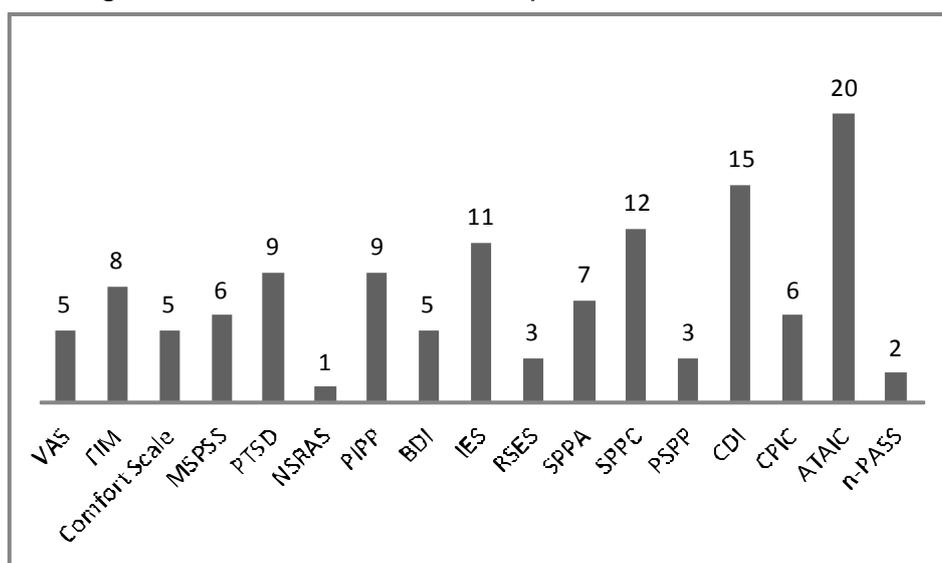
La investigación se realizó en Octubre de 2010, en bases electrónicas, de estudios no publicados (Dissertations and Theses (ProQuest); Bibliografía Nacional Portuguesa “on line”) y de estudios publicados en bases de datos como: CENTRAL, Cochrane Central Register of Controlled Trials; CINAHL, Cumulative Index to Nursing and Allied Health Literature; EMBASE, Medical Library Association; ISI Web of Knowledge; MEDLINE, U.S. National Library of Medicine; PsychINFO, abstract database of the psychological literature e SciELO, Scientific Electronic Library Online. Las palabras utilizadas fueron: *newborn, child, infant, adolescente, youth, scales, ordinal variables, parametric tests, non parametric teste*.

Se analizaron 240 artículos científicos. Tras una lectura de títulos y resúmenes decidimos incluir en el estudio 189 de estos artículos. Tras una lectura más minuciosa de los mismos, se excluyeron 62 artículos que no cumplían íntegramente los criterios de inclusión. Pasaron así al análisis exploratorio de las metodologías de investigación utilizadas 127 artículos.

RESULTADOS

Basándonos en el objetivo de identificar si los investigadores utilizan técnicas estadísticas paramétricas o no paramétricas cuando los instrumentos utilizados están compuestos por variables ordinales y analizar si se cumplen los presupuestos inherentes a la utilización de estas técnicas, se efectuó un análisis para distinguir lo que emerge en relación al instrumento utilizado y su caracterización por los autores, al tamaño de la muestra y a su relación con el número de ítems (variables) del instrumento y tipo de test estadísticos aplicados en los diferentes estudios. Los datos se extrajeron de los 127 artículos seleccionados para una matriz de datos, especialmente construida para este estudio, cuya parrilla incluía todos los aspectos en análisis, siendo después presentados gráficamente para una mejor visualización de los resultados. Estos artículos tenían como instrumento de recogida de datos una o dos entre las diecisiete escalas de evaluación presentadas, las cuales estaban compuestas por escalas únicas (un conjunto de ítems) o con varias subescalas (varios conjuntos de ítems) constituidas por variables ordinales. Así, en la figura 1 presentamos los instrumentos más utilizados: la State-Trait Anxiety Inventory for Children, STAIC (20); la Child Depression Inventory, CDI (15); la Self-Perception Profile for Children, SPPC (12); la Impact of Event Scale, IES (11); la Premature Infant Profile, PIPP y la Posttraumatic Stress Disorder scale, PTSD (9); la Functional Independence Measure, FIM (8); la Self-Perception Profile for Adolescents, SPPA (7); la Children's Perception of Interparental Conflict Scale, CPIC y la Multidimensional Scale of Perceived Social Support , MSPSS (6); la Beck Depression Inventory, BDI, la Comfort Scale, CS y la Visual Analogue Scale, VAS (5); la Rosenberg Self-Esteem Scale, RSES y la PSPP (3); la Neonatal Pain, Agitation and Sedation Scale, N-PASS (2) y la Neonatal Skin Risk Assessment Scale, NSRAS (1).

Figura 1 – Número de Artículos por escala de evaluación



Instrumento

Al hacer un análisis más pormenorizado de los artículos en lo que concierne al tipo de instrumento, verificamos que los artículos presentan de forma clara y explícita el tipo de instrumento (escalas) utilizado por los investigadores, el número de ítems de que están compuestas las escalas y también si estamos ante una escala única (un solo conjunto de ítems) o si las escalas están compuestas por subescalas (varios conjuntos de ítems), si estos conjuntos están equilibrados numéricamente o no y el número de puntos de la escala de medida (3, 4, 5, 7, 10 puntos).

En la mayoría de los casos se percibe, sin recurrir a otro tipo de consulta, la composición de los instrumentos, de las subescalas y así saber si existen diferencias en los conjuntos de ítems. En lo que respecta al número de ítems por escala, en los instrumentos analizados, existen instrumentos compuestos por diferentes números de ítems y estos oscilan entre 4 y 51 ítems. Se verifica también que pueden estar compuestos solo por una única escala o hasta 10 subescalas, variando estas en número de subescalas y número de ítems en cada una de ellas, lo que no siempre permite caracterizarlas con alguna uniformidad. En lo relativo a la escala de medida, los instrumentos analizados en este estudio presentaban diferentes escalas de medida que varían entre tres y diez (c.f. Tabla I).

Tabla I – Artículos científicos por tipo de escala de evaluación/psicométrica

Escalas de Evaluación (Psicométricas)	Nº Ítems	Sub-Escalas (nº de sub-escalas * nº de ítems)	Escala de Medida	Total de Artículos
N-PASS	5	-	3	2
STAIC	40	2(2*20)	3	20
CPIC	48	9(3*4; 1*5; 4*6;	3	6
	40	1*7)		
		7(1*4; 1*5; 4*6;		
		1*7)		
CDI	27	-	3	15
PSPP	36	-	4	3
SPPC	36	6*6	4	12
SPPA	51	10 (9*5; 1*6)	4	7
	45	9 (9*5)		
	36	6 (6*6)		
	20	4 (4*5)		
RSES	10	-	4	3
	42	2 (1*8; 1*34)		
IES	15	2 (1*7; 1*8)	4	11
	22	3(2*7; 1*8)		
BDI	21	-	4	5
PIPP	4	-	4	9
NSRAS	6	-	4	1
PTSD	20	-	5	9
MSPSS	12	3 (3*4)	5	6
Comfort Scale	8	-	5	5

FIM	30	5(1*2; 1*3; 2*4;	7	8
	18	2*5; 1*7)		
VAS	10	3(1*6; 2*3; 3*2)	10	5
TOTAL				127

Muestra y Test Estadísticos

La elección de la población que vamos a inquirir está íntimamente ligada al objeto de estudio, a los objetivos y al método de investigación. Los estudios que tienen una lógica extensiva asociada (usando lo preguntado por cuestionario como instrumento) y como tal una estrategia de investigación cuantitativa exigen cuidados con el tamaño de la muestra. Si alguna de las muestras fuera de pequeño tamaño (<30), debemos verificar la normalidad de la variable dependiente en esa muestra, a través de un test de normalidad (como ejemplo el *Test de Shapiro Wilks*) ya que en el caso contrario el Teorema del Límite Central, la distribución de la media muestral es aproximadamente normal.

Otros autores se refieren al uso de las «escalas», instrumentos compuestos por varias preguntas/afirmaciones o ítems que tienen como finalidad evaluar una determinada variable o característica²¹, siendo un tipo de instrumento que al exigir que se hagan transformaciones de variables, y aplicados algunos tests estadísticos, importa que el número de inquiridos sea más elevado, mínimo de cinco individuos por ítem y lo ideal sería 10 individuos por ítem de la escala (ejemplo para la aplicación de la escala SPPC con 36 ítems debería ser considerada una muestra entre 180 y 360 y la STAIC con 40 ítems debería ser considerada un tamaño de muestra de por lo menos 200 individuos, siendo lo ideal 400) teniendo también en cuenta el tipo de test estadísticos que van a ser aplicados y las variables independientes utilizadas. Al analizar los test estadísticos aplicados en los diferentes estudios, podemos decir que de los ocho artículos publicados en lengua portuguesa, tres presentaban simultáneamente tests paramétricos y no paramétricos (t-Student, ANOVA, MANOVA, Mann Whitney-U Test) y cinco presentaban solo tests paramétricos. Los restantes 119 se presentan distribuidos de la siguiente forma: seis usaron test no paramétricos (Mann Whitney-U Test) y ciento trece utilizan tests paramétricos (t-Student, ANOVA, MANOVA, ANCOVA) siendo que en once de estos estudios se utilizaron simultáneamente test no paramétricos.

Verificamos en la literatura científica que algunos autores parten del principio de que el análisis de varianza exige que sean observados los presupuestos de la normalidad (la variable dependiente debe tener una distribución normal en todos los grupos) y de la homocedasticidad (las varianzas deben ser iguales en todos los grupos), pero no siempre los artículos científicos expresan este presupuesto. En realidad, lo que verificamos es que en la mayoría de los artículos analizados no aparece explícito si se verificaron los presupuestos de las metodologías utilizadas y tampoco aparece ninguna justificación para la utilización de los tests paramétricos o no paramétricos. De ahí que no siempre se consigue hacer una evaluación rigurosa de los presupuestos enunciados por el hecho de estos estar omitidos en los artículos.

Como no podemos analizar todos los instrumentos, seleccionamos la escala más usada en los artículos revisados, la State-Trait Anxiety Inventory for Children (STAIC)⁽²⁰⁻³⁹⁾, que aparece en veinte de los artículos analizados. Verificamos que esta está compuesta por dos subescalas con 20 ítems cada una, cumpliendo el presupuesto de que las subescalas deberían tener un número de ítems semejante. Sin embargo emerge en el análisis el incumplimiento de los presupuestos relacionados con el tamaño de la muestra (solo en 30% de los estudios la muestra es superior a 200 individuos) y con los tests estadísticos aplicados no emerge la justificación del uso de test paramétricos o no paramétricos (son aplicados test paramétricos en todos los estudios siendo que en 15% se usan simultáneamente test no paramétricos).

La STAIC se utilizó en cuatro estudios experimentales, resultando que en uno de los estudios se usaron test paramétricos y no paramétricos y el tamaño de la muestra en estudio era en el grupo experimental $n=43$ y en el grupo de control $n=20$; otros dos estudios presentaban muestras con tamaños iguales para el grupo experimental y control ($n=30$ y $n=35$ respectivamente) y en el cuarto estudio los grupos eran $n=284$ y $n=220$ (experimental y control, respectivamente). En los restantes dieciséis estudios se verificó que el tamaño de la muestra varió entre $n=20$ y $n=3467$ y que en la mayoría de los estudios se usaron test paramétricos. Solo en tres estudios se usaron simultáneamente test paramétricos y no paramétricos, en uno de los experimentales ($n=43$ y $n=20$, grupos experimental y de control) y en dos de los otros ($n=29$ y $n=2280$). Lo que nos permite afirmar que los presupuestos de los mínimos de la muestra no se cumplieron en la mayoría de los estudios, basados en el presupuesto del tamaño de la muestra (muestra de gran tamaño), el presupuesto del número de participantes por ítem y el tipo de test estadísticos aplicados. Como las muestras no presentan las dimensiones exigidas, ni presentan los mismos tamaños, lo que puede llevar a una heterogeneidad de las matrices de covarianza lo que se podrá reflejar en un aumento de la probabilidad de error de tipo I y el poder del test estadístico puede disminuir⁽¹¹⁾. En cuanto a los test estadísticos aplicados no emerge ninguna justificación.

CONCLUSIONES

En este análisis exploratorio de datos, se quiso identificar el tipo de técnicas estadísticas utilizadas por los investigadores (artículos científicos ligados al área de salud del niño, adolescente y joven adulto con y sin deficiencia) cuando los instrumentos utilizados están compuestos por variables ordinales y analizar si se cumplen los presupuestos inherentes a la utilización de las técnicas estadísticas con este tipo de variables y con muestra de pequeño y gran tamaño. Se analizaron 127 artículos que cumplían los criterios de inclusión, teniendo como primera intención de análisis el tipo de instrumento (escala única o dividida en subescalas) con escalas ordinales de 3, 4, 5, 7 y 10 puntos. Como segunda intención de análisis el tamaño de la muestra y como tercera intención de análisis los test estadísticos aplicados.

Del análisis exploratorio de los artículos, concluimos que muchos no traen expreso algunos de los presupuestos que gostaríamos analizar en profundidad o porque los autores no los incluyeron en el artículo o no pudieron ser trabajados (ex: distribución, homogeneidad de las varianzas, transformación de las variables para poder ser tratadas como continuas). Respecto a la normalidad de la distribución de las variables ni todos los artículos muestran si fueron aplicados test ni cuales (algunos dicen que se aplicó el *Kolmogorov-Smirnov*), lo mismo sucedió con la homogeneidad de las

varianzas (algunos refieren el test de *Levene*). De esta forma no queda claro para el lector lo que estuvo en el origen de la toma de decisión de optar por test paramétricos o no paramétricos.

El hecho de estar, en algunos de los estudios analizados, ante muestras de pequeño tamaño donde son usadas estadísticas multivariantes, podría decirse que los investigadores deberían haber usado la inferencia no paramétrica, dado que esta no exige la especificación de un modelo poblacional, al contrario de lo que sucede en la inferencia paramétrica y porque en algunos de los casos no se verifican los presupuestos referidos anteriormente. Pero esto no se verifica tampoco, en estos estudios, pues en su mayoría son usados test paramétricos. Los investigadores optaron por aplicar test paramétricos basándose en estudios de simulación que han sugerido que algunos test no paramétricos son tan sensibles como los paramétricos a la violación de algunas condiciones como la homocedasticidad en la inflación de los errores tipo I y tipo II optando por eso por test paramétricos debido a su firmeza. Pero tampoco hay una explicitación clara en los artículos analizados.

Para algunos investigadores que trabajan en estas áreas donde no siempre es fácil medir, pues muchas veces no se consigue tener la certeza de que el instrumento mida, lo que es supuesto medir, de forma válida, fiable y sensible, se torna importante la valorización de lo que ha sido investigado con este tipo de variables para dar consistencia al análisis de datos de esta naturaleza y así podremos continuar investigando en el sentido de contribuir con evidencia científica para que el rigor crezca y los artículos científicos presenten de forma expresa todos los presupuestos analizados.

REFERENCIAS

1. Hill M, Hill A (2000), *Investigação por Questionário*, Lisboa, Edições Sílabo.
2. Marôco J (2010), *Análise Estatística: Com PASW (ex-PSS)*, Pêro Pinheiro, Report Number, Lda.
3. Marôco J (2009), *Avaliação das qualidades psicométricas de uma escala*, Lisboa, ISPA.
4. Pontes ACF (2005), *Análise de Variância multivariada com utilização de testes não-paramétricos e componentes principais baseados em matrizes de postos* [Internet]. São Paulo, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, [citado em 2010 Dez. 29]. Disponible em: http://www.lce.esalq.usp.br/tadeu/AntonioPontes_tese.pdf
5. Zimmerman DW(2000), "Statistical significance levels of nonparametric tests biased by heterogeneous variances of treatment groups", *Journal of General Psychology*, 127, pp.354-364.
6. Urdan TC (2010), *Statistics in Plain English*, 3rd ed. New York, Routledge, Taylor & Francis Group.
7. Carifio J, Perla R (2007) "Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Response Formats and their Antidotes", *Journal of Social Sciences*, 3(3), pp.106-116.
8. Carifio J, Perla R (2008), "Resolving the 50-year debate around using and misusing Likert scales. *Medical Education*", 42, pp.1150-1152.
9. Tabachnick BG, Fidell LS (2007), *Using multivariate statistics*, New York, Pearson Education Inc.
10. Sheskin DJ. (2007), *Handbook of Parametric and Nonparametric Statistical Procedures*, 4th ed. Boca Raton, Florida, Chapman & Hall/CRC.

11. Finch H (2005), "Comparison of the Performance on Nonparametric and Parametric MANOVA Test Statistics when Assumption are Violated", *Methodology*, 1(1), pp.27-38.
12. Jamieson S (2004), "Likert scales: how to (ab)use them", *Medical Education*, 38, pp.1217-1218.
13. Wang S-T, Wang, CJ, Yu M-L, Huang, C-C (1999), "Bridging the Gap between the Pros and Cons in Treating Ordinal Scales from an Analysis Point of View", *Nursing Research*, July/August, pp.226-229.
14. Knapp TR (1990), "Treating Ordinal Scales as Interval Scales: An Attempt to Resolve the Controversy" *Nursing Research*, March/Jun, pp.121-123.
15. Knapp TR (1993), "Treating Ordinal Scales As ordinal scales", *Nursing Research*, May/Jun, pp.184-186.
16. Mitchell J (1986), "Measurement Scales and Statistics: A Clash of Paradigms", *Psychological Bulletin*, 100(3), pp.398-407.
17. Weiss DJ. (1986), "The Discriminating Power of Ordinal Data", *Journal of Social Behavior and Personality*, 1(3), pp. 381-389.
18. Bollen KA (1989), *Structural Equations with Latent Variables*. Series in Probability and Mathematical Statistics, New York, Wiley.
19. Pell G (2005), "Uses and misuses ok Likert scales", *Medical Education*, pp.39:970.
20. Rodrigues MJB (2008), *Reações Emocionais e Percepções da Criança ao Conflito Parental* [Internet]. Porto: Instituto Superior de Ciências Biomédicas de Abel Salazar, Universidade do Porto; [citado 2010 dez. 29]. Disponível em: <http://repositorioaberto.up.pt/bitstream/10216/19395/2/ReacesEmocionaisPercepoCria naConflitoParental.pdf>
21. Arslan N, Büyükgebiz B, Öztürk Y, Akay AP. (2003), "Depression and anxiety in chronic hepatitis B: effect of hepatitis B virus infection on psychological state in childhood", *The Turkish Journal of Pediatrics*, 45, pp.26-28.
22. Bakib O, Erdogana A, Kantarcic O, Akisikd G, Kayaalpa L, Yalcinkayab C (2004), "Anxiety and depression in children with epilepsy and their mothers", *Epilepsy & Behavior*, 5, pp.958–964.
23. Bas M, Aşçi F, Karabudak E, Kiziltan G (2004), "Eating Attitudes and their Psychological correlates among Turkish Adolescents", *Adolescence*, 39, pp.593-599.
24. Bilgiç A, Yılmaz S, Tiraş S, Deda G, Kiliç EZ. (2006), "Depression and Anxiety Symptom Severity in a Group of Children with Epilepsy and Related Factors", *Turkish Journal of Psychiatry*, 17(3), pp.165-172.
25. Derryberry D, Reed MA (2002), "Anxiety-Related Attentional Biases and Their Regulation by Attentional Control", *Journal of Abnormal Psychology*, 111(2), pp.225–236.
26. Hobbie WL, Stuber M, Meeske K, Wissler K, Rourke MT, Ruccione K, et al. (2000), "Symptoms of Posttraumatic Stress in Young Adult Survivors of Childhood Cancer", *Journal of Clinical Oncology*, 18(24), pp.4060-4066.
27. Fernandez-Berrocal P, Alcaide R, Extremera N, Pizarro D (2006), "The Role of Emotional Intelligence in Anxiety and Depression among Adolescents", *Individual Differences Research*, 4(1), pp.16-26.
28. Huizinga GA, Visser A, Van der Graaf WTA, Hoekstra HJ, Klip EC, Pras E, et al. (2005) "Stress response symptoms in adolescent and young adult children of parents diagnosed with cancer", *European Journal of Cancer*, 41, pp.288-295.
29. O'Dea JA, Abraham S (2000), "Improving the Body Image, Eating Attitudes, and Behaviors of Young Male and Female Adolescents: A New Educational Approach that Focuses on Self- Esteem", *International Journal of Eating Disorders*, 28(1), pp.43-57.

30. O'Dea JA, Abraham S (2004), "Onset of Disordered eating attitudes and behavior in early adolescence: interplay of pubertal status, gender weight and age", *Adolescence*, 39, pp.155-165.
31. Muris P, Merckelbach H, Schmidt H, Tierney S (1999), "Disgust sensitivity, trait anxiety and anxiety disorders symptoms in normal children", *Behaviour Research and Therapy*, 37, pp.953-961.
32. Muris P, Merckelbach H, Van Brakel A, Mayer B, Van Dongen L (1999), "The Screen for Child Anxiety Related Emotional Disorders (Scared): Relationship with Anxiety and Depression in Normal Children", *Personality and Individual Differences*, 24(4), pp.451-456.
33. Muris P, Steerneman M, Merckelbach H, Meesters C (1996), "The role of parental fearfulness and modeling in children's fear", *Behaviour Research and Therapy*, 34(3), pp.265-268.
34. Von Weiss RT, Rapoff MA, Varni JW, Lindsley CB, Olson NY, Madson KL, et al. (2002), "Daily Hassles and social support as predictors of adjustment in children with pediatric rheumatic disease", *Journal of Pediatric Psychology*.; 27(2), pp.155-165.
35. Telze EH, Mogg K, Bradley BP, Mai X, Ernst M, Pine, DS, et al. (2008), "Relationship Between Trait Anxiety, Prefrontal Cortex, and Attention Bias to Angry Faces in Children and Adolescents", *Biological Psychology*, 79(2), pp. 216–222.
36. Vidanovic S, Andelkovic V (2006), "Ego development and the anxiety of gifted Adolescents", *Philosophy, Sociology and Psychology*, 5(1), pp.87-102.
37. Vila G, Hayder R, Bertrand C, Falissard B, De Blic J, Mouren-Simeoni M-C, et al. (2003), "Psychopathology and Quality of Life for Adolescents With Asthma and Their Parents", *Psychosomatics*, 44(4), pp.319-328.
38. Vila G, Porche M, Mouren-Simeoni M-C (1999), "An 18-Month Longitudinal Study of Posttraumatic Disorders in Children Who Were Taken Hostage in Their School", *Psychosomatic Medicine*, 61, pp.746–754.
39. Wolbeek M, Van Doornen LJP, Kavelaars A, Heijnen CJ (2006). "Severe fatigue in adolescents: A common phenomenon", *Pediatrics*, 117, pp.1078-1086.

ISSN 1695-6141

© [COPYRIGHT](#) Servicio de Publicaciones - Universidad de Murcia