



REVISIONES

Análise estatística de escalas ordinais. Aplicações na Área da Saúde Infantil e Pediatria

Análisis estadístico de escalas ordinales. Aplicaciones en el Área de Salud infantil y Pediatría
Statistical analysis of ordinal scales. Field applications of Pediatrics and Child Health

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

*Professora Coordenadora. Departamento de Saúde da Criança e Jovem. Escola Superior de Enfermagem de Lisboa E-mail: malicecurado@gmail.com **Professora Auxiliar. Departamento de Métodos Matemáticos. Faculdade de Motricidade Humana, CIPER. Professor Auxiliar. Departamento de Estatística. Instituto Superior de Psicologia Aplicada. Portugal.

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

Palabras clave: escalas; muestra; análisis estadístico; salud infantil; pediatria

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

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

RESUMEN

La investigación en diferentes áreas del conocimiento implica un gran número de variables que pueden ser evaluadas con técnicas univariantes y multivariantes y tests 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.

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.

INTRODUÇÃO

A investigação nas diferentes áreas do conhecimento envolve análise de variáveis através de diferentes técnicas estatísticas dependendo da validação dos pressupostos inerentes a cada uma delas, do tipo de variáveis e da dimensão amostral. O processo de investigação inclui o conceito de variável e a forma de medida que o investigador utiliza. Em ciências sociais e humanas e nas ciências da saúde, temos muitas vezes que utilizar estratégias que nos permitam quantificar o não observável. É neste contexto que emergem as «escalas de avaliação»⁽¹⁾ compostas por variáveis qualitativas com uma escala de medida ordinal que permitem estimar um constructo e que o investigador, muitas vezes, trata como uma variável contínua⁽²⁾. Os itens qualitativos que constituem estas escalas de medida podem apresentar-se com os mais diversos números de classes de pontuação (3, 5, 7), havendo uma grande variabilidade na utilização dos testes estatísticos. Como refere Marôco³ uma das necessidades mais frequentes em análise estatística passa pela comparação de parâmetros populacionais tendo por base amostras aleatórias, e em que se pretende testar se o tratamento, intervenção ou manipulação teve um efeito significativo na variável resposta, no caso univariado, ou na matriz de variáveis resposta, no caso multivariado. As metodologias estatísticas inferenciais usadas para este tipo de análises podem ser de cariz paramétrico ou não paramétrico.

Se os seus requisitos forem cumpridos o investigador pode usar a estatística paramétrica (mais potente do que a não paramétrica)^(3,4). Em caso contrário, a tradição de análise recomenda o uso da estatística não paramétrica, que não impõe a especificação de uma família de distribuições para a distribuição amostral.

Tradicionalmente, nas ciências sociais e humanas, a estatística não paramétrica tem sido utilizada em alternativa à estatística paramétrica quando os seus pressupostos de aplicação não se verificam. Contudo, estudos de simulação⁽⁵⁾ têm sugerido que alguns testes não paramétricos (e.g. Mann-Whitney e Kruskal-Wallis) são tão sensíveis quanto os paramétricos à violação de algumas condições, nomeadamente a homocedasticidade, e que a não verificação destas condições conduz geralmente à inflação dos erros tipo I e tipo II⁽³⁾. Assim sendo, a utilização da estatística não paramétrica em alternativa à paramétrica pode não ser a melhor estratégia de análise. Por outro lado, é frequente, nas ciências sociais, ver variáveis qualitativas ordinais analisadas com recurso à estatística paramétrica. Urdan⁽⁶⁾ questiona a validade dos

métodos paramétricos quando aplicados a variáveis ordinais e refere que as variáveis ordinais são “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⁾. A ausência de recomendações claras relativamente ao uso da estatística paramétrica versus não paramétrica é ainda mais pronunciada no caso multivariado. Com frequência, nas ciências sociais e humanas a natureza latente de um grande número de variáveis de interesse obriga à utilização de escalas sociométricas e psicométricas. Estes instrumentos de medida são geralmente constituídos por um conjunto de itens ordinais, assumidamente manifestações das variáveis latentes que, no domínio das ciências sociais e humanas não é possível observar directamente. De uma forma geral, este conjunto de itens são sujeitos a transformação lineares de forma a operacionalizar a medida, mas esta prática tem sido questionada por uns e defendida por outros^(7,8).

A análise multivariada de dados engloba várias técnicas que passam pela análise de variância multivariada (MANOVA), de componentes principais, de agrupamentos, de correlação canónica, discriminante e factorial. São vários os autores que consideram que a MANOVA tem vantagens relativamente à utilização de várias ANOVAs, das quais salientam a habilidade de medir os vários ângulos do problema, podendo em alguns casos melhorar a potência do teste e apresentar um valor de erro tipo I mais reduzido⁽⁹⁾.

Em muitos estudos, o uso de dados multivariados é relevante se a entidade medida tem uma variedade de componentes conceptuais, como a proficiência numa habilidade cognitiva, como a leitura ou a matemática, e se o investigador pretender comparar simultaneamente grupos nessas componentes. De notar que em alguns casos onde respostas múltiplas são recolhidas para dois ou mais grupos de sujeitos, as hipóteses de interesse são univariadas por natureza e, nesse caso, cada variável de resposta deve ser vista isoladamente⁽³⁾.

Para que a MANOVA produza conclusões válidas, existem postulados que devem ser cumpridos e que são muito similares àqueles que são requeridos para o uso da ANOVA: (1) as observações devem ser independentes, (2) o vector de variáveis resposta deve ter distribuição normal multivariada para as diferentes populações, e (3) as matrizes de covariâncias das populações devem ser homogéneas^(10,3). Perceber a performance da MANOVA quando os pressupostos não são cumpridos, tem sido o objectivo de muitos investigadores. Finch⁽¹¹⁾ encontrou alguns casos em que a violação do pressuposto de normalidade tem um impacto limitado na taxa de erro de tipo I, referindo também, alguns estudos que têm mostrado que o poder da MANOVA decresce substancialmente quando as variáveis do vector de resposta são platicúrticas. Estes resultados são similares com o que foi encontrado em investigações sobre a potência da ANOVA.

Algumas pesquisas indicam que quando os grupos não variam em tamanho, a heterogeneidade das matrizes de covariância tem um efeito muito pequeno no erro de tipo I e na potência do teste, mas quando as amostras não apresentam as mesmas dimensões, a heterogeneidade das matrizes de covariância pode causar um aumento do erro de tipo I e o poder do teste estatístico pode diminuir. Estes resultados são mais evidentes à medida que as «desigualdades» nas matrizes de covariâncias e nas dimensões das amostras se vão acentuando⁽¹¹⁾. Perante os aspectos apresentados e porque em ciências sociais e humanas e nas ciências da saúde temos muitas vezes que utilizar estratégias que nos permitam quantificar o não observável, emergem as

escalas, compostas por variáveis qualitativas com uma escala de medida ordinal e que o investigador, muitas vezes, trata como se o constructo fosse de uma variável contínua (escalas, testes psicológicos, inventários). Porém o problema surge pelo facto de na ausência de um factor de calibração consensual (*gold standard*) nunca conseguirmos ter a certeza de que o instrumento está a medir o que é suposto medir, de forma válida, fiável e sensível⁽²⁾.

Ao analisarmos as variáveis qualitativas, cuja escala de medida é ordinal, não podemos deixar de falar de Rensis Likert por todo o trabalho desenvolvido por este investigador em torno das variáveis ordinais. As escalas de tipo Likert são das mais usadas em diversos campos da investigação e particularmente na área da saúde (enfermagem e educação médica), psicologia e outras ciências sociais. Quanto à classificação do tipo destas variáveis não verificamos a existência de dúvidas, são variáveis qualitativas ordinais, com escalas de medida que podem apresentar diferentes formas (ex. das mais comuns 3, 4, 5, 6, 7, 10 pontos), o mesmo não podemos dizer quando através de operações algébricas (soma ou média) com os conjuntos de itens, transformamos a variável ordinal a um nível de medida intervalar (*Summated scales*) obtendo *scores* referentes a constructo teóricos. Neste ponto existe uma grande controvérsia entre os que são favoráveis a esta transformação e os que são desfavoráveis ^{(7,8) (12-17)}.

Nem todos os autores estão de acordo que se adicionem os itens, pois não conseguem compreender que, por exemplo, ao somar «concordo completamente» com «discordo totalmente», o que o investigador soma é o grau e não o conceito e, consideram que as variáveis são ordinais e o seu tratamento estatístico tem que privilegiar essa medida e não outra⁶. Ainda existem outros autores que consideram que esta operacionalização pode ser feita se os itens das escalas tiverem pelo menos 5 pontos, na escala de medida, e forem consideradas como intervalares ou quase intervalares, tendo subjacente um constructo contínuo^(18,2). Os investigadores que usam estes tipos de escala referem os estudos que valorizam a sua utilização sem prejuízo da investigação e recomendam que devemos estar atentos: à escala de medida, ao número de itens, se a escala é constituída por subescalas (se estas têm um número de itens iguais ou diferentes) e à dimensão da amostra.

No que concerne aos testes estatísticos, as opiniões voltam a divergir relativamente à sua aplicação, pois alguns autores consideram que com variáveis ordinais somente podem ser usados testes não paramétricos outros consideram que após a transformação das variáveis (*summated scales*) devem ser usados os paramétricos. Pell⁽¹⁹⁾ é um dos autores que concorda com este pressuposto dizendo que os grupos de itens podem ser analisados com estatística paramétrica desde que se verifiquem os pressupostos dos testes com todos os benefícios da sua robustez. Carifo e Perla⁽⁸⁾ reforçam esta ideia, dizendo que é perfeitamente adequado usar técnicas paramétricas com este tipo de variáveis.

Na literatura científica verificamos que muitos investigadores, no caso univariado, estudam a: normalidade da distribuição das variáveis (usando por exemplo o teste de *Kolmogorov-Smirnov*) e verificam a homogeneidade das variâncias (usando por exemplo o teste de *Levene*) com o objectivo de decidirem se optam pelos testes paramétricos (e.g., teste t ou ANOVA) ou pelos não paramétricos (e.g., U-Mann Whitney). Porém também verificamos que alguns investigadores têm esta prática, mas perante uma situação em que as variáveis não possuem uma distribuição normal, consideram que os testes paramétricos “(...) ANOVA, MANOVA, t-Student

são robustas mesmo quando as distribuições sob estudo apresentam enviesamento ou achatamento considerável (...)” (p.170)⁽²⁰⁾ e a sua decisão passa pelos testes paramétricos em detrimento dos não paramétricos. Se por um lado os testes paramétricos exigem que a forma de distribuição amostral seja conhecida (ex: Normal), os testes não paramétricos não têm esse tipo de exigência. Então, porque não se utilizam sempre os testes não paramétricos? Esta é uma questão que pode surgir. Alguns autores justificam este aspecto dizendo que a potência de teste e a probabilidade de rejeitar, correctamente, a hipótese nula é maior nos testes paramétricos do que nos testes não paramétrico. Para Marôco⁽³⁾ “(...) os testes não paramétricos devem apenas usar-se quando não existe alternativa, i. e., quando não é possível validar as condições de aplicação dos testes paramétricos ou quando as variáveis não são quantitativas” (p.199)⁽³⁾.

Da reflexão sobre as diferentes questões apresentadas e analisadas com este tipo de variáveis e porque no nosso quotidiano somos confrontados diariamente com estes problemas, resolvemos analisar de forma mais sistematizada o que emerge na literatura científica. Esta análise exploratória das metodologias de investigação vai no sentido de compreender qual o plano de análise dos dados, uma vez que na área das ciências sociais e humanas, da saúde e motricidade humana, áreas que utilizam este tipo de variáveis, os investigadores muitas vezes se interrogam se estão a seguir o melhor caminho. Assim, este estudo tem como objectivos: identificar o tipo de análise estatística (testes paramétricos ou não paramétricas) utilizada pelos investigadores quando os instrumentos utilizados são compostos por variáveis ordinais (com 3, 4, 5, 7 e 10 pontos na escala de medida) e analisar se são cumpridos os pressupostos inerentes à utilização das diferentes técnicas estatísticas.

METODOLOGIA

Este estudo começou com uma revisão sistemática da literatura (publicações periódicas em três idiomas, português, espanhol e inglês), onde foram incluídos os artigos científicos que tinham como instrumento de recolha de dados uma escala de avaliação/psicométrica. No contexto deste estudo definimos como critérios de inclusão, que estas escalas: 1) fossem compostas por variáveis ordinais, 2) tivessem uma escala de medida de 3, 4, 5, 7, 10 pontos; 3) fossem escalas usadas na área da saúde infantil e pediatria, podendo abranger desde o recém-nascido ao jovem adulto (com e sem deficiência motora), e 4) usassem na análise de dados testes paramétricos ou não paramétricos.

Estratégias de pesquisa

A pesquisa realizou-se em Outubro de 2010, em bases electrónicas, de estudos não publicados (Dissertations and Theses (ProQuest); Bibliografia Nacional Portuguesa “on line”) e de estudos publicados em bases de dados 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. As palavras utilizadas foram: *newborn, child, infant, adolescente, youth, scales, ordinal variables, parametric tests, non parametric teste*.

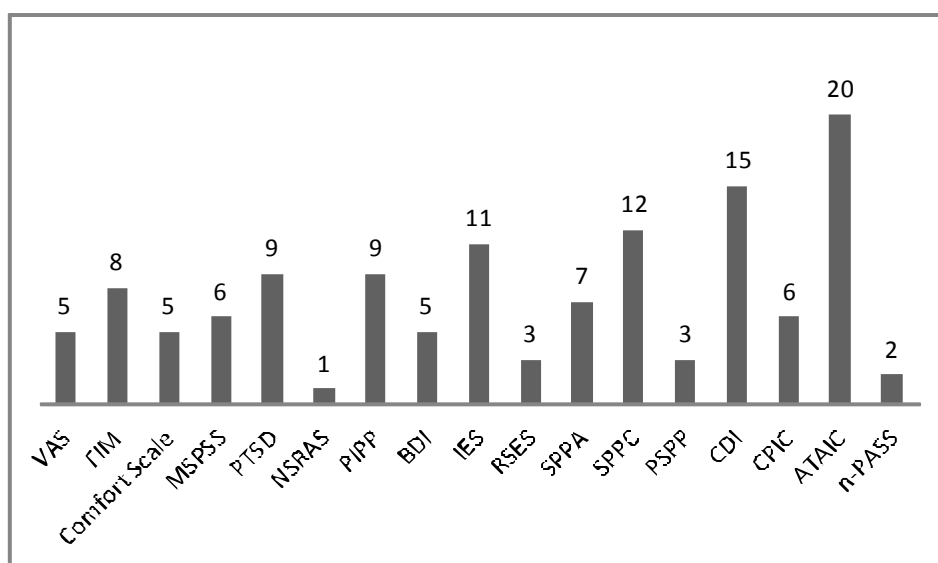
Foram analisados 240 artigos científicos. Após uma leitura de títulos e resumos decidimos pela inclusão no estudo de 189 desses artigos. Com uma leitura mais

minuciosa dos mesmos, foram excluídos 62 artigos que não cumpriam na íntegra os critérios de inclusão. Passaram assim, à análise exploratória das metodologias de investigação utilizadas 127 artigos.

RESULTADOS

Tendo por base o objectivo de identificar se os investigadores utilizam técnicas estatísticas paramétricas ou não paramétricas quando os instrumentos utilizados são compostos por variáveis ordinais e analisar se são cumpridos os pressupostos inerentes à utilização destas técnicas, será efectuada uma análise para salientar o que emerge relativamente ao instrumento utilizado e sua caracterização pelos autores, ao tamanho da amostra e ao seu relacionamento com o número de itens (variáveis) do instrumento e o tipo de testes estatísticos aplicados nos diferentes estudos. Os dados foram extraídos dos 127 artigos seleccionados para uma matriz de dados, especialmente construída para este estudo e cuja grelha incluía todos os aspectos em análise, sendo depois apresentados graficamente para uma melhor visualização dos resultados. Estes artigos tinham como instrumento de recolha de dados uma ou duas entre as dezassete escalas de avaliação apresentadas, as quais eram compostas por escalas únicas (um conjunto de itens) ou com várias subescalas (vários conjuntos de itens) constituídas por variáveis ordinais. Assim, na figura 1 apresentamos os instrumentos mais utilizados: a State-Trait Anxiety Inventory for Children, STAIC (20); a Child Depression Inventory, CDI (15); a Self-Perception Profile for Children, SPPC (12); a Impact of Event Scale, IES (11); a Premature Infant Profile, PIPP e a Posttraumatic Stress Disorder scale, PTSD (9); a Functional Independence Measure, FIM (8); a Self-Perception Profile for Adolescents, SPPA (7); a Children's Perception of Interparental Conflict Scale, CPIC e a Multidimensional Scale of Perceived Social Support, MSPSS (6); a Beck Depression Inventory, BDI, a Comfort Scale, CS e a Visual Analogue Scale, VAS (5); a Rosenberg Self-Esteem Scale, RSES e a PSPP (3); a Neonatal Pain, Agitation and Sedation Scale, N-PASS (2) e a Neonatal Skin Risk Assessment Scale, NSRAS (1).

Figura 1 – Número de Artigos por escala de avaliação



Instrumento

Ao fazermos uma análise mais pormenorizada dos artigos no que concerne ao tipo de instrumento, verificamos que os artigos apresentam de forma clara e explícita o tipo de instrumento (escalas) utilizado pelos investigadores, o número de itens por que são compostas as escalas e também se estamos perante uma escala única (um só conjunto de itens) ou se as escalas são composta por subescalas (vários conjuntos de itens), se estes conjuntos são equilibrados numericamente ou não e o número de pontos da escala de medida (3, 4, 5, 7, 10 pontos).

Na maioria dos casos consegue-se perceber, sem recorrer a outro tipo de consulta, a composição dos instrumentos, das subescalas e assim saber se existem diferenças nos conjuntos de itens. No que diz respeito ao número de itens por escala, nos instrumentos analisados, existem instrumentos compostos por diferentes números de itens e estes oscilam entre quatro e cinquenta e um itens. Verifica-se também que podem ser compostos apenas por uma única escala ou até 10 subescalas, variando estas em número de subescalas e número de itens em cada uma delas, o que nem sempre permite caracterizá-las com alguma uniformidade. Relativamente à escala de medida, os instrumentos analisados neste estudo apresentavam diferentes escalas de medida que varia entre três e dez (c.f. Tabela I).

Tabela I – Artigos científicos por tipo de escala de avaliação/psicométrica

Escalas de Avaliação (Psicométricas)	Nº Itens	Sub-Escalas (nº de sub-escalas * nº de itens)	Escala de Medida	Total de Artigos
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

Amostra e Teste Estatísticos

A escolha da população que vamos inquirir está intimamente ligada ao objecto sob estudo, aos objectivos e ao método de investigação. Os estudos que têm uma lógica extensiva associada (usando o inquérito por questionário como instrumento) e como tal uma estratégia de investigação quantitativa exigem cuidados com a dimensão da amostra. Se alguma das amostras for de pequena dimensão (<30), devemos verificar a normalidade da variável dependente nessa amostra, através de um teste de normalidade (como exemplo o *Teste de Shapiro Wilks*) já que no caso contrário o Teorema do Limite Central que a distribuição da média amostral é aproximadamente normal.

Outros autores referem-se ao uso das «escalas», instrumentos compostos por várias questões/afirmações ou itens que têm como finalidade avaliar uma determinada variável ou característica²¹, como sendo um tipo de instrumento que ao exigir que sejam feitas transformações de variáveis, e aplicados alguns testes estatísticos, importa que o número de inquiridos seja mais elevado, no mínimo de cinco indivíduos por item e o ideal seria 10 indivíduos por item da escala (exemplo para a aplicação da escala SPPC com 36 itens deveria ser considerada uma amostra entre 180 e 360 e a STAIC com 40 itens deveria ser considerada uma dimensão de amostra de pelo menos 200 indivíduos sendo o ideal 400) tendo também em conta o tipo de testes estatísticos que vão ser aplicados e as variáveis independentes utilizadas. Ao analisarmos os testes estatísticos aplicados nos diferentes estudos, podemos dizer que dos oito artigos publicados em língua portuguesa, três apresentavam simultaneamente testes paramétricos e não paramétricos (t-Student, ANOVA, MANOVA, Mann Whitney-U Test) e cinco apresentavam apenas testes paramétricos. Os restantes 119 apresentam-se distribuídos da seguinte forma: seis usaram testes não paramétricos (Mann Whitney-U Test) e cento e treze utilizam testes paramétricos (t-Student, ANOVA, MANOVA, ANCOVA) sendo que em onze destes estudos foram utilizados simultaneamente testes não paramétricos.

Verificamos na literatura científica que alguns autores partem do princípio de que a análise de variância exige que sejam observados os pressupostos da normalidade (a variável dependente deve ter uma distribuição normal em todos os grupos) e da homocedasticidade (as variâncias devem ser iguais em todos os grupos), mas nem sempre os artigos científicos têm expresso este pressuposto. Na realidade, o que verificamos é que na maioria dos artigos analisados não aparece explícito se foram verificados os pressupostos das metodologias utilizadas e também não aparece qualquer justificação para a utilização dos testes paramétricos ou não paramétricos. Daí nem sempre se conseguir fazer uma avaliação rigorosa dos pressupostos enunciados pelo facto de estes estarem omissos nos artigos.

Como não podemos analisar todos os instrumentos, seleccionámos a escala mais usada nos artigos revistos, a State-Trait Anxiety Inventory for Children (STAIC)⁽²⁰⁻³⁹⁾, que aparece em vinte dos artigos analisados. Verificamos que esta é composta por duas subescalas com 20 itens cada, cumprindo o pressuposto de que as subescalas deveriam ter um número de itens semelhante. No entanto emerge na análise o incumprimento dos pressupostos relacionados com a dimensão da amostra (apenas em 30% dos estudos a amostra é superior a 200 indivíduos) e com os testes estatísticos aplicados não emerge a justificação do uso de testes paramétricos ou não paramétricos (são aplicados testes paramétricos em todos os estudos sendo que em 15% se usa simultaneamente testes não paramétricos).

A STAIC foi usada em quatro estudos experimentais, sendo que num dos estudos foram usados testes paramétricos e não paramétricos e a dimensão da amostra sob estudo era no grupo experimental $n=43$ e no grupo de controlo $n=20$; dois outros estudos apresentavam amostras com dimensões iguais para o grupo experimental e controlo ($n=30$ e $n=35$ respectivamente) e no quarto estudo os grupos eram $n=284$ e $n=220$ (experimental e controlo, respectivamente). Nos restantes dezasseis estudos verificou-se que o tamanho da amostra variou entre $n=20$ e $n=3467$ e que na maioria dos estudos foram usados testes paramétricos. Apenas em três estudos foram usados simultaneamente testes paramétricos e não paramétricos, num dos experimentais ($n=43$ e $n=20$, grupos experimental e de controlo) e em dois dos outros ($n=29$ e $n=2280$). O que nos permite dizer que os pressupostos dos mínimos da amostra não foram cumpridos na maioria dos estudos, tendo por base o pressuposto da dimensão da amostra (amostra de grande dimensão), o pressuposto do número de participante por item e o tipo de testes estatísticos aplicados. Como as amostras não apresentam as dimensões exigidas, nem apresentam as mesmas dimensões, o que pode levar a uma heterogeneidade das matrizes de covariância o que se poderá reflectir num aumento da probabilidade do erro de tipo I e o poder do teste estatístico pode diminuir⁽¹¹⁾. Quanto aos testes estatísticos aplicados não emerge qualquer justificação.

CONCLUSÕES

Nesta análise exploratória de dados, pretendeu-se identificar o tipo de técnicas estatísticas utilizadas pelos investigadores (artigos científicos ligados à área de saúde da criança, adolescente e jovem adulto com e sem deficiência) quando os instrumentos utilizados são compostos por variáveis ordinais e analisar se são cumpridos os pressupostos inerentes à utilização das técnicas estatísticas com este tipo de variáveis e com amostra de pequena e grande dimensão. Foram analisados 127 artigos que cumpriam os critérios de inclusão, tendo como primeira intenção de análise o tipo de instrumento (escala única ou dividida em subescalas) com escalas ordinais de 3, 4, 5, 7 e 10 pontos. Como segunda intenção de análise a dimensão da amostra e como terceira intenção de análise os testes estatísticos aplicados.

Da análise exploratória dos artigos, concluímos que muitos não trazem expresso alguns dos pressupostos que gostaríamos de analisar em profundidade ou porque os autores não os incluíram no artigo ou poderão não ter sido trabalhados (ex: distribuição, homogeneidade das variâncias, transformação das variáveis para poderem ser tratadas como contínuas). No que diz respeito à normalidade da distribuição das variáveis nem todos os artigos mostram se foram aplicados testes nem quais (alguns dizem que foi aplicado o *Kolmogorov-Smirnov*), o mesmo acontecendo para a homogeneidade das variâncias (alguns referem o teste de

Levene). Desta forma não fica claro para o leitor o que esteve na origem da tomada de decisão de se optar por testes paramétricos ou não paramétricos.

O facto de estarmos, em alguns dos estudos analisados, perante amostras de pequena dimensão onde são usadas estatísticas multivariadas, poderia dizer-se que os investigadores deveriam ter usado a inferência não paramétrica, visto esta não exigir a especificação de um modelo populacional, ao contrário do que acontece na inferência paramétrica e porque em alguns dos casos não se verificam os pressupostos referidos anteriormente. Mas isso não se verifica também, nestes estudos, pois na sua maioria são usados testes paramétricos. Os investigadores optaram por aplicar testes paramétricos baseando-se, em estudos de simulação que têm sugerido que alguns testes não paramétricos são tão sensíveis quanto os paramétricos à violação de algumas condições como a homocedasticidade na inflação dos erros tipo I e tipo II optando por isso, por testes paramétricos devido à sua robustez. Mas também não há uma explicitação clara nos artigos analisados.

Para alguns investigadores que trabalham nestas áreas onde nem sempre é fácil medir, pois muitas vezes não se consegue ter a certeza de que o instrumento está a medir, o que é, suposto medir, de forma válida, fiável e sensível, torna-se importante a valorização do que tem sido investigado com este tipo de variáveis para que seja dada consistência à análise de dados desta natureza e assim podermos continuar a investigar no sentido de contribuirmos com evidência científica para que o rigor esteja em crescendo e os artigos científicos apresentem de forma expressa todos os pressupostos analisados.

REFERÊNCIAS

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]. Disponível 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, Asçi 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