

Análise de variáveis não diretamente observáveis: influência na tomada de decisão durante o processo de investigação

ANALYSIS OF VARIABLES THAT ARE NOT DIRECTLY OBSERVABLE: INFLUENCE ON DECISION-MAKING DURING THE RESEARCH PROCESS

ANÁLISIS DE VARIABLES NO DIRECTAMENTE OBSERVABLES
SU INFLUENCIA EN LA TOMA DE DECISIONS DURANTE EL PROCESO DE INVESTIGACIÓN

Maria Alice Santos Curado¹, Júlia Teles², João Marôco³

RESUMO

A dimensão da amostra, o tipo de variáveis, o seu formato de medida, a construção dos instrumentos de recolha de dados válidos e fiáveis, são aspectos a ter em consideração no processo de investigação. No âmbito das ciências sociais, da saúde e especificamente na área de enfermagem, os instrumentos de recolha de dados são muitas vezes compostos por variáveis componentes ou indicadores que dão origem a variáveis latentes ou não observáveis diretamente, daí a importância da decisão relativa à forma como são medidas (escala ordinal, *Likert* ou de tipo *Likert*). As escalas psicométricas são exemplos de instrumentos, pelo tipo de variáveis que as integram, que podem trazer problemas de medida e de análise estatística (testes paramétricos versus não paramétricos). Assim o investigador quando usa estas variáveis deve respeitar alguns pressupostos baseados em estudos de simulação ou em recomendações fundamentadas na evidência científica, de forma a tomar a melhor decisão.

DESCRIPTORIOS

Pesquisa em enfermagem
Coleta de dados
Medidas
Escalas
Psicometria

ABSTRACT

The sample dimension, types of variables, format used for measurement, and construction of instruments to collect valid and reliable data must be considered during the research process. In the social and health sciences, and more specifically in nursing, data-collection instruments are usually composed of latent variables or variables that cannot be directly observed. Such facts emphasize the importance of deciding how to measure study variables (using an ordinal scale or a Likert or Likert-type scale). Psychometric scales are examples of instruments that are affected by the type of variables that comprise them, which could cause problems with measurement and statistical analysis (parametric tests versus non-parametric tests). Hence, investigators using these variables must rely on suppositions based on simulation studies or recommendations based on scientific evidence in order to make the best decisions.

DESCRIPTORS

Nursing research
Data collection
Measures
Scales
Psychometrics

RESUMEN

El tamaño de la muestra, el tipo de variables, su medida y la construcción de instrumentos para la recogida de datos válidos y fiables son aspectos a considerar en el proceso de investigación. En el ámbito de las ciencias sociales, de la salud y particularmente en el área de enfermería, los instrumentos para la recogida de datos son muchas veces compuestos por variables componentes o indicadores que originan variables latentes o no observables directamente, lo que muestra la importancia de decidir cuidadosamente cómo se miden (escala ordinal, Likert o de tipo Likert). Las escalas psicométricas son ejemplos de instrumentos, por el tipo de variables que lo componen, que pueden traer problemas de medición y de análisis estadístico (test paramétricos versus no paramétricos). Así, el investigador cuando usa estas variables, debe respetar algunos supuestos basados en estudios de simulación o recomendaciones basadas en la evidencia científica, lo que permite una mejor toma de decisiones.

DESCRIPTORIOS

Investigación en enfermería
Recolección de datos
Medidas
Escalas
Psicometría

¹ Professora Coordenadora, Escola Superior de Enfermagem de Lisboa, Portugal. acurado@esel.pt ² Professora Auxiliar, Faculdade de Motricidade Humana, Universidade Técnica de Lisboa, Portugal ³ Professor Associado, Instituto Universitário de Ciências Psicológicas, Sociais e da Vida, Lisboa, Portugal

INTRODUÇÃO

Ao longo do processo de investigação o investigador torna-se perito no seu objeto de estudo e nos métodos e técnicas que vai utilizar. Passa por vários estádios e obrigatoriamente depara-se com o conceito de variável e com o pressuposto da sua medição⁽¹⁾. Porém medir em saúde pode apresentar duas facetas, uma mais ligada à medida global da população (índices, taxas, etc.) e outra mais ligada à medida individual (seja ela objetiva ou subjetiva). Este tipo de medida objetiva e subjetiva emerge, respetivamente, de dois tipos de variáveis: as diretamente observáveis, relacionadas com características físicas e biológicas (e.g., peso, altura, temperatura corporal, ph, hemoglobina), e como tal mais fáceis de medir; e as não diretamente observáveis (e.g., desenvolvimento motor oral, dor, satisfação, bem-estar e saúde, capacidades de desempenho de atividades de vida diária, estresse, *burnout*)^(2,3) cuja medição é feita através das suas manifestações, e por isso mais difíceis de medir. A ideia subjacente ao conceito de medida destas variáveis passa pelo facto de ser possível avaliar, atitudes, comportamentos, angústias, opiniões sobre saúde e bem-estar físico e psicológico dos indivíduos através de um processo de autodescrição ou de heteroavaliação^(4,5), o que se tem revelado como uma mais-valia na avaliação da saúde individual. É neste contexto que emergem as *escalas de avaliação*⁽³⁾ compostas por variáveis qualitativas com um formato de medida ordinal, cuja transformação numérica (soma, média, etc...) permite estimar o construto latente que está subjacente, que é uma manifestação das variáveis (itens) presentes na escala^(2,3,6).

Um dos maiores problemas associado a estes instrumentos (e.g. testes psicológicos, escalas, inventários) e aos construtos que eles procuram medir é que, não havendo um factor de calibração consensual (*gold standard*), não se consegue ter a certeza de que, o instrumento está a medir de forma válida, fiável e sensível o que está no seu objetivo medir⁽⁷⁾.

O dilema de não haver um factor de calibração da medida de constructos psicológicos, sociológicos, de saúde, etc., torna-se assim único pelo fato de não se conseguir observar diretamente a variável latente⁽⁶⁾. Nas áreas em que se privilegia o uso de medidas quantitativas, medir características físicas torna-se mais acessível, pelo facto de existirem instrumentos padronizados que permitem fazer essas medições. Ao contrário, a medição de características ligadas ao comportamento humano tem sempre implícita a opinião do indivíduo e, como tal, pode levar a um acréscimo no erro de medida⁽⁷⁾.

A área das ciências da saúde e especificamente em enfermagem, os investigadores e os profissionais confrontam-se, demasiadas vezes, com problemas desta natureza, fazendo a sua discussão a dois níveis. Um ao nível da investigação empírica em que se discutem metodologias e análise de dados ordinais (instrumentos e

formato de medida das variáveis, amostras, testes estatísticos); e outro ao nível clínico, em que os profissionais de saúde têm dificuldade na escolha de instrumentos de observação e avaliação para usar nas suas práticas. Pois, ao pesquisarem a melhor evidência científica, no sentido de replicarem estudos ou de traduzirem e validarem estatisticamente *escalas de avaliação*, deparam-se com uma grande diversidade destes instrumentos, com diferentes formatos de medida e com diferentes formas de análise estatística, o que dificulta a sua tomada de decisão. Todos estes aspectos têm feito parte de um quotidiano e de uma vivência profissional e de investigação, em que muitas vezes é difícil determinar qual o melhor caminho, quer na construção do projeto de investigação, quer quando se selecionam ou constroem os instrumentos de recolha de dados e se planeia a sua análise.

Acerca desta temática, as discussões e as questões são muitas e, por vezes, de difícil resposta. Destacamos algumas das questões mais comuns com que nos deparamos: O formato de medida dos itens deve ter um número par ou ímpar? Se for ímpar, há o risco dos respondentes escolherem frequentemente o ponto neutral (ancoragem)? Quantas classes de medida ou qual o formato de medida por item (3,4,5,6,7,...10)? O formato de medida tem implicações nos testes e análises estatísticas? Se o instrumento for constituído por subescalas, é necessário estar atento ao número de itens por subescala (equilíbrio numérico dos itens que constituem subescalas ou dimensões)? Podem ser feitos, por exemplo, cálculos de médias e desvios padrão (item a item ou somente nos itens somados)? Devemos usar estatística univariada ou multivariada? Testes paramétricos ou não paramétricos? Aquando da utilização dos testes devem ser tidos em conta os seus pressupostos? A robustez dos testes tem um papel importante na escolha do teste a utilizar? A dimensão da amostra influencia a opção por um determinado tipo de teste?

Algumas respostas; às questões enunciadas, emergem de forma clara na literatura científica, outras continuam a levantar algumas dúvidas e necessitam de mais investigação (e.g. simulação estatística com variáveis ordinais). Assim com este trabalho, pretendemos analisar alguns aspetos a ter em conta quando se utilizam de *escalas de avaliação* compostos por variáveis não diretamente observáveis (formato de medida ordinal) e alertar para a importância da dimensão da amostra e dos testes estatísticos a usar com este tipo de variáveis.

As variáveis e a medida

A variável é um receptáculo de características, qualidades ou quantidades que de alguma forma transporta a informação de um determinado fenómeno descritivo. A informação que é fornecida pelas variáveis em estudo é fundamental para o investigador e para o analista de dados. Porém essa informação e a sua qualidade irão depender da forma como as variáveis são quantificadas e da qualidade de sua mensuração, e mais especificamente do erro experimental que lhe esteja associado⁽²⁻³⁾.

São vários os autores que classificam as variáveis, em função da informação que transportam, como quantitativas ou qualitativas. As variáveis quantitativas são aquelas cujo formato de medida permite a ordenação e a quantificação de diferenças entre elas, podendo ser intervalares ou de razão. As variáveis intervalares ou consideradas como tal (e.g., Quociente de Inteligência, Psi 20) assumem valores quantitativos, assim como as variáveis de razão (e.g., comprimento do braço, altura, perímetro cefálico, peso) que diferem das anteriores por possuírem zero absoluto⁽²⁻⁶⁾. As variáveis qualitativas são medidas em escalas que indicam a presença de categorias de classificação discreta dos dados, exaustivas e mutuamente exclusivas. Podem ser nominais (e.g., gênero, estado civil) e ordinais (e.g., escalão de risco, nível de satisfação). Nas escalas de medida ordinal, as variáveis (itens) são medidas em classes discretas, entre as quais se verifica uma ordem (apresentam uma relação descritível mas não quantificável)⁽²⁻³⁾.

Medir é um processo de observação e registro de informação sobre a forma de atributos que reflete as qualidades ou as quantidades⁽⁷⁾, ou seja, atribuem-se números a objetos ou indivíduos seguindo um determinado conjunto de regras. Porém a mensuração de dimensões físicas e químicas em que se usam instrumentos com calibração consensual (e.g., balança, pipeta, termómetro) é diferente de medir sem padrão de calibração, como acontece muitas vezes em determinadas áreas do conhecimento.

As ciências exatas trabalham com variáveis observáveis e manipuláveis directamente ou acontecimentos manifestos, enquanto as ciências sociais e humanas e da saúde usam, muitas vezes, variáveis que não são observáveis nem manipuláveis directamente⁽⁶⁾, as chamadas variáveis latentes. A construção de instrumentos com variáveis desta natureza tem sido usada há longa data com o objectivo de medir grandezas que não são mensuráveis directamente. Uma das referências mais antigas é a escala de avaliação do *brilho das estrelas* (escala com seis pontos usada por Hiparco, 150 anos AC)⁽⁸⁾.

A utilização destas escalas teve uma grande expansão ao longo do século XX, com grande aceitação nas ciências sociais e humanas e também em áreas onde, apesar de se usarem mais variáveis quantitativas, se usam também este tipo de instrumentos com variáveis qualitativas que funcionam independentemente na investigação, ou como complemento às variáveis quantitativas. Algumas das escalas mais conhecidas são a Escalas de Likert⁽⁹⁾, a Escala de Thurstone⁽¹⁰⁾ e Escala de Guttman⁽¹¹⁾. Construir instrumentos com variáveis qualitativas, cujo objecto de mensuração não é directamente observável, dificulta o trabalho do investigador e a dificuldade continua aquando da análise de dados.

Rensis Likert foi um dos investigadores que trabalhou de forma sistematizada com este tipo de variáveis⁽⁹⁾. A metodologia de Likert é uma das mais usadas nos mais diversos campos da investigação e particularmente na área da psicologia, saúde e educação médica. Os estudos deste

autor⁽⁹⁾ preconizam um método específico de construção de escalas que utilizam afirmações que permitem que pessoas com diferentes opiniões e diferentes pontos de vista respondam diferenciadamente. Na construção deste tipo de escalas de medida por item, o autor considerou a utilização de um número ímpar de pontos por item, em que o ponto central é considerado neutro e os extremos são opostos e simétricos. Delas emergiram outro tipo de escalas de medida, chamadas de tipo Likert, em que apesar de os itens possuírem um formato de medida ordinal, ou não existe ponto neutro (ponto central) ou os extremos não são opostos nem simétricos⁽⁷⁾. Após estudos de fiabilidade e análise dos diferentes itens, o autor sugeriu que a medida de atitude, comportamento ou outra, fosse o resultado da soma das valorações dos itens elegidos (escalas adicionadas)⁽⁹⁾.

A *Escala de Thurstone* é constituída por itens com diferentes pesos, em que o respondente deve manifestar o seu acordo ou desacordo e a atitude do sujeito é medida tendo por base a média ponderada dos itens em acordo⁽¹⁰⁾. A *Escala Cumulativa de Guttman* é também constituída por itens, em que o respondente tem que dar o seu acordo ou desacordo, mas cuja organização é feita de forma hierárquica, pelo que a construção do instrumento exige que a ordenação dos itens seja efectuada do menos favorável para o mais favorável. Assim se o respondente concorda com um item, isso implica que está em concordância com os anteriores⁽¹¹⁾. Estes dois tipos de escalas, apesar de apresentarem diferenças relativamente às de Likert, endereçam o mesmo problema, da mensuração de variáveis latentes, não observáveis directamente.

A construção de escalas de avaliação, psicométricas ou sociométricas, começou a ser feita por investigadores da área das ciências sociais, sobretudo da psicologia, sendo depois também apropriada pelas ciências da saúde. Esses instrumentos⁽⁹⁻¹²⁾, têm gerado grande discussão, sobretudo nas questões ligadas à forma de medir as variáveis. Assim, quando se pretende analisar construtos recorrendo a este tipo de escalas é fundamental definir a base conceptual que embasa o constructo e o suporte empírico que a escala ou instrumento de medida dá ao construto⁽⁷⁾. Pois a operacionalização do conceito passa pela especificação do mesmo nas suas várias componentes (indicadores), uma vez que o conceito corresponde mais a um conjunto complexo de fenómenos do que a um fenómeno simples e directamente observável. Os indicadores (variáveis ou itens) são chamados de conceitos operacionais e classificatórios que assumem vários valores, passíveis de ser medidos (formato de medida ordinal) (e.g. conceito de burnout, avaliado pela Maslach Burnout Inventory que permite observar a resposta prolongada, no tempo, ao estresse e que é composta quinze itens ordinais que refletem estados emocionais e sentimentos, organizados em três dimensões compostas por 5,4,6 itens respetivamente com um formato de medida ordinal (com 7 classes) e que somados permitem avaliar as dimensões: exaustão emocional, despersonalização, redução da realização pessoal).

Os investigadores que trabalham com variáveis, com formato de medida ordinal, *Likert* ou de tipo *Likert*, muitas vezes analisadas como se fossem quantitativas, sabem da controvérsia⁽¹³⁾ que o seu tratamento estatístico suscita. Estes autores partem do princípio que estas variáveis que compõem as escalas usadas na área da saúde e da enfermagem, na psicologia e sociologia originam muitas vezes scores, que são tratados como variáveis latentes do tipo intervalar⁽²⁻⁶⁾. Estas variáveis latentes, por vezes, apresentam problemas com as qualidades métricas da medida pois, por um lado, podem não medir o constructo que era suposto medir (isto é, não ter validade) e, por outro lado, o formato de medida pode ser pouco consistente (pouco fiável), podendo levar a um aumento dos erros na análise dos dados. No entanto, desde que feita usando os itens adicionados, sua análise possibilita tratá-las como intervalares⁽³⁾, o que é justificado empiricamente, por estudos de simulação⁽¹⁴⁻¹⁵⁾, e conceptualmente:

Likert scales (collections of items) as opposed to individual Likert items are not ordinal in character, but rather are interval in nature and, thus, may be analyzed parametrically with all the associated benefits and power of these higher levels of analyses⁽¹⁵⁾.

Apesar da controvérsia relacionada com a análise dessas variáveis como se fossem quantitativas, pois se em teoria elas não o são, na prática elas são tratadas como se o fossem. Existem estudos que defendem a sua utilização, justificando que os resultados são de confiança desde que se usem escalas com classes de medida de pelo menos cinco pontos e que a distribuição das classes seja próxima da distribuição normal^(6,16). As escalas obtidas por meio da soma ou da média destas variáveis são tratadas como variáveis de tipo intervalar, possibilitando o uso de testes paramétricos e não paramétricos:

It is, therefore, as the intervalist contend, perfectly appropriate to summarize the ratings generated from Likert scales using means and standard deviations, and it is perfectly appropriate to use parametric techniques like Analysis of Variance to analyze Likert scales⁽¹⁵⁾.

Assim podemos dizer que a transformação das variáveis manifestas (soma, média dos itens,...) que irão permitir estimar o construto (variável latente) que a escala procura medir (operacionalizar) emerge no centro da discussão.

Os investigadores mais tradicionais não se revêem na transformação de itens ordinais para obter uma medida do constructo. Na falta de uma medida precisa (régua), podemos operacionalizar uma medida contínua (linha com 100 mm) usando uma escala ordinal com sete pontos ancorados como formato de resposta para afirmações de atitude ou frases derivadas de um diferencial semântico⁽¹⁷⁾. Produzem-se assim dados que empiricamente são conjuntos de itens sujeitos a transformação lineares e em carácter intervalares, de forma a podermos operacionalizar a medida⁽¹⁸⁻²²⁾. Como tal podem ser analisados através de estatística paramétrica desde que sejam verificados os

pressupostos deste tipo de testes, com todos os benefícios do uso destes pelo facto de serem considerados mais potentes⁽¹⁵⁻¹⁶⁾.

Contrastando com o que vem sendo exposto, em 2004, o assunto foi abordado num artigo no *Journal of Social Sciences* cuja ênfase centrava-se em abusos na utilização de instrumentos com escalas de Likert e na escolha de metodologia para análise de resultados. *The response categories in Likert scales have a rank order, but the intervals between values cannot be presumed equal*⁽²³⁾. Esta afirmação, porém, é contestada por outros autores⁽¹⁷⁻²¹⁾ que referem-se a estes aspetos no artigo e à literatura consultada como estando impregnados de mal-entendidos e erros grosseiros que advêm, de uma informação baseada em mitos, não verdades e erros conceptuais acerca das escalas de Likert. Quando se parte de uma premissa errada e não se entende ou ainda não se está familiarizado com as fontes primárias dificilmente se chega a uma teoria^(17,21). *Historically, there has been debate between those who maintain the ordinalist (rank order) and intervalist views in Likert scales*⁽²³⁾.

No que concerne à construção dos instrumentos, as escalas de avaliação, aparecem na literatura científica de diferentes formas, apresentando um número de classes de medida (formato de medida dos itens) que oscilam entre dois e dez pontos. No caso dos instrumentos compostos por um conjunto de subescalas, estas nem sempre apresentam uma construção homogênea, no que concerne ao número de itens que as compõem, encontramos na literatura subescalas que agrupam mais ou menos o mesmo número de itens (e.g., um instrumento com duas subescalas com vinte itens cada)⁽²³⁾ ou um instrumento com cinco dimensões, com 2,3,4,5,7 itens respetivamente⁽²⁴⁾.

A construção deste tipo de instrumentos deve seguir alguns fundamentos no que concerne à qualidade e a quantidade dos itens (variáveis). Quanto ao formato de medida, a preocupação não deve estar centrada na questão do par ou ímpar, mas sim no número de classes por item. Se as opções forem entre dois (e.g. 0=Não e 1=Sim), três (e.g. 1=Insatisfeito, 2=Nem satisfeito nem insatisfeito, 3=Satisfeito), cinco (e.g. 1=Discordo totalmente, 2=Discordo, 3=Indeciso, 4=Concordo, 5=Concordo totalmente) e sete (e.g. 1=Nunca, 2=Quase nunca, 3=Algumas vezes, 4=Regularmente, 5=Bastantes vezes, 6=Quase Sempre, 7=Sempre), importa considerar dois aspectos. Um relacionado com o número de possibilidades que o respondente tem e o outro com a sensibilidade dos itens. Assim, quanto maior o número de classes do item, maior a possibilidade dada ao respondente, o que se vai reflectir na sensibilidade dos itens, isto é, na capacidade que o item apresenta para discriminar indivíduos estruturalmente diferentes^(7,25). Itens com maior número de classes geralmente são mais sensíveis e dão maior garantia de credibilidade aos resultados da análise estatística^(3,7,17).

Algumas escalas são compostas por subescalas ou dimensões. Nessas situações é aconselhável que cada subescala ou dimensão tenha entre cinco e dez itens e no mínimo deve conter três⁽⁷⁾. O não cumprimento deste pressuposto pode levar a problemas relacionados com a fiabilidade (a escala mede o construto de forma consistente e reproduzível, sujeitos com a mesma característica, ou o mesmo sujeito em momentos equivalentes, apresentam o mesmo valor na medida) e com a validade (se a escala mede o que era presumível medir)^(5,7,26).

A validade reporta-se a, diferentes aspectos como: 1) a validade de conteúdo (traduz o grau de concordância de um painel de especialistas e avalia se os itens são representativos do domínio que a escala vai avaliar); 2) a validade de constructo (i.e. mede o que se pretende medir), que pode ser convergente (os itens que fazem parte do construto estão correlacionados), discriminante (os itens não se correlacionam com outros que medem outro construto), de critério (a operacionalização de um construto está de acordo com um critério previamente estabelecido)^(5,7,16). Assim, a heterogeneidade na construção dos instrumentos, também pode trazer problemas relativos à validade e à fiabilidade da medida que se está a tentar operacionalizar.

O alfa de Cronbach é usado com alguma frequência pelos investigadores para estimar a fiabilidade (consistência interna dos itens). No entanto, pode apresentar valores elevados quando afetados por determinados fatores como: o número de itens da escala (quanto maior o seu número, em princípio, maior o valor do alfa); a variabilidade intra e inter-itens (quanto menor a variabilidade das respostas intra-sujeito e maior a variabilidade das respostas inter-sujeito, maior o valor do alfa); a homogeneidade de variâncias inter-itens (quando existe homogeneidade de variâncias inter-itens, maior o valor do alfa)^(3,27); e a dimensão da amostra (quanto maior a dimensão da amostra maior o valor do alfa). Alguns desses aspectos exigem que antes de calcular o alfa, standardizem-se as observações usando as correlações médias entre itens (covariâncias standardizadas) para corrigir a sobrestimação do alfa provocada pelas covariâncias heterogêneas inter-itens⁽²⁷⁾.

Dimensão da amostra

O processo de investigação passa por diferentes estádios, dos quais já salientámos a importância das variáveis e, neste caso específico, das variáveis não observáveis directamente e da forma como são medidas. Importa ainda discutir a questão de «onde vão ser medidas», ou seja, se vão ser medidas na população ou numa amostra dessa população e de que forma vão ser medidas. Trabalhar com a população teórica (todos os elementos, casos, eventos, objetos, indivíduos, agregáveis e sobre os quais queremos generalizar) é quase impossível em investigação por isso se recorre a grupos mais restritos (população acessível, população de estudo), dos quais vamos determinar a amostra (subconjunto da população que vai ser usado no estudo para, seja representativa da população)⁽²⁾.

Em ciências da saúde e de enfermagem, é muito difícil (e em algumas situações impossível) aceder a populações, sendo mais fácil usar uma amostra, a qual pode ser representativa da população (se for utilizado algum tipo de amostragem probabilística) ou pode não representar convenientemente as características da população (se for utilizado uma amostragem não probabilística)⁽²⁾.

No domínio em análise, é de relevante importância a dimensão da amostra. Este é um dos problemas mais discutidos na literatura científica e em discussões entre investigadores. Quando se quer fazer inferência estatística, dependendo do tipo de testes estatísticos que vão ser usados, existem pressupostos aos quais o investigador deve obedecer, de modo a não comprometer a validade dos resultados obtidos. Porém, o tamanho da amostra pode ser uma questão difícil de contornar, sobretudo em áreas como a da saúde, em que muitas vezes o número de indivíduos é muito pequeno (e.g., doenças raras, situações de saúde específicas).

Nesse caso, há investigadores que optam por obter uma amostra, com uma dimensão tão grande quanto possível, dentro do número de indivíduos disponíveis, ficando muitas vezes, mesmo assim, perante uma amostra de dimensão tão pequena que não podem fazer uso das técnicas multivariadas. Outros optam por determinar a dimensão mínima da amostra necessária para efetuar uma análise estatística adequada, usando para tal *regras de polegar (rules of thumb)*⁽³⁾. Estas são baseadas na experiência da investigação e têm como pressuposto de base a complexidade das técnicas de análise (univariadas, bivariadas ou multivariadas). Há os que fundamentam a sua decisão em estudos de análise estatística da potência dos testes de modo a garantir que a amostra tem dimensão adequada⁽³⁾.

Quando estamos a trabalhar com instrumentos, como os que temos vindo a discutir, se as amostras são muito pequenas podem ocorrer erros de estimação⁽¹²⁾, sobretudo se a relação entre o número de sujeitos e o número de itens é baixa e também se a amostra não é representativa da população pretendida, o que se verifica na prática, em muitas situações. Assim, a dimensão da amostra deve garantir os objectivos e a qualidade da investigação, devendo o investigador respeitar algumas regras de polegar e adequar o tamanho da amostra segundo o teste estatístico que pretende aplicar, a potência pretendida para o teste, o tamanho do efeito (effect size) que se pretende detectar e o valor do nível de significância^(2,3).

O número de participantes está muitas vezes relacionado com o tipo de estudo e as metodologias de análise de dados. Se a análise passa por técnicas de análise estatística como a análise da regressão e a análise fatorial exploratória devem manter-se no mínimo cinco observações por variável no modelo^(3,6,25,30). Este pressuposto advém da necessidade de existência de variabilidade para estimar os parâmetros do modelo. No caso de estarmos perante um modelo de equações estruturais, onde os *dados* do modelo

são as variâncias-covariâncias não redundantes entre as variáveis, a dimensão da amostra deve ser maior. Para que haja garantia de variabilidade para estimar os parâmetros do modelo, o número de observações por cada variável manifesta deve ser entre dez e quinze^(3,6,12,29,30).

Nas ciências sociais e da saúde, os investigadores têm usado estatística não paramétrica em alternativa à estatística paramétrica, com variáveis ordinais, quando os pressupostos desta não se verificam (distribuição normal, homogeneidade de variâncias, independência das observações entre grupos). Contudo, estudos de simulação 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 de aplicação, o que se traduzem na inflação dos erros de tipo I e II, o que não favorece a opção de usar testes não paramétricos.

A ausência de recomendações claras relativamente ao uso da estatística paramétrica versus não paramétrica com variáveis ordinais, sobretudo no caso multivariado tem feito com que os investigadores invistam cada vez mais em estudos de simulação estatística. Alguns estudos salientam a importância da análise multivariada, quando a entidade a medir tem várias componentes conceptuais e o investigador pretender comparar simultaneamente grupos nessas componentes (e.g. Análise da variância multivariada). A MANOVA oferece vantagem sobre múltiplas ANOVAs, pois permite medir os vários ângulos do problema, aumentar a potência (em alguns casos), e diminuir o erro de tipo I⁽²⁾, mas exige o cumprimento de postulados como: (i) as observações devem ser independentes, (ii) o vetor de variáveis resposta deve ter distribuição normal multivariada para as diferentes populações, e (iii) as matrizes de covariâncias das populações devem ser homogêneas⁽²⁾, o que nem sempre se verifica na prática.

CONSIDERAÇÕES FINAIS

As considerações finais que iremos apresentar reportam a importância a dar às variáveis e sua medida e à dimensão da amostra quando trabalhamos com variáveis que não são observáveis directamente. Procuraremos também deixar algumas recomendações teóricas para serem usadas por jovens investigadores na área das ciências da saúde e, especificamente na enfermagem, e nas ciências sociais.

A análise dos aspetos teóricos e da investigação com variáveis não directamente observáveis permite-nos dizer que as tomadas de decisão devem ser ponderadas no sentido de diminuir a margem de erro, seja ele de medida ou de análise de dados. Pois a construção e aplicação de instrumentos compostos por estas variáveis é complexa e nem sempre compreendida.

Os estudos que seguem a teoria de Likert utilizam instrumentos cujas variáveis (itens) apresentam um número ímpar de classes de medida (cinco pontos) e

a medida dos itens não é feita individualmente mas é obtida pela soma das avaliações dos itens (score), justificando a denominação de *escalas adicionadas (sum-mated scales)*. Esses scores podem ser tratados estatisticamente como variáveis de tipo intervalar e, em sua análise, pode ser equacionada a utilização de testes paramétricos e não paramétricos.

A discussão sobre o número de classes da variável (item) ser par (e.g. 4, 6) ou ímpar (e.g. 5, 7, 9) é considerada irrelevante. O mesmo não pode ser dito quando se refere à sua classificação. Para que o formato de medida seja chamado de escala ou item de Likert, o seu número deve ser ímpar e o ponto central é um ponto neutro cujos extremos são opostos e simétricos. As escalas que usem um formato de medida sem ponto neutro e cujos extremos não são opostos nem simétricos devem ser identificadas como *do tipo Likert*, o que se torna importante, sob o ponto de vista do rigor, quando se escreve a metodologia de um trabalho de investigação.

Relativamente às dimensões que compõem a escala, a sua heterogeneidade pode originar problemas de fiabilidade e validade, pelo que as recomendações apontam para um mínimo de três itens por dimensão e o ideal seria entre cinco e vinte itens. Quanto ao número de classes por item, seja ele, par ou ímpar, alguns autores valorizam os instrumentos cujas variáveis apresentam um maior número de classes (cinco ou mais) em detrimento dos que apresentam três ou quatro classes, pois consideram que quanto maior for o número de classes, melhor a possibilidade de resposta dos participantes e melhor a qualidade da soma dos itens; o que vai refletir-se em sua sensibilidade e fiabilidade. Daí a importância de, quando da procura ou construção de um instrumento, ter em conta que o formato de medida tenha pelo menos cinco classes. Relativamente ao tipo de testes estatísticos, verificámos que a frequente utilização de testes paramétricos, mesmo quando não se cumprem as condições de aplicação, deve-se ao facto de alguns investigadores considerarem que ambos os testes são sensíveis à violação dos pressupostos e os paramétricos são mais potentes do que os não paramétricos.

Se a análise passa por técnicas usuais de análise estatística como a análise da regressão e a análise factorial exploratória devem manter-se no mínimo cinco observações por variável no modelo. Mas, se optarmos por uma análise de equações estruturais, em que os dados são as correlações (ou as variâncias e covariâncias) entre as variáveis manifestas, teremos que manter um número mínimo de dez a quinze observações por variável no modelo.

A evidência científica mostra que a dimensão da amostra influencia de forma determinante a análise de dados. Assim, ao reportar os resultados de uma investigação, em que foram usados testes de hipóteses, o investigador deve incluir, além do nível de significância, uma medida de dimensão de efeito, a potência do teste e, dependendo da

análise efetuada, intervalos de confiança para as estimativas dos parâmetros.

Como as amostras de pequena dimensão podem originar resultados estatisticamente não significativos em situações em que não têm significância prática e as de grande dimensão podem, ao contrário, originar resultados estatisticamente significativos, mesmo que sem significância prática, importa por isso que estejam expressos os resultados referentes quer à significância estatística quer à significância prática.

A presença de resultados opostos nestes dois tipos de significância pode dever-se à dimensão da amostra, ao erro de tipo I e à potência de teste. Assim, na fase de planeamento da investigação, os investigadores menos experientes devem ter atenção, por um lado, à escolha e a construção do instrumento (único ou com subescalas), ao número de itens (variáveis) que o compõem e ao formato de medida que lhe está associado. Por outro lado, devem estar atentos à dimensão da amostra (e.g. ter em atenção o número de observações por item) e decidir as técnicas estatísticas a utilizar com base nestes dois fatores.

Importa realçar que na área das ciências da saúde e especificamente na disciplina de enfermagem deve haver uma atenção redobrada no que respeita à dimensão da amostra, sobretudo quando estamos perante amostras muito pequenas (doenças raras).

Algumas das dúvidas iniciais vão continuar por esclarecer, pois trabalhar com variáveis não diretamente observáveis é um problema real e ainda há muito trabalho a desenvolver para que mais recomendações possam emergir da pesquisa. Fica por isso a sugestão para que os profissionais de saúde e outros profissionais que usam este tipo de variáveis em investigação e na prática de cuidados se envolvam em projetos de investigação conjuntos. Por um lado, para que equipas multiprofissionais investiguem e discutam estas problemáticas sob diferentes pontos de vista e contribuam para o esclarecimento das dúvidas que ainda persistem. Por outro lado, para que cada vez mais, os instrumentos usados nas práticas de cuidados sejam fundamentados na melhor evidência científica. Pois só através do conhecimento que emerge da investigação e da sua apropriação pelos profissionais, se faz a otimização dos resultados em saúde e se favorecem as *boas práticas*.

REFERÊNCIAS

1. Watty AD, Lecumberri López J. La importancia de medir. Vet Méx [Internet]. 1997 [citado 2013 mar. 15];28(1):69-72. Disponible en: <http://www.medigraphic.com/pdfs/vetmex/vm-1997/vm971n.pdf>
2. Marôco J. Análise estatística com o PASW (SPSS Statistics). Pêro Pinheiro: Report Number; 2010.
3. Hill MM, Hill A. Investigação por Questionário. Lisboa: Sílabo; 2009.
4. Agresti A. Categorical data analysis. New Jersey: John Wiley & Sons; 2002.
5. Anastasi A. Psychological testing. New York: Macmillan; 1990.
6. Marôco J. Análise de equações estruturais: fundamentos teóricos, software e aplicações. Pêro Pinheiro: Report Number; 2010.
7. Marôco J. Avaliação das qualidades psicométricas de uma escala. Lisboa: Manuscrito; 2009.
8. Lodge M. Magnitude scaling: quantitative measurement of opinions. Beverly Hills: Sage; 1981.
9. Likert R. A technique for the measurement of attitudes. Arch Psychol. 1932;22(140):1-50.
10. Thurstone LL. Attitudes can be measured. Am J Sociol. 1928;33(4):529-54.
11. Gutman L. The basis for Scalogram analysis. In: Stouffer SA. Measurement and prediction. New York: Wiley; 1950. v. 4.
12. Stevens SS. On the theory of scales of measurement. Science. 1946;103(2684):667-80.
13. Urdan TC. Statistics in plain English. London: Laurence Erlbaum Associates; 2005.
14. Glass GV, Peckham PD, Sanders JR. Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. Rev Educ Res. 1972; 42(3):237-88.
15. Carifio J, Perla R. Resolving the 50-year debate around using and misusing Likert scales. Med Educ. 2008;42(12):1150-2.
16. Worthington R, Whittaker T. Scale development research: a content analysis and recommendations for best practices. Couns Psychol. 2006;34(6):806-38.
17. Carifio J, Perla R. Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert Scales and Likert response formats and their antidotes. J Soc Sci. 2007;3(3):106-16.
18. Pell G. Use and misuse of Likert scales [letter]. Med Educ. 2005;39(9):970.
19. Gaito J. Measurement scales and statistics: resurgence of an old misconception. Psychol Bull. 1980;87(3):564-7.

-
20. Knapp TR. Treating ordinal scales as interval scales: an attempt to resolve the controversy. *Nurs Res.* 1990;39(2):121-3.
 21. Knapp TR. Treating ordinal scales as ordinal scales. *Nurs Res.* 1993;42(3):184-6.
 22. Wang ST, Yu ML, Wang CJ, Huang CC. Bridging the gap between the pros and cons in treating ordinal scales from an analysis point of view. *Nurs Res.* 1999;48(4):226-9.
 23. Jamieson S. Likert scales: how to (ab) use them. *Med Educ.* 2004;38(12):1217-8.
 24. Wilson FC. Analysis of intensive outpatient neuro-rehabilitation outcomes using FIM+ FAM (UK). *NeuroRehabilitation.* 2009;24(4):377-82.
 25. Gliem JA, Gliem RR. Calculating, interpreting, and reporting Cronbach's Alpha reliability coefficient for Likert-type scales. In: Paper Presented at the Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education; 2003; Columbus, OH. Columbus: Ohio State University; 2003.
 26. Pasquali L. Psychometrics. *Rev Esc Enferm USP [Internet].* 2009 [cited 2013 Mar 19]; 43(n.spe):992-9. Available from: http://www.scielo.br/pdf/reeusp/v43nspe/en_a02v43ns.pdf
 27. Marôco J, Garcia-Marques T. Qual a fiabilidade do alfa de Cronbach? Questões antigas e soluções modernas? *Lab Psicol.* 2006;4(1):65-90.
 28. Kahn J. Factor analysis in counseling psychology research, training, and practice: principles, advances and applications. *Couns Psychol.* 2006;34(5):684-718.
 29. Curado MAS, Teles J, Marôco J. Análise estatística de escalas ordinais: aplicações na área da saúde infantil e pediatria. *Enferm Global.* 2013;(30):446-57.
 30. Zimmerman DW. Statistical significance levels of nonparametric tests biased by heterogeneous variances of treatment groups. *J Gen Psychol.* 2000;127(4):354-64