



Corra que o survey vem aí. Noções básicas para cientistas sociais¹

Run, the survey is coming. Basics for social scientists

Ranulfo Paranhos, Dalson Britto Figueiredo Filho, Enivaldo Carvalho da Rocha y José Alexandre da Silva Junior

Resumen.

O que é, para que serve e como se faz uma pesquisa de *survey*? O principal objetivo desse trabalho é apresentar uma introdução à técnica de *survey* na pesquisa empírica em Ciências Sociais. O foco repousa sobre a compreensão intuitiva de noções básicas de *survey*, seleção da amostra, construção e aplicação dos questionários e análise dos resultados. Metodologicamente, sintetizamos as principais recomendações da literatura. Além disso, utilizamos dados do *Quality of Government Institute* para demonstrar como o teste de associação do qui-quadrado e a técnica de análise de componentes principais podem ser utilizados para analisar dados de *survey*.

Palavras-chaves: pesquisa de *survey*; questionários; métodos quantitativos.

Abstract.

What is, which it does and how we do a survey research? The principal aim of this paper is provide an introduction to survey technique in social science empirical research. The focus relies on the intuitive comprehension of basic survey concepts, sample selection, construction and application of the questionnaires and results analysis. On methodological grounds, we summarize some core literature recommendations. In addition, we employ *Quality of Government Institute* data to show how chi-square association test and principal component analysis can be used to analyze survey data.

Keywords: survey research; questionnaires; quantitative methods.

¹ Para os propósitos desse trabalho, o grau de complexidade matemática foi minimizado. Leitores interessados em abordagens mais sofisticadas devem seguir as referências bibliográficas. Muitos dos exemplos utilizados aqui foram retirados do curso Aspectos Cognitivos da Metodologia de *Survey*, ofertado pelo curso de Metodologia Quantitativa (MQ) da Universidade Federal de Minas Gerais (UFMG). Agradecemos à professora Solange Simões pelo dedicado treinamento. Eventuais imprecisões são integralmente creditadas aos autores. Essa pesquisa contou com aporte financeiro da CAPES e do CNPq.

Everyone takes surveys. Whoever makes a statement about human behavior has engaged in a survey of some sort.

Andrew Greeley

Introdução

Nosso principal objetivo é apresentar a lógica e as potencialidades da técnica de survey na pesquisa empírica em Ciências Sociais. Para Babbie (2005), “pesquisa de survey se refere a um tipo particular de pesquisa social empírica, mas há muitos tipos de survey. O termo pode incluir censos demográficos, pesquisas de opinião pública, pesquisas de mercado sobre preferências do consumidor, estudos acadêmicos sobre preconceito, estudos epidemiológicos, etc.” (Babbie, 2005: 95). O foco repousa sobre a compreensão intuitiva de noções básicas de survey, seleção da amostra, construção e aplicação de questionários e análise dos resultados. Nosso público alvo contempla não só estudantes de graduação, pós-graduação e pesquisadores em Ciências Sociais, mas também alunos de outras áreas e profissionais interessados em aplicar surveys. Metodologicamente, sintetizamos as principais recomendações da literatura especializada. Além disso, utilizamos dados do Quality of Government Institute para ilustrar as possibilidades de análise empírica dos dados, enfatizando o papel de duas diferentes técnicas: (1) teste de associação do qui-quadrado e (2) análise de componentes principais.

Mas por que um artigo sobre survey? De acordo com Brady (2000),

surveys were used in about 10% of the articles published between 1991 and 1995 in the American Political Science Review and in about 15% of those published in the American Journal of Political Science. No other method for understanding politics is used more, no other method has so consistently illuminated political science theories with political facts (Brady, 2000: 47).

Dada as limitações espaciais inerentes a um artigo, não é possível contemplar todos os aspectos relevantes do planejamento e execução de uma pesquisa de *survey*.² No entanto, dada a importância do assunto e a escassez de literatura mais aplicada em formato de artigo, é oportuno apresentar os fundamentos dessa técnica para que estudantes de graduação, pós-graduação e pesquisadores possam aplicá-la em seus desenhos de pesquisa.

O restante do artigo está dividido da seguinte forma: a próxima seção apresenta a lógica da inferência estatística, enfatizando as vantagens de se trabalhar com amostras. A segunda parte descreve os principais cuidados que os cientistas sociais devem ter na formulação do questionário (instrumento). Depois disso, apresentamos exemplos práticos de como o teste de associação de qui-quadrado e a técnica de análise de componentes principais podem ser utilizadas para analisar dados de *survey*. Por fim, a última seção sumariza nossas principais conclusões.

2. Amostras,³ aleatoriedade e a lógica da inferência.

² Por exemplo, não discutiremos as vantagens e limitações associadas aos diferentes tipos de amostras. Similarmente, iremos negligenciar o debate sobre *complex sample designs*. Também não abordaremos a função dos pesos para corrigir problemas de representatividade nem o efeito de diferentes desenhos amostrais sobre a estimação da variância amostral. Leitores interessados nesses aspectos devem consultar Lee e Forthofer (2006), Levy e Stolte (2000) e Kiecolt e Nathan (1985).

³ Quando o tamanho da população é maior do que 100.000 deve-se utilizar a fórmula para populações infinitas:

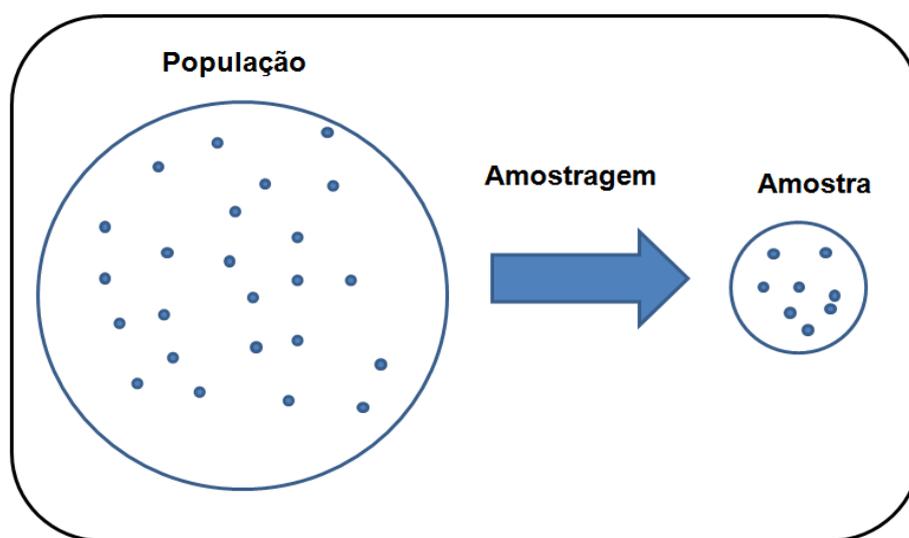
$$n = \frac{\sigma^2 * p * q}{E^2}$$

Em que n é o tamanho da amostra, σ representa o nível de confiança escolhido pelo pesquisador (em geral 90%, 95% ou 99%), p representa a proporção das características pesquisadas no universo (%), q representa a

Existem duas principais razões para utilizar amostras na pesquisa científica: (1) economia de tempo e (2) economia de recursos. Em geral, catalogar informações sobre o universo pode inviabilizar alguns desenhos de pesquisa. O conceito de população diz respeito à totalidade de indivíduos/unidades, enquanto que a amostra refere-se a uma parte da população. Por exemplo, suponha que uma pesquisa tem como objetivo examinar a intenção de voto em um determinado candidato à Presidência do Brasil. Logisticamente, não faz sentido entrevistar todos os eleitores brasileiros. A pesquisa seria demasiadamente onerosa e demorada. Além disso, o esforço computacional necessário para trabalhar com amostras é menor do que aquele empregado para analisar grandes bases de dados. Dessa forma, sempre que as amostras forem corretamente selecionadas, as inferências produzidas serão precisas, confiáveis e detalhadas.

Para Moore e McCabe (2003), “the idea of sampling is to study a part in order to gain information about the whole” (Moore e McCabe, 2003: 225). Para King, Keohane e Verba (1994), “inference is the process of using the facts we know to learn about facts we do not know. The facts we do not know are the subjects of our research questions, theories, and hypotheses. The facts we do know are our (quantitative or qualitative) data or observations” (King, Keohane e Verba, 1994: 46). Em particular, a inferência estatística é uma ferramenta essencial ao desenvolvimento do conhecimento científico. Independente da área de estudo, é exatamente a inferência estatística que permite utilizar informações limitadas sobre os fatos/fenômenos conhecidos para fazer inferências válidas a respeito de fatos/fenômenos desconhecidos. Na pesquisa de *survey*, a utilização da inferência estatística é fundamental para explorar questões de pesquisa pouco conhecidas, descrever fenômenos, testar hipóteses teoricamente orientadas e, até mesmo, fazer previsões. No entanto, para que as conclusões realizadas a partir de dados amostrais sejam representativas dos parâmetros populacionais é necessário obedecer alguns critérios. A figura abaixo ilustra a relação entre população, amostragem e amostra.

Figura 1: População, amostragem e amostra⁴



proporção do universo que não possui a característica de interesse ($q = 1 - p$) e E representa o erro de estimação permitido (em geral, 2%, 3% ou 4%).

Quando o tamanho da população é menor do que 100.000 deve-se utilizar a fórmula para populações finitas:

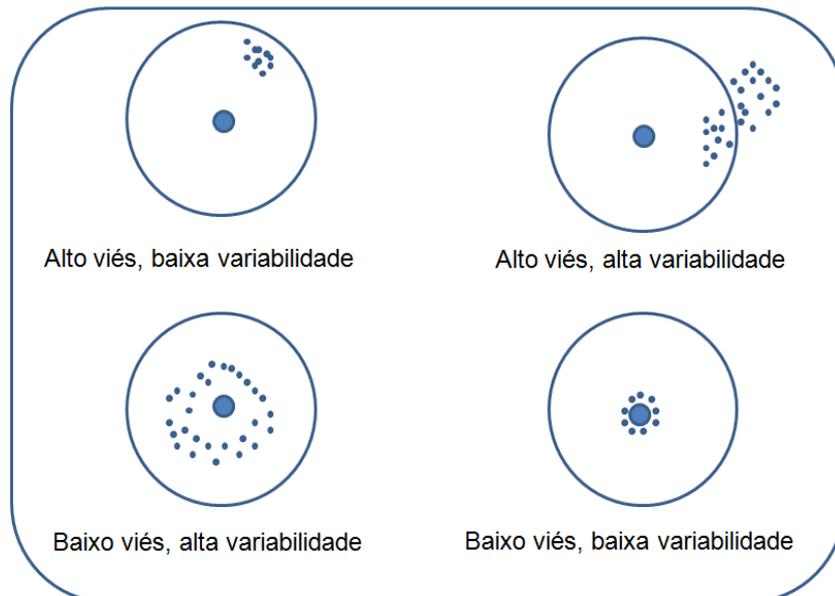
$$n = \frac{\sigma^2 * p * q * N}{E^2 (N - 1) + \sigma^2 * p * q}$$

Em que n é o tamanho da amostra, σ representa o nível de confiança escolhido pelo pesquisador (em geral 90%, 95% ou 99%), p representa a proporção das características pesquisadas no universo (%), q representa a proporção do universo que não possui a característica de interesse ($q = 1 - p$), N representa o tamanho da população e E representa o erro de estimação permitido (em geral, 2%, 3% ou 4%).

⁴ Uma maneira intuitiva de pensar sobre o conceito de amostra é imaginar uma colher retirada de um prato de sopa. Não é necessário tomar toda a sopa para saber se ela está boa de sal, temperada ou consistente. Toda a informação para fazer uma inferência válida para a população (sopa) está contida na própria amostra retirada pela colher.

Moore e McCabe (2003) afirmam que “a simple random sample (SRS) of size n consists of n individuals from the population chosen in such a way that every set of n individuals has an equal chance to be in the sample actually selected” (Moore e McCabe, 2003: 250).⁵ Para que as estimativas amostrais sejam representativas dos parâmetros populacionais, é necessário garantir a aleatoriedade da amostra. Apenas amostras aleatórias garantem que o princípio da equiprobabilidade será obedecido, ou seja, todos os indivíduos da população terão a mesma chance de participar da amostra. Uma vantagem fundamental da seleção aleatória dos casos é garantir a qualidade das estimativas. Qualidade no sentido de assegurar não-viesamento e baixa variabilidade. A figura abaixo ilustra as qualidades de uma boa estimativa.

Figura 2: Qualidades de uma boa estimativa



Uma estimativa é não-viesada quando ela nem sobrestima nem subestima sistematicamente o valor do parâmetro populacional. A eficiência diz respeito à variabilidade da estimativa. Quanto maior a variabilidade, menor a precisão. A melhor estimativa é aquela não tendenciosa e com a menor variabilidade possível (figura do canto inferior direito).

3. Por que fazer um survey?⁶

Para Babbie (2005),

há provavelmente tantas razões diferentes para se fazer *surveys* quanto há *surveys*. Um político pode encomendar um visando a sua eleição. Uma empresa de marketing pode fazer um *survey* visando vender mais sabonetes de marca X. Um governo pode fazer um *survey* para projetar um sistema de trânsito de massa ou para modificar um programa de bem-estar social (Babbie, 2005: 95).

Ou seja, a pesquisa de *survey* tem múltiplas finalidades. Sempre que o pesquisador estiver interessado em identificar opiniões, atitudes, valores, percepções, etc., ele pode empregar o *survey* como técnica de coleta de dados. Comparativamente, é durante o período eleitoral que esse tipo de pesquisa ganha maior visibilidade. Mas nada impede que o *survey* seja utilizado para auxiliar gestores governamentais na formulação de políticas públicas e informar o planejamento estratégico

⁵ Para Babbie (2005), “a amostragem aleatória simples [AAS] é o método de amostragem básico suposto pelos cálculos estatísticos do *survey*” (Babbie, 2005: 135).

⁶ Para uma introdução à metodologia de *survey* ver Converse e Presser (1986). Para uma abordagem mais avançada ver Saris e Gallhofer (2007) e Fowler (1995). Reproduzimos aqui as principais recomendações de Babbie (2005), além de incluir nas notas de rodapé sugestões de literatura sobre o assunto.

de empresas. Academicamente, a pesquisa de *survey* tem três principais funções: (1) exploração; (2) descrição; e (3) explicação.

A pesquisa exploratória é realizada na ausência de teoria sobre o que se deseja investigar. Ela é particularmente adequada quando o pesquisador não tem uma questão de pesquisa bem definida. A principal consequência disso é a ausência de hipóteses teoricamente orientadas. A perspectiva exploratória pode identificar padrões empiricamente inesperados, o que em um segundo momento pode informar o refinamento do desenho de pesquisa. Para Jaeger e Halliday (1998), “*exploratory approaches to research can be used to generate hypothesis that latter can be tested with confirmatory approaches (...) the end goal of exploratory research, though, is to gain new insights, from which new hypotheses might be developed*” (Jaeger e Halliday, 1998: 564).

Por sua vez, a perspectiva descritiva procura literalmente descrever as principais características de um determinado fenômeno. Para Babbie (2005),

surveys são frequentemente realizados para permitir enunciados descritivos sobre alguma população, isto é, descobrir a distribuição de certos traços e atributos. Nestes, o pesquisador não se preocupa com o porquê da distribuição observada existir, mas com o que ela é (Babbie, 2005: 96).

Por fim, a pesquisa de *survey* pode adotar uma perspectiva explicativa/confirmatória.⁷ Aqui o pesquisador está interessado em testar hipóteses teoricamente orientadas. A partir da literatura sobre o assunto, deve-se identificar uma questão de pesquisa. Depois disso, o pesquisador deve propor a hipótese de trabalho.

Em relação aos tipos de *survey*, a literatura identifica dois principais grupos de estudos: (1) transversais e (2) longitudinais.⁸ Em um *survey* do tipo transversal “os dados são colhidos, num certo momento, de uma amostra selecionada para descrever alguma população maior na mesma ocasião. Tal *survey* pode ser usado não só para descrever, mas também para determinar relações entre variáveis” (Babbie, 2005: 101). Os estudos longitudinais, por sua vez, permitem que o pesquisador examine um determinado fenômeno ao longo do tempo. Existem três principais formas de implementar um estudo longitudinal: (1) estudos de tendência; (2) estudos de coorte⁹ e (3) estudos de painel.

No estudo de tendência “uma população pode ser amostrada e estudada em ocasiões diferentes. Ainda que pessoas diferentes sejam estudadas em cada *survey*, cada amostra representa a mesma população” (Babbie, 2005: 102). O exemplo típico dessa opção metodológica são as pesquisas eleitorais. Várias pesquisas são realizadas ao longo do período eleitoral com o objetivo de descrever as preferências da população de eleitores. A cada nova pesquisa, uma nova amostra é selecionada, mas a representatividade da amostra em relação à população é mantida.

Além dos estudos de tendência, existem ainda os estudos de coorte. Para Babbie (2005), “um estudo de coorte focaliza a mesma população específica cada vez que os dados são coletados, embora as amostras estudadas possam ser diferentes” (Babbie, 2005: 102). Por exemplo, em um estudo sobre egressos, suponha que o pesquisador deseja investigar a coorte de formandos do curso de Ciência Política da Universidade Federal de Pernambuco em 2013. Todos os casos da

⁷ Para Jaeger e Halliday (1998), “*with confirmatory research the hypotheses are in the Introduction and the inferences evaluating those hypotheses are in the Discussion. For exploratory research, the Introduction merely states the novelty of the biological situation and makes clear why the novelty is important to elucidate; the Discussion then poses the hypotheses gained from the data analyzed, and these hypotheses are themselves the inferences*” (Jaeger e Halliday, 1998: 565).

⁸ Para Menard (2008), “*in a purely cross-sectional design, data are collected on one or more variables for a single period of time. In longitudinal research, data are collected on one or more variables for two or more time periods, allowing at least measurement of change and possibly explanation of change*” (Menard, 2008: 3). Para mais informações sobre estudos longitudinais ver Menard (2002; 2008).

⁹ Define-se coorte como o conjunto de indivíduos/unidades de análise que experimentaram um mesmo evento. Os estudos de coorte, por sua vez, são estudos observacionais em que os indivíduos são classificados/selecionados segundo o *status* de exposição a um determinado evento. Esse tipo de estudo é bastante recorrente em epidemiologia para “avaliar a incidência da doença em determinado período de tempo. Os estudos de coorte também podem ser utilizados para avaliar os riscos e benefícios do uso de determinada medicação” (Oliveira e Parente, 2010: 115).

amostra (formandos) experimentaram um evento em comum (conclusão do curso), por isso, são considerados uma coorte.

Por fim, nos estudos de painel a mesma unidade de análise é entrevistada em períodos diferentes do tempo. Para Babbie (2005), “estudos de painéis envolvem a coleta de dados, ao longo do tempo, da mesma amostra de respondentes, que se chama painel” (Babbie, 2005: 103). Por exemplo, o pesquisador pode acompanhar diferentes famílias no tempo e observar os seus padrões de consumo. A unidade de análise é a família. O objetivo do estudo é analisar como o padrão de consumo varia ao longo do tempo.¹⁰ A tabela abaixo sumariza as funções e os tipos da pesquisa de survey.

Quadro 1: Funções e tipos da pesquisa de survey

Funções			Tipos	
Exploração	Descrição	Explicação	Transversais	Longitudinais
Indicado para explorar fenômenos/ objetos desconhecidos.	Adequado para descrever as características de fenômenos/ objeto. Serve como fonte de informações para outros estudos.	Ideal para testar hipóteses teoricamente orientadas.	Exemplo típico de pesquisa de survey. É mais barato e mais frequentemente utilizado não só na academia, mas também em pesquisas de mercado.	Tendência Coorte Painel

Fonte: elaboração dos autores a partir de Babbie (2005)

Depois de definir as principais funções e tipos da pesquisa de survey, o próximo passo é apresentar o planejamento de um survey. A próxima seção sumariza essas informações.

4. O planejamento de um survey

Esquemáticamente, o planejamento de uma pesquisa de survey deve seguir sete etapas: (1) identificação da questão de pesquisa; (2) elaboração do instrumento; (3) definição da equipe e treinamento dos aplicadores; (4) pré-teste do instrumento; (5) coleta dos dados; (6) tabulação dos dados e (7) análise dos dados.

O primeiro passo para realizar um survey é identificar claramente uma questão de pesquisa (Saris e Gallhofer, 2007). Apenas depois disso o pesquisador deve se preocupar com a elaboração do questionário. Quanto mais bem definida a questão de pesquisa, menor é a dificuldade na construção do instrumento. Além disso, o pesquisador deve deixar claro qual é o objetivo da pesquisa de survey: exploratória, descritiva ou explicativa.

Depois disso, o próximo passo é elaborar o questionário. A elaboração do instrumento é um trabalho complexo e requer treinamento específico. Mesmo pesquisadores experientes se deparam com problemas na sua construção.¹¹ Dada a amplitude do tema, não é possível apresentar todas as peculiaridades que envolvem a elaboração de um questionário. No entanto, é possível apresentar algumas recomendações mais gerais. Primeiramente, o pesquisador deve identificar o tipo de questão (valores, crenças, opiniões, conhecimento, comportamento, background, contextuais, etc.).

¹⁰ Apesar de ser o mais adequado para realizar inferências causais, o estudo de painel é mais difícil de ser implementado. Uma primeira limitação diz respeito aos recursos financeiros. Em geral, é muito dispendioso manter uma pesquisa dessa natureza. Uma segunda limitação diz respeito à perda de indivíduos (observações) ao longo do tempo. De acordo com Ribas e Soares (2010), existem duas principais causas para o desgaste de um painel: (1) mudança de endereço das pessoas da amostra e (2) recusa de entrevista. Quanto maior a duração da pesquisa, maior é a chance de atrito. Quando ele não é aleatório, isso representa um grave problema para as inferências.

¹¹ Converse e Presser (1986) apresentam uma introdução bastante pedagógica aos principais desafios de elaboração do questionário.

Outra decisão importante diz respeito às questões fechadas ou abertas, à utilização de probes,¹² à inclusão da opção “não sei”, à formatação das escalas, etc. Cada procedimento é importante para garantir a confiabilidade das informações.

Outro elemento que deve ser observado com cuidado é a linguagem. Ela deve ser a mais simples possível. Para Babbie (2005), “em geral, você deve supor que os entrevistados irão ler os itens rapidamente e dar respostas rápidas, você deve fornecer itens claros e curtos que não serão mal interpretados sob tais condições” (Babbie, 2005: 193). Por exemplo, não questione: “A sua família é gerenciada segundo uma linha matriarcal ou patriarcal?”. Prefira: “Quem é o chefe da família?”. Devemos evitar também jargões, termos técnicos e/ou complexos. Devemos evitar também perguntas com duplo negativo,¹³ assim como perguntas dúbias.¹⁴

Depois de finalizar o instrumento, o próximo passo é definir a equipe de trabalho e treinar os aplicadores. O treinamento deve ser exaustivo, compreendendo todas as questões do questionário, inclusive aquelas consideradas mais simples.¹⁵ Para Babbie (2005), “o número de entrevistadores requeridos é determinado com base no (1) número de entrevistas a serem feitas, (2) tempo médio necessário para cada entrevista, (3) tempo programado para toda a operação de entrevistas e (4) número de entrevistadores qualificados disponíveis” (Babbie, 2005: 270). É importante que os aplicadores colem as informações de forma padronizada, ou seja, deve-se maximizar a homogeneidade do comportamento dos entrevistadores com o objetivo de minimizar as distorções sobre o padrão de respostas coletadas. A própria pesquisa de survey já apresenta peculiaridades inerentes que podem distorcer a confiabilidade das informações.¹⁶ Em geral, deve-se ter um ou mais coordenadores de campo disponíveis para dirimir eventuais dúvidas dos aplicadores. Para Babbie (2005), “apesar de uma pessoa ser responsável pela supervisão de toda a operação, ela pode receber assistência de uma equipe de supervisores. Como regra básica geral, basta um supervisor para cada dez entrevistadores” (Babbie, 2005: 270). Cabe ao coordenador, entre outras funções, garantir que as unidades amostrais geográficas selecionadas sejam devidamente visitadas e/ou substituídas, seguindo critérios pré-definidos. Como os aplicadores recebem por questionário aplicado, é importante também estabelecer alguma forma de checagem dos questionários. A experiência acumulada revela que existem diferentes formas de aumentar artificialmente o número de questionários efetivamente aplicados.¹⁷

Tão importante quanto a experiência técnica dos entrevistadores é assegurar também que a equipe seja coesa e confiável. É extremamente relevante estabelecer uma relação de confiança com os entrevistadores de modo que qualquer eventual problema seja devidamente reportado e dirimido. Muitas vezes a pesquisa enfrenta problemas inesperados que, na ausência de uma forte coesão da equipe, pode reduzir a confiabilidade dos dados coletados e, conseqüentemente, dos resultados da pesquisa. Por exemplo, em pesquisas que dependem de alguma espécie de repasse de recursos, é comum o repasse atrasar e a pesquisa ficar temporariamente sem recursos. Se a equipe de entrevistadores e digitadores resolver parar de trabalhar, isso pode comprometer fortemente a continuidade e conclusão da pesquisa.

¹² De acordo com Silva (2013), o probe funciona como um aprofundamento da questão, sem induzir o entrevistado às respostas, mas ajudando a entender e responder à questão proposta.

¹³ Por exemplo, “o uso da maconha não deveria ser descriminalizado. Você concorda ou discorda?” Aqui, a utilização do duplo negativo além de confundir o entrevistado pode gerar um padrão de resposta exatamente contrário ao esperado, ou seja, o entrevistado é a favor da liberalização, mas dada a formatação da questão, vai responder que é contrário. Uma opção seria perguntar: você é a favor da descriminalização da maconha?

¹⁴ Por exemplo, “com que frequência você visita seu país?” Aqui, a falta de acento na palavra “país”, pode confundir o entrevistado, minando a qualidade das respostas.

¹⁵ Para Babbie (2005), “se os entrevistadores não estiverem familiarizados com o questionário, a pesquisa sofrerá e será posto um peso injusto sobre o entrevistado. A entrevista provavelmente tomará mais tempo do que o necessário e, em geral, será incômoda” (BABBIE, 2005: 263).

¹⁶ Existe uma vasta literatura que mostra que diferenças sutis no instrumento geram diferentes padrões de resposta. Por exemplo, Anderson, Silver e Abramson (1988) examinam a influência da raça do entrevistador/aplicador. Bishop et al. (1988) comparam as respostas produzidas por questionários auto administrados versus aqueles feitos por telefone. Fox e Tracy (1993) analisam as questões sensíveis. Rasinski (1989) examina o efeito da ordem das palavras sobre o padrão de respostas.

¹⁷ Por exemplo, em uma ocasião o coordenador de campo surpreendeu o aplicador sentado na calçada, respondendo os questionários.

Depois de definir e treinar a equipe de trabalho de campo deve-se realizar o pré-teste do instrumento.¹⁸ Essa fase é fundamental não só para dar mais confiança aos aplicadores, mas principalmente para aperfeiçoar o instrumento.¹⁹ De acordo com Babbie (2005),

O estudo-piloto deve envolver a administração de um instrumento de pesquisa o mais idêntico possível ao planejado para o *survey* final. Mas, se for usado um procedimento de postagem muito elaborado, pode não ser viável produzir o número necessário para o estudo-piloto e depois um lote revisado para o *survey* final. O questionário do estudo-piloto deve conter todas as questões planejadas, com a redação, o formato e a sequência que o pré-teste indicou serem os melhores. O estudo-piloto não deve ser um veículo para testar novos itens ainda não pré-testados (Babbie, 2005: 311).²⁰

Em geral, o coordenador da pesquisa identifica que questões funcionaram melhor e quais geraram dúvidas entre os respondentes.²¹ É possível ainda produzir uma estimativa da média de tempo de aplicação do questionário, o que é uma informação valiosa durante a execução da coleta de dados.

Depois de coletar os dados, o próximo passo é realizar a sua tabulação. Comumente, a tabulação acontece através da digitação dos questionários em um banco de dados. O advento dos *tablets* tem gradativamente substituído o papel dos digitadores, maximizando a quantidade de questionários aplicados e reduzindo o tempo de duração da pesquisa. No entanto, muitos pesquisadores ainda não tem recursos para trabalhar com esses equipamentos. Antes disso, porém, é necessário realizar alguma forma de checagem no sentido de verificar se o questionário foi corretamente preenchido. Em caso afirmativo, o questionário é digitado. Em caso negativo, deve-se avaliar qual foi erro cometido e definir o melhor curso de ação. É importante também assegurar que os digitadores tenham alguma experiência. Isso aumenta a velocidade da digitação e minimiza eventuais erros analíticos.

Depois de devidamente tabulados, o procedimento final é analisar os dados. A depender do objetivo principal da pesquisa, existem diferentes técnicas estatísticas que podem ser utilizadas. De acordo com Brady (2000), “*modern statistical techniques and computer technology make it possibly for survey researchers to apply elegant data reduction methods that summarizes trends and locate important anomalies*” (Brady, 2000: 47). O pesquisador também deve informar qual foi o pacote estatístico utilizado para produzir as análises, disponibilizar o banco de dados em algum repositório e reportar as rotinas utilizadas. Nesse trabalho, enfatizamos a aplicação e interpretação de duas principais ferramentas: (1) teste de associação do qui-quadrado e (2) análise de componentes principais. Computacionalmente, utilizamos o *Statistical Package for Social Sciences* e o STATA. O banco de dados está publicamente disponível no endereço eletrônico: <http://www.gog.pol.gu.se/data/>. A próxima seção descreve, sumariamente, a lógica subjacente a essas duas técnicas, bem como informa ao leitor como realizar a interpretação das saídas (*outputs*) computacionais.

¹⁸ Para Babbie (2005), “todo manual de pesquisa aconselha fazer algum tipo de teste do desenho de pesquisa antes da pesquisa maior. Os argumentos a favor do pré-teste são convincentes. Ninguém quer investir grandes quantias de dinheiro e muito trabalho num grande projeto e não alcançar os objetivos da pesquisa devido a algum erro imprevisto” (Babbie, 2005: 302).

¹⁹ Existem diferentes formas de administrar o instrumento: entrevistas pessoais, telefone, correio, e-mail e internet. Cada uma delas apresenta vantagens e limitações. É importante que o pesquisador esteja ciente disso antes de optar pela forma de aplicação.

²⁰ Segundo Simões e Pereira (2007), “no nosso entendimento o pré-teste é um elemento central para a validade dos dados tanto no que concerne a verificar e aprimorar a operacionalização das questões de pesquisa e dos conceitos –ou seja, na passagem da teoria para a empiria– quanto na garantia da correspondência de significados entre pesquisador e entrevistados” (Simões e Pereira, 2007: 247).

²¹ Para Babbie (2005), “todo instrumento ou só uma parte dele pode ser pré-testado. Talvez a preocupação básica seja a replicabilidade de um conjunto de perguntas. Se for assim, você pode concentrar a atenção em vários pré-testes desta parte do questionário e em modificações dela. Embora isso seja legítimo, lembre que o contexto no qual as perguntas aparecem afeta sua recepção. Portanto, se vale a pena fazer testes iniciais de partes do questionário, é melhor acrescentar um ou mais pré-testes de todo o instrumento” (Babbie, 2005: 305).

5. Análise de dados: qui-quadrado e análise de componentes principais

5.1. Teste de qui-quadrado²²

O teste de qui-quadrado (χ^2) é uma técnica não paramétrica de análise de dados.²³ Como em toda técnica estatística é necessário satisfazer alguns pressupostos, entre eles é possível destacar a aleatoriedade da amostra e a independência das observações. A aleatoriedade se justifica já que é requisito básico para realizar inferências confiáveis. A independência das observações postula que a ocorrência de um caso não pode influenciar a ocorrência de outro. Para fins de pesquisa, existem diferentes aplicações desse teste (Gravetter e Wallnau, 2004). Em Ciência Política, os pesquisadores geralmente estão interessados em analisar o grau de associação entre variáveis categóricas em amostras independentes.²⁴ Fundamentalmente, o teste compara a frequência observada com os valores esperados assumindo que não há associação entre as variáveis. Quanto maior o valor do qui-quadrado, maior é a diferença entre os valores observados e os esperados, sugerindo associação na distribuição de frequências das variáveis. A análise se baseia em uma tabela cruzada que sumariza a distribuição de frequências para as diferentes categorias das variáveis de interesse. O quadro abaixo apresenta um exemplo de questão de pesquisa em que é possível utilizar a técnica de qui-quadrado.

Quadro 2: Exemplo de Qui-quadrado

Questão de pesquisa	Variáveis
Existe relação entre independência do judiciário e fraude eleitoral?	Duas variáveis categóricas. Judiciário independente (sim ou não) e Fraude eleitoral (sim ou não)

Enquanto a hipótese nula postula que não existe relação entre independência do judiciário e a ocorrência de fraude eleitoral (H_0), a hipótese alternativa (H_a) aposta que existe dependência na distribuição das frequências dessas variáveis, ou seja, que existe associação entre elas. Teoricamente orientado, o pesquisador postula que judiciários mais independentes devem reduzir a frequência de fraude eleitoral. Ou seja, espera-se uma relação negativa entre as variáveis. Depois de realizado o teste, o pesquisador deve analisar não só o valor do qui-quadrado, mas também a significância estatística (p-valor) dos resultados observados. Um p-valor não significativo ($\alpha > 0,05$) sugere independência estatística entre as variáveis, ou seja, não é possível rejeitar a hipótese nula.

²² Para uma discussão sobre a origem histórica do qui-quadrado ver Plackett (1983). Mirkin (2001) apresenta onze diferentes definições do coeficiente de qui-quadrado. Para uma introdução intuitiva ver Dancey e Reidy (2005). Para trabalhos mais avançados ver Blalock (1979), Garson (1976), Iverson (1979) e Agresti e Agresti (1977). Para uma introdução a análise de dados nominais ver Reynolds (1984). Para uma introdução a análise de dados ordinais ver Hildebrand, Laing e Rosenthal (1977).

²³ Algebricamente,

$$\chi^2 = \frac{(fo - fe)^2}{fe}$$

O teste compara a distribuição observada (f_o) em uma determinada amostra com a distribuição esperada (f_e) assumindo equiprobabilidade de ocorrência dos casos. Quanto maior a diferença entre os valores observados e os esperados, maior é o valor do qui-quadrado. Na maior parte dos livros de estatística aplicada, as diferentes modalidades de qui-quadrado aparecem tanto na seção de relação entre variáveis quanto na seção de comparação de grupos. Isso porque o pesquisador pode comparar se existem diferenças nas frequências de dois grupos distintos, tratando o grupo como variável independente. Por exemplo: a) homens fumam mais do que mulheres?; b) a taxa de suicídio é maior entre protestantes do que entre católicos?; c) governos presidencialistas são mais passíveis de colapso institucional do que governos parlamentaristas?; d) homens ocupam mais cargos gerenciais do que mulheres? O teste de associação do qui-quadrado é ideal para examinar questões de pesquisa dessa natureza.

²⁴ Para Reynolds (1984), "the chi-square for independence provides a standard for deciding whether two variables are statistically independent. The test consists of four parts: (1) the null hypothesis (H_0) that the variables are statistically independent; (2) expected frequencies derived under the assumption that the null hypothesis is true; (3) a comparison of these expected values with the corresponding observed frequencies; and (4) judgment about whether or not the differences between expected and observed frequencies would have arisen by chance" (Reynolds, 1984: 15).

Contrariamente, um p-valor significativo ($\alpha < 0,05$) sugere dependência estatística entre as variáveis. Além disso, é possível analisar o tamanho do efeito. Em tabelas cruzadas de 2x2 (quando ambas as variáveis são dicotômicas), o coeficiente *phi* (ϕ) indica a força da associação entre as variáveis de modo que quanto mais perto de 1, maior é o nível de associação entre as variáveis.²⁵

Tabela 1: Tabulação cruzada entre independência do judiciário e fraude eleitoral

Judiciário independente	Fraude Eleitoral (%)		Total
	Sim	Não	
Sim	8 (12,5)	56 (87,5)	90 (100)
Não	25 (27,8)	65 (72,2)	64 (100)

* Qui-quadrado de Pearson 5,185 (p-valor=0,023)²⁶

** Correção de continuidade 4,317 (p-valor = 0,038)

*** Coeficiente Phi -0,183 (p=valor = 0,023)²⁷

O percentual deve ser analisado no sentido da variável independente, no caso, a independência do judiciário. Dos 90 países com judiciário independente, em 12,5% dos casos observou-se fraude eleitoral, totalizando oito observações. Quando não existe independência, esse percentual passa para 27,8%, perfazendo 25 casos. O teste de qui-quadrado sugere que existe uma associação estatisticamente significativa entre independência do judiciário e fraude eleitoral ($\chi^2 = 5,185$ e p-valor = 0,023). O coeficiente *phi* (ϕ), por sua vez, informa a força da correlação: (*phi* = -0,183). Ao conferir interpretação substantiva aos resultados de nosso exemplo, devemos concluir que países com judiciários independentes tendem a apresentar menor incidência de fraude eleitoral.

5.2 Análise de componentes principais²⁸

A principal função das diferentes técnicas de análise fatorial é reduzir uma grande quantidade de variáveis observadas em um número menor de fatores.²⁹ Mas o que são fatores?

²⁵ Em tabelas maiores do que 2x2 deve-se reportar a estatística Cramer's V. Ela é calculada a partir da divisão da raiz quadrada do qui-quadrado pelo número de casos (tamanho da amostra), controlando pelo número de linhas e colunas da tabela. Assim como o coeficiente Phi, varia entre 0 e 1. Quanto mais perto de zero, menor é o nível de associação entre as variáveis. Quanto mais perto de 1, maior é o nível de associação entre as variáveis.

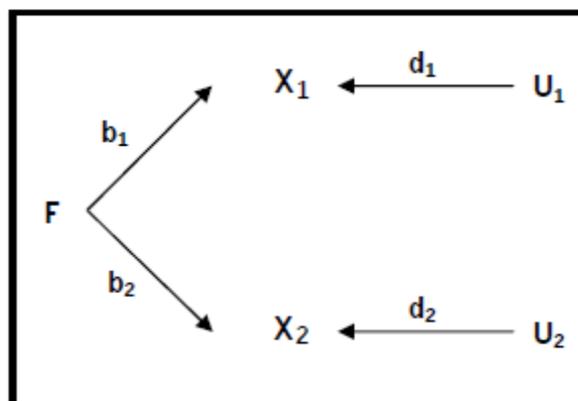
²⁶ O pesquisador pode utilizar também a razão de verossimilhança do qui-quadrado. De acordo com Reynolds (1984), ainda que assumam valores diferentes, essas estatísticas são consistentes e produzem as mesmas conclusões.

²⁷ O coeficiente Phi é calculado a partir da raiz quadrada da divisão do valor do qui-quadrado pelo tamanho da amostra. Quanto maior o seu valor, maior é o grau de associação entre as variáveis. Para Reynolds (1984), "one reason for not using the good-of-fit chi square as a measure of association is that its numerical magnitude depends partly on the size of the sample. Dividing chi square by n corrects for n and leads to a popular measure of association - phi-squared" (Reynolds, 1984: 42).

²⁸ Essa seção foi elaborada a partir do artigo "Visão além do alcance: uma introdução à análise fatorial", veiculado pela *Revista Opinião Pública* (Figueredo Filho e Silva Júnior, 2010). Para trabalhos clássicos sobre análise fatorial ver Harman (1967), Rummel (1967; 1970), Cooper (1983) e Bartholomew (1984). Para textos introdutórios ver Kim e Mueller (1978a; 1978b), Zeller e Carmines (1980), Decoster (1998) e Costelo e Osborne (2005). Para uma abordagem mais aprofundada ver Tabachnick e Fidell (2007), Lawley e Maxwell (1973), Isogawa e Okamoto (1980), Yalcin e Amemiya (2001) e Bollen e Arminger (1991). Para análise fatorial de dados *missing* ver Mackelprang (1970) e Ligny et al. (1981), para análise fatorial de dados categóricos ver Bartholomew (1980) e Vermunt e Magidson (2004). Para aplicações práticas utilizando o SPSS ver Dancey e Reidy (2004), Pallant (2007) e Ho (2006). Para uma introdução em português ver Hair et al. (2005). Para diferentes aplicações ver Grumm (1963), Roberts (1987), Dunn, Schneck e Lawson (1973), Bonjean e Browning (1969) e Slatin (1974).

Hair *et al.* (2005) definem fator como a combinação linear das variáveis (estatísticas) originais. A figura abaixo ilustra a relação entre variáveis diretamente observadas e os seus respectivos fatores.

Figura 3: Modelo das vias para duas variáveis, modelo de um fato



Fonte: elaboração dos autores a partir de Asher (1983)

Na figura acima, X_1 e X_2 são variáveis observadas: X_1 é causado por F e por U_1 . Da mesma forma, X_2 é causado por F e por U_2 . Na medida em que F é comum a X_1 e X_2 ele é considerado um fator comum. Contrariamente, tanto U_1 quanto U_2 são considerados fatores únicos já que são restritos a X_1 e X_2 , respectivamente (Asher, 1983). De acordo com Kim e Mueller (1978), "a análise fatorial se baseia no pressuposto fundamental de que alguns fatores subjacentes, que são em menor número que as variáveis observadas, são responsáveis pela covariação entre as variáveis" (Kim e Mueller, 1978: 12). Nesse exemplo F , U_1 e U_2 são considerados fatores (não podem ser diretamente observados) enquanto que X_1 e X_2 são as variáveis que o pesquisador pode observar diretamente.³⁰ A literatura identifica duas principais modalidades de análise fatorial: exploratória e confirmatória (Tabachnick e Fidell, 2007). A análise fatorial exploratória (AFE) geralmente é utilizada nos estágios embrionários da pesquisa, no sentido de literalmente explorar os dados. Nessa fase, procura-se explorar a relação entre um conjunto de variáveis, identificando padrões de correlação. Além disso, a AFE pode ser utilizada para criar variáveis independentes ou dependentes que podem ser empregadas posteriormente em modelos de regressão. Por sua vez, a análise fatorial confirmatória (AFC) é utilizada para testar hipóteses. Nesse caso, o pesquisador guiado por alguma teoria testa em que medida determinadas variáveis são representativas de um conceito/dimensão.

Nesse trabalho, utilizamos análise de componentes principais para criar uma medida de secularização a partir dos dados do *World Value Survey*, que foram compilados pelo *Quality of Government Institute* (QoG). A variável secularização tem média zero e desvio padrão igual a um. Testamos a hipótese de que existe uma correlação positiva entre secularização e controle da corrupção.

²⁹ Para Babbie (2005), a "análise fatorial é usada para descobrir padrões de variações nos valores de diversas variáveis, essencialmente pela geração de dimensões artificiais (fatores) que se correlacionam altamente com diversas das variáveis reais" (Babbie, 2005: 418).

³⁰ King (2001) adverte que "um erro comum consiste em ver as variáveis observadas como causas do fator. Isso é incorreto. O modelo correto tem variáveis dependentes observáveis como funções dos fatores subjacentes e não (King, 2001: 682). Ou seja, por mais intuitivo que seja acreditar que as variáveis observadas causam o fator, a interpretação correta é justamente o posto: o fator é um construto (dimensão) comum entre as variáveis. Para a diferença na interpretação entre fatores e componentes ver Tabachnick e Fidell (2007).

Quadro 3: Análise de componentes principais

Questão de pesquisa	Variáveis
Qual a correlação entre secularização e controle da corrupção?	Religiosidade (wvs_rel), Índice de Autonomia (wvs_auton), Aborto é justificável (wvs_abort), Orgulho nacional (wvs_proud), Respeito pela autoridade (wvs_auth). Todas as variáveis são ordinais. Variam entre 1 e 10 e serão utilizadas para criar a medida de secularização. Por sua vez, o controle da corrupção é uma estimativa da percepção em que os governos conseguem controlar atividades corruptas

Tabela 2: Teste de adequação da amostra

Teste Kaiser-Meyer-Olkin de adequação da amostra		0,820
Teste de Esfericidade de Bartlett	Qui-quadrado	336,190
	GI	10
	P-valor	0,000

O teste de esfericidade da amostra indica se os dados são adequados à técnica de análise de componentes principais. O teste Kaiser-Meyer-Olkin (KMO) varia entre 0 e 1. Quanto maior, maior é a adequabilidade dos dados. O teste de esfericidade de Bartlett (BTS) foi estatisticamente significativo ($p\text{-valor} < 0,000$) com um qui-quadrado aproximado de 336,190, reforçando a ideia de que os dados são apropriados. A tabela abaixo sumariza as comunalidades.

Tabela 3: Comunalidades

Variável	Extração
Aborto é justificável	0,753
Respeito pela autoridade	0,600
Índice de autonomia	0,857
Orgulho nacional	0,590
Religiosidade	0,839

Método de extração: Análise de Componentes Principais

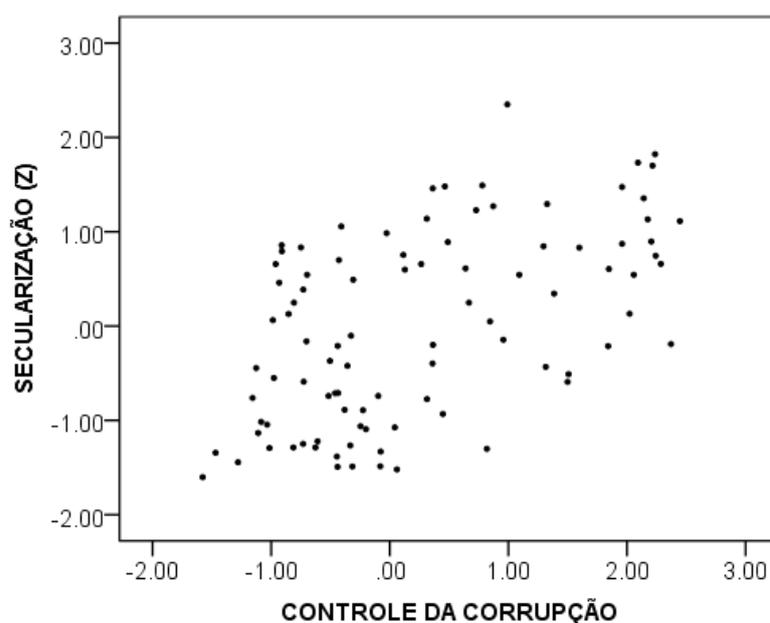
De acordo com Schawb (2007), as comunalidades representam a proporção da variância para cada variável incluída na análise que é explicada pelos componentes extraídos. Quanto mais próximo de um, maior é o grau de associação. Em geral, variáveis com comunalidades abaixo de 0,5 devem ser excluídas da análise. Em nosso exemplo, todas as variáveis estão acima desse patamar. Comparativamente, os indicadores de autonomia (0,857) e de religiosidade (0,839) estão fortemente correlacionados com o componente de secularização. A tabela abaixo sumariza a variância total explicada pelo componente extraído.

Tabela 4: Variância total explicada

Autovalor inicial			Extraction Sums of Squared Loadings		
Total	% de Variância	Acumulado %	Total	% de Variância	Acumulado %
3,638	72,750	72,750	3,638	72,750	72,750
0,542	10,843	83,594			
0,505	10,104	93,698			
0,177	3,538	97,236			
0,138	2,764	100,000			

Método de extração: Análise de Componentes Principais

A literatura sugere que o componente extraído deve carregar, pelo menos, 60% da variância total das variáveis observadas (Hair et al., 2005). Em nosso exemplo, o componente 1 apresentou autovalor de 3,638 e carregou 72,75% de toda a variância. Esse componente será retido e posteriormente utilizado para verificar a correlação entre corrupção e secularização.

Gráfico 1: Correlação entre secularização e controle de corrupção

Existe uma correlação positiva (0,547) e estatisticamente significativa ($p\text{-valor} < 0,000$) entre secularização e o controle da corrupção ($n=94$). Dessa forma, deve-se rejeitar a hipótese nula de que essas variáveis são estatisticamente independentes ($r=0$) em função da hipótese de trabalho ($r \neq 0$).

Conclusão

Esse trabalho apresentou uma introdução intuitiva às potencialidades de *survey*. Nosso principal objetivo foi munir os estudantes de graduação, pós-graduação e pesquisadores menos familiarizados com essa técnica de ferramentas básicas. O próximo passo deve ser a capacitação técnica através de cursos especializados e/ou disciplinas específicas. De acordo com Brady (2000), “with these strong capacities for data collection, accurate data magnification, data reduction, and capitalizing on events and manipulations, survey methods is an extraordinarily powerful approach to studying social world” (Brady, 2000: 47). Concordamos com o professor Brady e acreditamos que

além de uma importante ferramenta analítica, a técnica de *survey* é pedagogicamente importante para ensinar metodologia de forma geral e métodos quantitativos, em particular. Esperamos com esse artigo facilitar a compreensão e principalmente a utilização do *survey* na pesquisa empírica em ciências sociais.

REFERENCIAS BIBLIOGRÁFICAS

- AGRESTI, A. and AGRESTI, B. F. (1977). Statistical analysis of qualitative variation. In: SCHUESSLER, K.F. (Ed.), *Sociological Methodology* 1978. Jossey-Bass, San Francisco, pp. 204-237.
- ANDERSON, B. A.; BRIAN D. S. and ABRAMSON, P. R. (1988). "The Effects of Race of the Interviewer on Measures of Electoral Participation by Blacks in SRC national Election Studies." *Public Opinion Quarterly*. 52: 53-83.
- ASHER, H. B. (1983). *Causal Modeling*. Beverly Hills, CA: Sage.
- BABBIE, E. (2005). *Métodos de Pesquisas em Survey*. Belo Horizonte-MG: Editora UFMG.
- BARTHOLOMEW, D. J. (1984). The foundations of factor analysis, *Biometrika*, 71, 221-232.
- BISHOP, G. F.; TUCHFABER, A. J.; OLDENDICK, R. W. (1986). "Options in fictitious issues: the pressure to answer survey questions". *Public Opinion Quarterly*, v. 50, n. 2, p. 240-250.
- BLALOCK, H. M. (1984). *Measurement in the social sciences: Theories and strategies*.
- BOLLEN, K. A. (1989). *Structural Equations with Latent Variables*. Wiley Series in Probability and Mathematical Statistics. Nova York: Wiley.
- BOLLEN, K. A. and ARMINGER, G. (1991). Observational Residuals in Factor Analysis and Structural Equation Models. *Sociological Methodology*, 21, 235-262.
- BONJEAN, C. M. and BROWNING, H. L. (1969). Toward Comparative Community Research: A Factor Analysis of United States Counties. *The Sociological Quarterly*, 10, 2, 157-176.
- BRADY, H. (2000), "Contributions of Survey Research to Political Science", *PS: Political Science and Politics*, vol. 33, pp. 47-57.
- CONVERSE, J. M. and PRESSER, S. (1986). *Survey Questions: Handcrafting the Standardized Questionnaire*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-063. Thousand Oaks, California. Sage Publications. 80 p.
- COOPER, J. C. B. (1983). Factor Analysis: An Overview. *The American Statistician*, 37, 2, 141- 147.
- COSTELLO, A. B. and OSBORNE, J. W. (2005). "Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis." *Practical Assessment Research & Evaluation*, 10, 7, 13-24.
- DANCEY, C. e REIDY, J. (2006). *Estatística Sem Matemática para Psicologia: Usando SPSS para Windows*. Porto Alegre, Artmed.
- DECOSTER, J. (1998). Overview of Factor Analysis. [Online] Disponível em: <<http://www.stat-help.com/notes.html>> Acesso em: 19 nov. 2012.
- DUNN, M. J.; SCHNECK, R. and LAWSON, J. (1973). "A Test of the Uni-Dimensionality of Various Political Scales through Factor Analysis: A Research Note." *Canadian Journal of Political Science / Revue Canadienne de Science Politique*, 6, 4, 664-669.
- FIGUEIREDO FILHO, D. B.; SILVA JÚNIOR, J. A. (2010). Visão além do alcance: uma introdução à análise fatorial. *Opinião Pública*, Campinas, v. 16, n. 1, p. 160-185, Junho.
- FOWLER, L. L. (1995). "Replications as Regulation." *PS. Political Science and Politics* 28 (3): 478-481.

- FOX, J. A., and TRACY, P. E. (1986). *Randomized Response: A Method for Sensitive Surveys*. Sage Publications, Beverly Hills.
- GARSON, G. D. (2009). *Statnotes: Topics in Multivariate Analysis*. [Online] Disponível em: <<http://faculty.chass.ncsu.edu/garson/PA765/statnote.htm>> Acesso em: 02 fev. 2013.
- GRAVETTER, F. J. & WALLNAU, L. B. (2004). *Statistics for the behavioral sciences*. Sixth Edition. Belmont, CA: Wadsworth
- GRUMM, J. G. (1963). "A Factor Analysis of Legislative Behavior." *Midwest Journal of Political Science*, 7, 4, 336-356.
- HAIR, Jr; BLACK, W. C; BABIN, B. J; ANDERSON, R. E e TATHAM, R. L. (2006). *Multivariate Data Analysis*. 6ª edition. Upper Saddle River, NJ: Pearson Prentice Hall.
- HARMAN, H. H. (1967). *Modern Factor Analysis*. 2ª edição. Chicago: University of Chicago Press.
- HILDEBRAND, D. K; LAING, J.D. and ROSENTHAL, H. (1977). *Prediction analysis of cross-classifications*. New York: Wiley.
- HO, R. (2006). Handbook of Univariate and Multivariate Data Analysis and Interpretation with SPSS. *Journal of Statistical Software*, July 2006, Volume 16, Book Review 4.
- ISOGAWA, Y and OKAMOTO, M. (1980). "Linear Prediction in the Factor Analysis Model." *Biometrika*, 67, 2, 482-484.
- IVERSON, G. R. (1979). Decomposing chi-square: A forgotten technique. *Sociological Methods Research*, 8, 143-157.
- JAEGER, R. G. and HALLIDAY, T. R. (1998). On confirmatory versus exploratory research. *Herpetologica*, 54 (supplement), 64-6.
- KIECOLT, K. J. and NATHAN, L. E. (1985). *Secondary analysis of survey data*. London: Sage Publications. ISBN 0-8039-2302-3
- KIM, J. and MUELLER, C. W. (1978a). *Factor analysis: Statistical methods and practical issues*. Beverly Hills, CA: Sage.
- _____ (1978b). *Introduction to factor analysis - what it is and how to do it*. Beverly Hills, CA: Sage.
- KING, G. (2001). How not to lie with statistics [Online] Disponível em: <<http://gking.harvard.edu/files/mist.pdf>> Acesso em: 18 fev. 2013.
- KING, G., KEOHANE, R. e VERBA, S. (1994). *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton. N.J.: Princeton University Press.
- KLINE, R. B. (2004). *Principles and Practice of Structural Equation Modeling*. Nova York: Guilford.
- LAWLEY, D. N and MAXWELL, A. E. (1973). "Regression and Factor Analysis." *Biometrika*, 60, 2, 331-338.
- LEE, E. S. e FORTHOFFER, R. N. (2006). *Analyzing Complex Survey Data*. Thousand Oaks, Sage.
- LEVY, P. S. and STOLTE, K. (2000). Statistical methods in public health and epidemiology: a look at the recent past and projections for the next decade. *Statistical Methods in Medical Research*, 9, 41-

55.

LIGNY, C. L.; NIEUWDORP, G. H. E; BREDERODE, W. K; HAMMERS, W. E; HOUWELINGEN, J. C. van. (1981). An Application of Factor Analysis with Missing Data. *Technometrics*, 23, 1, 91-95.

MACKELPRANG, A. J. (1970). "Missing Data in Factor Analysis and Multiple Regression." *Midwest Journal of Political Science*, 14, 3, 493-505.

MENARD, S. (2002). *Longitudinal Research*. 2d ed. Thousand Oaks, CA: SAGE, 2002.

_____ (Ed.) (2008). *Handbook of longitudinal research: Design, measurement, and analysis*. Amsterdam: Elsevier.

MOORE, D. S. and McCABE, G. P. (2003). *Introduction to the Practice of Statistics*, W.H. Freeman and Company.

MIRKIN, B. (2001). Eleven Ways to Look at the Chi-Squared Coefficient for Contingency Tables, *The American Statistician*, 55, no. 2, 111-120.

OLIVEIRA, M. A. P. e PARENTE, R. C. M. (2010). Estudos de Coorte e de Caso-Controle na Era da Medicina Baseada em Evidência. *Bras. J. Video-Sur.*, July/Sept.

PALLANT, Julie. (2007). *SPSS Survival Manual*. Open University Press.

PLACKETT, R.L. (1983). Karl pearson and the Chi-squared test. *Int. Statist. Rev.* 51, 59-72.

RASINSKI, K. A. (1989). "The Effect of Question Wording on Public Support for Government Spending." *Public Opinion Quarterly* 53: 388-94.

REYNOLDS. H. T. (1984). *Analysis of nominal data* (2nd ed). Beverly Hills, CA: Sage Publications.

RIBAS, R. P.; SOARES, S. S. D. (2010). O atrito nas pesquisas longitudinais: o caso da Pesquisa Mensal de Emprego (PME/IBGE). *Estudos Econômicos*, São Paulo, v. 40, p. 213-244.

RUMMEL, R. J. (1970). *Applied Factor Analysis*. Evanston: Northwestern University Press, 1970.

SARIS, W. E. and GALLHOFER, I. N. (2007). *Design, Evaluation, and Analysis of Questionnaires for Survey Research*. Hoboken, NJ: Wiley.

SIMÕES, S. e PEREIRA, M. A. M. (2007). "A arte e a ciência de fazer perguntas: aspectos cognitivos da metodologia de survey e a construção do questionário". In: AGUIAR, Neuma. (Coord.). *Desigualdades sociais, redes de sociabilidade e participação política*. Belo Horizonte: Editora UFMG, 2007, p. 249 -69.

SCHAWB, A. J. (2007) *Eletronic Classroom*. Disponível em: <<http://www.utexas.edu/ssw/eclassroom/schwab.html>> Acesso em: 22 out. 2012.

SLATIN, G. T. (1774). "A Factor Analytic Comparison of Ecological and Individual Correlations: Some Methodological Implications." *The Sociological Quarterly*, 15, 4, 507-520.

TABACHNICK, B. and FIDELL, L. (2007). *Using multivariate analysis*. Needham Heights: Allyn & Bacon.

VERMUNT, J. K.; MAGIDSON, J. (2004). "Factor Analysis with categorical indicators: A comparison between traditional and latent class approaches". In: VAN DER ARK, A.; CROON, M.A. e SIJTSMA, K. *New Developments in Categorical Data Analysis for the Social and Behavioral Sciences*. Mahwah: Erlbaum.

YALCIN, I and AMEMIYA, Y. (2001). Nonlinear Factor Analysis as a Statistical Method. *Statistical Science*, 16, 3, 275-294.

ZELLER, R. A. & CARMINES, E. G. (1980). *Measurement in the social sciences: The link between theory and data*. Cambridge: Cambridge University Press.

Sites pesquisados:

<http://www.gog.pol.gu.se/data/> Acesso em: 10 jan.2013

<http://davidakenny.net/cm/causalm.htm> Acesso em: 18 jan.2013

<http://www2.gsu.edu/~mkteer/semfaq.html> Acesso em: 22 jan.2013

Autores.

Ranulfo Paranhos.

Universidade Federal de Alagoas (UFAL) / Universidade Federal de Pernambuco (UFPE). Brasil.
Professor do Instituto de Ciências Sociais de Universidade Federal de Alagoas (ICS/UFAL),
Doutorando e Mestre em Ciência Política pelo Departamento de Ciência Política da Universidade
Federal de Pernambuco (DCP/UFPE). Brasil.

E-mail: ranulfoparanhos@me.com

Dalson Britto Figueiredo Filho.

Universidade Federal de Pernambuco (UFPE). Brasil.
Professor do Departamento de Ciência Política da Universidade Federal de Pernambuco (DCP/UFPE),
Doutor e Mestre em Ciência Política pelo Departamento de Ciência Política da Universidade Federal
de Pernambuco (DCP/UFPE). Brasil

E-mail: dalsonbritto@yahoo.com.br

Enivaldo Carvalho da Rocha.

Universidade Federal de Pernambuco (UFPE). Brasil.
Professor do Departamento de Ciência Política da Universidade Federal de Pernambuco (DCP/UFPE),
Pós-doutorando do Departamento de Ciência Política da Universidade Federal de Minas Gerais
(DCP/UFMG) e Doutor em Estatística pela Universidade de São Paulo (USP). Brasil.

E-mail: enivaldocrocha@gmail.com

José Alexandre da Silva Junior

Universidade Federal de Goiás (UFG) / Universidade Federal de Pernambuco (UFPE). Brasil.
Professor do Curso de Ciências Sociais de Universidade Federal de Goiás (UFG), Doutor e Mestre em
Ciência Política pelo Departamento de Ciência Política da Universidade Federal de Pernambuco
(DCP/UFPE). Brasil.

E-mail: jasjunior2007@yahoo.com.br <mailto:jymena@terra.com.pe>

Citado.

PARANHOS, Ranulfo; DALSON BRITTO FIGUEIREDO FILHO; ENIVALDO CARVALHO DA ROCHA y José
Alexandre DA SILVA JUNIOR (2013). "Corra que o survey vem aí. Noções básicas para cientistas
sociais". *Revista Latinoamericana de Metodología de la Investigación Social - ReLMIS*. N° 6. Año 3.
Oct. 2013 - Marzo 2014. Argentina. Estudios Sociológicos Editora. ISSN 1853-6190. Pp. 07 - 24.
Disponível em: <http://www.relmis.com.ar/ojs/index.php/relmis/article/view/74>

Plazos.

Recibido: 10 / 03 / 2013. Aceptado: 10 / 06 / 2013.