



O problema da amostragem no contexto da Educação Estatística Crítica

Celso Ribeiro **Campos**

Pontifícia Universidade Católica de São Paulo

Brasil

crcampos@pucsp.br

Cileda Queiroz e Silva **Coutinho**

Pontifícia Universidade Católica de São Paulo

Brasil

cileda@pucsp.br

Resumo

Este artigo é concebido no contexto do universo das pesquisas eleitorais. Nosso objetivo é investigar o processo de amostragem de um instituto de pesquisa e discutir seus conceitos no âmbito da Teoria da Amostragem. Na linha da Educação Estatística Crítica e do letramento estatístico, vamos complementar esta investigação com uma proposta de atividade pedagógica acerca do tema de amostragem, fazendo uso da estratégia didática da modelagem matemática. Investigamos uma pesquisa do instituto IBOPE, um dos maiores do Brasil. Constatamos alguns problemas metodológicos: equívoco na informação da margem de erro; estágio não probabilístico (amostragem por quotas); e falta de informação sobre os não respondentes. A atividade pedagógica sugerida contém uma pesquisa a ser feita pelos alunos e um debate sobre a problemática da metodologia e sobre os aspectos políticos das eleições presidenciais.

Palavras chave: metodologia de pesquisa, letramento estatístico, educação estatística crítica, modelagem matemática, eleição presidencial.

Introdução

O período eleitoral no Brasil é costumeiramente uma época de debates, de apresentação de propostas de melhorias para o país, de discussões ideológicas e de, principalmente, acusações, disputas e animosidades entre os políticos e entre os eleitores. O que deveria ser um momento profícuo para discussão de ideias passa a ser palco para veiculação de notícias falsas, de acusações sem provas e até, como no caso recente da eleição presidencial de 2018 no Brasil, tentativa de homicídio de um dos candidatos ao cargo maior da nação.

Outro item que parece ser um combustível para essa fogueira eleitoral são as pesquisas de intenção de voto, que na reta final das eleições protagonizam intensas discussões, definem estratégias e encaminham acordos para um hipotético segundo turno das eleições.

As pesquisas são igualmente objetos de acaloradas discussões e acusações de viesamento ou falseamento de resultados. Na medida em que um candidato se vê desfavorecido no resultado de uma pesquisa, parece ser mais fácil culpar o agente pesquisador do que refletir sobre seu comportamento e/ou suas estratégias.

É nesse contexto que concebemos este artigo, ou seja, o universo das pesquisas eleitorais. Nosso objetivo é investigar o processo de amostragem de um grande instituto brasileiro de pesquisas e discutir seus conceitos e metodologia no âmbito da Teoria da Amostragem.

Na linha da Educação Estatística Crítica e do letramento estatístico, entendemos que esta investigação poderá compor um material profícuo para o trabalho em sala de aula, na medida em que vamos complementar o trabalho com uma proposta de atividade didática acerca do tema de amostragem, fazendo uso da estratégia pedagógica da modelagem matemática.

Quadro teórico estatístico

Segundo Witte & Witte (2005, p. 154), “Qualquer conjunto completo de observações (ou potenciais observações) pode ser caracterizado como uma população”. Qualquer subconjunto de observações obtidas com base em uma população pode ser caracterizado como amostra.

Em aplicações típicas da estatística inferencial, o tamanho da amostra é pequeno em relação ao tamanho da população. Menos de 1% de todos os domicílios dos EUA estão incluídos na pesquisa mensal do *Bureau of Labor Statistics* para estimar a taxa atual de desemprego. Embora cerca de 4.000 eleitores, no máximo, tenham sido selecionados em uma pesquisa eleitoral para candidatura à presidência da república, recentemente realizada pelo Gallup, as previsões têm se mostrado surpreendentemente precisas [...] (WITTE & WITTE, 2005, p. 155).

Existem várias técnicas de amostragem, que podem ser classificadas em probabilísticas e não probabilísticas. Para que seja possível utilizar técnicas da estatística inferencial para a análise dos resultados de uma pesquisa, é necessário que a amostragem seja probabilística. Entre as diversas modalidades de amostragem probabilística, vamos destacar quatro: a amostragem aleatória simples (AAS), a amostragem estratificada, a amostragem por conglomerados e a amostragem sistemática.

Witte & Witte (op. cit.) destacam que “uma amostra é aleatória se, em cada estágio da amostragem, o processo de seleção garantir que todas as observações remanescentes na população possuem iguais chances de virem a ser incluídas na amostra”.

Quando a população é dividida em grupos (subpopulações), que em relação ao objeto pesquisado têm comportamento homogêneo dentro de si e heterogêneo entre os grupos, então é recomendável que a amostragem seja feita por meio de um critério de estratificação. As subpopulações representam os estratos, que devem ser quantificados como proporções da população. A amostra deve ter representantes de todos os estratos, na mesma proporção em que eles ocorrem na população. A seleção dos elementos dentro de cada estrato deve ser aleatória.

Embora alguns países desenvolvidos adotem a AAS para suas pesquisas de intenção de voto, no Brasil predomina a estratificação devido à grande desigualdade social da população. Os estratos comumente utilizados nas pesquisas de intenção de voto são: gênero, idade, nível de instrução, renda e região geográfica.

Quando a população é dividida em subpopulações (conglomerados) que apresentam comportamento homogêneo entre si, é possível sortear um conglomerado e fazer a pesquisa toda ali. Também é possível sortear diversos conglomerados e neles fazer uma estratificação.

Quando a população já se encontra ordenada, é possível estabelecer um intervalo de seleção aleatório, isto é, se, digamos, queremos uma amostra com 10% da população, então devemos selecionar um em cada 10 indivíduos.

Métodos não probabilísticos correspondem a amostragens em que há uma escolha deliberada dos elementos da amostra. Em geral, nesses casos não é possível generalizar os resultados das pesquisas para a população. Existem basicamente dois tipos de amostragem não probabilística, a intencional (por conveniência) e a acidental (por quotas).

A amostragem é intencional quando se faz a inferência com grupos que estão mais acessíveis, mais à mão, tais como pessoas da vizinhança, do trabalho, família e amigos, pessoas provenientes de listas de *mailing* e grupos de redes sociais, por exemplo.

A amostragem é acidental quando se infere pessoas que passam pelo pesquisador em um determinado local e um determinado momento. Normalmente esse tipo de pesquisa pode ser feito com a divisão da população em estratos, compondo quotas para o pesquisador preencher. Desse forma, a amostragem acidental é disfarçada de amostragem estratificada e, como as pessoas são abordadas na rua sem uma seleção específica, é também disfarçada de aleatório, embora não o seja. Sobre esse tipo de pesquisa, More et al. (2016) afirma que sempre redundante em um estudo viesado que, sistematicamente, favorece determinados resultados,

Os resultados de uma pesquisa não probabilística são locais, ou seja, são válidos para o universo pesquisado e não podem ser generalizados para a população como um todo.

Moore (op. cit., p. 163) aponta a melhor forma de fazer a amostragem:

Em uma amostra de resposta voluntária, as pessoas escolhem se respondem. Em uma amostra de conveniência, o entrevistador faz a escolha. Em ambos os casos, a escolha pessoal produz viés. A solução do estatístico é deixar que o acaso impessoal escolha a amostra. Uma amostra escolhida ao acaso não permite nem favoritismo por quem faz a amostra, nem autosseleção por parte de quem responde. Escolher uma amostra ao acaso ataca o viés, atribuindo a todos os indivíduos a mesma chance de serem escolhidos. Rico o pobre, jovem ou velho, negro ou branco, todos têm a mesma chance de estar na amostra.

Outra questão levantada por Moore (op. cit.) é a porcentagem de não resposta, ou seja, de indivíduos selecionados para participar da pesquisa e que se recusam a responder. Tal porcentagem, se alta, segundo o autor, pode provocar viés nos resultados da pesquisa.

Em resumo, os principais problemas relacionados aos resultados de pesquisas amostrais estão ligados à metodologia adotada, ou seja, amostra voluntária, acidental ou de conveniência, além de viés de não resposta. Os problemas relacionados à confiabilidade dos resultados de pesquisas feitas por meio de amostragens não se resumem a isso, mas esses casos já são suficientes para as análises que vamos tecer sobre as pesquisas eleitorais do Brasil.

Segundo Levine et al. (2016), sendo π a proporção desconhecida da população, sua estimativa é a proporção amostral p , tal que a estimativa do intervalo de confiança de π é dada por:

$$\pi = p \pm Z_{\alpha/2} \cdot \sqrt{\frac{p \cdot (1-p)}{n}}$$

Nessa equação, α é o nível de significância correspondente ao nível de confiança de $(1-\alpha)\%$. Nesse caso, $\alpha/2$ é a proporção em cada cauda da curva normal que fica fora do nível de confiança. Em pesquisas de intenção de voto, costuma-se utilizar um nível de confiança de 95%, que corresponde a um $Z_{\alpha/2} = 1,96$, conforme a tabela da curva normal (LEVINE et al., op. cit.).

Como a proporção amostral p pode variar, $p(1 - p)$ pode assumir diferentes valores. As pesquisas de intenção de voto costumam adotar o valor máximo possível para esse produto, que ocorre quando $p = 0,5$. Assim, se considerarmos uma amostra de tamanho $n = 2000$, temos a seguinte margem de erro ao nível de confiança de 95%:

$$ME = 1,96 \times \sqrt{\frac{0,5 \cdot (1 - 0,5)}{2000}} = 1,96 \times 0,01118 = 0,0219$$

Esse resultado corresponde aproximadamente a uma margem de erro de 2,2%. Para obter exatos 2% de margem de erro, é necessário que a amostra seja de 2.401 observações. Essa margem de erro (2%) é a que é comumente adotada pelos institutos de pesquisa.

Quadro teórico de Educação Estatística

A ideia de letramento estatístico nos remete à noção de alfabetização funcional em Estatística, ou seja, a capacidade de ler, escrever, interpretar criticamente os conceitos e os resultados estatísticos, argumentar usando terminologia própria, explicar seus pontos de vista, etc. Gal (2002) advoga que:

[...] é proposto que o termo ‘letramento estatístico’ se refere amplamente a dois componentes inter-relacionados, principalmente (a) a capacidade das pessoas de interpretar e avaliar criticamente informações estatísticas, argumentos relacionados a dados ou fenômenos estocásticos, os quais podem ser encontrados em contextos diversos e quando relevante; (b) sua capacidade de discutir ou comunicar suas reações a tais informações estatísticas, tais como sua compreensão do significado da informação, suas opiniões sobre as implicações dessa informação ou suas preocupações quanto à aceitabilidade de dados e conclusões (GAL, 2002, pp. 2-3)

Em Coutinho e Campos (2018), defendemos que o letramento pode ser estimulado nos estudantes por meio de exemplos de atividades contextualizadas que favoreçam o engajamento dos alunos, a mobilização de conhecimentos, habilidades e atitudes. Esses exemplos devem ser extraídos da realidade e construídos com base em notícias veiculadas na mídia ou a partir de temas de interesse dos alunos, tais como, levantamento de preços, dados relativos a esportes, dados populacionais, etc.

Em Campos et al. (2011, pp. 25-26), destacamos uma forma de estimular o desenvolvimento do letramento estatístico:

Dar aos estudantes a oportunidade de produzir os próprios dados e encontrar os resultados básicos ajuda-os a tomar as rédeas de seu próprio aprendizado. Também promove a habilidade de assumir a responsabilidade de resolver seus problemas, como eles terão que fazer em seu ambiente de trabalho. É possível solicitar aos estudantes que não apenas colem os seus dados, mas, igualmente, elaborem as variáveis que irão compor seus questionários. Isso os ajuda a descobrir ou determinar métodos e técnicas por si próprios.

Assim, pretendemos construir uma atividade didática para favorecer o desenvolvimento do letramento estatístico por meio do uso da Estatística como evidência para assuntos e argumentos encontrados em sua vida diária como consumidores de informações veiculadas pelas diversas mídias. Essa atividade que vamos mostrar está ligada à problemática das pesquisas de opinião a respeito das eleições no Brasil.

Sobre a educação estatística crítica

Os princípios da Educação Crítica e da Educação Matemática Crítica, apresentados nos trabalhos de Freire (1970), Giroux (1997) e Skovsmose (2004), foram adaptados por Campos (2016), que compôs a chamada educação estatística crítica, cujos princípios são:

- i) Contextualizar os dados de um problema estatístico, preferencialmente utilizando dados reais;
- ii) Encorajar a interpretação e análise dos resultados obtidos;
- iii) Inserir os assuntos tratados em um contexto social/político e promover discussões sobre os problemas levantados.

Além disso, em Campos (2016) encontramos outras características importantes da educação estatística crítica, tais como: a promoção de uma educação problematizadora, dialógica e que estimula a criatividade e a reflexão do aluno; a valorização dos aspectos políticos envolvidos na educação, tanto em relação ao processo educativo como em relação aos conteúdos disciplinares; o desenvolvimento dos relacionamentos sociais, o combate às posturas alienantes dos alunos e a defesa da ética e da justiça social; a democratização do ensino; etc.

Com isso, defendemos que devemos conjugar a ideia de ensinar Estatística sempre acompanhada de um objetivo maior de promover a criticidade e o engajamento dos estudantes em questões políticas, sociais, econômicas ou ambientais que fazem parte de sua realidade como cidadãos, de modo a encorajar o desejo de justiça social em um ambiente democrático e desalienado.

Metodologia

Neste trabalho de investigação, vamos analisar uma pesquisa de intenção de voto para presidente da república do Brasil, realizada pelo instituto IBOPE. Nossa intenção é investigar sua metodologia, comparando-as com o que foi exposto no nosso quadro teórico de Estatística.

Adicionalmente, a atividade didática que vamos sugerir segue a estratégia da modelagem matemática, que se articula bem com as ideias de letramento estatístico e de educação estatística crítica.

Resultados

A pesquisa analisada foi feita pelo IBOPE entre os dias 22 e 24 de setembro. Os resultados são mostrados na figura 1.

Sobre a metodologia da pesquisa, encontramos as seguintes informações:

O modelo de amostragem utilizado é o de conglomerados em 3 estágios.

No primeiro estágio os municípios são selecionados probabilisticamente através do método PPT (Probabilidade Proporcional ao Tamanho), tomando os eleitores que votaram nas últimas eleições (votantes) como base para tal seleção.

No segundo estágio são selecionados os conglomerados: setores censitários, com PPT (Probabilidade Proporcional ao Tamanho) sistemático. A medida de tamanho é o número de votantes dos setores.

Finalmente, no terceiro estágio é selecionado em cada conglomerado um número fixo de votantes segundo cotas de variáveis [...]

Foram realizadas 2000 entrevistas em 126 municípios. A margem de erro máxima estimada é de 2 pontos percentuais para mais ou para menos sobre os resultados encontrados no total da amostra. O nível de confiança utilizado é de 95% (IBOPE, 2018, p. 2)

O método PPT não é descrito em nenhuma das fontes que consultamos. Em uma busca na internet, encontramos referências no website do INPAD (2012) (Instituto Nacional de Ciência e Tecnologia para Políticas Públicas do Álcool e Outras Drogas). Vimos que a probabilidade

proporcional ao tamanho (PPT) é um método de seleção aleatória no qual são atribuídos pesos a cada município de acordo com a sua população, de forma que se um município A tem população x e outro município B tem população $2x$, a chance de o município B ser selecionado é duas vezes maior que a do município A. Isso compõe o primeiro estágio da pesquisa do IBOPE. No segundo estágio, são designados os conglomerados ou setores censitários, ou seja, regiões dos municípios selecionados no primeiro estágio. Considerando a ponderação chamada de PPT, a seleção das regiões a serem pesquisadas é feita pelo método sistemático. O terceiro estágio é o mais sensível, no qual dentro de cada conglomerado é selecionado um número fixo de votantes segundo cotas de variáveis. As variáveis, nesse caso, são os estratos, que o IBOPE discrimina como sexo, idade, instrução e renda.

INTENÇÃO DE VOTO PARA PRESIDENTE DA REPÚBLICA
(Estimulada - Uma opção)
(%)

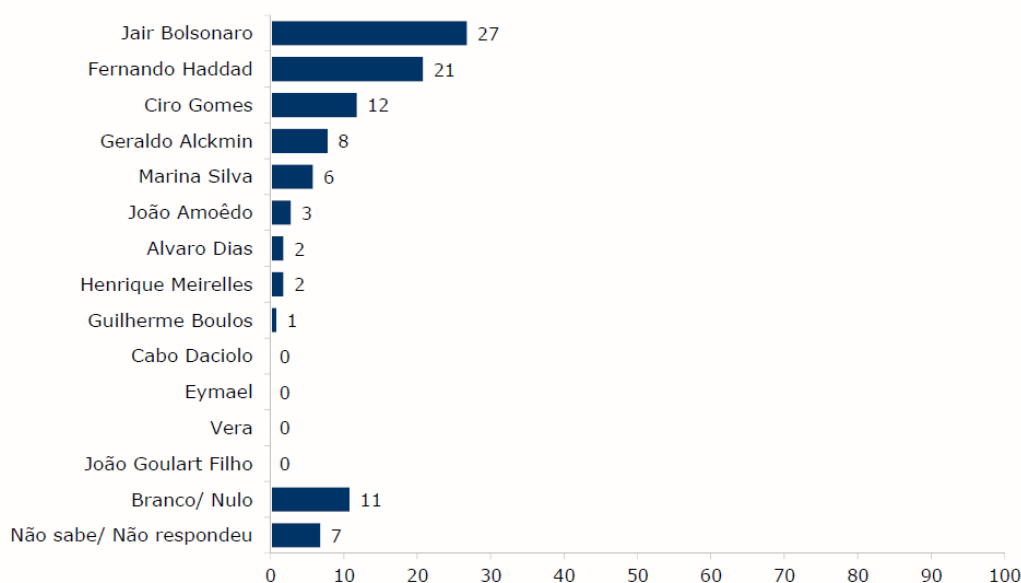


Figura 1: Resultado geral da pesquisa estimulada divulgada em 26/09/2018. Fonte: IBOPE (2018)

Não há qualquer informação no documento do IBOPE sobre como é feita a abordagem ao eleitor pesquisado. Como não há um estágio de sorteio de residências, podemos supor que o pesquisador aborda pessoas na rua, buscando preencher suas quotas. Esse tipo de amostragem compõe um estágio não probabilístico de pesquisa, no qual o pesquisador escolhe quem vai pesquisar, produzindo naturalmente um viés. Como já dissemos no quadro teórico (MOORE et al., 2016), tanto no caso em que o pesquisador faz a escolha de quem entrevistar quanto na situação em que a pessoa escolhida decide se quer responder ou não, há um viés de resposta que pode comprometer o resultado da pesquisa.

A pesquisa do IBOPE falha principalmente nos seguintes aspectos: não compõe amostras probabilísticas em seu terceiro estágio ao fazer o levantamento por quotas; não divulgam a correta margem de erro, que é de 2,2% ao invés de 2% e não informa a porcentagem de não resposta.

De acordo com Moore (2016), o levantamento por quotas corresponde a um planejamento amostral ruim que resulta em um erro sistemático, gerando viés nos resultados. Talvez o viés não

seja tão dramático (e possivelmente o instituto de pesquisa aposte nisso) dado que as duas primeiras etapas do processo de amostragem são feitas com métodos probabilísticos.

Proposta didática

No âmbito do assunto tratado neste artigo, pensamos em propor uma atividade didática baseada na modelagem matemática para trabalhar o tema de pesquisas de intenção de votos.

Nossa sugestão é válida tanto para o ensino fundamental como para o ensino médio e o superior. A ideia é dividir a classe em grupos (os próprios alunos formam os grupos) e pedir para eles obterem dados a respeito da intenção de voto junto a pessoas próximas, ou seja, família, amigos, vizinhos, etc. Depois de obtidos os dados, o grupo deve preparar um relatório descrevendo a amostra e contendo uma tabela e um gráfico com os resultados obtidos. Posteriormente, na data combinada para a entrega do relatório, o professor, com a ajuda dos alunos, junta todos os relatórios e mostra um resultado unificado. Aos alunos pode ser atribuída uma meta de entrevistar 20 pessoas cada. Se assim for feito, uma sala de 40 alunos produzirá uma amostra de 800 pessoas.

Sugerimos comparar os resultados unificados com as pesquisas publicadas pelos institutos para começar a discutir os princípios de metodologia de pesquisa. É esperado que o resultado dos alunos apresente um viés causado pela amostragem por conveniência (não probabilística), o que deve ser evidenciado pelo professor.

Por fim, a atividade pode gerar uma conversa sobre política e o professor pode pedir para os alunos fazerem campanha para seus candidatos ou até simular um debate. Isso é uma oportunidade para discutir quais são os problemas do país e quais são as propostas dos candidatos para enfrenta-los.

Evidentemente, essa atividade é indicada para períodos próximos a eleições, que no Brasil ocorrem a cada dois anos. Em épocas outras, poderá ser difícil encontrar pesquisas de opinião que fomentem tal debate político na mídia.

Considerações finais

Neste artigo, tínhamos estabelecido como objetivo investigar o processo de amostragem de uma pesquisa eleitoral e discutir seus conceitos à luz da teoria da amostragem estatística. Examinamos a metodologia, comparando os procedimentos adotados com o que nosso quadro teórico sugeria e notamos três problemas, a saber: equívoco na designação da margem de erro; adoção de um estágio não probabilístico na pesquisa; e omissão sobre não respondentes.

Outro objetivo deste estudo era propor uma atividade didática sobre o tema, o que fizemos adotando a estratégia da modelagem matemática. Sugerimos comparar os resultados dos alunos com pesquisas publicadas na mídia para evidenciar a problemática da amostragem. Por fim, tratamos de valorizar a Educação Estatística Crítica quando propusemos um debate político sobre os candidatos mencionados nas pesquisas.

Acreditamos que tal atividade didática favorece o desenvolvimento do letramento estatístico na medida em que sugerimos avaliar criticamente as informações das pesquisas, discutir e comunicar as interpretações (GAL, 2002), além de se trabalhar uma atividade contextualizada, favorecendo o engajamento dos estudantes (COUTINHO & CAMPOS, 2018).

Desta feita, acreditamos ter contribuído para a disseminação de conhecimentos relevantes para a compreensão da realidade brasileira em um momento importante de definição política caracterizado pela eleição presidencial.

Bibliografia

- Campos, C. R. (2007). *Educação estatística: uma investigação acerca dos aspectos relevantes à didática da Estatística em cursos de graduação*. 242 f. Tese (Doutorado em Educação Matemática). Rio Claro: UNESP.
- Campos, C. R. (2016). *Towards critical statistics education – theory and practice*. Saarbrücken/Alemanha: Lambert Academic Publishing.
- Campos, C. R.; Wodewotzki, M. L. L.; Jacobini, O. R. (2011). *Educação estatística – teoria e prática em ambientes de modelagem matemática*. Belo Horizonte: Autêntica.
- Coutinho, C. Q. S.; CampOS, C. R. (2018). Perspectivas em didática e educação estatística e financeira: reflexões sobre convergências entre letramento matemático, matemacia, letramento estatístico e letramento financeiro. In: G. P. Oliveira (org.), *Educação Matemática – epistemologia, didática e tecnologia*, pp. 143-180. São Paulo: Livraria da Física.
- Freire, P. *Pedagogia do oprimido*. Rio de Janeiro: Paz e Terra, 1970.
- Gal, I. (2002). Adult's Statistical literacy: meanings, components, responsibilities. *International Statistical Review*, 70(1), 1-25.
- Giroux, H. A. (1997). *Os professores como intelectuais: rumo a uma pedagogia crítica*. Trad. Daniel Bueno. Porto Alegre: Artes Médicas.
- IBOPE. (2018). *Pesquisa de opinião pública sobre assuntos políticos/administrativos - Job 1152*. Registro no TSE: BR-04669/2018 Ibope Inteligência. Disponível em: <http://www.ibopeinteligencia.com/noticias-e-pesquisas/jair-bolsonaro-mantem-lideranca-e-diferenca-de-6-pontos-percentuais-com-relacao-a-fernando-haddad/>. Acesso em 29/09/2018.
- INPAD. (2012). II LENAD – *Levantamento Nacional de Álcool e Drogas*. Disponível em: <https://inpad.org.br/lenad/sobre-o-lenad-ii/metodologia/>. Acesso em 29/09/2018.
- Levine, D. M.; Stephan, D. F.; Szabat, K. A. (2016). *Estatística – teoria e aplicações*. Rio de Janeiro, LTC.
- Moore, D. S.; William, I. N.; Fligner, M. A. (2016). *A Estatística básica e sua prática*. 6ª ed. Rio de Janeiro, LTC.
- Skovsmose, O. (2004). *Educação Matemática Crítica – A Questão da Democracia*. 2ª ed. Campinas: Papirus.
- Witte, R. S.; Witte, J. S. (2005). *Estatística*. 7ª ed. Rio de Janeiro: LTC.?