

MÉTODOS QUANTITATIVOS E PESQUISA CONTÁBIL: UM ESTUDO DE CASO RELACIONADO A PEQUENAS AMOSTRAS DE DADOS

QUANTITATIVE METHODS AND ACCOUNTING RESEARCH: A CASE STUDY RELATED TO SMALL DATA SAMPLES

Carlos Roberto Souza Carmo

Mestre em Ciências Contábeis (PUC-SP)
Universidade Federal de Uberlândia (FACIC-UFU)
carlosjj2004@hotmail.com

Adriano Dawison de Lima

Doutor em Agronomia - Energia na Agricultura (UNESP-BOTUCATU)
Universidade de Uberaba (UNIUBE)
adawison@bol.com.br

Resumo:

A presente pesquisa teve por objetivo analisar como certa metodologia quantitativa aplicada ao processo de amostragem, denominada de processo de reamostragem ou método *bootstrap*, pode auxiliar na solução de problemas relacionados às pequenas amostras de dados utilizadas em estudos de natureza científica, na área das ciências contábeis. Para tanto, a partir de um estudo de caso real, além da aplicação do processo de reamostragem pelo método de *bootstrap*, foram utilizadas estatísticas descritivas baseadas em medidas de tendência central e de dispersão, gráficos, e, ainda, testes de normalidade, para a validação do conjunto de evidências coletadas neste estudo. Ao final, foi demonstrado que a aplicação do método de *bootstrap* pode ser caracterizada como uma alternativa ao processo de amostragem nos estudos de problemas científicos de natureza contábil cuja amostra de dados pode ser considerada pequena em relação à respectiva população, e, ainda, foram observados, testados e confirmados dois dos pressupostos apresentados anteriormente por estudiosos da metodologia de amostragem de *bootstrap*.

Palavras-chave: Amostragem. *Bootstrap*. Métodos quantitativos aplicados.

Abstract:

The present research had as objective to analyze how a certain quantitative methodology applied to the sampling process, denominated of resampling process or bootstrap method, can help in the solution of problems related to small data samples used in scientific studies, in the area of accounting sciences. To do so, based on a real case study, besides the application of the resampling process by the bootstrap method, we used descriptive statistics based on measures of central tendency and dispersion, graphs, and, also, tests of normality, for the Validation of the evidence set collected in this study. In the end, it was demonstrated that the application of the bootstrap method can be characterized as an alternative to the sampling process in the studies of scientific problems of an accounting nature whose data sample can be considered small in relation to the respective population, Test and confirm two of the assumptions previously presented by scholars of the bootstrap sampling methodology.

Keywords: Sampling. Bootstrap. Quantitative methods applied.

- Resumo do processo editorial:**
- a) Submissão em: 10/01/2017.
 - b) Envio para avaliação em: 11/03/2017.
 - c) Término da avaliação em: 17/03/2017.
 - d) Correções solicitadas em: 20/04/2017.
 - e) Recebimento da versão ajustada em: 15/05/2017.
 - f) Aprovação final em: 06/06/2017.

Abstract:

Basic sanitation services are considered relevant for the promotion of public health and socioeconomic development. This study aims to analyze the correlations between efficiency in the allocation of public resources in basic sanitation and health, education, income and urbanization rate in Minas Gerais municipalities. For this, the statistical methods of Data Envelopment Analysis and Spearman Correlation were used. The results point to a low efficiency in the allocation of resources in basic sanitation for most of the municipalities analyzed. Access to the sewage service is the most critical among the variables that make up in the study. Only in municipalities with a high level of efficiency is it possible to observe a positive correlation between sanitation and health development. Based on the results, it is concluded that there is a need for improvements in the management of resources used in basic sanitation. In addition, in order to improve access to basic sanitation services, intersectoral policies are needed, especially with health and education, since, as noted, efficient municipalities in basic sanitation show greater municipal development in the health area..

Keywords: Basic sanitation. Data Envelopment Analysis. Municipal Development.

1 Introdução

Em linhas muito gerais, costuma-se dizer que as diferenças básicas entre as pesquisas qualitativas e aquelas pesquisas de caráter predominantemente quantitativo residem no objeto de estudo, no número de variáveis e/ou observações realizadas, e, na metodologia analítica adotada, bem como, no binômio “profundidade *versus* amplitude” do estudo de determinado objeto em si.

Segundo esse raciocínio, ainda em linhas muito gerais, afirmar-se-ia que os estudos científicos de caráter quantitativo avaliam números, ao passo que, os estudos qualitativos avaliam conceitos.

Em relação ao número de variáveis e/ou observações realizadas pode-se dizer que nos estudos qualitativos existe a predominância de uma quantidade de variáveis e/ou observações muito elevada, ao passo que, nos estudos qualitativos, esse número é bem mais reduzido.

Em relação à metodologia analítica, parece evidente que os estudos de natureza quantitativa valem-se de métodos quantitativos aplicados, normalmente, estatística, sendo que, os estudos qualitativos tendem a utilizar outras metodologias.

Em função de tais diferenças, também pode-se afirmar que quanto maior a amplitude de determinado estudo, como acontece com estudos de natureza quantitativa, menor tende a ser sua profundidade, ou o inverso, que também verdadeiro, quanto menor sua amplitude, maior tende a ser a profundidade, como normalmente acontece com os estudos de caráter qualitativo.

Contudo, será que ao combinar um objeto de estudo científico de natureza qualitativa com métodos de análise quantitativa, poder-se obter profundidade e amplitude em elevada escala? A resposta não é muito difícil. Ou seja, ao admitir que esses dois fatores não sejam mutuamente excludentes, contudo, admitindo-se também que a presença de um tende a reduzir a intensidade do outro, pode-se buscar certa dose de equilíbrio, considerada satisfatória de acordo com o respectivo objeto de estudo.

Aplicando-se tal raciocínio à ciência contábil, que se situa em uma área do conhecimento social aplicado, pois, apesar do seu objeto de estudo ser o patrimônio das entidades, mensurável monetariamente, ele sofre alterações a partir das ações humanas, pode-

se vislumbrar não só aquela possibilidade levantada no parágrafo anterior, mas, considerar tal possibilidade como uma necessidade.

Isso é, uma vez que a contabilidade tem por objeto de estudo o patrimônio das entidades, cuja mensuração implica na utilização de metodologias quantitativas, porém, tal patrimônio tem suas mutações geradas periodicamente a partir da tomada de decisões humanas, a busca por aquele equilíbrio parece evidentemente necessária.

Por outro lado, se a utilização de métodos quantitativos proporciona amplitude, contudo, diminui a profundidade dos estudos, devido à elevada quantidade de observações demandadas, o problema em questão se deslocaria fatalmente para a quantidade de dados a serem levantados (observações) acerca de determinado fenômeno de estudo.

A temática relacionada ao problema do número de observações demandado no processo de análise quantitativa de determinado objeto de estudo, que é tratado pela estatística como amostragem, diz respeito à identificação do comportamento acerca de um todo (universo maior de determinado objeto de estudo, denominado de população) a partir de uma parcela menor desse todo (denominado de amostra).

Nesse sentido, o processo de escolhas de observações para composição da amostra de uma pesquisa, ou simplesmente amostragem conforme denomina a ciência estatística, é fator crítico, sendo que, nem sempre um grande número de observações é garantia de sucesso acerca do estudo do comportamento de determinado grupo de objetos, indivíduos, fenômenos, etc.

Por outro lado, ao considerar que quanto maior a quantidade de observações integrantes de uma amostra de pesquisa, maior a possibilidade de obter sucesso no estudo de determinada população, deve-se admitir também que, quanto menor o número de observações, mais limitadas tendem a ser as possibilidades de se obter sucesso no delineamento do perfil da população de determinado objeto de estudo.

Ainda deve ser destacado, como fator agravante, o fato de que a obtenção de amostras muito numerosas pode ser economicamente inviável quando o objeto de estudo apresenta um alto custo para sua obtenção ou é de difícil mensuração. Em outro sentido, a realização de estudos a partir de critérios de amostragem mais simples ou em pouca quantidade pode levar a estimativas imprecisas.

Ao considerar o patrimônio das entidades, enquanto objeto de estudo da ciência contábil, adiciona-se ainda uma variável agravante a mais àquele contexto exposto no parágrafo anterior; isto é, o sigilo relacionado às informações das empresas dificulta ainda mais a obtenção de amostras mais numerosas para a realização de estudos de natureza contábil.

Nesse contexto, a presente pesquisa tem por objetivo analisar como metodologia quantitativa aplicada ao processo de amostragem, denominada de processo de reamostragem ou método *bootstrap*, pode auxiliar na solução de problemas relacionados às pequenas amostras de dados utilizadas em estudos de natureza científica, na área das ciências contábeis.

Para tanto, formulou-se o seguinte questionamento de pesquisa: como o processo de reamostragem pelo método *bootstrap* pode auxiliar na solução de problemas relacionados a pequenas amostras de dados utilizadas em estudos de natureza científica, na área das ciências contábeis?

A partir daquele questionamento direcionador, buscou-se constituir a plataforma teórica sob a qual foi conduzida esta pesquisa, em que, foi descrita e analisada a metodologia de reamostragem pelo método *bootstrap*, originando-se a segunda seção deste artigo. Na sequência, conforme relatado na terceira seção deste trabalho, foi identificado e proposto um objeto de estudo de natureza científica, cuja amostra de dados pudesse ser considerada pequena, em relação à respectiva população de dados, e, no qual, sua solução fosse

desenvolvida a partir da utilização de métodos quantitativos aplicados. A quarta seção deste artigo, subdivida em duas seções secundárias, foi destinada à análise dos dados e apresentação dos resultados, respectivamente. Finalmente, na quinta e última seção, foram apresentadas as considerações acerca de todo este processo de investigação científica, bem como, suas limitações e sugestões para a sua continuidade.

2 Referencial Teórico

A forma como um conjunto de dados é distribuída constitui-se em um fator direcionador da natureza dos testes e metodologias analíticas a serem empregados nos estudos quantitativos.

Existem várias formas de distribuição de dados em torno da respectiva média populacional e/ou amostral, considerando ainda o respectivo desvio padrão. Mas, dentre tais formas de distribuição, a mais conhecida é distribuição normal ou simétrica.

Stevenson (1981, p. 136) observa que “as distribuições normais ocupam posição proeminente tanto na estatística teórica como na aplicada, por várias razões”, e, ainda, faz especial destaque aos seguintes motivos:

[...]com bastante frequência, elas representam, com boa aproximação, as distribuições de frequência observadas de muitos fenômenos naturais e físicos. Outra razão é que as normais servem como aproximação de probabilidades binomiais, quando n é grande.

Todavia, o motivo mais importante da proeminência da distribuição normal é que as distribuições tanto das médias como das proporções em grandes amostras tendem a ser distribuídas normalmente, o que tem relevante implicação na amostragem (STEVENSON, 1981, p. 136).

[...]

Outra característica importante é que uma distribuição normal fica completamente especificada por dois parâmetros: sua média e seu desvio padrão. Em outras palavras, existe uma única distribuição normal para cada combinação de uma média e um desvio padrão. Diferentes combinações de média e desvio padrão originam curvas normais distintas. Como médias e desvios padrões são medidos em escala contínua, segue-se que o número de distribuições normais é ilimitado (STEVENSON, 1981, p. 137).

[...]

Uma consequência importante do fato de uma curva normal poder ser completamente especificada por sua média e seu desvio padrão é que a área sob a curva entre um ponto qualquer e a média é função somente do número de desvios padrões que aquele ponto dista da média. Esta é a chave que nos permite calcular probabilidades para a curva normal (STEVENSON, 1981, p. 138).

Conforme dito inicialmente, o formato assumido por uma curva representativa da distribuição de uma série de dados é um dos direcionadores acerca da categoria de testes a serem utilizados no processo de análise quantitativa. Nesse sentido, destacam-se duas categorias básicas, ou seja, os testes paramétricos e os não paramétricos.

Para adoção dos testes paramétricos, assume-se como premissa básica o fato da variável estudada assumir uma distribuição de frequência que se comporte segundo a distribuição da curva normal, também chamada de distribuição simétrica ou, ainda, curva de Gauss, entre outros fatores (STEVENSON, 1981).

Quando o formato da representação gráfica de uma distribuição de frequência assume um comportamento não definido, adotam-se os testes não paramétricos.

Conforme observam Reis e Ribeiro Júnior (2007), os testes paramétricos são considerados mais rigorosos devido às pressuposições necessárias para sua validação. Para

Callegari-Jacques (2003), os testes não paramétricos são aplicáveis a distribuições de frequência de comportamento (formato gráfico) desconhecido.

A despeito da preferência pela adoção de testes de natureza paramétrica devido a sua robustez, Stevenson (1981, p. 308) observa que, além de serem fáceis de aplicar e intuitivamente compreensíveis, os testes não paramétricos “[...] podem, pois, ser usados quando as hipóteses exigidas por outras técnicas não são satisfeitas, ou quando não é possível verificar essas hipóteses (em razão do pequeno tamanho da amostra).”

Ou seja, resgata-se aqui o problema relacionado ao tamanho da amostra a ser utilizada para a pesquisa e compreensão de determinado objeto de estudo, cujo método analítico escolhido é de natureza quantitativa.

Stevenson (1981) observa que as técnicas de amostragem têm por objetivo permitir a coleta de um grupo de dados (amostra) de forma a se conhecer o comportamento de um universo de dados (população).

Contudo, conforme destaca Praciano-Pereira (2008, p.104), “em geral se quer obter mais informações, mas obviamente existe um custo associado à quantidade de informações: mais pessoas levantando dados, mais tempo de busca, mais tempo de processamento.” Logo, ainda, segundo Praciano-Pereira (2008, p.104), “[...] precisamos de métodos mais eficientes para conseguir o melhor resultado com uma quantidade menor de informações [...]”.

Introduzido por Efron (1979), o método de amostragem de *bootstrap* busca se aproximar da distribuição original de uma população de dados, por meio de uma amostra finita, a partir de replicações dos elementos dessa amostra, obtendo-se quantas amostras adicionais (ou amostra de *bootstrap*) se desejar.

O método de *bootstrap* consiste na reamostragem de dados a partir de uma amostra mestre considerada relativamente pequena, ou melhor, ele consiste em replicar os elementos de uma amostra base, por meio da reamostragem, contudo, mantendo as características daquela amostra inicial (base).

Segundo Hesterberg *et al* (2003), a amostra mestre serve como parâmetro base acerca do comportamento da população da qual foi extraída, e, a amostra de *bootstrap*, obtida a partir do processo de reamostragem, passa a ser considerada a amostra de estudo, permitindo dessa forma avaliar qual o perfil da curva de distribuição dos dados da população, a partir da estimativa de uma quantidade muito elevada de amostras.

Considerando que a forma assumida por uma distribuição de dados em torno da respectiva média populacional e/ou amostral pode ser definida a partir do respectivo desvio padrão, o método de *bootstrap* mensura tanto esse desvio quanto outros parâmetros pelo princípio denominado *plug-in*.

Hesterberg *et al* (2003) afirmam que, segundo o princípio do *plug-in*, todos os parâmetros (média, mediana, desvio, etc.) da população são obtidos a partir do perfil da amostra originalmente extraída, ou amostra mestre, assumindo-se tal amostra base como população, e, a partir reamostragens, replicam-se seus parâmetros obtendo quantas amostras forem necessárias para o processo de análise.

A principal vantagem oriunda da aplicação do princípio do *plug-in* reside no fato dos parâmetros da distribuição amostral serem obtidos diretamente a partir dos dados da amostra mestre, o que, mediante o processo de replicação amostral, pode conferir precisão às estimativas de probabilidade baseada em grandes quantidades de dados amostrais, não correndo o risco dos erros de imprecisão oriundos da aplicação do Teorema do Limite Central à amostras consideradas pequenas.

Apesar de permitir produzir quantas observações se desejar a partir do processo de reamostragem pelo método de *bootstrap*, Montgomery, Peck e Vining (2001) ponderam que a quantidade ideal de reamostras, ou “amostras de *bootstrap*” conforme já dito, pode ser

identificada a partir do momento em que a variação do desvio padrão do parâmetro em análise se estabilizar, sendo importante observar que tal processo de reamostragem deve ser realizado mediante a reposição de observações a partir de seleção aleatória.

Já Effron e Tibshirani (1993) afirmam que a quantidade ideal de reamostras não precisa ser superior a 200 replicações amostrais aleatórias para identificar o desvio padrão em torno da média de um conjunto de observações de determinado dado ou variável em estudo. Sendo que, uma quantidade de reamostras superior a 200 replicações só seria demandada caso se desejasse conhecer o intervalo de confiança da amostra de *bootstrap* (EFFRON; TIBSHIRANI, 1993).

A despeito da utilidade do processo de amostragem pelo método de *bootstrap*, poder-se-ia imaginar que a *internet* seria solução para o problema relacionado ao número de observações de determinado amostra de dados, especialmente nas pesquisas envolvendo questionários para a coleta de dados *on-line*, como é o caso do estudo cuja proposta metodológica para solução será apresentada mais adiante na presente investigação.

Contudo, devido à baixa taxa de resposta desse tipo de pesquisa para coleta de dados (EVANS; MATHUR, 2005; ILIEVA; BARON; HEALEY, 2002; MALHOTRA, 2004; SCHONLAU; FRICKER JR.; ELLIOT, 2001; WILSON; LASKEY, 2003), novamente, depara-se com problemas relacionados às amostras com um número muito reduzido de observações, conforme já descrito inicialmente neste referencial teórico.

Dessa forma, espera-se que a aplicação do método de *bootstrap* em um estudo de caso real de natureza científica, cuja amostra de dados pode ser considerada pequena em relação à respectiva população, possa caracterizar-se como uma alternativa ao processo de amostragem nos estudos de problemas científicos de natureza contábil.

3 Metodologia

A presente seção foi dividida em quatro seções secundárias, de forma que a primeira delas foi destinada à descrição das informações referente ao estudo de caso único utilizado com base desta pesquisa. A segunda seção secundária foi destinada à descrição do processo de amostragem pelo método de *bootstrap* (reamostragem). A terceira seção descreve os procedimentos utilizados para análise dos dados levantados a partir do caso estudado nessa investigação e do respectivo processo de reamostragem. E, finalmente, a quarta seção secundária foi destinada à caracterização geral da presente pesquisa científica.

3.1 Descrição do Caso Estudado

Ao longo dos meses de agosto, setembro e outubro de 2012, foi realizada uma pesquisa junto aos profissionais de contabilidade com registro no Conselho Regional de Contabilidade do estado de Minas Gerais (CRC-MG) que, entre outros fatores, buscou avaliar como os contadores daquele estado percebiam sua familiaridade e sua necessidade de conhecimentos relacionados às *Internacional Financial Reporting Standards* (IFRS).

Mediante mensagem de e-mail enviada aos mais de 30.000 contadores do estado de Minas Gerais integrantes da base de dados dos registros do CRC-MG, e, ainda, a partir de um questionário composto de 7 partes e várias questões, hospedado na plataforma Google Docs®, foi solicitado aos respondentes que indicassem, entre outras informações, sua percepção quanto ao seu grau de familiaridade com a temática relacionada às IFRS, e, seu grau de necessidade de conhecimentos relacionados às IFRS. Sendo que, ambas as percepções deveriam ser expressas a partir de uma nota de 0 (zero) a 10 (dez), em que 0 (zero) indicava nenhuma familiaridade/necessidade e 10 (dez) total familiaridade/necessidade.

A pesquisa realizada em 2012 não contemplou somente as respostas dos contadores, mas, também, técnicos, professores e até alunos dos cursos de Ciências Contábeis daquele estado, conforme resumo descrito pela Tabela 1.

Tabela 1 - Perfil profissional dos integrantes da amostra da pesquisa realizada em 2012

Descrição	Frequência	
	(Absoluta)	(Relativa)
Contador com registro no Cadastro Nacional de Auditores Independentes (CNAI)	11	6,4%
Contador sem registro no Cadastro Nacional de Auditores Independentes (CNAI)	87	50,9%
Técnico em contabilidade	34	19,9%
Estudante	23	13,5%
Professor	16	9,4%
Total	171	100,0%

Fonte: elaborado pelos autores, com base nos dados da pesquisa.

Contudo, para o exercício a ser apresentado na presente investigação a título de estudo de caso, serão consideradas somente as respostas dos contadores (registro no CNAI), portanto, 98 profissionais (11 contadores com registro no CNAI + 87 contadores sem registro no CNAI).

Nesse sentido, considerando uma população superior a 30.000 contadores com registro no CRC-MG naquela época, a amostra em questão não atingiu nem 0,5% (meio por cento) da população de profissionais contadores, o que pode caracterizá-la como uma amostra muito pequena.

Conforme pode ser visto na Tabela 2, vale observar que, em termos médios, as respostas dos profissionais cujos dados integraram a amostra do caso estudado nesta pesquisa expressam coerência acerca da sua percepção.

Tabela 2 – Estatísticas descritivas da amostra pertencente ao estudo de caso desta pesquisa

Estatísticas descritivas	Percepção de familiaridade	Percepção de necessidade
Frequência absoluta	98	98
Mínimo	1	1
Máximo	9	10
Amplitude	8	9
Média	3,28	6,64
Desvio Padrão	2,07	1,97
Mediana	3	7

Fonte: elaborado pelos autores, com base nos dados da pesquisa.

Ou seja, ainda segundo as informações resumidas na Tabela 2, as médias de familiaridade (nota média = 3,28) e necessidade (nota média = 6,64) são inversas, indicando que diante da baixa familiaridade daqueles profissionais, eles reconhecem uma maior necessidade, sendo ambas as informações relacionadas aos respectivos conhecimentos sobre IFRS.

3.2 Descrição dos Procedimentos Utilizados para Realização da Amostragem pelo Método de *Bootstrap*

A partir da amostra base constituída pelas notas atribuídas pelos 98 respondentes às suas percepções, devidamente separadas em dois grupos de acordo com as respectivas naturezas (familiaridade e necessidade), foram geradas 1.000 (mil) amostras, cada uma com 100 (cem) observações.

Cada uma daquelas observações foi composta por uma das notas atribuídas pelos respondentes da pesquisa, cuja extração ocorreu a partir de sorteio empregando geração de

números aleatórios entre 1 e 98, com o auxílio da função <ALEATÓRIOENTRE> disponível no Microsoft Excel® (MS Excel®).

Ou seja, cada número gerado aleatoriamente entre 1 e 98, inclusive, indicava a observação a ser extraída de cada uma daquelas duas séries de dados contendo as notas dos respondentes indicando sua percepção de familiaridade e sua percepção de necessidade, ambas relacionadas à conhecimentos de IFRS.

Nesse sentido, a amostra de *bootstrap* foi composta a partir da geração de 100.000 (cem mil) possibilidades de notas para cada um daqueles dois grupos de respostas (familiaridade e necessidade), disposta em 1.000 (mil) amostras de 100 (cem) observações cada uma, porém, todas respeitando aleatoriamente as notas atribuídas por aqueles 98 respondentes da pesquisa original.

Na sequência, foi calculada a média de cada uma daquelas 1.000 (mil) amostras com 100 (cem) observações cada. E, a partir daí, essas 1.000 (mil) médias foram tabuladas em 20 intervalos iguais de 0,5 em 0,5 pontos (de 0,00 até 0,50; a partir de 0,50 até 1,00; a partir de 1,00 até 1,50; ...; a partir de 9,00 até 9,50; e, a partir de a partir de 9,50 até 10,00), identificando-se a frequência absoluta e relativa de ocorrência de cada nota média, de forma a se avaliar a respectiva probabilidade de ocorrência.

Adicionalmente, foram analisados os dados referentes à mediana e ao desvio padrão e, ainda, foram aplicados testes de normalidade àquela série de 1.000 (mil) observações referentes às médias apuradas, de forma a avaliar a possibilidade de existência de distribuição simétrica (normal), e, assim, viabilizar a aplicação de testes paramétricos para a realização de inferências a partir da amostra formada pelo método de *bootstrap*.

3.3 Descrição dos Procedimentos de Análise dos Dados

Além dos testes de normalidade, foram levantadas estatísticas descritivas básicas relacionadas à média, mediana e desvio padrão, e, ainda, realizada a análise de probabilidade de ocorrência com base em frequência absoluta e relativa, bem como sua interpretação gráfica.

Conforme já dito na seção anterior, os testes de normalidade têm por objetivo avaliar a possibilidade da existência de distribuição simétrica (normal), de forma a viabilizar a aplicação de testes paramétricos para a realização de inferências a partir da amostra formada pelo método de *bootstrap*.

O levantamento de estatísticas descritivas básicas relacionadas à média, mediana e desvio padrão têm por finalidade avaliar o comportamento das medidas de tendência central e o grau de dispersão dos valores observados a partir amostra de *bootstrap*.

A análise baseada em frequência absoluta e relativa, e sua interpretação gráfica, têm por objetivo avaliar qual a probabilidade de ocorrência de determinados valores observados na amostra obtida pelo método de *bootstrap*.

Todos esses procedimentos de análise descritos foram aplicados à amostra composta por aquelas 1.000 (mil) observações geradas para a amostra inicial de *bootstrap*, e, ainda, foram aplicados somente às 200 (duzentas) primeiras observações daquela amostra.

A partir daí, foram realizadas análises nos desvios padrões observados como meio para avaliar comparativamente o que propõem Efron e Tibshirani (1993) e Montgomery, Peck e Vining (2001).

Ou seja, tentou-se avaliar se, no caso estudado nessa pesquisa, bastaria ser produzida uma quantidade não superior a 200 replicações amostrais aleatórias para identificar o comportamento do desvio padrão em torno da média (EFFRON; TIBSHIRANI, 1993) dos conjuntos de observações referentes à percepção dos respondentes acerca da sua familiaridade e sua necessidade de conhecimentos relacionados às IFRS, ou, até que ponto (quantidades de

observações da reamostragem pelo método de *bootstrap*) o desvio padrão se estabilizaria (MONTGOMEY; PECK; VINING, 2001) e, nesse caso, avaliar também se tal quantidade de observações era inferior àquelas 200 observações sugeridas por Effron e Tibshirani (1993).

3.4 Caracterização Geral do Estudo

Diante do que foi exposto acerca de todo o proceder metodológico, o presente estudo pode ser caracterizado como uma pesquisa empírica de natureza quantitativa, apoiada em um estudo de caso único e em métodos quantitativos aplicados.

4 Análise dos Dados e Apresentação dos Resultados

Esta seção foi dividida em duas seções secundárias. Na primeira, foi abordada a aplicação do método de *bootstrap* e a avaliação da amostra gerada para as notas referentes à familiaridade de conhecimentos relacionados às IFRS, e, na segunda, obviamente, foi abordado o mesmo processo, porém, aplicado às notas referentes à percepção da necessidade de conhecimentos relacionados às IFRS por parte dos respondentes da pesquisa alvo do estudo de caso realizado no processo de investigação científica aqui relatado.

4.1 Aplicação do Método de *Bootstrap* e Avaliação da Amostra Gerada para as Notas Referentes à Familiaridade de Conhecimentos Relacionados às IFRS

Com o auxílio da função <ALEATÓRIOENTRE>, disponível no MS Excel®, foi realizado o sorteio de números aleatórios entre 1 e 98, que indicavam o dado (nota) a ser extraído da amostra base.

Esse processo foi realizado 100.000 vezes gerando 1.000 (mil) grupos de amostras, cada um com 100 (cem) observações. E, a partir disso, foram identificadas as medidas de tendência central e dispersão descritas na Tabela 3.

Tabela 3 – Medidas de tendência central e de dispersão da amostra de *bootstrap* com 1.000 observações (n=1.000)

Intervalos	Média		Mediana		Desvio padrão	
	Frequência observada (Absoluta)	Frequência observada (Relativa)	Frequência observada (Absoluta)	Frequência observada (Relativa)	Frequência observada Absoluta	Frequência observada Relativa
0,0 — 1,0	0	0%	0	0%	0	0%
1,0 — 1,5	0	0%	0	0%	0	0%
1,5 — 2,0	0	0%	82	8%	321	32%
2,0 — 2,5	0	0%	32	3%	679	68%
2,5 — 3,0	86	9%	873	87%	0	0%
3,0 — 3,5	787	79%	8	1%	0	0%
3,5 — 4,0	127	13%	5	1%	0	0%
4,0 — 4,5	0	0%	0	0%	0	0%
4,5 — 5,0	0	0%	0	0%	0	0%
4,0 — 5,5	0	0%	0	0%	0	0%
5,5 — 6,0	0	0%	0	0%	0	0%
6,0 — 6,5	0	0%	0	0%	0	0%
6,5 — 7,0	0	0%	0	0%	0	0%
7,0 — 7,5	0	0%	0	0%	0	0%
7,5 — 8,0	0	0%	0	0%	0	0%
8,0 — 8,5	0	0%	0	0%	0	0%
8,5 — 9,0	0	0%	0	0%	0	0%
9,0 — 9,5	0	0%	0	0%	0	0%
9,5 — 10,0	0	0%	0	0%	0	0%

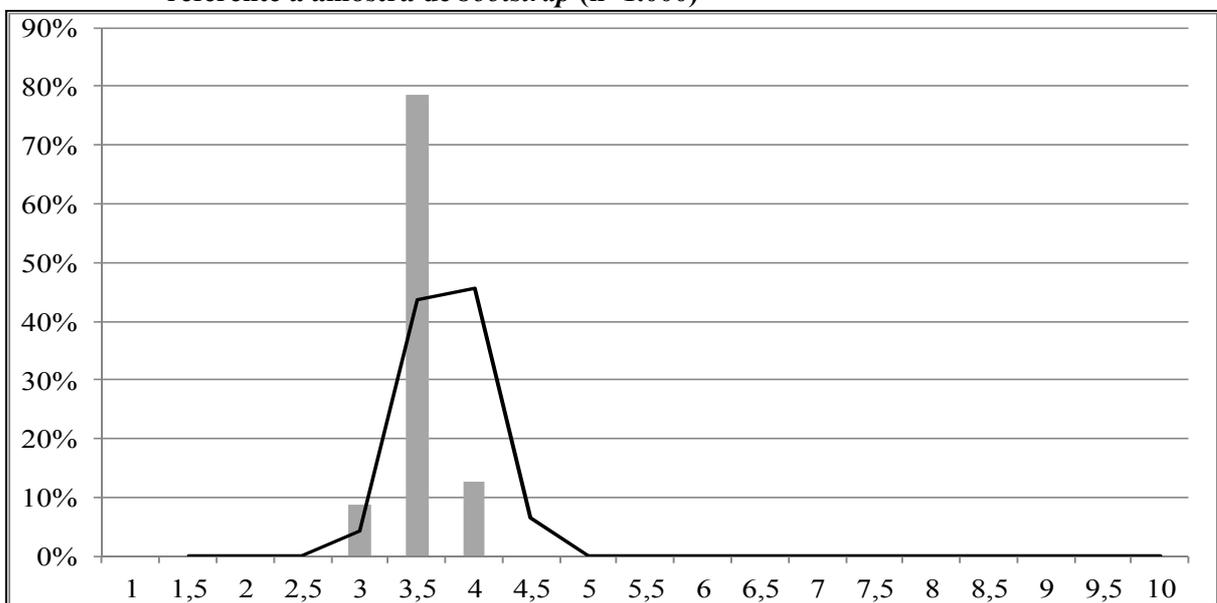
Fonte: elaborado pelos autores, com base nos dados da pesquisa.

Conforme pode ser visto na Tabela 3, a frequência relativa observada indica a probabilidade de ocorrência de cada uma daquelas medidas de tendência central e de dispersão referentes às 1.000 (mil) reamostras compostas por 100 (cem) observações cada uma.

Nesse sentido, observou-se que existe uma probabilidade de 79% de que a média das observações geradas pela amostra de *bootstrap* se situe ente as notas 3,0 e 3,5. Já a mediana apresentou uma probabilidade de 87% de se situar ente as notas 2,5 e 3,00. O desvio padrão sinalizou uma probabilidade de 68% de se situar entre 2,0 e 2,5.

Ao analisar graficamente o comportamento da média observada para série de dados referentes à amostra de *bootstrap*, levantada com base naquelas 98 notas integrantes da amostra base, constatou-se uma forte concentração em torno da nota média de 3,5, conforme pode ser visto na Figura 1.

Figura 1 – Representação gráfica do comportamento da média observada na série de dados referente à amostra de *bootstrap* (n=1.000)



Fonte: elaborado pelos autores, com base nos dados da pesquisa.

A análise da Figura 1 indica a possibilidade de ocorrência de uma distribuição de natureza simétrica, ou normal, para a série de dados da amostra de *bootstrap*. E, de fato, tal possibilidade foi confirmada mediante a aplicação dos testes de normalidade de Anderson-Darling (valor-p = 0,164580952, sign. do valor-p = 0,9412), Kolmogorov-Smirnov (valor-p = 0,044687306, sign. do valor-p = 0,4266), Shapiro-Wilk (valor-p = 0,997690159, sign. do valor-p = 0,9929), e, Ryan-Joiner (valor-p = 0,998951142, sign. do valor-p = 0,9939).

Essa evidência serviria para direcionar qual a natureza dos testes que poderiam ser aplicados para a realização de inferências com base na amostra de *bootstrap*, ou seja, seriam utilizados testes de natureza paramétrica.

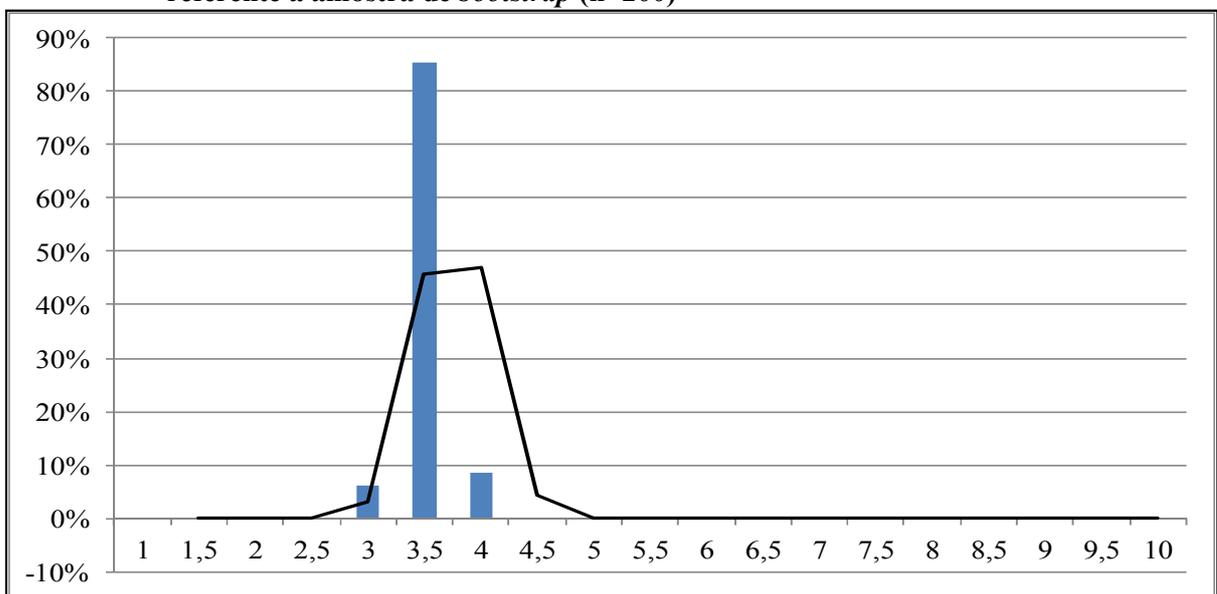
Comparativamente, foram realizados os mesmos procedimentos analíticos relatados até aqui, porém, aplicados somente a um total de 200 (duzentas) reamostras com 100 (cem) observações cada, conforme sugerido por Effron e Tibshirani (1993).

E, conforme pode ser visto tanto na Tabela 4 quanto na Figura 2, as mesmas evidências apuradas a partir da série de reamostras com 1.000 (mil) observações foram constatadas na série de *bootstrap* com apenas 200 observações, confirmando o que foi proposto por Effron e Tibshirani (1993).

Tabela 4 – Medidas de tendência central e de dispersão da amostra de *bootstrap* com 200 observações (n=200)

Intervalos	Média	
	(Absoluta)	(Relativa)
0,0 — 1,0	0	0%
1,0 — 1,5	0	0%
1,5 — 2,0	0	0%
2,0 — 2,5	0	0%
2,5 — 3,0	12	6%
3,0 — 3,5	171	86%
3,5 — 4,0	17	9%
4,0 — 4,5	0	0%
4,5 — 5,0	0	0%
4,0 — 5,5	0	0%
5,5 — 6,0	0	0%
6,0 — 6,5	0	0%
6,5 — 7,0	0	0%
7,0 — 7,5	0	0%
7,5 — 8,0	0	0%
8,0 — 8,5	0	0%
8,5 — 9,0	0	0%
9,0 — 9,5	0	0%
9,5 — 10,0	0	0%

Fonte: elaborado pelos autores, com base nos dados da pesquisa.

Figura 2 – Representação gráfica do comportamento da média observada na série de dados referente à amostra de *bootstrap* (n=200)

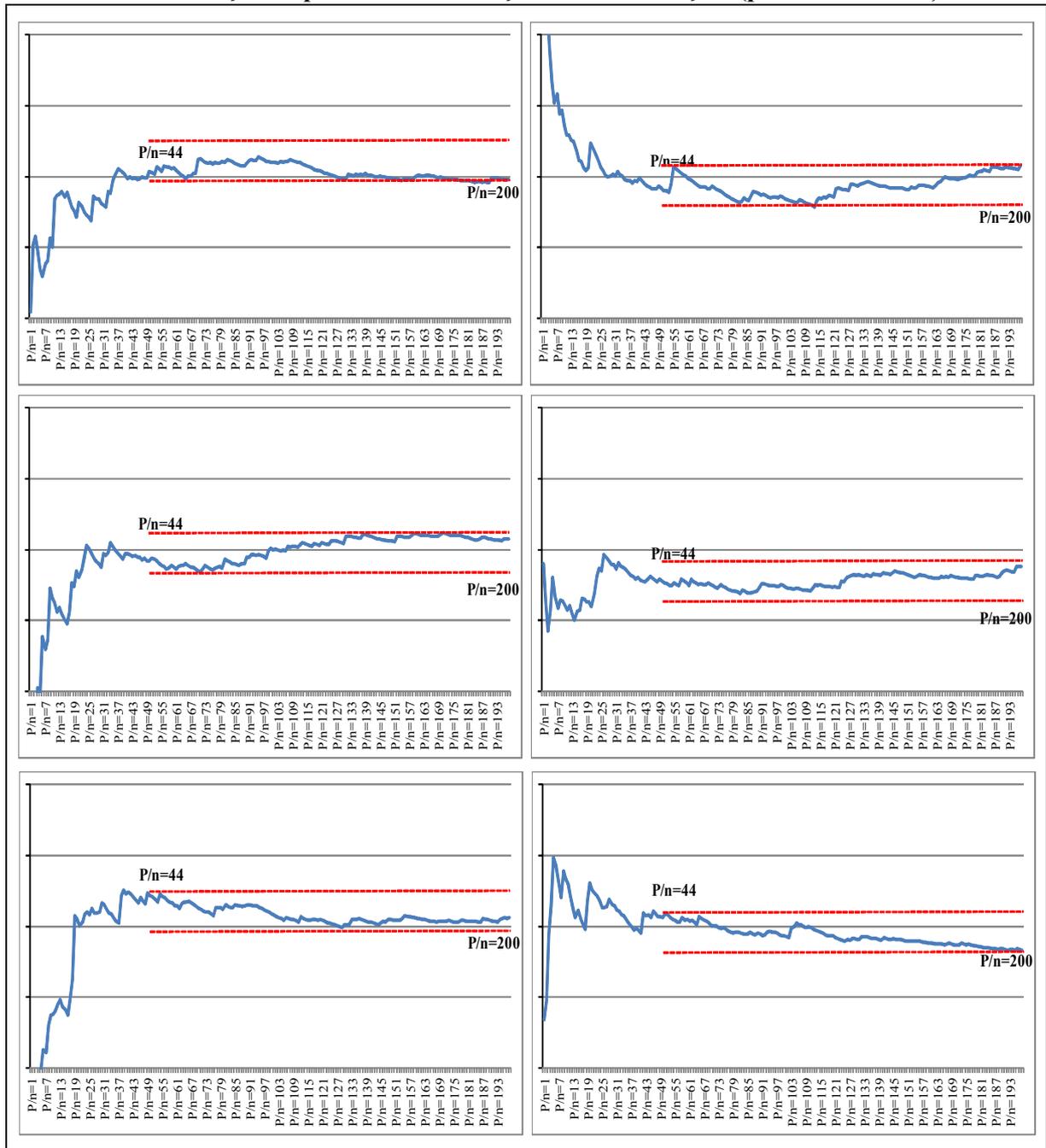
Fonte: elaborado pelos autores, com base nos dados da pesquisa.

Na tentativa de avaliar em que ponto (quantidade de observações da reamostragem pelo método de *bootstrap*) o desvio padrão se estabilizaria, conforme proposto por Montgomery, Peck e Vining (2001), foi calculado o desvio padrão da média a cada ingresso de uma nova observação naquela série de 200 (duzentas) observações da reamostragem realizada pelo método de *bootstrap*. Sendo que, neste estudo, o ponto em questão foi denominado de “ponto de inflexão”, e, ele serve para indicar, graficamente, o ponto a partir do qual o

comportamento gráfico da série de dados representativa do respectivo desvio padrão tende a se apresentar menos volátil.

Conforme pode ser visto na Figura 3, no caso das observações referentes à percepção dos respondentes da pesquisa acerca da sua familiaridade com as IFRS, a partir de 44 (quarenta e quatro) observações (n=44), o desvio padrão tende a apresentar uma menor variação.

Figura 3 – Representação gráfica do comportamento do desvio padrão em torno da média para identificação do ponto de estabilização das sua variações (ponto de inflexão)



Fonte: elaborado pelos autores, com base nos dados da pesquisa.

Essa evidência indica que, segundo a amostra original do caso estudado nessa investigação, não seria necessário realizar o processo de reamostragem de 1.000 (mil) séries

contendo 100 (cem) observações cada, portanto, 100.000 (cem mil) observações. Ou seja, bastaria realizar uma reamostragem em torno de 44 (quarenta e quatro) séries de 100 (cem) observações cada, e, portanto, compor uma amostra de *bootstrap* em torno de 4.400 (quatro mil e quatrocentas) observações.

Nesse sentido, pode-se afirmar que, no caso da amostra original do caso estudado nessa investigação, o parâmetro proposto por Montgomery, Peck e Vining (2001) seria o mais indicado para se identificar quantas observações deveriam integrar a amostra de *bootstrap*.

4.2 Aplicação do Método de *Bootstrap* e Avaliação da Amostra Gerada para as Notas Referentes à Necessidade de Conhecimentos Relacionados às IFRS

Todos os procedimentos de sorteio, com o auxílio da função <ALEATÓRIOENTRE> disponível no MS Excel®, realizados para a amostra base referente às notas atribuídas à familiaridade dos respondentes relação aos seus conhecimentos sobre IFRS, também foram aplicados à amostra base referente à percepção daqueles respondentes sobre sua necessidade acerca de conhecimentos relacionados às IFRS.

De forma semelhante, também foi aplicado o método de reamostragem de *bootstrap* e foram geradas 100.000 (cem mil) observações separadas em 1.000 (mil) grupos de amostras, cada um com 100 (cem) observações. Desse ponto em diante, também foram identificadas as respectivas medidas de tendência central e dispersão, segundo o que está descrito na Tabela 5.

Tabela 5 – Medidas de tendência central e de dispersão da amostra de *bootstrap* com 1.000 observações (n=1.000)

Intervalos	Média		Mediana		Desvio padrão	
	Frequência observada (Absoluta)	Frequência observada (Relativa)	Frequência observada (Absoluta)	Frequência observada (Relativa)	Frequência observada (Absoluta)	Frequência observada (Relativa)
0,0 — 1,0	0	0%	0	0%	0	0%
1,0 — 1,5	0	0%	0	0%	0	0%
1,5 — 2,0	0	0%	0	0%	654	65%
2,0 — 2,5	0	0%	0	0%	346	35%
2,5 — 3,0	0	0%	0	0%	0	0%
3,0 — 3,5	0	0%	0	0%	0	0%
3,5 — 4,0	0	0%	0	0%	0	0%
4,0 — 4,5	0	0%	0	0%	0	0%
4,5 — 5,0	0	0%	0	0%	0	0%
4,0 — 5,5	0	0%	0	0%	0	0%
5,5 — 6,0	1	0%	38	4%	0	0%
6,0 — 6,5	230	23%	18	2%	0	0%
6,5 — 7,0	742	74%	932	93%	0	0%
7,0 — 7,5	27	3%	7	1%	0	0%
7,5 — 8,0	0	0%	5	1%	0	0%
8,0 — 8,5	0	0%	0	0%	0	0%
8,5 — 9,0	0	0%	0	0%	0	0%
9,0 — 9,5	0	0%	0	0%	0	0%
9,5 — 10,0	0	0%	0	0%	0	0%

Fonte: elaborado pelos autores, com base nos dados da pesquisa.

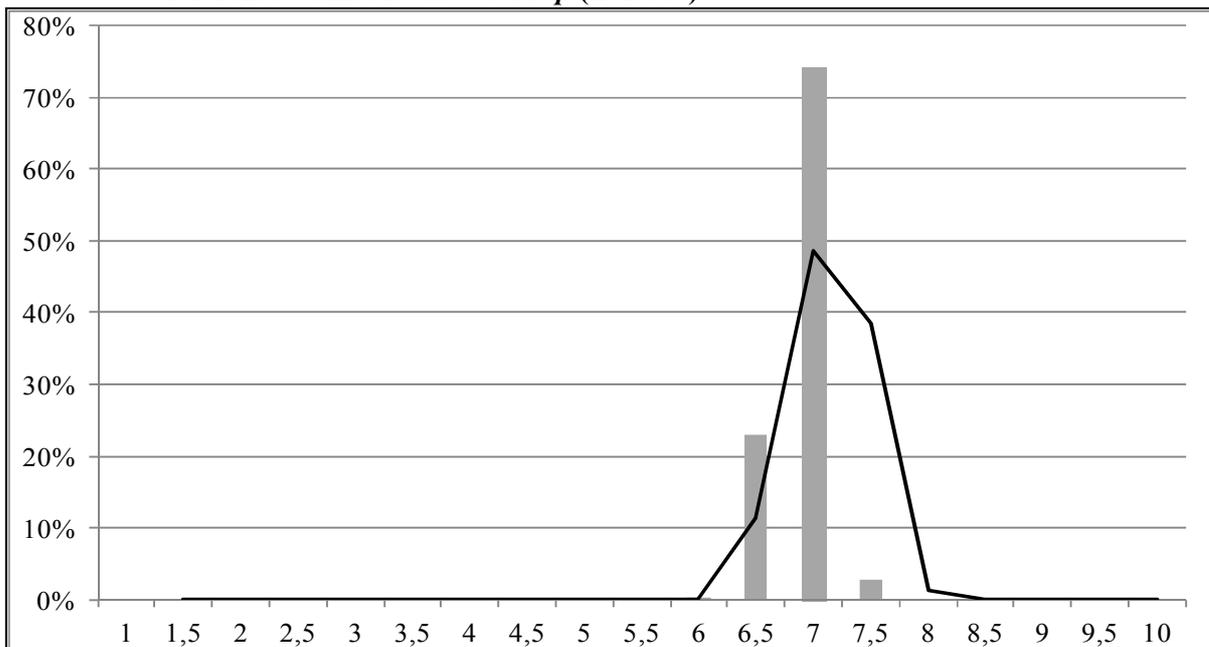
Conforme pode ser visto na Tabela 5, a frequência relativa observada indica a probabilidade de ocorrência de cada uma daquelas medidas de tendência central e de dispersão referentes às 1.000 (mil) reamostras compostas por 100 (cem) observações cada uma.

Nesse sentido, observou-se que existe uma probabilidade de 74% de que a média das observações geradas pela amostra de *bootstrap* se situe ente as notas 6,5 e 7,0. Já a mediana

apresentou uma probabilidade de 93% de se situar também ente as notas 6,5 e 7,0. O desvio padrão sinalizou uma probabilidade 65% de se situar entre 1,5 e 2,0.

Ao analisar graficamente o comportamento da média observada para série de dados referente à amostra de *bootstrap*, levantada com base nas 98 notas atribuídas pelos respondentes da pesquisa indicado sua percepção acerca da sua necessidade de conhecimentos relacionados às IFRS, foi constatada uma forte concentração em torno da nota média de 7,0, conforme pode ser visto na Figura 4.

Figura 4 – Representação gráfica do comportamento da média observada na série de dados referente à amostra de *bootstrap* (n=1.000)



Fonte: elaborado pelos autores, com base nos dados da pesquisa.

A análise do comportamento gráfico demonstrado na Figura 4 indica a possibilidade de ocorrência de uma distribuição normal para a série de dados da amostra de *bootstrap*. Tal possibilidade também foi confirmada mediante a aplicação dos testes de normalidade de Anderson-Darling (valor-p = 0,49685266, sign. do valor-p = 0,2102), Kolmogorov-Smirnov (valor-p = 0,06112818, sign. do valor-p = 0,0664), Shapiro-Wilk (valor-p = 0,992208207, sign. do valor-p = 0,3646), e, Ryan-Joiner (valor-p = 0,995877805, sign. do valor-p = 0,2663).

De forma semelhante ao que aconteceu na amostra de *bootstrap* referente à percepção de familiaridade, a constatação da ocorrência de uma distribuição normal na série de dados que compôs a amostra de *bootstrap* referente à percepção de necessidade serviria para indicar a possibilidade de se utilizar testes da natureza paramétrica no processo de inferência.

De forma comparativa, todos os procedimentos analíticos relatados até aqui foram aplicados a um total de apenas 200 (duzentas) observações conforme sugerido por Effron e Tibshirani (1993).

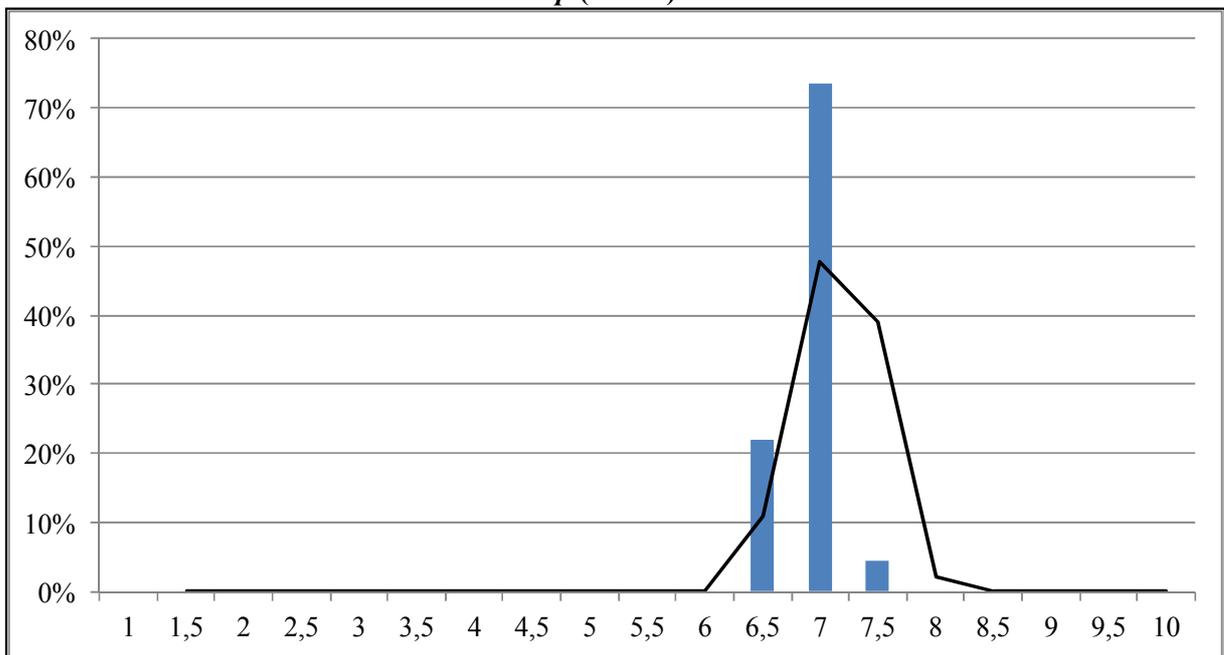
Segundo o que pode ser visto tanto na Tabela 6 quanto na Figura 5, as mesmas evidências apuradas a partir da série de reamostras com 1.000 (mil) observações também foram constatadas na série de *bootstrap* com apenas 200 observações, e, novamente, foi confirmada a proposta de Effron e Tibshirani (1993).

Tabela 6 – Medidas de tendência central e de dispersão da amostra de *bootstrap* com 200 observações (n=200)

Intervalos	Média	
	(Absoluta)	(Relativa)
0,0 — 1,0	0	0%
1,0 — 1,5	0	0%
1,5 — 2,0	0	0%
2,0 — 2,5	0	0%
2,5 — 3,0	0	0%
3,0 — 3,5	0	0%
3,5 — 4,0	0	0%
4,0 — 4,5	0	0%
4,5 — 5,0	0	0%
4,0 — 5,5	0	0%
5,5 — 6,0	0	0%
6,0 — 6,5	44	22%
6,5 — 7,0	147	74%
7,0 — 7,5	9	5%
7,5 — 8,0	0	0%
8,0 — 8,5	0	0%
8,5 — 9,0	0	0%
9,0 — 9,5	0	0%
9,5 — 10,0	0	0%

Fonte: elaborado pelos autores, com base nos dados da pesquisa.

Figura 5 – Representação gráfica do comportamento da média observada na série de dados referente à amostra de *bootstrap* (n=200)



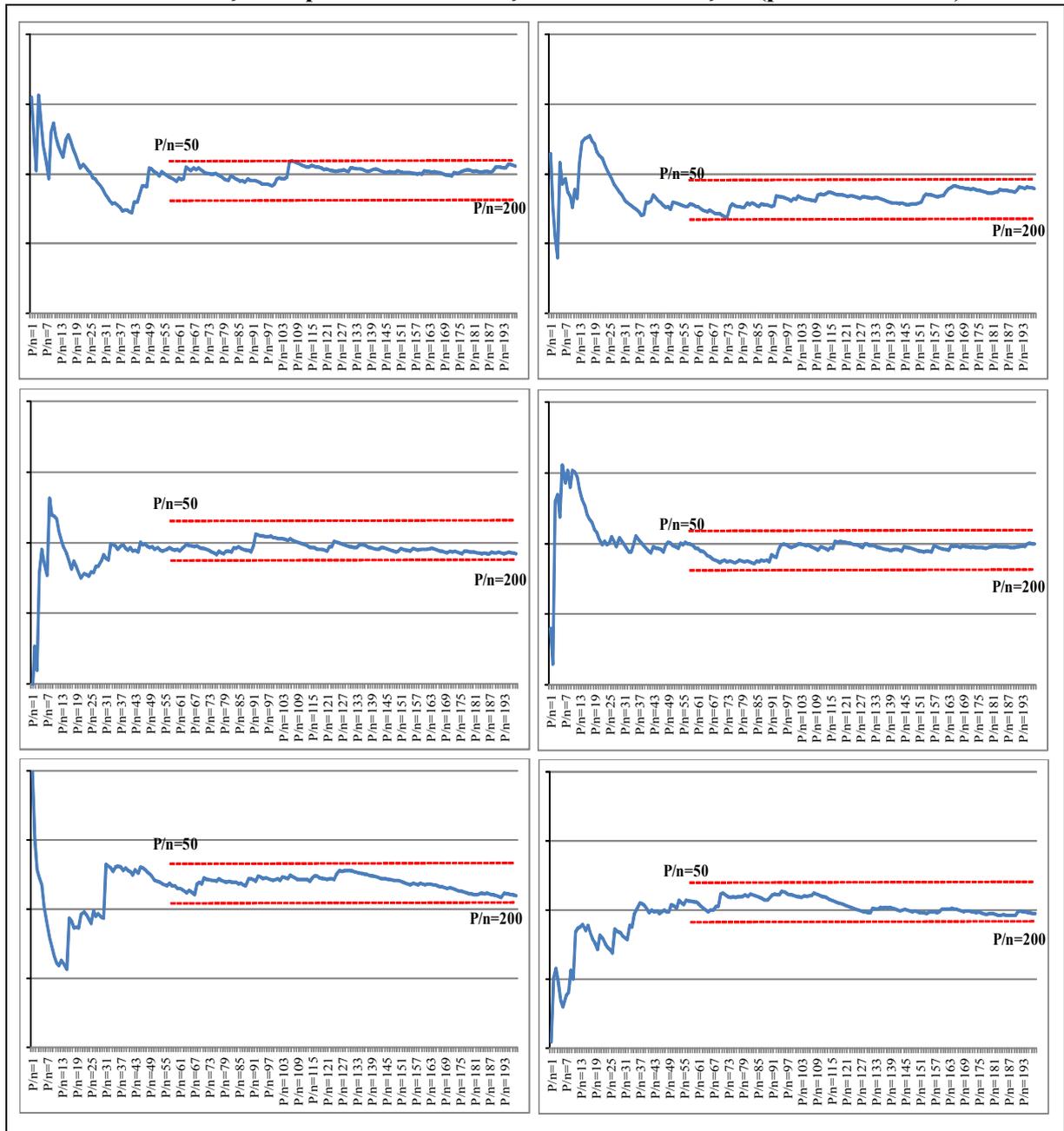
Fonte: elaborado pelos autores, com base nos dados da pesquisa.

No intuito de identificar aquilo que foi denominado neste estudo como “ponto de inflexão”, isto é, o ponto referente à quantidade de observações da amostra de *bootstrap* em que desvio padrão passa a se estabilizar, conforme propuseram Montgomey, Peck e Vining (2001), novamente, foi calculado o desvio padrão da média a cada ingresso de uma nova

observação naquela série de 200 (duzentas) observações da reamostragem realizada pelo método de *bootstrap*.

Conforme pode ser visto na Figura 6, no caso das observações referentes à percepção dos respondentes da pesquisa acerca da sua necessidade de conhecimentos relacionados às IFRS, em torno de 50 (cinquenta) observações ($n=50$), o desvio padrão tende a apresentar uma menor variação.

Figura 6 – Representação gráfica do comportamento do desvio padrão em torno da média para identificação do ponto de estabilização das suas variações (ponto de inflexão)



Fonte: elaborado pelos autores, com base nos dados da pesquisa.

Essa evidência indica que, segundo o perfil amostra base, novamente não seria necessário realizar o processo de reamostragem de 1.000 (mil) séries contendo 100 (cem) observações cada, portanto, 100.000 (cem mil) observações. Isto é, bastaria realizar uma

reamostragem em torno de 50 (cinquenta) séries de 100 (cem) observações cada, e, portanto, compor uma amostra de *bootstrap* em torno de 5.000 (cinco mil) observações.

5 Considerações Finais

A partir de um estudo de caso real de natureza científica em que sua amostra de dados foi considerada pequena em relação à respectiva população (menos de 0,5% [meio por cento] da população), foi demonstrado que a aplicação do método de *bootstrap* pode ser caracterizada como uma alternativa ao processo de amostragem nos estudos de problemas científicos de natureza contábil.

Adicionalmente, dois dos pressupostos apresentados anteriormente por estudiosos da metodologia de amostragem de *bootstrap* puderam ser observados e confirmados neste estudo.

Em relação ao que propuseram Efron e Tibshirani (1993), foi observado que uma série de dados amostrais identificados pelo método de *bootstrap* composta por até 200 observações já permite identificar as características de comportamento médio da amostra.

Além disso, conforme o perfil da amostra base do caso apresentado neste estudo, foi observado que, segundo a proposta de Montgomery, Peck e Vining (2001), a identificação do número de observações a partir do qual o desvio padrão da amostra de *bootstrap* tende a se estabilizar é um bom parâmetro para a definição da quantidade de reamostragens a serem realizadas para composição daquela amostra. Sendo que, no presente estudo, tal ponto (ponto de inflexão) indicou uma quantidade observações muito inferior àquela proposta por Efron e Tibshirani (1993).

Como principal limitação dos resultados alcançados nesta investigação, destaca-se o fato das evidências coletadas estarem diretamente relacionadas ao perfil da amostra base utilizada no estudo de caso apresentado, pois, em função disso, tais resultados não permitem generalizações.

Por outro lado, deve-se levar em conta o rigor metodológico-analítico empregado no processo de reamostragem em si, bem como, nos respectivos testes estatísticos utilizados para validação das evidências coletadas. Isso, por sua vez, indica uma possível alternativa para a solução dos problemas de estudos de natureza contábil cujas amostras podem ser consideradas relativamente pequenas em relação às respectivas populações de dados.

Diante do exposto, espera-se que os resultados desse estudo possam ser somados aos achados científicos de outros estudos de natureza correlata, e, assim, se contribua para o debate relacionado ao processo de amostragem em estudos de natureza científica em geral, e, em especial, nos estudos científicos relacionados às ciências contábeis.

Referências

CALLEGARI-JACQUES, S. M.. **Bioestatística**: princípios e aplicações. Porto Alegre: Artmed, 2003.

EFRON, B. Bootstrap methods: another look at the jackknife. **The Annals of Statistics**, Stanford, v.7, n.1, p.1-26, jan.1979. Disponível em:
<https://projecteuclid.org/download/pdf_1/euclid.aos/1176344552> Acesso em: 11 out. 2016.

EFRON, B.; TIBSHIRANI, R. J.. **An introduction to the bootstrap**. New York: Chapman & Hall, 1993.

EVANS, J. R.; MATHUR, A. The value of online survey. **Internet Research** (Emerald Group Publishing Limited), Bingley (UK), v. 15, n. 2, p. 195-219, 2005. Disponível em: <<http://www.emeraldinsight.com/doi/pdfplus/10.1108/10662240510590360>>. Acesso em: 05 nov. 2016.

HESTERBERG, T.; MOORE, D. S.; MONAGHAN, S.; CLIPSON, A.; EPSTEIN, R. Bootstrap methods and permutation tests. In: _____. **The practice of business statistics: using data for decisions**. New York: W.H. Freeman, 2003.

ILIEVA, J.; BARON, S.; HEALEY, N. M. **Online surveys in marketing research: pros and cons** (Working Paper). Manchester (UK): Graduate School of Business of the Manchester Metropolitan University, 2001. Disponível em: <https://www.researchgate.net/publication/27398278_On-line_surveys_in_international_marketing_research_pros_and_cons>. Acesso em: 06 nov. 2016.

MALHOTRA, N. **Pesquisa de marketing: uma orientação aplicada**. 3. ed. Porto Alegre: Bookman, 2004.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to linear regression analysis**. 3rd ed. New York: Wiley, 2001.

PRACIANO-PEREIRA, T.. **Cálculo numérico computacional**. Sobral: UeVA, 2008.

REIS, G. M.; RIBEIRO JÚNIOR, J. I. Comparação de testes paramétricos e não paramétricos aplicados em delineamentos experimentais. In: SIMPÓSIO ACADÊMICO DE ENGENHARIA DE PRODUÇÃO, 3., 2007, Viçosa-MG. **Anais...** Viçosa-MG: UFV, 2007. Disponível em: <<http://www.saepr.ufr.br/wp-content/uploads/2007-3.pdf>>. Acesso em: 11 out. 2016.

SCHONLAU, M.; FRICKER JR., R. D.; ELLIOT, M. N. **Conducting research surveys via e-mail and the web**. Santa Monica-CA (EUA): Rand Corporation, 2001. Disponível em: <http://www.rand.org/pubs/monograph_reports/MR1480.html> Acesso em: 06 nov. 2016.

STEVENSON, W. _J., **Estatística aplicada à administração**. São Paulo: Harbra, 1981.

WILSON, A.; LASKEY, N. Internet-based marketing research: a serious alternative to traditional research methods? **Journal of Marketing Intelligence & Planning** (Emerald Group Publishing Limited), Bingley (UK),v. 21, n. 2, p.79 – 84, 2003. Disponível em: <<http://www.emeraldinsight.com/doi/pdfplus/10.1108/02634500310465380>>. Acesso em: 05 nov. 2016.