

Métodos e Técnicas Quantitativas em Contabilidade e Finanças: uma Análise sob a Ótica do Software R

Sabrina Amélia de Lima e Silva

Belo Horizonte - MG

Mestre em Administração e Doutoranda em Administração, Finanças pela UFMG
silva.saamelia@gmail.com**Naiara Leite dos Santos Sant' Ana**

Belo Horizonte - MG

Doutoranda em Administração/Finanças pela UFMG
Mestre em Administração pela Universidade Federal de Letras
Professora Assistente da UFJF
naiara.leite@ufjf.edu.br**João Paulo Calembó Batista Menezes**

Teófilo Otoni - MG

Professor Auxiliar da UFVJM

Mestre em Administração pela Universidade de Lisboa (Lisboa - Portugal, CEP: 1600-276)

Doutorando em Administração, Finanças pela UFMG
joao.calemb@gmail.com**Jéssica Santos de Paula**

Belo Horizonte - MG

Bacharel em Ciências Atuariais e Graduanda em Estatística pela UFMG

santosdepaula.jessica@gmail.com

Caio Peixoto Chain

Lavras - MG

Mestre em Administração e Doutorando em Administração, Finanças pela UFLA

caiochain@hotmail.com

Resumo

As pesquisas em contabilidade e finanças adotaram ao longo dos últimos cinquenta anos fundamentos do funcionalismo positivista, fazendo uso de métodos e técnicas quantitativas. Em uma filosofia normativa, tal proposição tornou-se possível por meio da contribuição advinda de áreas como estatística e matemática. Deste modo, pesquisadores podem fazer uso de softwares estatísticos que facilitem e viabilizem a obtenção de resultados mais consistentes e robustos, em pesquisas com foco quantitativo ou qualitativo-quantitativo. Este trabalho visa, por um estudo exploratório, seguido de uma análise bibliométrica, catalogar os três métodos e técnicas de pesquisas quantitativas mais utilizadas em contabilidade e finanças de 2013 a 2015. Em seguida, apresentam-se as técnicas mais empregadas por meio dos scripts no programa R. Identificaram-se trinta e oito artigos publicados no período, por meio de consulta à plataforma Spell com palavra-chave “finanças”, sendo os métodos mais frequentes: estatística descritiva, análise de correlação e regressão linear. Espera-se que os resultados dos scripts contribuam para desenvolvimento e avanço das pesquisas em contabilidade, administração e áreas afins.

Palavras-chave: Métodos e técnicas de pesquisa quantitativa. Contabilidade. Finanças. Programa R. Produção científica.

Abstract

Research in accounting and finance over the past 50 years have adopted foundations of positivist functionalism, using predominantly methods and quantitative techniques in a normative philosophy, can be explained

¹ UFMG - Universidade Federal de Minas Gerais. Av. Pres. Antônio Carlos, 6627 - Pampulha, Belo Horizonte - MG - CEP: 31270-901.

² UFJF - Universidade Federal de Juiz de Fora. Rua José Lourenço Kelmer, S/n - Martelos, Juiz de Fora - MG CE:36036-330.

³ UFVJM - Universidade Federal dos Vales do Jequitinhonha e Mucuri. Rodovia MGT 367 - Km 583, 5000 - Alto da Jacuba, Diamantina - MG - CEP: 39100-000.

⁴ UFLA - Universidade Federal de Lavras. Av. Doutor Sylvio Menicucci, 1001 - Kennedy, Lavras - MG - CEP: 37200-000.

by the arising of areas such as statistical and mathematical contribution. Thus, researchers can make use of statistical software to facilitate and enable getting more consistent and robust results, in research with quantitative or qualitative-quantitative focus. Thus, this work aims for an exploratory study, followed by a bibliometric analysis, catalog the three methods and quantitative research techniques most commonly used in accounting and finance in the last three years, since 2013 to 2015. Then we present the most used techniques by routines and scripts in R. Identified if the program 38 articles published in the period, from Spell data base, with keyword "finance" and the most common methods: descriptive statistics, correlation analysis and linear regression. It is expected that the results of scripts contribute to the development and advancement of research in accounting, business and related fields.

Keywords: methods and quantitative research techniques. Accounting. Finances. Program R. Scientific production.

1. Introdução

O uso de *softwares* aliados com pacotes estatísticos assume relevante papel na ciência desde o desenvolvimento e aplicação de métodos até a análise e interpretação dos resultados. Um entrave na difusão do uso destes programas é, por vezes, o custo elevado de aquisição dos mesmos e sendo assim, cresce a demanda e incentivo ao uso dos chamados *softwares* livres.

Em contabilidade e finanças a utilização de recursos eletrônicos representa um papel importante, que permitiu uma mudança no perfil das pesquisas científicas na área, já que ao longo dos últimos 50 anos, com a adoção dos fundamentos do funcionalismo positivista, os cientistas passaram a utilizar predominantemente métodos e técnicas quantitativas, em uma filosofia normativa. Embora essa mudança de perfil do pesquisador esteja sendo observada, deparamo-nos com um fator limitante, qual seja: uma baixa oferta de disciplinas com enfoque quantitativo nas grades curriculares dos cursos de graduação em ciências contábeis oferecidas pelas Universidades Federais no país, em especial naquelas localizadas em cidades interioranas. Desse fato decorre a falta de contato desses acadêmicos com *softwares* estatísticos. Ocorre que para esta alteração de perfil científico, em um ambiente que exige predominantemente a utilização da estatística e matemática, a aplicação de *softwares* estatísticos tornou-se indispensável. Neste contexto evolucionário, tais aportes estabeleceram e constituíram as pesquisas em Finanças, consolidando as "Finanças Modernas".

Dentre os diversos *softwares* estatísticos disponíveis no mercado, destacam-se, como apresentado, aqueles denominados *softwares* livres (*free software*) ou de código aberto (*open source*), em que a licença de uso não se restringe a venda ou distribuição do *software*. Nesta linha, há um projeto denominado R Project que disponibiliza gratuitamente o programa

R, que permite a realização de testes estatísticos variados, estatística linear e não-linear, modelagem, análise de séries temporais, classificação, *clustering*, dentre outras.

Ligges (2007) pondera que o R é uma linguagem de alto nível, gratuito e com código-fonte aberto que permite modificações ou ainda implantação de novos procedimentos e comandos desenvolvidos a qualquer momento por qualquer usuário. Importante ressaltar que o programa R tem suas origens em uma linguagem anterior, denominada S (BECKER, CHAMBERS, & WILKS, 1988). Foi criado como um projeto de pesquisa e está sob constante desenvolvimento por um grupo denominado R Core Team.

O presente trabalho teve como objetivo listar os métodos e técnicas quantitativas que foram utilizadas em artigos em contabilidade e finanças nos últimos três anos (2013-2015) consultados na plataforma *Spell*, na sequência selecionar e apresentar as três técnicas mais recorrentes nos trabalhos científicos por meio das rotinas e scripts no programa R. Espera-se que, por meio desse artigo, mais pesquisadores possam fazer uso ou se interessarem pelo *software* R, ademais se espera estimular os vários tipos de publicações em Contabilidade e Finanças, visto que é uma área a evoluir em relação à quantidade e à qualidade das publicações.

A justificativa para a consecução desse trabalho encontra-se na pouca produção científica na área, assim como no acesso limitado de conhecimento e tecnologias que facilitarão e instigarão pesquisas em Contabilidade e Finanças. Esse artigo pretende provocar e aguçar os alunos a se enveredarem no caminho da pesquisa ainda na graduação. Deve-se acrescentar que o destaque dado a Teoria de Finanças nesse artigo se justifica pelo fato da completude e aproximação observada entre as áreas Contabilidade e Finanças, que são explicitadas por meio das publicações científicas.

A estrutura do artigo é composta pela Introdução, constando a contextualização do tema, objetivo, justificativa e estrutura; em seguida tem-se o referencial teórico que foi fracionado conforme as seguintes temáticas: Evolução da Teoria de Finanças; Introdução e Interface do *software* R e por último apresentam-se os trabalhos com propostas semelhantes. Na seção 3 apresentaram-se os métodos e técnicas expostos na pesquisa; na seção 4 encontram-se as análises dos resultados e por último tem-se a conclusão.

2. Revisão de Literatura

Esta seção busca fornecer o embasamento teórico motivador do desenvolvimento deste trabalho ao apresentar a evolução da teoria de finanças, em que se nota uma mudança no paradigma com o advento da utilização de técnicas quantitativas. Apresenta-se em seguida o *software* R, utilizado na produção dos *scripts* deste estudo e, finaliza-se esta seção com a apresentação de trabalhos contendo propostas semelhantes à encontrada neste.

2.1 Evolução da Teoria de Finanças

A evolução da teoria de finanças, atualmente denominada de “finanças modernas” foi marcada por um período transitório na década de cinquenta, em que se dividiu a teoria tradicional da chamada teoria moderna (IQUIAPAZA, AMARAL, & BRESSAN, 2009).

Anteriormente os pesquisadores baseavam-se em experiências práticas para analisar o comportamento dos investidores e empresas, tendo estas como sugestões para tomada de decisão, não tratando os resultados de forma universal, fornecendo um viés positivista e não universal. A forma de escolher eficazmente os ativos financeiros era baseada na compra de ativos subavaliados e venda daqueles que estivessem sobre-avaliados, fixando apenas a análise da subavaliação dos ativos.

A denominada Teoria moderna foi marcada pela mudança de paradigma causada pelo trabalho de Modigliani e Miller (1958), tracejando o caminho funcionalista da linha caracterizada por uma maior proximidade com as ciências naturais e abordagem mais normativa, universal e exata, sendo os resultados obtidos nos Estados Unidos, berço das inovações, exportados para outros mercados com diferentes graus de investimento.

Pondera-se que é no caminho funcionalista que a teoria de finanças se baseou, sendo predominante na sociologia acadêmica, na teoria organizacional e nas pesquisas acadêmicas de finanças. Como colocado pelos autores o mundo de finanças é tratado como um lugar de realidade concreta onde o indivíduo assume um papel passivo, sendo o seu comportamento determinado pelo ambiente econômico (IQUIAPAZA, AMARAL, & BRESSAN, 2009).

2.2 Introdução ao R

R é uma linguagem e ambiente para computação estatística e gráficos que fornece uma ampla variedade de estatística e técnicas gráficas altamente extensíveis. O programa está disponível como *software* livre sob os termos da GNU - *General Public License* em forma de código-fonte do *Free Software Foundation*. Ele compila e roda em uma ampla variedade de plataformas UNIX e os sistemas semelhantes (incluindo FreeBSD e Linux), Windows e MacOS.

O ambiente R é uma integração de instalações de *software* para manipulação de dados, cálculo e exibição gráfica. Possui uma linguagem de programação eficaz e bem desenvolvida que inclui condicionais, loops, funções recursivas definidas pelo usuário e recursos de entrada e saída.

Desenvolvido a fim de fornecer suporte e outras inovações em computação estatística, acredita-se que o R se tornou uma ferramenta madura e valiosa que proporciona um ponto de referência para os indivíduos ou empresas que visam apoiar ou interagir com a comunidade de desenvolvimento R.

Destaca-se ainda que o R não é apenas um programa estatístico, uma vez que permite operações matemáticas, manipulação de vetores e matrizes, banco de dados,

dentre outros. Um dos pontos fortes do R é a facilidade com que gráficos bem delineados e de alta qualidade para impressão podem ser produzidos com possibilidade de inclusão de fórmulas e símbolos matemáticos quando necessário.

O R é também altamente expansível com o uso dos pacotes, que são bibliotecas para funções específicas ou áreas de estudo específicas. Um conjunto de pacotes é incluído com a instalação do *software*, mas muitos outros estão disponíveis na rede de distribuição do R (em inglês *CRAN*).

2.3 Trabalhos com Propostas Semelhantes

Esta seção objetiva apresentar trabalhos que tangenciam a utilização de *softwares* estatísticos em pesquisas acadêmicas principalmente nas áreas de contabilidade, finanças e administração.

O trabalho de Silva Filho e Souza (2013) teve como escopo a apresentação de passos iniciais do uso do programa R para alunos iniciantes, justificado pelos autores por ser um *software* de interface não tão amigável com os usuários. Segundo tais autores, o *software* R é bastante utilizado no meio acadêmico, sendo considerado um dos melhores aplicativos voltados para a obtenção de estatísticas. A sua utilização por áreas do conhecimento como finanças e contabilidade tem crescido principalmente após o desenvolvimento de pacotes como *Rcmdr – R Commander*, que tornam o R uma ferramenta mais amistosa para os alunos.

Segundo Fávero (2013), a utilização de modelos que envolvam dados longitudinais em contabilidade e finanças tem sido recorrente. Todavia, verifica-se pouco cuidado quanto aos critérios para adoção de um modelo em detrimento de outro. O artigo deste autor possui a finalidade de discutir a técnica estatística de dados em painel, assim como a apresentação das principais rotinas para os modelos no *software* Stata.

Fávero (2013) coloca que a existência de rotinas de *softwares* estatísticos pode propiciar uma melhor relação entre teoria e prática, além de facilitar a implementação dos modelos em pesquisas futuras.

Desempenhando papel crescente e importante em quase todas as fases da pesquisa humana, a Estatística influencia atualmente, diversos setores, tais como: economia, administração de empresas, agricultura, educação, política, psicologia, medicina, engenharia, comércio etc. (TOLEDO e OVALLE, 1995). O campo de aplicação da Estatística tem-se ampliado consideravelmente, devido, principalmente à necessidade de se tomar decisões rápidas e, na medida do possível seguras.

Destacando a aplicabilidade do método estatístico especialmente nos campos de contabilidade e finanças, Werkema (1996) destaca que importantes ferramentas estatísticas podem ser empregadas para uma realização mais eficiente das atividades de gerenciamento de uma empresa. Neste sentido, o trabalho de Milagre (2001), propõe-se a discutir a estatística e uma proposta de ensino para os cursos de administração de empresas, tangenciando o tema deste trabalho.

3. Métodos e Técnicas Adotados na Pesquisa

Trata-se de uma pesquisa qualitativa, caracterizada quanto ao tipo como “Descritiva”, foi realizada por meio do método indutivo e a técnica utilizada foi a bibliográfica. Ludke e Andre (1986) conceituam um trabalho qualitativo como sendo aquele que faz uso de técnica descritiva, desenvolve de forma natural e flexível e analisa a realidade de forma contextualizada e complexa.

A pesquisa descritiva é aplicada quando se busca identificar, relatar e comparar características de determinada população ou fenômeno e o estabelecimento de relações entre variáveis Beuren (2013). Conforme Marconi e Lakatos (2010) em uma pesquisa indutiva parte-se de dados particulares para uma verdade geral ou universal por meio da inferência. Segundo Gil (2010) a pesquisa bibliográfica tem como base materiais prontos, elaborados, como livros e artigos.

3.1 População, Amostra e Coleta de Dados

A população foi constituída por cento e um artigos disponível na plataforma *Spell* para os anos 2013, 2014 e 2015, a palavra-chave de busca utilizada foi “Finanças”. A amostra foi constituída por tinta e oito artigos, pois o interesse dos pesquisadores estava voltado para “artigos da área finanças com metodologia quantitativa”.

3.2 Análise de Dados

A análise de dados ocorreu por meio da técnica de análise bibliométrica. Como apresentado no tópico 3.1, dos cento e um artigos encontrados por meio da estratégia de busca descrita anteriormente, apenas tinta e oito estavam de acordo com o objetivo da pesquisa por se tratarem de trabalhos com metodologias quantitativas.

Decorrida essa primeira etapa partiu-se para a investigação da metodologia apresentada em cada um desses 38 artigos, juntamente com a separação de produção científica por ano, foi elaborada uma planilha contendo o título do artigo, o ano de publicação e a metodologia apresentada.

Na sequência, procedeu-se com a análise da frequência das técnicas metodológicas mais recorrentes, para que com isso fosse possível a definição das três metodologias mais utilizadas, dada a amostra. Por último foram apresentadas as rotinas do *software* R para as metodologias mais utilizadas, definidas anteriormente.

4. Apresentação e Análise dos Resultados

Conforme proposto, em decorrência do levantamento dos dados, foram identificados dezoito artigos publicados em 2015, dezesseis em 2014 e quatro em 2013, totalizando trinta e oito artigos. A seguir apresentamos os principais métodos e técnicas identificados, quais sejam, estatística descritiva, análise de correlação e regressão linear (simples e múltipla).

4.1 Estatística Descritiva

Este artigo objetiva apresentar os principais métodos e técnicas quantitativas que foram utilizadas em artigos em finanças nos últimos três anos por meio de *scripts* no *software* R. Anteriormente à apresentação destes, deve-se proceder com o cálculo de estatísticas descritivas visando maior conhecimento acerca dos dados a serem utilizados. Assim, após coleta de dados, uma fase introdutória à análise resume-se às estatísticas descritivas. Este tópico apresenta como programar essas principais medidas no R, por meio de funções pré-existentes.

Quando se tratam de variáveis quantitativas, os dados podem ser resumidos sob a forma de distribuições de frequência ou por medidas descritivas. Medidas descritivas são formas de expressar a informação trazida pelos dados e dividem-se em duas categorias: medidas de posição e medidas de dispersão. As medidas de posição indicam a posição global dos dados na escala de valores possíveis e as mais comuns são a média, mediana e moda. As medidas de dispersão apresentam o quanto os dados se distribuem em relação à média, e as mais comuns são desvio-padrão, variância, coeficiente de variação e amplitude.

Quanto às medidas de posição, a média pode ser calculada pela forma conceitual, ou seja, procedendo-se com a soma dos elementos e dividindo pela quantidade deles, ou utilizando a função *mean*. Pode-se também calcular a média condicionada com o comando *mean(dados>1)*, por exemplo, onde são considerados apenas os dados maiores que 1. Esta condição pode ser útil quando se deseja, por exemplo, a média por sexo ou escolaridade. Outra medida de posição é a mediana, apurada pelo comando *median*.

Para as medidas de dispersão, *sd* representa o desvio-padrão enquanto *var* retorna a variância dos dados. Tem-se ainda a função *quantile*, que retorna quaisquer quantis que se deseja na amostra. O comando *summary* dá um resumo de algumas funções que foram apresentadas anteriormente: *min(x)*, *max(x)*, *median(x)*, *mean(x)*, *quantile(x,0.25)* e *quantile(x,0.75)*. A amplitude, pode então ser calculada pela diferença entre *max(x)* e *min(x)*. Por fim, o coeficiente de variação pode ser calculado por *sd(x)* dividido por *mean(x)* e multiplicado por 100 para apresentar um resultado em percentual.

Alguns gráficos são fundamentais na análise descritiva. Apresentam-se a seguir, três gráficos fundamentais neste quesito: histograma, gráfico de barras e gráfico de caixas. São reconhecidos no R pelos nomes *hist*, *barplot* e *boxplot*.

Um histograma divide uma série de dados em diferentes classes igualmente espaçadas e mostra a frequência de valores em cada classe. Em um gráfico, o histograma mostra diferentes barras, com bases iguais e amplitudes relativas às frequências dos dados em cada classe. O eixo das ordenadas, portanto, mostra a frequência relativa de cada classe e o eixo das abscissas os valores e intervalos das classes. A sintaxe é apresentada na Figura 1 e um exemplo ilustrativo no Gráfico 1.

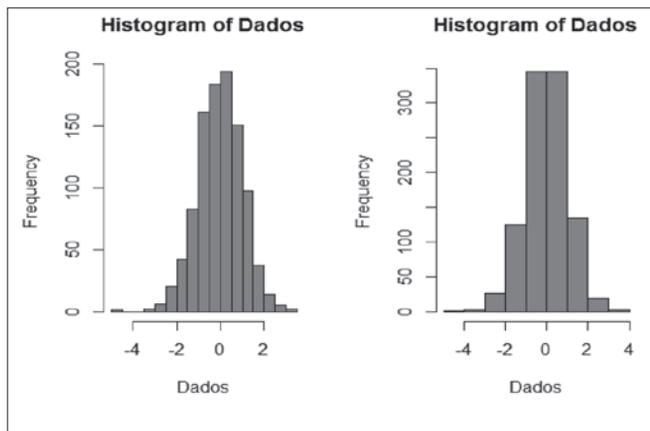
Figura 1: Script para construção de histograma no R

```

17 # Histograma
18 hist(dados, nclass=p) # p é o número de classes do histograma
19
20 # Exemplo
21 par(mfrow=c(1,2))
22 hist(dados, nclass=12, col=2)
23 hist(dados, nclass=6, col=4)
    
```

Fonte: Os autores. (2016)

Gráfico 1: Histograma para os dados trabalhados com 12 e 6 classes



Fonte: Os autores. (2016)

A função *barplot()* produz gráfico de barras, onde cada barra representa a medida de cada elemento de um vetor, ou seja, as barras são proporcionais com a “dimensão” do elemento. A sintaxe geral da função está na Figura 2 com um exemplo no Gráfico 2.

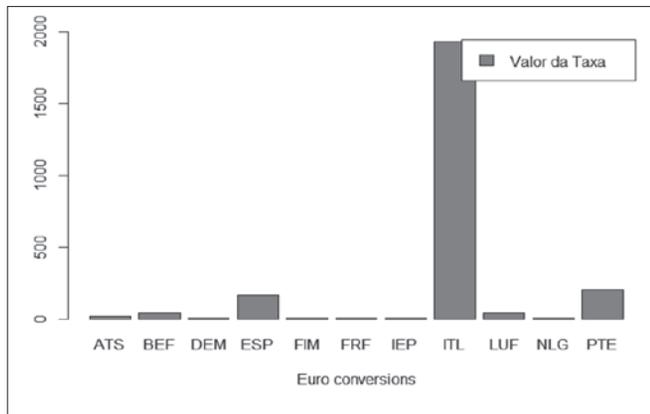
Figura 2: Script para construção de gráfico de barras no R

```

25 # Gráfico de Barras
26 barplot(x, col = "", legend.text = "", xlab = "", ylab = "")
27
28 # Exemplo
29 barplot(euro, col = "red", legend.text = "Valor da Taxa",
30         xlab = "Euro conversions")
    
```

Fonte: Os autores. (2016)

Gráfico 2 - Gráfico de barras para o valor da taxa de conversão do Euro



Fonte: Os autores. (2016)

Na sintaxe apresentada, temos que *x* é o vetor ou arquivo de dados; *col* define-se a cor de exibição do gráfico de barras; *legend.text* refere-se à legenda do gráfico; *xlab* e *ylab* atribuem nome das grandezas expressas nos eixos *x* e *y*, respectivamente. O exemplo no gráfico mostra as taxas de conversão entre as diversas moedas e o euro nos países da união europeia.

Como último dos gráficos mais relevantes, o boxplot é um gráfico que possibilita representar a distribuição de um conjunto de dados com base em alguns de seus parâmetros descritivos, a saber, a mediana e os quartis. Ele permite avaliar a simetria dos dados e a sua dispersão. É especialmente recomendado para a comparação de dois ou mais conjuntos de dados correspondentes às categorias de uma variável qualitativa.

O comando utilizado no R é o *boxplot()*. Este comando possui vários argumentos. Utilizando o comando *help(boxplot)* é possível obter maiores informações. A Figura 3 apresenta a sintaxe para plotar dois boxplots no mesmo gráfico ou em gráficos diferentes. O Gráfico 3 apresenta um exemplo de boxplot produzido no R.

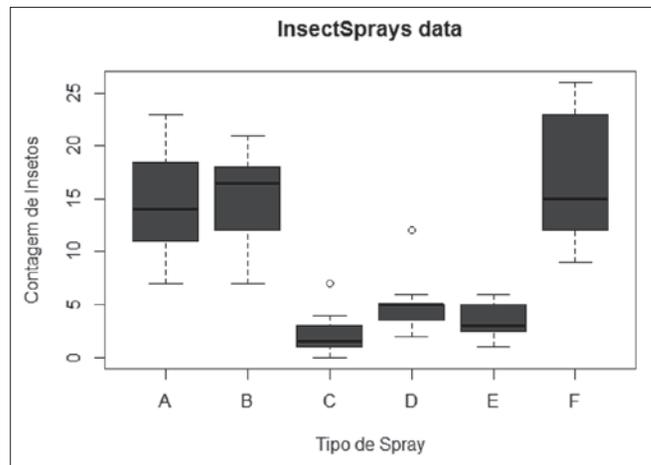
Figura 3: Script para construção de boxplot no R

```

31 # Boxplot
32 x = c(5,5,5,13,7,11,11,9,8,9)
33 y = c(11,8,4,5,9,5,10,5,4,10)
34 boxplot(x,y) # para plotar no mesmo gráfico (comparação)
35 boxplot(x); boxplot(y) # para plotar em gráficos diferentes
36
37 # Exemplo
38 boxplot(count=spray, data = InsectSprays, xlab = "Tipo de Spray",
39         ylab = "Contagem de Insetos", main = "Insectsprays data",
40         col = "darkgreen")
    
```

Fonte: Os autores. (2016)

Gráfico 3: Exemplo de Boxplot para contagem de insetos e tipo de spray no R



Fonte: Os autores. (2016)

4.2 Métodos e Técnicas Mais Encontrados

4.2.1 Análise de Correlação

Por vezes precisamos avaliar o grau de relacionamento entre duas ou mais variáveis, pois se deseja saber o quanto uma variável interfere no resultado de outra. As técnicas associadas à Análise de Correlação

representam uma ferramenta fundamental de aplicação nas Ciências Sociais e do comportamento.

Um dos métodos visuais mais usados para a investigação de pares de dados são os diagramas de dispersão. Geometricamente, um diagrama de dispersão é simplesmente uma coleção de pontos num plano cujas duas coordenadas cartesianas são os valores de cada membro do par de dados. Este gráfico permite examinar os dados no que se refere à ocorrência de tendências podendo ser lineares ou não. Permite ainda visualizar agrupamentos de uma ou mais variáveis, mudanças de espalhamento de uma variável em relação à outra e verificar a ocorrência dos valores discrepantes.

Para melhor entendimento da Análise de Correlação, apresenta-se um exemplo prático elaborado a partir de um banco de dados disponível no R, chamado "longley". Esse banco de dados possui sete variáveis econômicas: inflação, PIB, número de empregados, número de pessoas nas forças armadas, população não institucionalizada maior de 14 anos, ano e número de pessoas empregadas, observadas anualmente entre 1947 e 1962.

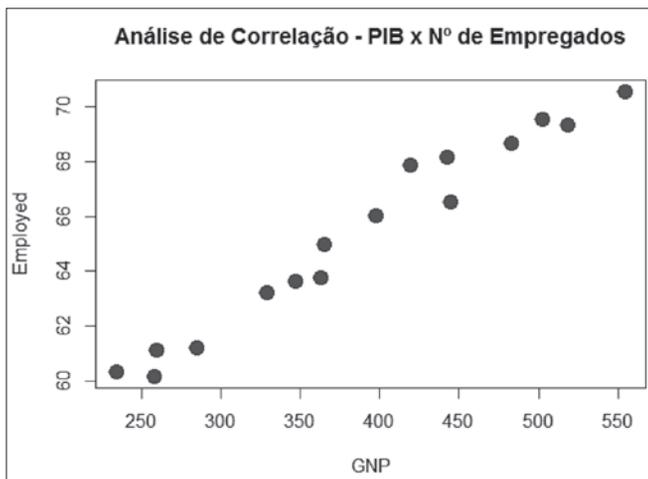
Deseja-se verificar se há correlação entre o PIB e o número de pessoas empregadas e entre o PIB e o número de pessoas desempregadas. Para avaliar o comportamento desses pares de variáveis, utiliza-se a análise gráfica através do Diagrama de Dispersão. Apresenta-se na Figura 4 o *script* do R e os gráficos resultantes.

Figura 4: Script para construção de diagrama de dispersão no R

```
43 # Análise de correlação
44 # Diagrama de Dispersão
45 # plot dos pares (x,y) num sistema de eixos coordenados
46 dados<-longley
47 attach(dados)
48 plot(GNP,Employed, col = "darkred", lwd = 8,
49      main = c("Análise de Correlação - PIB x N° de Empregados"))
50 plot(GNP,unemployed, col = "darkred", lwd = 8,
51      main = c("Análise de Correlação - PIB x N° de Desempregados"))
```

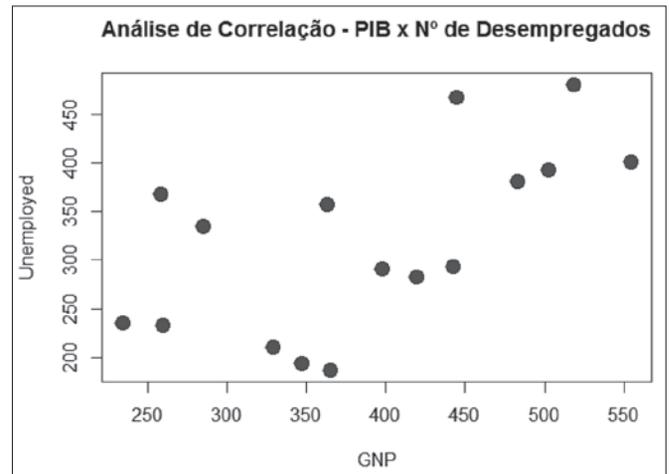
Fonte: Os autores. (2016)

Gráfico 4: Exemplo de Gráfico de Dispersão no R



Fonte: Os autores. (2016)

Gráfico 5: Exemplo de Gráfico de Dispersão no R



Fonte: Os autores. (2016)

Conforme ilustrado no Gráfico 4, verifica-se uma relação positiva forte entre as variáveis, em que quanto maior o número de empregados, maior será o PIB. Não obstante, analisando o Gráfico 5 não é possível afirmar se há relação entre o PIB e número de desempregados.

Visando quantificar a correlação indicada visualmente, pode ser calculado o coeficiente de correlação, que é uma medida do grau de relação linear entre duas variáveis quantitativas. Esse coeficiente é geralmente denominado por r e pode assumir valores entre -1 e 1, onde variáveis relacionadas positivamente como demonstrado no Gráfico 4 assumem coeficientes positivos e variáveis relacionadas negativamente, possuem coeficientes negativos.

O R fornece a possibilidade de fazer testes de correlação com base nos métodos de Pearson, Kendall ou Spearman. A partir desses testes é possível avaliar o coeficiente de correlação e inferir se este é significativo ou não.

A função *cor.test()* testa a correlação entre variáveis quantitativas contínuas x e y utilizando o coeficiente de correlação de Pearson ou de Kendall ou de Spearman. Através do argumento *alternative*, é possível alterar a hipótese alternativa do teste de correlação. Se for bi-caudal, por exemplo, deve-se utilizar *alternative = "two-sided"*, o que significa testar a hipótese de que a correlação é diferente de zero, já que esta pode assumir valores positivos ou negativos. O R apresenta então o resultado da correlação e ainda se esta é significativa ou não, conclusão que pode ser obtida a partir do *p-value* apresentado.

Figura 5: Resultados da correlação pelo método de Pearson

```
> cor.test(GNP,Employed, alternative = "two.sided", method = "pearson")

Pearson's product-moment correlation

data: GNP and Employed
t = 20.374, df = 14, p-value = 8.364e-12
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9519914 0.9944238
sample estimates:
 cor
0.9835516

> cor.test(GNP,Unemployed, alternative = "two.sided", method = "pearson")

Pearson's product-moment correlation

data: GNP and Unemployed
t = 2.8376, df = 14, p-value = 0.01317
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1549766 0.8464304
sample estimates:
 cor
0.6042609
```

Fonte: Os autores. (2016)

O primeiro teste apresentado na Figura 5 diz respeito ao PIB e ao número de empregados e, confirmando as análises realizadas graficamente, há uma forte correlação linear positiva, com r igual a 0,9835 e P -valor próximo a zero e, portanto, inferior ao nível de significância (estabelecido pelo pesquisador, geralmente de 1%, 5% ou 10%) e rejeitando a hipótese de correlação nula.

No segundo teste, é possível observar que o PIB e o número de desempregados não são fortemente correlacionados, como esperado pela análise gráfica. Nesse caso, r é igual a 0,6042, indicando relação positiva, e o P -valor igual a 0,01317, que ao nível de 1% de significância ainda não seria significativo. Portanto, a hipótese de que a correlação entre as variáveis é nula não é rejeitada.

4.2.2 Regressão Linear Simples e Múltipla

Os modelos de regressão linear são muito utilizados em análise estatística, pois quantificam a relação entre uma variável dependente e demais variáveis independentes ou explicativas. No R, a estrutura para implantação de modelos simples ou múltiplos é relativamente simples e feita pela função *lm* (linear model).

Modelos lineares são representados por: $lm(y \sim x_1 + x_2 + \dots + x_n)$ - função para fazer modelo de regressão linear, sendo que $lm(y \sim x_1)$ é um modelo simples e $lm(y \sim x_1 + x_2)$ é um modelo múltiplo. A variável y é dependente e a variável x é independente. Para observar os resultados do modelo, deve-se utilizar a função *summary()*. Para a regressão múltipla a alteração se dá pela inclusão de novas variáveis explicativas. Para a regressão múltipla a alteração se dá pela inclusão de novas variáveis explicativas.

A partir dos dados utilizados no exemplo apresentado anteriormente para Análise de Correlação, é criado um modelo de regressão linear simples onde se procura explicar o PIB em função das variáveis: número de desempregados, número de pessoas nas forças armadas, população não institucionalizada maior de 14 anos e número de pessoas empregadas. Os resultados são apresentados na Figura 6.

Figura 6: Exemplo de Resultados da Regressão Linear Simples

```
> modelo<-lm(GNP~Unemployed+Armed.Forces+Population+Employed)
> modelo<-lm(GNP~Unemployed+Armed.Forces+Population+Employed)
> summary(modelo)

Call:
lm(formula = GNP ~ Unemployed + Armed.Forces + Population + Employed)

Residuals:
   Min       1Q   Median       3Q      Max
-6.7850 -4.3097 -0.6786  3.3328  8.8336

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.339e+03  3.512e+01 -38.137 4.87e-13 ***
Unemployed   3.430e-04  4.055e-02  0.008  0.99340
Armed.Forces  8.245e-02  2.955e-02  2.790  0.01759 *
Population   9.338e+00  1.504e+00  6.211 6.62e-05 ***
Employed     9.321e+00  2.397e+00  3.888  0.00253 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.683 on 11 degrees of freedom
Multiple R-squared:  0.9976, Adjusted R-squared:  0.9967
F-statistic: 1145 on 4 and 11 DF, p-value: 2.513e-14
```

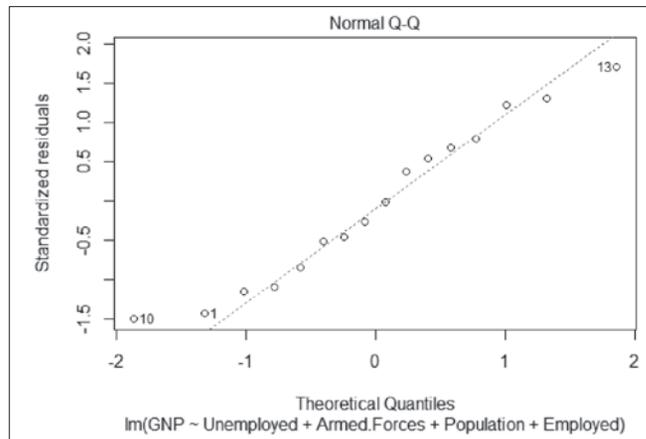
Fonte: Os autores. (2016)

A função *summary(modelo)* exibe os coeficientes de regressão para cada variável, seus respectivos erros-padrão, estatística de teste e P -valor, representado por $Pr(>|t|)$. A significância de cada variável no modelo é avaliada pelo P -valor que deve ser inferior ao nível de significância estabelecido. No exemplo apresentado, apenas a variável número de desempregados não é significativa no modelo.

Além dessas estatísticas de análise individual das variáveis, é apresentada a Estatística F, os graus de liberdade, o R^2 , o R^2 ajustado e o P -valor do modelo, que são ferramentas para avaliar a significância do modelo, considerando todas as variáveis incluídas. No exemplo, o R^2 ajustado é de 0,9967, o que significa que 99,67% do PIB podem ser explicados pelas variáveis independentes do modelo, com P -valor igual a $2,513e^{-14}$, sendo, portanto, um modelo significativo.

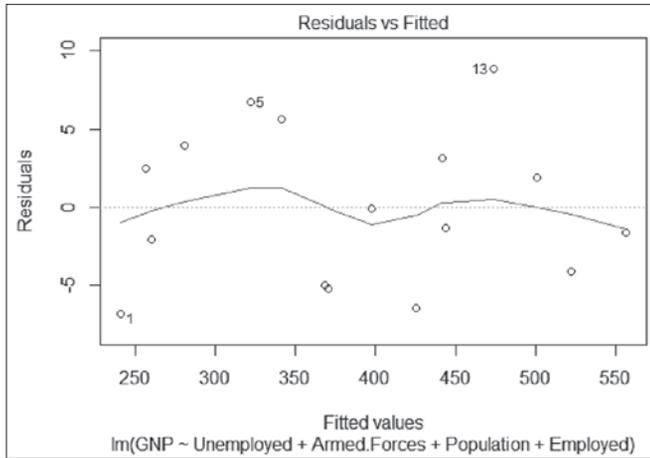
Após ajuste do modelo, é necessário analisar se os resíduos gerados na regressão linear são independentes, com distribuição Normal de média zero e variância constante. Através do comando *plot()*, é possível visualizar alguns gráficos que permitem essas análises.

Gráfico 6: Gráfico de Normalidade para resíduos da regressão ajustada no R



Fonte: Os autores. (2016)

Gráfico 7: Gráfico de resíduos e valores ajustados para a regressão linear simples



Fonte: Os autores. (2016)

No Gráfico 6, a linha pontilhada refere-se a dados hipotéticos gerados a partir de uma distribuição normal. Logo, quando mais aderentes forem os resíduos à essa linha, maior a evidência de que são normais. Porém, apenas a análise gráfica pode não ser suficiente para atestar que os dados são normais e para isso, pode-se realizar o teste de *Shapiro-Wilk*.

O Gráfico 7 é um Diagrama de Dispersão criado com os valores ajustados pelo modelo de regressão no eixo horizontal e os resíduos obtidos no eixo vertical. Através dele observamos que há evidências de que os resíduos possuem média zero e variância constante, porém, a média exata pode ser verificada através do comando `summary(residuals(modelo))` e, para a variância pode ser realizado o teste de *Breusch-Pagan*.

Além da análise de resíduos, é preciso verificar se há colinearidade e multicolinearidade entre as variáveis independentes utilizadas no modelo de regressão. Variáveis fortemente relacionadas ou que podem ser escritas como combinação linear das demais, podem inferir nos resultados e levar o pesquisador a conclusões equivocadas.

A existência de colinearidade pode ser observada através da matriz de correlações das variáveis explicativas do modelo. No R, a matriz pode ser obtida através do comando `cor()` e quanto mais próximos os valores forem de 1, maior o indício de que há colinearidade entre as variáveis. A matriz de correlações do exemplo estudado está apresentada na Figura 7.

Figura 7: Exemplo de Matriz de correlações – Análise de Colinearidade

	GNP	Unemployed	Armed.Forces	Population	Employed
GNP	1.0000000	0.6042609	0.4464368	0.9910901	0.9835516
Unemployed	0.6042609	1.0000000	-0.1774206	0.6865515	0.5024981
Armed.Forces	0.4464368	-0.1774206	1.0000000	0.3644163	0.4573074
Population	0.9910901	0.6865515	0.3644163	1.0000000	0.9603906
Employed	0.9835516	0.5024981	0.4573074	0.9603906	1.0000000

Fonte: Os autores. (2016)

No exemplo, há indícios da existência de colinearidade entre as variáveis população não institucionalizada maior de quatorze anos e número de pessoas empregadas. Pelo princípio da parcimônia, que estabelece que o modelo deve possuir o menor número de parâmetros possível, não se deve incluir variáveis fortemente correlacionadas já que ambas podem dar a mesma contribuição à explicação do modelo. Portanto, o pesquisador deve optar por uma das variáveis a permanecer no modelo e realizar novamente a análise de regressão e a análise de resíduos.

A multicolinearidade pode ser avaliada através do VIF – *Variance Inflation Factor*. O VIF mede o grau em que cada variável independente é explicada pelas demais variáveis independentes. É dado por: $VIF = \frac{1}{1-R^2}$, em que R^2 é o coeficiente de determinação da regressão de cada uma das variáveis explicativas. No R, é preciso instalar o pacote “*HH*” e em seguida, obter o VIF pelo comando `vif()`. Geralmente, VIF indica problemas de multicolinearidade quando é maior que 10. Os resultados para o exemplo estudado estão apresentados na Figura 8.

Figura 8: Exemplo de VIF – Análise de Multicolinearidade

```
> vif(modelo)
unemployed Armed.Forces Population Employed
6.668482 1.964550 50.814189 32.923691
```

Fonte: Os autores. (2016)

A partir da análise do VIF, conclui-se que há problemas de multicolinearidade com as variáveis população não institucionalizada maior que 14 anos e número de empregados. Dessa forma, seria necessário excluir essas variáveis do modelo e realizar novamente a análise de regressão.

Diante dos resultados apresentados, cabe recomendar que as análises de colinearidade e multicolinearidade sejam realizadas na elaboração do modelo, antes de realizar a análise de regressão e posterior análise de resíduos.

4.2.3 Regressão em Painel

Dados em Painel referem-se à observação de diversos indivíduos ($i = 1, \dots, n$), cada um observado em diversos pontos no tempo ($t = 1, \dots, T$) e envolve uma análise mais orientada à perspectiva *cross-section*. A heterogeneidade entre as unidades é o foco central da análise e uma vantagem notável desta abordagem é a possibilidade de modelar diferenças de comportamento entre indivíduos.

Existem três estruturas possíveis em painel: Modelo Pooled, Modelo de Efeitos Fixos e Modelo de Efeitos aleatórios. A Figura 9 mostra a estrutura de cada uma destas modelagens:

Figura 9: Script para dados em painel por três ajustes distintos no R

```
67 # Painel
68 summary(mqp<-plm(InvestativoFixo ~ EstoqueCapital + ValorMercado,
69 data = Dados, model = "pooling"))
70 summary(mfe<-plm(InvestativoFixo ~ EstoqueCapital + ValorMercado,
71 data = Dados, model = "within"))
72 summary(mre<-plm(InvestativoFixo ~ EstoqueCapital + ValorMercado,
73 data = gpdf, model = "random"))
```

Fonte: Os autores. (2016)

A estrutura apresentada na Figura 9, através do comando *summary*, já retorna o valor dos coeficientes, valores de *t*, probabilidades, os graus de liberdade, o R^2 e o R^2 ajustado para cada uma das regressões. *Pooling* se refere ao modelo pooled, *within* a efeitos fixos e *random* a efeitos aleatórios. Ressalta-se que a base denominada no *script* de Dados deve estar organizada em formato painel, com a disposição adequada de linhas e colunas.

A análise de regressão em painel segue os mesmos pressupostos da análise de regressão linear, sendo, portanto, necessário analisar os resíduos, colinearidade e multicolinearidade. Além disso, é preciso verificar se existe autocorrelação de primeira ordem nos resíduos do modelo através do Teste de *Durbin-Watson*.

5. Conclusões

Estudos que visam a caracterização da produção científica são relevantes por ajudarem a comunidade acadêmica a identificar as lacunas teóricas e metodológicas existentes. Ao apresentar métodos e técnicas que podem ser utilizados em Contabilidade e Finanças e mostrar os caminhos a serem percorridos por meio dos *scripts* em um *software* de uso livre, pensa-se atingir o objetivo maior da ciência que é o da propagação de conhecimento.

Devido ao elevado volume de informações a serem disponibilizadas periodicamente por órgãos públicos e privados, é esperado que pesquisas em contabilidade e finanças venham a fazer uso de modelos estatísticos a serem implementados via *scripts*. Para tanto, é necessário, que a aplicação seja acompanhada de

rigor metodológico e cautela quando da análise dos resultados.

No presente artigo, procurou-se identificar as principais metodologias presentes nos artigos publicados entre 2013 e 2015 na área de contabilidade e finanças. Conforme proposto, em decorrência do levantamento dos dados, foram identificados dezoito artigos publicados em 2015, dezesseis em 2014 e quatro em 2013, totalizando trinta e oito artigos, respondendo ao seu objetivo inicial por meio da apresentação dos três métodos e técnicas de pesquisas mais utilizados dado a sua amostra, quais sejam: estatística descritiva, análise de correlação e regressão linear (simples e múltipla). Em seguida apresentaram-se os *scripts* correspondentes na seção de resultados.

Como sugestões para trabalhos futuros têm-se: entender a pesquisa para as metodologias qualitativas que mais estão sendo utilizadas em Contabilidade e Finanças; ampliar o período de análise para os dez últimos anos; coletar os artigos para análise nos principais periódicos da área; apresentar *scripts* de mais métodos e técnicas, pelo menos as cinco mais utilizadas, sejam elas qualitativas ou quantitativas e por último observar as temáticas que têm sido exploradas em contabilidade e finanças nos últimos anos, e por último buscar compreender as razões que levam à reduzida oferta de disciplinas com enfoque quantitativo nos cursos de graduação em ciências contábeis.

Espera-se, portanto, que essa discussão tenha sua importância acrescida em contabilidade e finanças, visto que a conjugação destas áreas apresenta vastas possibilidades de pesquisa.

Referências Bibliográficas

FÁVERO, Luiz Paulo Lopes. Dados em painel em contabilidade e finanças: teoria e aplicação. **Brazilian Business Review**, Vitória, 2013, v. 10, n. 1, pp. 131-156, 2013.

GIL, Antônio Carlos. **Como elaborar projetos de pesquisa**. [S.l.]: Atlas, 2010.

IQUIAPAZA, Robert Aldo; AMARAL, Hudson Fernandes; BRESSAN, Aureliano Angel. Evolução da pesquisa em finanças: epistemologia, paradigma e críticas. **Organizações & Sociedade**, Salvador, 2009, v. 16, n. 49, pp. 351-370, 2009.

LIGGES, Uwe; CRAWLEY, Michael John. The R Book. **Statistical Papers**, England, 2007, v. 50, n. 2, pp. 445-446, 2007.

LUDKE, Menga; ANDRE, Marle EDA. **Pesquisa em educação: abordagem qualitativa**. São Paulo: EPU, 1986.

MARCONI, Marina de Andrade; LAKATOS, Eva Maria. **Metodologia do trabalho científico: procedimentos básicos, pesquisa bibliográfica, projeto e relatório, publicações e trabalhos científicos**, v. 7. São Paulo: Atlas, 2010.

MILAGRE, Robson Amaral. **Estatística: uma proposta de ensino para os cursos de Administração de Empresas**. 200 1. 160 f. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-graduação em Engenharia de Produção, UFSC, 2001.

MODIGLIANI, Franco; MILLER, Merton Howard. The cost of capital, corporation finance and the theory of investment. **The American Economic Review**, Boston, 48, n. 3, pp. 261-297, 1958.

SILVA FILHO, Augusto Souza da; SOUZA, Alessandra Caetano de. Estatística aplicada a Administração usando o software R. **Ciências Gerenciais**, Londrina, v. 17, n. 26, pp. 175-193, 2013.

TOLEDO, Luciano Geraldo; OVALLE, Ivo Izidoro. **Estatística Básica**. 2. ed. São Paulo: Atlas, 1995.

WERKEMA, Maria Cristina C. **Como Estabelecer Conclusões com Confiança: entendendo a inferência estatística**. Belo Horizonte: Fundação Christiano Ottoni; Escola de Engenharia (UFMG), 1996.