

**A ESTATÍSTICA COMO FERRAMENTA PARA AUXILIAR A TOMADA DE  
DECISÃO EM UMA PEQUENA EMPRESA: O USO DA REGRESSÃO LINEAR  
MÚLTIPLA E O SOFTWARE R**

**STATISTICS AS A TOOL TO SUPPORT DECISION-MAKING IN A SMALL  
COMPANY: THE USE OF MULTIPLE LINEAR REGRESSION AND SOFTWARE R**

Maria Clara de Melo Oliveira<sup>1</sup>

Eder Angelo Milani<sup>2</sup>

Josney Freitas Silva<sup>3</sup>

**RESUMO**

As empresas de pequeno porte utilizam o empirismo e a intuição nas tomadas de decisões para alcançarem os resultados desejados. Com isso, muitas organizações não atingem suas metas por falta de recursos estatísticos alinhados com um sistema de gerenciamento de dados. O objetivo deste trabalho foi utilizar o *Software R* para o desenvolvimento de análises descritivas, de regressão e de correlação; estimar valores da variável resposta selecionada pelo pesquisador, com base nos valores conhecidos ou fixados das variáveis regressoras e mostrar aos empreendedores a importância da Estatística como ferramenta para auxiliar o administrador na tomada de decisão. O estudo obteve os dados de uma academia, retirados do livro “Análise Multivariada” (CUNHA; COELHO, p. 136, 2011). Nesse sentido, as informações foram utilizadas a fim de aplicar os recursos proporcionados pela estatística, como a média, a mediana, a variância, o desvio-padrão e a regressão linear múltipla, por meio do desenvolvimento tecnológico. Portanto, as análises permitiram concluir que as

---

1 Graduanda do Curso de Administração na Universidade do Estado de Minas Gerais – UEMG, Frutal, MG, Brasil. E-mail: melo clara9898@yahoo.com

2 Professor Doutor na Universidade Federal de Goiás – UFG, Goiânia, GO. E-mail: edermilani@ufg.br

3 Professor Doutor na Universidade do Estado de Minas Gerais – UEMG, Frutal, MG. E-Mail: josney.silva@uemg.br

ferramentas estatísticas aliadas a um mecanismo de inovação colaboram para uma melhor compreensão dos dados e agilizam o processo decisório de uma empresa.

Palavras-Chave: Tomada de decisão. Regressão linear múltipla. Pequena empresa.

### **ABSTRACT**

*The small businesses use empiricism and intuition in decision-making to achieve the desired results. Thereby, many organizations do not reach their goal because of the lack of statistical resources aligned with a data management system. The objective of this work was to use the Software R to develop descriptive, regressive and correlative analysis; estimate values of the dependent variable selected by the researcher, based on known or fixed values of the independent variables and show entrepreneurs the importance of Statistics as a tool to assist the manager in decision-making. The study obtained data from an academy, taken from the book "Multivariate Analysis" (Cunha e Coelho, p. 136, 2011). In this sense, the information was used in order to apply the resources provided by statistics, such as mean, median, variance, standard deviation and multiple linear regression, through technological development. Therefore, the analysis concluded that statistical tools combined with an innovation mechanism contribute to a better understanding of data and speed up the decision-making process of a business.*

*Keywords: Decision-making. Multiple linear regression. Small business.*

## **1 INTRODUÇÃO**

Com a evolução da Estatística e o avanço no processamento de dados, as ferramentas estatísticas estão se tornando cada vez mais indispensáveis para a tomada de decisão de empresas de grande porte, tais como: bancos, corretoras de seguro, empresas de marketing, empresas ligadas ao mercado financeiro, entre outras.

A tomada de decisão em pequenas empresas, geralmente, se dá de forma empírica, intuitiva baseada, muitas vezes, na experiência de seus administradores. Nestes casos, nem sempre se obtém os melhores resultados para a empresa,

podendo acarretar até mesmo o aumento da taxa de mortalidade de empresas, principalmente nos primeiros anos de vida desses empreendimentos.

Segundo Cunha e Coelho (2011, p. 132), existe, nas empresas, a “necessidade de descrever e prever o comportamento de certas variáveis importantes para a tomada de decisões, tais como: custos, despesas, resultados.” Entretanto, como dito anteriormente, as decisões em pequenas empresas ocorrem de forma empírica e intuitiva, gerando resultados nem sempre satisfatórios.

Este trabalho se justifica nas Diretrizes Curriculares Nacionais do Curso de Administração, em seu art. 4º, dispondo sobre as competências e habilidades que a formação profissional deve possibilitar aos seus egressos. Dentre elas, destaca em seu inciso I, que o Curso de Administração deve:

reconhecer e definir problemas, equacionar soluções, pensar estrategicamente, introduzir modificações no processo produtivo, atuar preventivamente, transferir e generalizar conhecimento e exercer, em diferentes graus de complexidade, o processo de decisão. BRASIL (2005, p. 2).

Para Anderson, Sweeney e Willians, a Estatística, em um sentido amplo, é a “arte e a ciência de se coletar, analisar, apresentar e interpretar dados” (2003, p. 19). E complementam que, para a Administração, uma grande razão para efetuar estas ações inerentes à ciência Estatística é “dar aos gerentes e tomadores de decisão um melhor entendimento do ambiente administrativo e econômico e assim tornar-lhes possível tomar decisões mais fundamentadas e melhores” (2003, p. 19).

Nesta linha de pensamento, e apesar de inserida num contexto de divergências conceituais que, aliás, é próprio do fazer científico, Ponte, Brocardo e Oliveira (2009, p. 105) defendem, assim como muitos outros autores, que a Educação Estatística deve se deslocar do campo do cálculo e da realização de tarefas para o processo de investigação.

O objetivo deste trabalho concentra-se em contribuir para a formação de recursos humanos por meio do uso de técnicas e tecnologias relacionadas à Estatística e que possam contribuir com o desenvolvimento empresarial. Diante do objetivo geral apresentado, os objetivos específicos constituem em utilizar o *Software R* para o desenvolvimento de análises descritivas, de regressão e de correlação,

estimar valores da variável resposta selecionada pelo pesquisador, com base nos valores conhecidos ou fixados das variáveis regressoras e mostrar aos tomadores de decisão a importância da Estatística dentro do ambiente empresarial.

De acordo com Milone (2004, p. 301), “o número de variáveis avaliadas no modelo de estudo é importante, uma vez que as abandonadas sempre reduzem a precisão das estimativas”. Desta forma, os autores supracitados reforçam que “como o poder de predição do modelo é diretamente proporcional ao número de variáveis significativas, quanto mais variáveis forem consideradas, mais precisos os resultados”. Neste caso, as técnicas aplicáveis são a regressão e a correlação múltiplas. E ainda Milone (2004, p.301) complementa e reforça que “como os fenômenos provocados por variáveis isoladas são raros, a regressão e a correlação múltiplas são sempre metodologias teóricas mais adequadas”.

Para a aplicação eficiente das técnicas de análise de regressão e correlação, uma ferramenta computacional interessante é o *Software R*. Considerado por Jelihovschi (2014) como a maior criação Estatística nos últimos vinte anos por explorar, de forma mais eficiente possível, todo poder computacional na criação de um *software* único para toda necessidade computacional que a Estatística necessitava. Trata-se de um *software* disponibilizado sob os termos da GNU - *General Public License* da *Free Software Foundation* na forma de código aberto.

## 2 REFERENCIAL TEÓRICO

### 2.1 As Micro e Pequenas Empresas no Universo Mercadológico

Segundo Lemes Júnior e Pisa (2010, p.43), não existe um único requisito para definir as pequenas empresas, seja no Brasil ou em qualquer lugar do mundo, pois as diferentes classificações são determinadas de país para país, entre regiões de uma mesma nação e, até mesmo, entre órgãos governamentais ou não. De acordo com Viapiana (2001, apud LEMES JÚNIOR E PISA, 2010, pg. 43), “é justamente a dificuldade de caracterização, devido aos diferentes critérios, tanto qualitativos quanto quantitativos, adotados pelos órgãos públicos, um dos maiores entraves ao se estudar as MPEs (Micro e Pequenas Empresas) no Brasil.” (LEMES JÚNIOR e PISA, 2010, p.43).

Para Dornelas (2005, p.18), é importante realizar um estudo aprofundado no conceito de empreendedorismo, sendo que a maioria dos negócios criados são de pequenos empreendedores. Dessa forma, esses empresários atuam geralmente de forma empírica e sem planejamento, fazendo com que isso afete o alto índice de mortalidade desse tipo de empresa que chegou a 45,8% no ano de 2008. Mas em comparação com ano de 2012 decaiu para 23,4%, segundo pesquisa do Sebrae (Serviço Brasileiro de Apoio às Micro e Pequenas Empresas, 2016). Mesmo com a diminuição do percentual de mortalidade, é preciso planejar, organizar, dirigir e controlar para que os processos inseridos dentro da organização se desenvolvam e assim aumente a durabilidade da empresa no mercado atual. E o mesmo ponto de vista é defendido por Lemes Júnior e Pisa que diz o seguinte:

O resultado é que as vendas diminuem, as pressões sobre o caixa aumentem e exigem mais dedicação ao curto prazo, transformando-se num círculo vicioso que pode obrigar a empresa a fechar as portas, em consequência, principalmente, de: falta de planejamento; carência de inovação; vendas insuficientes; caixa negativo e inadimplência (2010, p.50).

No Brasil, a participação das micro e pequenas empresas no PIB (Produto Interno Bruto) possui o menor índice quando comparados aos outros países. Segundo a Micro e Pequenas Empresas em Foco apud Lemes Júnior e Pisa (2010, p.46), a nação brasileira possui apenas 20%, contra 64% da Espanha, 40% do Canadá, 35% da Alemanha e 27% da França.

O governo, para garantir a sobrevivência e reduzir a informalidade das MPEs, criou algumas políticas que são apresentadas por meio da Constituição Federal de 1988, da Lei nº 123/2006 (Estatuto da MPEs) e Lei nº 8666/1993 – Licitações e Contratos da Administração Pública.

Para Everton Junior (2017, p.23), as empresas têm logrado êxito e elevado crescimento em número não somente em virtude das leis proporcionadas a elas, mas também, pela capacidade de se multiplicar em relação ao empreendedorismo, à busca de oportunidade e às necessidades de fazer algo que renda dinheiro. E diante desses fatores, as empresas estão aumentando a sua participação nos mercados mais populosos e, juntamente, às regiões com maiores recursos financeiros.

As incubadoras possuem um outro papel fundamental que é o processo de inovação tecnológica nas MPEs existentes. Isso colabora para a diminuição do índice de mortalidade, uma vez que mais da metade das micro, pequenas e médias empresas encerram suas atividades até o terceiro ano de sua existência, segundo o Sebrae. Conforme apresentam Lemes Júnior e Pisa (2010, p.29).

Segundo Longenecker et al. (2007, p.162), “avanços tecnológicos aplicados aos negócios são catalisadores importantes para o rápido crescimento de empresas caseiras.” Para complementar essa ideia, Lemes Júnior e Pisa (2010, p.44) dizem que com o alto índice da taxa de desemprego e o crescimento da informalidade, as MPEs estão investindo em computadores e softwares para transmitir aos clientes uma imagem de competência, modernidade e confiabilidade. Essa modernização tecnológica também é defendida por Dornelas (2005, p.108), determinando que ao utilizar um software em seu negócio, o empreendedor agiliza o seu processo dentro da organização.

De acordo com o Everton Junior (2017, p.11), a vulnerabilidade das pequenas empresas está associada com as características intrínsecas do negócio, ou seja, esse problema tem relação com dependência familiar ainda existente dentro das empresas, pois mais ou menos a metade das empresas possuem essa característica demonstrada acima e assim gera uma confusão no caixa das organizações que pode atender muito mais aos interesses da família do que aos da empresa. Mas, mesmo com todas essas questões que precisam ser melhoradas, as MPEs representam mais de 96% do número de empresas formais, conforme Everton Junior (2017, p.23).

Diante do mapa apresentado por Everton Junior (2017, p.24), conclui-se que há uma relação entre o número da população e a quantidade de MPEs, ou seja, quanto maior forem os habitantes de uma determinada região, maior será o número de micro e pequenas empresas. De acordo com as informações vistas, pode dizer que a concentração desse tipo de organização é nas regiões Sul, Sudeste e Nordeste.

Conforme as tecnologias aplicadas nas MPEs, é importante que elas estejam associadas à Estatística, pois é por meio da análise matemática que a empresa saberá se está tendo lucro ou prejuízo. De acordo com Longenecker et al. (2007, p.365), as ferramentas e técnicas da gestão da qualidade total envolve-se com os métodos estatísticos, servindo assim para o controle da qualidade. A utilização dessas medidas

faz com que as empresas economizem com as taxas de consultoria cobradas na criação de um programa eficaz. Por meio dessas análises estatísticas, é possível estabelecer limites de tolerância nas variações decorrentes do acaso. Mas se ocorrer algum desvio dentro do limite permitido, o controlador de qualidade percebe que existe algum problema e, automaticamente, recorre a solução. Com o aumento da tecnologia, os recursos estatísticos de controle também estão em processo de evolução nas pequenas empresas.

## 2.2 As ferramentas estatísticas

Segundo Spiegel (1993, p.75), quando o conjunto de dados é organizado em ordem de grandeza, o número central ou a média aritmética dos dois valores centrais que proporciona duas partes iguais dentro do conjunto é a mediana.

Segundo a Equipe Estatcamp (2014), a regra dos 5 itens divide a distribuição em 4 partes iguais e proporciona as seguintes informações:

- Valor mínimo: menor valor de uma amostra;
- 1° Quartil: deixa 25% das observações abaixo e 75% acima;
- 2° Quartil: deixa 50% das observações abaixo e 50% acima, ou seja, é o valor da mediana;
- 3° Quartil: deixa 75% das observações abaixo e 25% acima;
- Valor máximo: maior valor de uma amostra.

Os quartis, para Anderson, Sweeney e Williams (2003, p. 85), podem ser representados pela seguinte fórmula:

$$i = \frac{q}{4} \times n \quad (1)$$

Onde  $q$  é igual a 1, 2 ou 3, e representa o quartil de interesse,  $i$  corresponde a posição no conjunto de dados onde se encontra o quartil procurado que deve estar em ordem crescente. Se  $i$  não for um número inteiro, o próximo inteiro maior que  $i$  denota a posição do quartil. Se  $i$  for um número inteiro, o quartil procurado é a média dos valores de dados que ocupam a posição  $i$  e  $i + 1$ .

De acordo Spiegel (1993, p.67), a média representa um valor típico de um conjunto de dados. Como ele tende a se dispor em um ponto central, ou seja, as médias são identificadas como medidas de tendência central. Com isso, A média aritmética, ou média, de um conjunto de N números  $x_1, x_2, \dots, x_n$  é representada por:

$$\bar{x} = \frac{x_1+x_2+x_3+\dots+x_n}{n} = \frac{\sum_{i=1}^n x_i}{n} = \frac{\sum x}{n} \quad (2)$$

O conjunto de dados numéricos, para Spiegel (1993, p.104), tende a dispersar em relação a um valor médio, chamado de variação, dispersão dos dados ou desvio padrão. Dessa forma, o desvio padrão, para Anderson, Sweeney e Williams (2003, p.92), de um conjunto N de números  $x_1, x_2, \dots, x_n$  é representado por s e definido por:

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \quad (3)$$

- $x_i - \bar{x}$  representa o desvio de cada um dos números  $x_i$  em relação à média  $\bar{x}$ .

Já a variância de um agrupamento de dados é apresentada como o quadrado do desvio padrão, ou seja, é o valor de  $s^2$ .

Para diferenciar o desvio padrão de uma população e o de uma amostra dela extraída, adota-se frequentemente o símbolo s para o último e  $\sigma^2$  para o primeiro caso. Com isso,  $s^2$  e  $\sigma^2$  representariam a variância da amostra e a da população, respectivamente.

Segundo Spiegel (1993, p.359), a correlação linear procura determinar uma equação linear entre as variáveis. O coeficiente de correlação analisa se duas variáveis possuem uma relação forte, fraca ou se não existe uma associação linear entre as variáveis (resposta e regressora). De acordo com Anderson, Sweeney e Williams (2003, p. 109), a fórmula é a seguinte:

$$r = \frac{\sum x_i y_i - (\sum x_i \sum y_i) / n}{\sqrt{\sum x_i^2 - (\sum x_i)^2 / n} \sqrt{\sum y_i^2 - (\sum y_i)^2 / n}} \quad (4)$$



A Tabela 1 mostra a classificação de cada intervalo dos valores com as suas respectivas interpretações:

**Tabela 1: Interpretação do Coeficiente de Correlação**

Valor de r (+ ou -)	Interpretação
0.00 a 0.19	Uma correlação bem fraca
0.20 a 0.39	Uma correlação fraca
0.40 a 0.69	Uma correlação moderada
0.70 a 0.89	Uma correlação forte
0.90 a 1.00	Uma correlação muito forte

Fonte: Shimakura (2006).

A distribuição de frequência, de acordo com Martins e Donaire (1990, p.75), significa uma série estatística específica em que os dados estão distribuídos em classes ou categorias junto com as frequências correspondentes. Com isso, as distribuições de frequências podem ser divididas em dois tipos: variável discreta ou variável contínua:

- Variável discreta: a variável adquire valores em pontos da reta real. Geralmente, esta variável assume valores inteiros.
- Variável contínua: a variável assume valores em intervalos da reta real e, com isso, não há possibilidade de enumerar todos os valores. Geralmente, esta variável provém de medições.

Assim, para realizar a distribuição dos dados, são necessárias as definições de algumas ferramentas, como:

- Dados brutos: aqueles que não foram numericamente organizados.
- Rol: os dados brutos em ordem de grandeza crescente ou decrescente.
- Limites de classe: os números extremos de cada classe, ou seja, limite inferior e o superior.
- Intervalo de classe: várias maneiras de apresentar o intervalo de classes: iguais ou diferentes entre si.
- Amplitude total ou "Range": a diferença entre o maior e o menor dado.
- Ponto médio das classes: é a média aritmética entre o limite superior e o limite inferior da classe.

A análise de regressão linear múltipla aborda a relação existente entre uma única variável resposta (critério) e várias variáveis regressoras (preditoras). Essas

variáveis regressoras denotam um peso em relação a previsão geral e colaboram na interpretação da influência de cada variável em realizar a previsão, embora a correlação entre essas variáveis dificultam a análise interpretativa<sup>4</sup>.

Quando o problema se relaciona com apenas uma variável regressora, o recurso estatístico é chamado de regressão linear simples. Já quando o problema apresenta duas ou mais variáveis regressoras, tem-se o nome de regressão linear múltipla. Com isso, três passos podem ser mencionados para a estimação de regressão entre variáveis regressoras e resposta<sup>5</sup>:

1. Estabelecimento de uma previsão de referência sem variável regressora, usando somente a média da variável resposta.
2. Previsão usando uma única variável regressora – regressão simples
3. Previsão usando diversas variáveis regressoras – regressão múltipla.

No estabelecimento de uma previsão de referência sem variável regressora deve ser realizada a média da variável resposta. A partir desse cálculo, é preciso fazer a soma de quadrados dos erros ( $SS_E$ ), ou seja, ela proporciona uma medida de previsão que varia conforme a quantia de erros de previsão<sup>6</sup>.

Um outro ponto importante é a previsão usando uma única variável regressora e nela atuará o coeficiente de correlação, para que se descubra o melhor valor de interação entre as variáveis. Além disso, para que haja uma maior eficiência na fórmula, o intercepto pode ser utilizado na fórmula geral<sup>7</sup>:

$$\hat{y} = b_0 + b_1 v_1 \tag{5}$$

- $b_0$  intercepto;
- $b_1$  coeficiente de correlação;
- $v_1$  valor da variável independente.

---

4 HAIR JR., 2009, p. 154.

5 Ibid., p.154.

6 Ibid., p.155.

7 Ibid., p.157.

Na parte da avaliação da precisão de previsão, a soma de quadrados dos erros ( $SS_E$ ) apresenta uma medida de erros de previsão e, além disso, é necessária a abordagem do sucesso de previsão, a qual pode ser chamada de soma de quadrados de regressão ( $SS_R$ ). A soma de ambas as medidas compõe a soma total de quadrados ( $SS_T$ )<sup>8</sup>:

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2, \text{ ou seja, } SS_T = SS_E + SS_R \quad (6)$$

- $\bar{y}$  média de todas as observações;
- $y_i$  valor da observação individual  $i$ ;
- $\hat{y}_i$  valor previsto da observação  $i$ .

O coeficiente de determinação ( $R^2$ ) está relacionado com a soma de quadrados de regressão dividida pela soma total de quadrados. Isso demonstra que quanto maior for o valor do coeficiente de determinação entre a variável resposta com às regressoras, compreender-se-á que a relação encontrada explicará a variação possível da variável resposta<sup>9</sup>.

A parte da regressão linear múltipla tem associação com a colinearidade (correlação entre duas variáveis regressoras) e com multicolinearidade (correlação entre três ou mais variáveis regressoras). Com isso, para maximizar a previsão de acordo com as variáveis regressoras, a análise deve conter variáveis regressoras com baixa multicolinearidade, porém com correlações elevadas em comparação com a variável resposta<sup>10</sup>.

Dessa forma, a fórmula geral da regressão linear múltipla é<sup>11</sup>:

$$y = b_0 + b_1 v_1 + \dots + b_n v_n + e \quad (7)$$

- $y$  variável resposta;
- $b_0$  valor constante;
- $b_1$  e  $b_n$  variação da variável resposta com a variação da variável regressora;
- $V_1$  e  $V_n$  variáveis regressoras;
- $e$  erro de previsão (resíduo).

---

8 Ibid., p.159.

9 Ibid., p.160.

10 Ibid., p.161.

11 Ibid., p.161.

### 3 Resultados

#### 3.1 Apresentação dos dados da academia

Para apresentar um exemplo de aplicação destes conceitos estatísticos com o uso do *Software R*, utilizou-se os dados apresentados por Cunha e Coelho (2011). Os principais comandos utilizados no R para a realização da análise estatística são apresentados no **Apêndice**. Abaixo segue a tabela com os dados da academia de ginástica:

**Tabela 2: Resultados da pesquisa sobre os gastos da academia**

Identificação do período	Gastos da academia	Consumo de Kilowatts	Horas de MOD	Número de alunos
1	350	6	10	100
2	400	8	14	110
3	470	12	16	110
4	550	10	26	98
5	620	15	24	112
6	380	7	12	95
7	290	6	13	75
8	490	9	21	124
9	580	11	20	126
10	610	13	24	116
11	560	12	23	99
12	420	14	12	104
13	450	11	19	108
14	510	12	19	108
15	380	9	11	98

Fonte: Cunha e Coelho (2011)

#### 3.2 Aplicação das variáveis da academia no *Software R*

As variáveis da academia serão: gastos da academia, consumo de Kilowatts, horas de MOD (mão de obra direta) e número de alunos. Dessa forma, a primeira variável é a resposta e as outras, as regressoras.

A Tabela 3 apresenta a regra dos cinco itens. Essa ferramenta colabora no entendimento básico em relação às variáveis da academia e ajuda na compreensão dos recursos estatísticos seguintes:

**Tabela 3: A regra dos cinco itens e a média**

Variáveis	Min.	1° Quartil	Mediana	Média	3° Quartil	Max.
<b>Gastos da academia</b>	290	390	470	470,7	555	620
<b>Consumo de Kilowatts</b>	6	8,50	11	10,33	12	15
<b>Horas de MOD</b>	10	12,5	19	17,6	22	26
<b>Número de alunos</b>	75	98,5	108	104,9	111	126

Fonte: próprios autores

A variância e o desvio padrão têm como função medir a distância dos seus dados em relação à média, ou seja, a empresa pode perceber o quanto dos seus padrões está saindo dos eixos. Já o coeficiente de correlação aborda a interação entre a variável resposta com a regressora, levando o administrador a conhecer a variável regressora que mais explica à dependente. Abaixo segue a tabela da variância, do desvio padrão e do coeficiente:

**Tabela 4: Variância e desvio padrão**

Variáveis	Variância	Desvio padrão	Coefic.C.
<b>Gastos da academia</b>	10106,67	100,53	-
<b>Consumo de Kilowatts</b>	7,81	2,79	0,76
<b>Horas de MOD</b>	28,83	5,37	0,89
<b>Número de alunos</b>	156,84	12,52	0,67

Fonte: próprios autores

Nas análises a seguir, será mostrada a distribuição de frequência em intervalos de classes iguais para cada variável. Esse recurso serve para verificar em qual intervalo de classes está concentrado a maior quantidade de dados e fazer comparações com anos anteriores. A Tabela 5 apresenta a frequência absoluta, frequência relativa e frequência percentual:

**Tabela 5: Frequência absoluta, relativa e percentual**

Gastos da academia				Horas de MOD			
Intervalo de classes	Fa	Fr	Fp	Intervalo de classes	Fa	Fr	Fp
[0; 500)	9	0,6	60%	[9; 20)	10	0,67	66,67%
[500;1000)	6	0,4	40%	[20;31)	5	0,33	33,33%
<b>Total</b>	15	1	100%	<b>Total</b>	15	1	100%

Consumo de Kilowatts				Número de alunos			
[1;10)	7	0,47	46,67%	[70; 110)	11	0,73	73,33%
[10;20)	8	0,53	53,33%	[110;150)	4	0,27	26,67%
<b>Total</b>	<b>15</b>	<b>1</b>	<b>100%</b>	<b>Total</b>	<b>15</b>	<b>1</b>	<b>100%</b>

Fonte: próprios autores

A regressão linear múltipla serve para explicar o quanto as variáveis regressoras determinam à resposta, isto é, elas demonstram qual a relação existente entre os dados. Para alcançar a fórmula geral da regressão linear múltipla, primeiro foi realizada a relação da variável resposta com as regressoras e o maior resultado de coeficiente de determinação foi a interação entre as três variáveis regressoras sem o intercepto. Na Tabela 6, são apresentados os coeficientes da regressão linear múltipla:

**Tabela 6: Coeficientes de regressão**

Variáveis	Coeficientes de regressão
<b>b2</b>	10,914
<b>b3</b>	11,520
<b>b4</b>	1,482
<b>Coeficiente de determinação: 0,9971</b>	

Fonte: próprios autores.

A fórmula geral da regressão linear múltipla:

$$\bullet \hat{y} = 10,914 x_2 + 11,520 x_3 + 1,482 x_4 \quad (8)$$

Com o coeficiente de determinação que neste caso é de 0,9971, isso significa que 99,71% da variação dos valores da variável resposta são explicados pelos valores das variáveis regressoras, ou seja, há uma interação muito forte entre esses dados.

#### 4 Considerações finais

Portanto, as variáveis foram utilizadas para a aprendizagem dos comandos no *software R* e para incentivar e explicar aos empresários como utilizar as ferramentas por meio desse recurso tecnológico. Com isso, os empreendedores conseguem alcançar um melhor resultado ao fazer a análise dos dados da empresa.

O *software R* proporciona aos pequenos empresários uma versatilidade de recursos estatísticos e assim os objetivos da empresa são adquiridos de forma mais eficiente e eficaz. Com essa tecnologia, os administradores observarão os pontos fortes e fracos da organização e, dessa forma, saberão em que aspecto eles terão de melhorar para obter mais rentabilidade, produtividade e menos perdas. Além disso, por ser um *Software* livre, as empresas de pequeno porte podem ter acesso facilitado, não sendo necessária a compra de *softwares* caros para auxílio na tomada de decisão, com base em análises estatísticas mais sofisticadas.

Com o auxílio do *Software R* e das ferramentas estatísticas é possível perceber que os dados da academia possuem uma interação forte entre as variáveis, principalmente entre os gastos da academia (resposta) e horas de mão de obra (regressora). Uma outra análise importante é a da regressão linear múltipla, pois ela permite a criação de uma expressão geral entre variáveis resposta e regressora para o descobrimento do valor da resposta no futuro.

Desse modo, os recursos estatísticos e tecnológicos permitem uma análise mais técnica das informações geradas pela empresa e assim podem alcançar suas metas de maneira mais célere e com melhor eficiência e eficácia.

## REFERÊNCIAS

ANDERSON, D.R.; SWEENEY, D. J.; WILLIAMS, T. A. **Estatística aplicada a administração e economia**. (Tradução de Luiz Sérgio de Castro Paiva) São Paulo: Pioneira Thomson Learning, 2003.

AQUINO, J. A. **R para cientistas sociais**. Ilhéus, BA: EDITUS, 2014. 157 p.: il. Disponível em:< [http://www.uesc.br/editora/livrosdigitais\\_20140513/r\\_cientistas.pdf](http://www.uesc.br/editora/livrosdigitais_20140513/r_cientistas.pdf). Acesso em: 10 ago. 2019.

BRASIL. **Resolução nº 4, de 13 de julho de 2005**. Institui as diretrizes curriculares nacionais do curso de administração, bacharelado e dá outras providências. Diário Oficial da União. Brasília, DF, nº 137, 19 jul. de 2005. Seção 01.

BRASIL. Presidência da República. Casa Civil. Subchefia para Assuntos Jurídicos. **CONSTITUIÇÃO DA REPÚBLICA FEDERATIVA DO BRASIL DE 1988**. Brasília, Diário Oficial da União, 05 out. 1988. Disponível em:<[http://www.planalto.gov.br/ccivil\\_03/constituicao/ConstituicaoCompilado.htm](http://www.planalto.gov.br/ccivil_03/constituicao/ConstituicaoCompilado.htm)>. Acesso em: 15 ago. 2019.

BRASIL. **Lei Complementar nº 123, de 14 de dezembro de 2006**. Brasília, Diário Oficial da União, 14 dez. 2006. Disponível

em:<<https://www2.camara.leg.br/legin/fed/leicom/2006/leicomplementar-123-14-dezembro-2006-548099-publicacaooriginal-63080-pl.htm>>. Acesso em: 15 ago. 2019.

BRASIL. Presidência da República. Casa Civil. Subchefia para Assuntos Jurídicos. **LEI Nº 8.666, DE 21 DE JUNHO DE 1993**. Brasília, Diário Oficial da União, 12 jun. 1993. Disponível em:<[http://www.planalto.gov.br/ccivil\\_03/leis/L8666compilado.htm](http://www.planalto.gov.br/ccivil_03/leis/L8666compilado.htm)>. Acesso em: 15 ago. 2019.

CUNHA, J. V. A.; COELHO, A. C. Regressão Linear Múltipla. In: CORRAR, L. J.; PAULO, E.; DIAS FILHO, J. M. **Análise Multivariada: para Cursos de Administração, Ciências Contábeis e Economia**. – 1.ed. – 3. reimpressão. - São Paulo: Atlas, 2011.

DORNELAS, J. C. A. **Empreendedorismo: Transformando ideias em negócios**. - 2.ed. – 6ª reimpressão – Rio de Janeiro: Elsevier, 2005.

EQUIPE ESTATCAMP. Estatística básica: **quartis**. Estatcamp - Consultoria em estatística e qualidade. - São Carlos: 2014. Disponível em:<<http://www.portalaction.com.br/estatistica-basica/23-quartis>>. Acesso em: 16 ago. 2019.

EVERTON JUNIOR, A. **MPE: avanços importantes para as micro e pequenas empresas 2017-2018**. – Rio de Janeiro: Confederação Nacional do Comércio de Bens, Serviços e Turismo, 2017. Disponível em:<[http://cnc.org.br/sites/default/files/arquivos/mpe\\_-\\_avancos\\_importantes\\_2017-2018.pdf](http://cnc.org.br/sites/default/files/arquivos/mpe_-_avancos_importantes_2017-2018.pdf)> Acesso em: 10 ago. 2019.

HAIR Jr., J.F. et al. **Análise multivariada dados**. Tradução Adonai Schlup San't Anna. – 6.ed. – Porto Alegre: Bookman, 2009.

ITANO, F.; SANTOS, S.M. **Tópicos de Estatística utilizando R**. Instituto de Matemática e Estatística – Universidade de São Paulo, 2007. Disponível em:<<https://cran.r-project.org/doc/contrib/Itano-descriptive-stats.pdf>>. Acesso em: 21 nov. 2017.

JELIHOVSCHI, E. **Análise Exploratória de Dados usando o R**. Ilheus-BA: IDITUS, 2014. Disponível em: [http://www.uesc.br/editora/livrosdigitais2/analiseexploratoria\\_r.pdf](http://www.uesc.br/editora/livrosdigitais2/analiseexploratoria_r.pdf) . Acesso em: 03 jan. 2017.

LANDEIRO, V.L. **Introdução ao uso do programa R**. Instituto Nacional de Pesquisas da Amazônia – Programa de Pós Graduação em Ecologia. Disponível em:<<https://cran.r-project.org/doc/contrib/Landeiro-Introducao.pdf>>. Acesso em: 12 nov. 2017.

LEMES JÚNIOR, A. B.; PISA, B. J. **Administrando micro e pequenas empresas**. – Rio de Janeiro: Elsevier, 2010.

LONGENECKER, J. G. et al. **Administração de pequenas empresas**. - 13.ed.- São Paulo: Thomson Learning, 2007.

MARTINS, G. A.; DONAIRE, D. **Princípios de estatística**. – 4.ed. – São Paulo: Atlas, 1990.



MILONE, G. **Estatística Geral e Aplicada**. São Paulo: Pioneira Thomson Learning, 2004.

PONTE, J. P.; BROCARD, J.; OLIVEIRA, H. Investigações em estatística. In: \_\_\_\_\_. **Investigações matemáticas em sala de aula**. Belo Horizonte: Autêntica, 2009.

SEBRAE. **Sobrevivência das Empresas no Brasil**. Portal Sebrae, 2016. Disponível em: <<https://m.sebrae.com.br/Sebrae/Portal%20Sebrae/Anexos/sobrevivencia-das-empresas-no-brasil-relatorio-apresentacao-2016.pdf>>. Acesso em: 16 ago. 2019.

SHIMAKURA, S. E. **Interpretação do coeficiente de correlação**. Disponível em <<http://leg.ufpr.br/~silvia/CE003/node74.html>>. Acesso em: 10 abr 2019.

SPIEGEL, Murray R. **Estatística**. Tradução e revisão técnica Pedro Cosentino – 3ª ed. – São Paulo: Makron Books, 1993.

## Apêndice

Abaixo apresenta a tabela com os comandos utilizados para o desenvolvimento das ferramentas estatísticas, baseando-se no Landeiro (2011), Jelihovschi (2014), Aquino (2014) e Itano e Santos (2007).

**Tabela 7: Comandos do Software R**

<b>Baixar e transferir os dados para o Software R</b>	
<b>Comando</b>	<p>1 - Baixar o <b>Software R</b> por meio do link: <a href="https://www.r-project.org/">https://www.r-project.org/</a>;</p> <p>2 - Copiar os dados no bloco de notas, que contém o formato "txt";</p> <p>3 - Instalar o pacote "descr" # "packages" – "Brazil"- "descr";</p> <p>4 - library(descr) # carrega o pacote "descr";</p> <p>5 - getwd() # mostra o diretório padrão. Para isso, o arquivo com os dados no formato "txt" deve ser salvo no "change dir" que se localiza no "file";</p> <p>6 - Nome dado ao conjunto de dados &lt;- read.table ("C:\\Users\\nome do usuário do computador\\local do arquivo\\nome da pasta\\nome do arquivo.txt") # localiza o local em que estão os dados;</p> <p>7 - Colocar novamente só o nome dado ao conjunto de dados # os dados serão mostrados no software R, ou seja, eles serão transportados do bloco de notas para o programa;</p> <p>8 - ls (nome dado ao conjunto de dados) # mostra nome das variáveis que, neste caso, são quatro: V1, V2, V3 e V4;</p> <p>9 - Nome da variável = nome dado ao conjunto de dados \$V1# V1 será nomeado.</p>

<b>Regra dos 5 itens e a média das variáveis</b>	
<b>Comando</b>	<code>summary(nome dados ao comando\$V1, V2, V3 ou V4)</code>
<b>Variância</b>	
<b>Comando</b>	<code>var(nome dado ao conjunto de dados\$V1, V2,V3 ou V4)</code>
<b>Desvio padrão</b>	
<b>Comando</b>	<code>sd(nome dado ao conjunto de dados\$V1, V2,V3 ou V4)</code>
<b>Coefficiente de correlação</b>	
<b>Comando</b>	<code>cor(nome dado ao conjunto de dados\$V4,nome dado ao conjunto de dados\$V1, V2 ou V3)</code>
<b>Frequência absoluta, relativa e percentual</b>	
<b>Comando</b>	<pre> a&lt;-c(nome dado ao conjunto de dados\$ coloca-se V1,V2,V3 ou V4) df&lt;-matrix(0,quantidade de classes,3) colnames(df)&lt;-c("fa", "fr", "fp") rownames(df)&lt;-c("[intervalos de classes]", "Total") tab.a&lt;-table(cut(a,breaks=c(os números que estão nos intervalos de classes))) df[1:2,1]&lt;-tab.a df[3,1]&lt;-length(a) for(i in 1: total de classes) (df[i, segunda classe ]&lt;-df[i, primeira classe]/length(a)) for(i in 1:3) (df[i,3]&lt;-df[i,2]*100) Df </pre>
<b>Regressão linear múltipla</b>	
<b>Comando</b>	<pre> a&lt;-nome dado ao conjunto de dados[,1] V2&lt;-nome dado ao conjunto de dados[,2] V3&lt;-nome dado ao conjunto de dados[,3] V4&lt;- nome dado ao conjunto de dados [,4] lm(a~V2, V3 ou V4) reg.mult&lt;-lm(a~V2, V3 ou V4) reg.mult summary(reg.mult) </pre> <p><b>Observação:</b> quando for necessário retirar o intercepto, basta colocar "0" depois de "a~".</p>

Fonte: próprios autores

*Recebido em 20/08/2019*

*Aprovado em 25/11/2019*