

SELEÇÃO DE SOFTWARE UTILIZANDO ANÁLISE DE DECISÕES E GERENCIAMENTO DE TECNOLOGIA E SISTEMAS DE INFORMAÇÃO

Danilo Augusto Sarti

Universidade de São Paulo, São Paulo, SP, Brasil

RESUMO

Este trabalho objetivou determinar a melhor alternativa, dentre cinco alternativas, para implementação de análises estatísticas através de software em uma empresa de sementes. A metodologia utilizada considerou custos relacionados ao processo, programação em R e sua relação com a decisão acerca do software estatístico da empresa através da ferramenta de análise de decisões (Howard, 1988; Howard, 2004) aplicada no contexto da gestão de T.I. Os resultados mostram redução de custos no processo estatístico e mudança na decisão sobre software estatístico a ser utilizado pela empresa. O trabalho contemplou apenas uma aplicação específica da empresa e a solução adotada pode ser melhorada com uso da plataforma Rstudio, além dos *packages Knitr* e *shiny*.

Palavras-chave: Análise de Decisões, Gerenciamento de sistemas de Informação, Open Source/R, Seleção de Software, TI no agronegócio.

1. INTRODUÇÃO

1.1. Objetivos:

O objetivo central do trabalho foi utilizar a ferramenta de análise de decisões proposta por Howard, 1988; Howard, 2004, descrita no referencial teórico, para selecionar a melhor alternativa para software, dentre cinco alternativas consideradas, para realização do teste de Tukey largamente utilizado por uma empresa. Quatro das alternativas eram programadas em R e uma alternativa envolvia o uso de um software comercial para a finalidade. A ferramenta descrita por Howard (1988; 2004) foi eleita por minimizar os problemas cognitivos e de outras ordens durante tomadas de decisão e descritos em Bell & Raiffa (1990); Hogarth, R (1991) e Plessner & Bletsch & Bletsch (2007).

Além do foco central anteriormente citado, outros objetivos foram buscados. Dentre eles, podemos destacar a quantificação dos custos relacionados ao processo de

Manuscript first received/Recebido em: 23/07/2013 Manuscript accepted/Aprovado em: 09/02/2015

Address for correspondence / Endereço para correspondência

Danilo Augusto Sarti, Ex-aluno LES/USP, Eng Agrônomo/USP Msc. Applied Economics/USP, Data scientist, Aluno especial do IMECC/Unicamp, Address for correspondence: Av 33 1999 Estádio 13501440 Rio Claro SP daniloasarti@gmail.com (preferred contact).

Published by/ Publicado por: TECSI FEA USP – 2015 All rights reserved.

análise em cada alternativa e quantificação da redução de custos verificada durante o processo de construção da base de decisão (conceito definido no referencial).

O trabalho também pretende ilustrar como uma função em R pode servir como interface eficiente e a influência do nível dessa interface na preferência dos usuários por diversas alternativas consideradas.

Objetivou-se também a ilustração da possibilidade do uso de plataformas livres em empresas, além de evidenciar como habilidades de programação em R impactaram a decisão em relação ao software mais adequado ao sistema de análise da empresa.

1.2 Justificativa:

O trabalho se justifica por ilustrar a resolução do problema de seleção de software de maneira formalizada. Essa formalização permite o registro das informações utilizadas no momento da tomada de decisão.

A ilustração do uso de análise de decisões em gerenciamento de sistemas de informação pode servir como base para decisões em escolhas de sistemas futuros ou, até mesmo, em outras áreas do conhecimento humano.

A redução de custo obtida no processo de resolução do problema foi significativa, ilustrando grande potencial para uso de plataformas *open source*, de forma criativa, em diversos setores, conferindo, dessa forma, competitividade às organizações. A questão dos custos é importante para qualquer negócio, especialmente para negócios nascentes, nos quais, em geral, são realizados investimentos iniciais cujo retorno é incerto.

O trabalho serve como justificativa para o argumento de que o uso de plataformas livres de forma criativa e eficiente em empresas e seu ensino em universidades pode reduzir a dependência tecnológica de organizações por plataformas com elevados custos de licença.

1.3 Da demanda e importância de software estatístico

Empresas que realizam pesquisa aplicada em agricultura são altamente dependentes de análises estatísticas de dados coletados em experimentos realizados. Este é o caso da indústria de melhoramento de plantas e produção de sementes que utilizam análises estatísticas para tomada de decisões em vários setores de sua atividade.

Usualmente, essas análises são realizadas por softwares que geram relatórios para o tomador de decisão ou para o analista envolvido no desenvolvimento de produtos. No processo do melhoramento genético de plantas, diversas são as análises utilizadas para responder questões acerca do desempenho produtivo e de outras características dos materiais vegetais ensaiados.

Em geral, os ensaios de produtos são realizados em delineamento estatístico apropriado como forma de responder às questões de interesse. Um tipo típico de análise que corresponde à grande parte do número de análises dos experimentos conduzidos são testes de comparação de médias, como o teste de Tukey.

A produção dessas análises incorre em custos para a companhia, principalmente custos devidos ao pagamento de licenças, em caso do uso de softwares

comerciais, além de custos relacionados ao processo de geração da análise em si. No caso de pequenas companhias tais custos podem ser expressivos e afetar a sua competitividade.

Dessa forma, as decisões relativas ao tipo de software a ser utilizado, a eficiência deste na resolução da aplicação específica, entre outros quesitos, passam a ser estratégicos para o gerenciamento de custos relacionados ao setor de tecnologia da informação da empresa.

1.4 Do potencial, popularidade e facilidades relacionadas a softwares livre e *open source*

Com o advento e consolidação de plataformas *open sources* e livres, sob licença GPL, o número de alternativas para software estatístico tem aumentado. Um bom exemplo é o software R.

Muenchen (2012) trata do aumento da popularidade do software R tanto na comunidade acadêmica quanto em aplicações em empresas e outras instituições. O mesmo autor demonstra o avanço do software e sua versatilidade através do número de artigos científicos que citam o uso do pacote estatístico R, número de mensagens em fóruns especializados em *data science*, entre outras fontes.

O uso de plataformas *open source*, livres e seguras é um atrativo em geral por dois motivos principais. Primeiramente, pode-se conhecer com detalhes os procedimentos realizados pelo software, uma vez que seu código é aberto, e, em geral, permite que usuários bem familiarizados com o software desenvolvam suas próprias aplicações. No caso do R, para conhecer o método utilizado por determinada função, basta digitar o nome da função no console obtendo-se assim o seu código fonte.

Outro fator favorável às plataformas *open source* é a redução de burocracia relacionada ao processo de obtenção de software pela empresa. Em geral, para a aquisição de sistemas que requeiram licença a empresa precisa iniciar um processo de licitação que envolve diversos departamentos, entre eles o de compras, o que pode tornar o processo moroso e retardar o acesso ao software, que se faz necessário muitas vezes com urgência. No caso de softwares livres o usuário pode baixar o sistema livremente em qualquer máquina.

2. REFERENCIAL TEÓRICO

2.1 Análise de Decisões e sua implementação

O referencial teórico de análise de decisões para escolha do melhor sistema estatístico para o sistema de análise da empresa foi a ferramenta de análise de decisões proposta em Howard (1988; 2004).

De acordo com esse autor, a busca por uma boa decisão deve conter uma caracterização eficiente das alternativas em relação ao problema de decisão, das consequências relacionadas a cada uma das alternativas, das preferências dos indivíduos relacionados ao problema e, em condições de incerteza, uma descrição adequada dos riscos envolvidos (Howard, 1988; Howard, 2004).

Esta ferramenta torna possível uma modelagem iterativa e compreensível do problema de decisão, além de tornar o processo facilmente comunicável entre pessoas de áreas técnicas e não técnicas envolvidas na decisão, minimizando os problemas

cognitivos descritos em Bell & Raiffa (1990); Hogarth, R (1991); Plessner & Bletsch & Bletsch, (2007) e Ariely, D (2013). Entre tais problemas descritos nas referências anteriores podemos citar a tendência dos tomadores de decisão a:

- a) **Ancorar nos primeiros pensamentos e considerações sobre o problema:** Fenômeno psicológico em que a mente dá peso desproporcional às primeiras informações recebidas sobre a problema a que a decisão se relacione.
- b) **Viés do status quo:** Tendência do tomador de decisões a preferir alternativas que mantenham o status quo estabelecido nas instituições e organizações.
- c) **Proteção de decisões prévias:** Tendência a se manter aspectos relacionados a decisões tomadas anteriormente, mesmo que tais aspectos sejam racionalmente prejudiciais ou tenham se comprovado falhos.
- d) **Ver o que se quer ver:** Viés cognitivo em que o tomador de decisão tende a buscar informação que comprove suas primeiras impressões sobre o problema.

Colocar o problema de forma equivocada, excesso de autoconfiança do tomador de decisão e negligenciar informações.

Os problemas anteriores podem ser resolvidos pelo uso da ferramenta proposta em Howard (2004) cujo objetivo é obter um conjunto de informações denominado de base de decisões. O processo geral da análise de decisões pode ser encontrado na Figura 1.

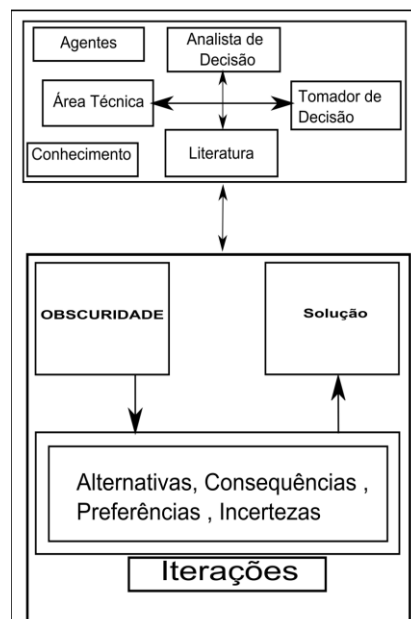


Figura 1 Dinâmica da Construção da Base de Decisões. Fonte: Autor. 1

Na Figura 1 podemos notar que o processo de análise da decisão e da construção da base de decisões começa com uma situação de obscuridade em relação ao problema de decisão. A partir da Figura 1, verificamos que os agentes das áreas técnicas relacionadas ao problema, o analista de decisão e o tomador de decisão agem de forma integrada complementando suas experiências com a literatura disponível sobre o problema. Através desse intercâmbio, surge o conhecimento para a formação da base de decisão que explicitará as alternativas, as consequências, as preferências dos envolvidos por cada alternativa e as incertezas relacionadas. O processo deve se

repetir iterativamente até que se chegue a uma solução para o problema ou que se tome uma decisão.

Em geral, em uma tomada de decisão parte-se de uma base inicial de decisão que evolui com a descrição iterativa das alternativas, consequências, preferências e riscos até que se chegue a um conjunto de informações que seja suficiente para embasar uma decisão.

2.2 Teoria Estatística Relacionada ao Problema

O tópico estatístico relacionado ao problema em questão era o da realização de comparação de médias entre tratamentos ensaiados em blocos casualizados. Relatórios estatísticos contendo este tipo de teste, embora simples e questionáveis, suportam decisões em diversos setores da empresa, desde o melhoramento genético até o registro de produtos no Ministério da Agricultura, tendo, portanto, impacto na legalização do desenvolvimento de produtos (MAPA, 2014).

2.3 Teste de Tukey

O teste de Tukey gerado pelo software a ser escolhido pela empresa consiste em uma técnica clássica em desenho e análise de experimentos (Montgomery, 2013), (Cox & Reid, 2000; Hinkelmann & Kempthorne, 2005; Mason et al., 2003).

Tal técnica consiste na comparação entre as médias de todos os $a*(a-1)/2$ pares de médias de tratamentos conduzidos em um experimento (Montgomery, 2013).

No caso em questão os experimentos eram realizados em blocos casualizados e as médias comparadas eram as médias de produtividade dos cultivares pré-comerciais da empresa. O teste baseia-se em um valor máximo para diferença entre duas médias de tratamentos para que tais tratamentos sejam considerados estatisticamente iguais, ou seja, estejam dentro de uma mesma classe, caso a distância entre tais médias seja menor que a distância determinada como máxima. Isso permite o agrupamento dos produtos desenvolvidos em classes de médias de produtividades (Montgomery, 2013).

3. TECNOLOGIA DA INFORMAÇÃO

O conhecimento de tecnologia da informação utilizado pode ser dividido em tópicos teóricos de programação, tópicos técnicos de programação em R e da estrutura e disponibilidade do software R e do software comercial.

3.1 Tópicos teóricos de programação

Os tópicos teóricos de programação envolvidos foram tópicos relativos à produção de algoritmos para resolução de problemas, além de temas gerais acerca da implementação dos procedimentos resolvidos por tais algoritmos.

Para interessados em maiores detalhes recomendamos a leitura de Cormen et al. (2009), Muller (2006) e Knuth (1997). Para um enfoque computacional e estatística sugerimos a leitura de Givens & Hoeting (2013).

3.2 Plataforma R: obtenção e breve histórico

O software R é uma plataforma *open source*, livre e disponível para todos os sistemas operacionais em www.r-project.org.

Trata-se de uma plataforma multiobjetivo, a qual permite a análise e visualização de dados e análises estatísticas em diferentes níveis de complexidade.

Além disso, pode ser utilizado também como plataforma de desenvolvimento de aplicações específicas totalmente inéditas ou que mesclam código inédito com aplicações já presentes no software.

O software conta com um módulo básico de instalação, que promove um ambiente básico de desenvolvimento, um console para realização de análises e uma interface gráfica. O módulo básico é complementado por packages adicionais para aplicações específicas em áreas básicas como matemática até aplicadas como ecologia, finanças e processamento geoespacial.

O software foi criado em 1995 e, atualmente, é gerenciado por um grupo de indivíduos denominado R-core e mantido pela fundação-R, a qual é suportada por diversas empresas de segmentos variados. De acordo com Muenchen (2012), é o software mais utilizado em competições mundiais de análise de dados como o Kaggle, para maiores detalhes acessar www.kaggle.com.

3.3 Programação em R

O R é um software que permite o uso de diversas técnicas de programação. Seu código fonte aproveita-se de rotinas bem estabelecidas em Fortran, C e na própria linguagem R.

Uma vantagem dessa linguagem é a possibilidade de programação de funções para realização de aplicações específicas inéditas, ou então, a programação de aplicações que permitam um uso mais amigável por parte de usuários com menos familiaridade com o software.

Interessados em programação podem acessar o item referente a tutoriais no site do projeto R, o qual disponibiliza grande quantidade de material para leitores de diferentes níveis e, em vários idiomas, entre eles o português. Introduções ao R podem ser encontradas em Teetor (2011); Micheaux & Drouilhet & Liquet (2013) e Gardener (2012).

4. METODOLOGIA

A seguinte metodologia visa à aplicação do referencial teórico apresentado anteriormente para a seleção da melhor alternativa de software estatístico pela empresa.

4.1 Caracterização do Problema

De acordo com Howard (1988; 2004), a primeira etapa para a construção da base de decisões e para o processo de análise de decisões é a caracterização do problema a ser decidido.

Neste trabalho, o problema é decidir-se entre uma de cinco alternativas para implementação do teste de comparação de médias utilizado pela empresa.

4.2 Alternativas Consideradas

Foram consideradas como alternativas para realização do teste requerido pela empresa quatro alternativas programadas em R, com diferentes abordagens e níveis de interface, além de uma alternativa de software comercial previamente conhecida por membros da equipe de análise e por isso sugerida.

As alternativas programadas em R foram denominadas de alternativa1-R até alternativa4-R, entretanto, a produção delas foi realizada iterativamente através do processo de construção da base de decisões a partir de um cenário inicial, que contava apenas com alternativa1-R e alternativa comercial, e que evoluiu até um cenário final com as cinco alternativas a serem consideradas para a decisão.

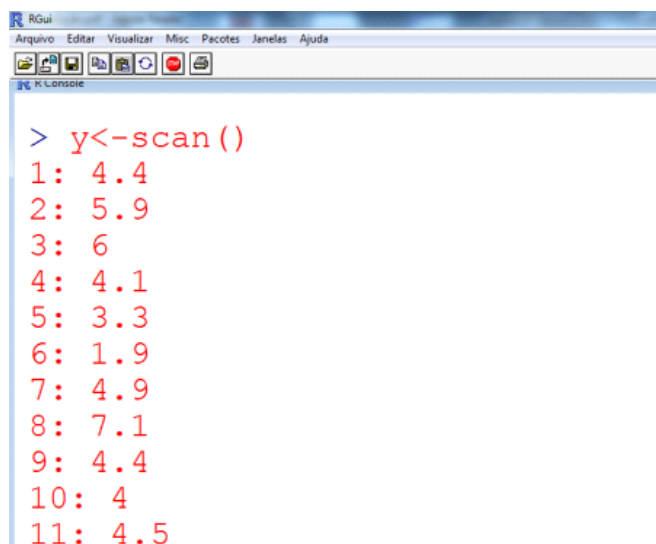
Na ilustração das alternativas e dinâmicas faremos menção apenas verbal à alternativa comercial e suas características, sem figuras ou *screenshots* que possam identificá-la e, de certa forma, prejudicar a sua marca ou seus distribuidores.

4.2.1 Alternativa 1-R:

A primeira alternativa desenvolvida em R era baseada na função *scan()* presente no módulo básico de instalação. Um aspecto importante dessa alternativa era a necessidade de digitação manual dos dados no console do R para que posteriormente fosse programado o teste requerido pela empresa.

A Figura2 ilustra a dinâmica geral da alternativa1-R. Por esta figura verifica-se que para a implementação do teste, ou seja, antes de sua programação o usuário precisa utilizar a função *scan()* e digitar todos os dados do experimento manualmente.

Como o volume de dados a serem analisados era da ordem de milhares, a digitação manual exigiu grande consumo de tempo.



```
> y<-scan()  
1: 4.4  
2: 5.9  
3: 6  
4: 4.1  
5: 3.3  
6: 1.9  
7: 4.9  
8: 7.1  
9: 4.4  
10: 4  
11: 4.5
```

Figura2

Dinâmica da Alternativa1-R. Fonte: Autor.

4.2.2 Alternativa Comercial

A alternativa comercial tratava-se de uma aplicação em software comercial, que requeria pagamento de licença de uso. Através dessa aplicação, um arquivo com os dados previamente coletados e organizados era aberto em interface bastante amigável ao usuário e o teste de comparação de médias era realizado com rotinas previamente estabelecidas e conhecidas por alguns membros da equipe de análise.

4.2.3 Alternativa2-R:

A alternativa2-R baseava-se no uso de uma função para importação de arquivos no formato *csv* (*comma-separated values*). Após a importação dos dados o analista poderia programar a análise no console do R, dispensando-se assim digitação manual dos dados.

Entretanto, os dados organizados nos arquivos *csv* deveriam estar no formato ilustrado na Figura 3. Tal figura representa a estrutura tabular dos dados em coluna de blocos, tratamento e produção, requerida pelo código programado para execução do teste de Tukey pelo software.

Embora essa alternativa resolva o problema relacionado à digitação manual de grande volume de dados, a estrutura do arquivo *.csv* era diferente da gerada pelo sistema de coleta de dados, representada na Figura 4, relativamente automatizado e já implementado na empresa.

1	bloco	trat	prod
2	1	1	4.4
3	1	2	3.3
4	1	3	4.4
5	1	4	6.8
6	1	5	6.3
7	1	6	6.4
8	2	1	5.9
9	2	2	1.9
10	2	3	4
11	2	4	6.6
12	2	5	4.9

Figura 3. Organização de Dados no arquivo utilizado pela Alternativa2-R. Fonte: Autor.

	A	B	C	D	E
1	tratamento	b1	b2	b3	b4
2	1	4.4	5.9	6	4.1
3	2	3.3	1.9	4.9	7.1
4	3	4.4	4	4.5	3.1
5	4	6.8	6.6	7	6.4
6	5	6.3	4.9	5.9	7.1
7	6	6.4	7.3	7.7	6.7

Figura 4. Estrutura gerada pelo sistema de coleta de dados e consolidado na empresa. Fonte: Autor.

4.2.4 Alternativa3-R:

Como decorrência do formato *csv* consolidado na empresa representado na Figura 4, foi desenvolvida a alternativa3-R. É interessante observar que esta estrutura tabular é diferente da representada na Figura 3 com impactos na estrutura do código que contemplasse a estrutura representada na Figura 4.

Esta alternativa3-R incluiu no código da análise a ser realizada procedimentos que permitiriam a análise dos dados importados nessa organização representada na

Figura 4. O mecanismo de importação dos dados era basicamente o mesmo da alternativa2-R.

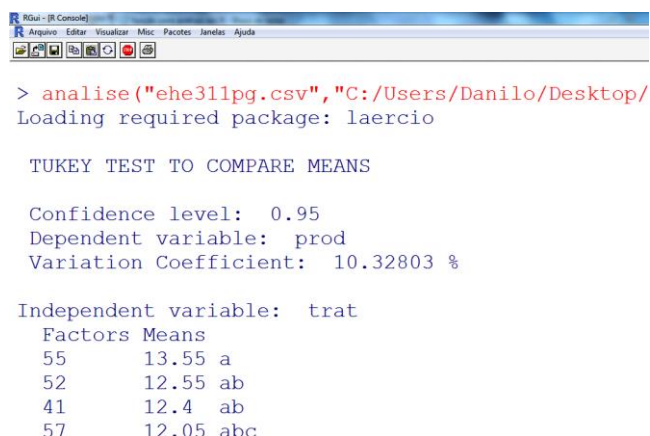
Após a importação do arquivo contendo os dados, o analista procedia à programação da análise para obtenção do teste de comparação de médias.

4.2.5 Alternativa4-R

A quarta alternativa desenvolvida em R era baseada em uma função programada. O intuito dessa função era permitir o uso das funções e do código programado no console para obtenção do teste de médias de forma mais amigável ao usuário final.

A função denominada de *analise* possui como argumentos o nome do arquivo *csv* que continha os dados organizados no formato consolidado pela empresa, além do diretório que continha o arquivo.

Após informar os dois argumentos anteriores, o analista apenas acionava a tecla *enter* e o relatório com o teste de médias era gerado automaticamente, sem necessidade de digitação manual de dados nem de programação de código no console do R. A dinâmica geral dessa alternativa está representada na Figura 5, na qual podemos ver que basta o usuário digitar o comando “analise” seguido do nome do arquivo e do diretório que contém o arquivo com os dados para que o resultado do teste seja mostrado. Nessa solução o código fonte utiliza um *package* denominado *laercio* e determina os níveis de confiança e o coeficiente de variação do experimento (Montgomery, 2013).



```

> analise("ehe311pg.csv", "C:/Users/Danilo/Desktop/
Loading required package: laercio

TUKEY TEST TO COMPARE MEANS

Confidence level: 0.95
Dependent variable: prod
Variation Coefficient: 10.32803 %

Independent variable: trat
Factors Means
55      13.55 a
52      12.55 ab
41       12.4 ab
57      12.05 abc

```

Figura 5. Dinâmica da alternativa4-R. Fonte: Autor.

4.3 Caracterização das Consequências

Na composição da base de decisões proposta em Howard (1988; 2004), é necessário a determinação das consequências em relação a cada uma das alternativas consideradas.

No presente trabalho a principal consequência analisada para cada alternativa foi o custo referente ao processo de análise, composto a partir da seguinte equação:

$$CA = CL / NA + CHF * Th$$

Equação

Custo associado a cada alternativa

Na qual,

CA= Custo da análise em dólares

1

CL= Custo de licença relacionada ao uso do software (paga por equipamento onde o software estava instalado) em dólares

CHF= Custo em horas do funcionário que realiza a análise

NA= Número de Análises Realizadas

Th= Tempo em horas requerido para análise

Maiores detalhes acerca da teoria econômica sobre custos podem ser encontrados em Silberberg & Suen (2001) e Stackelberg (2014).

4.4 Caracterização das Preferências

A caracterização de preferências pode assumir diferentes níveis de complexidade no contexto de análise de decisões. No caso em questão tentou-se adotar uma metodologia de aplicação facilitada no contexto de uma instituição empresarial.

A captação da preferência por cada alternativa procedeu-se da seguinte forma. O teste de comparação de médias foi realizado em cada uma das alternativas com cada membro da equipe em separado para minimização do viés de opinião.

A seguir cada membro, também em condições sigilosas, era convidado a dar uma nota de 0 a 10 para a solução proposta como forma de ilustrar sua preferência pela alternativa. Nesse processo, os membros da equipe eram convidados a julgar a alternativa sem necessariamente se embasar em experiências passadas, como forma de minimizar a tendência dos indivíduos que possuíam experiência anterior com a alternativa comercial a privilegiarem-na.

5. RESULTADOS

5.1.1 Construção da base de decisão para o problema em questão

No estágio inicial das discussões acerca do problema de escolha da melhor alternativa para realização do teste de médias a base de decisões contava com duas alternativas: alternativa comercial e alternativa1-R. A composição da base de decisões nesse momento pode ser resumida na Tabela 1.

Alternativa	Consequência em termos de custos	Preferência Associada	Outras Características
Comercial	\$122,08 / análise	8	Não requeria conhecimentos de programação, envolvia interface amigável e rotinas prontas, requer pagamento de licença
Alternativa1-R	\$7,60 /análise	7	Requeria digitação manual de dados no console e conhecimentos de programação, não requer pagamento de licença

Tabela 1
Estágio inicial da base de decisões para o problema de escolha de software estatístico. Fonte: Autor.

Nesse estágio inicial a decisão foi pela utilização do software comercial, entretanto, após o conhecimento das possibilidades de desenvolvimento de aplicações no software R a empresa concordou em prosseguir com o desenvolvimento de novas alternativas, desde que o processo não envolvesse custos e ficasse a cargo do analista de decisões que possuía conhecimento da plataforma R.

A partir daí a base de decisões passou para um segundo estágio com o desenvolvimento e inclusão da alternativa2-R descrita no item referente à metodologia. Dessa forma, a base de decisões passou a conter a configuração representada na Tabela 2.

Alternativa	Consequência em termos de custos	Preferência Associada	Outras Características
Comercial	\$122,08 / análise	8	Não requeria conhecimentos de programação, envolvia interface amigável e rotinas prontas, requer pagamento de licença
Alternativa1-R	\$7,60 /análise	7	Requeria digitação manual de dados no console e conhecimentos de programação, não requer pagamento de licença
Alternativa2-R	\$4,2/análise	8,5	Permitia a importação de dados em arquivos <i>csv</i> , entretanto, o formato do arquivo não correspondia ao consolidado na empresa

Tabela

2

Segundo Estágio da Base de Decisão para o Problema. Fonte: Autor.

Diante da evidência de resultados iniciais devidos ao desenvolvimento de uma nova alternativa em R decidiu-se proceder mais uma rodada de produção de alternativas e verificação de seus custos, preferências e características.

Primeiramente, a alternativa2-R teve o código alterado para que pudesse ser utilizada a estrutura consolidada de arquivo *csv* utilizada pela empresa, contendo os dados a serem importados para análise no software, originando-se a alternativa3-R. Após isso, o processo evolui para o desenvolvimento da alternativa4-R baseada numa função que permitia uma aplicação para realização da análise requerida mais amigável ao usuário.

Dessa forma, a base de decisões passou a conter a configuração explicitada na Tabela 3.

Alternativa	Consequência em termos de custos	Preferência Associada	Outras Características
Comercial	\$122,08 / análise	8	Não requeria conhecimentos de programação, envolvia interface amigável e rotinas prontas, requer pagamento de licença
Alternativa1-R	\$7,60 /análise	7	Requeria digitação manual de dados no console e conhecimentos de programação, não requer pagamento de licença
Alternativa2-R	\$4,22/análise	8,5	Permitia a importação de dados em arquivos <i>csv</i> , entretanto, o formato do arquivo não correspondia ao consolidado na empresa, além de envolver a programação da análise no console
Alternativa3-R	\$4,22	9	Idem ao anterior, entretanto, permitia o uso do arquivo <i>.csv</i> consolidado na empresa
Alternativa4-R	\$0,9	9,5	Baseava-se em uma função que requeria apenas a digitação do nome do arquivo <i>csv</i> que continha os dados a serem analisados e o diretório onde o arquivo se encontrava

Tabela 3
Constituição Final da Base de Decisões para o problema em questão. Fonte: Autor.

5.1.2 Decisão Tomada

Diante da base de decisões construída em processo iterativo de desenvolvimento de alternativas representada na Tabela 3, a empresa decidiu por adotar a alternativa4-R como padrão para realização do teste de médias.

A alternativa4-R foi escolhida por ter maior nota de preferência, menor custo/análise e ser de uso amigável por usuários que não possuíam conhecimentos avançados de programação em computadores e linguagem R.

A escolha da alternativa permitiu também a abertura da possibilidade para novas aplicações das áreas de *Data Science* e *Machine Learning* aos dados coletados pela empresa devido à grande compatibilidade da plataforma R com aplicações dessa natureza.

Dessa forma, além de realizar o teste de médias, outros tipos de análises mais avançadas e específicas seriam permitidos, fato que agradou bastante os envolvidos na gestão do sistema de aquisição e análise de dados da empresa.

6. DISCUSSÃO

6.1 Impactos dos conhecimentos de programação em R na decisão final

Pode-se verificar também como resultado da construção da base de decisão que as habilidades de programação em R tiveram impactos na decisão final acerca da alternativa mais apropriada para a finalidade da análise requerida e, portanto, de software a ser utilizado pela empresa.

O principal impacto foi a mudança da decisão em favor da alternativa comercial para alternativa4-R, eliminando-se, assim, a necessidade de pagamentos de licenças relacionadas ao software comercial.

Assim, evidencia-se a forma como conhecimentos técnicos de desenvolvimento de aplicações em plataformas livres e/ou *open source* podem reduzir a dependência tecnológica de organizações.

Além disso, este resultado também permite a recomendação do ensino desse tipo de plataforma em universidades e cursos técnicos, como forma de gerar capital humano com o conhecimento técnico que permita redução de dependência tecnológica nas organizações.

6.2 Evolução do nível de interface entre as alternativas desenvolvidas em R e seus impactos em termos de custo e preferência

Durante o processo de desenvolvimento de alternativas para implementação do teste de médias ocorreu também uma evolução no nível da interface entre as alternativas e os usuários. Essa evolução no nível de interface relacionou-se com o a evolução da preferência pelas alternativas desenvolvidas representada na Figura 6. O maior nível de interface foi conseguido na alternativa4-R.

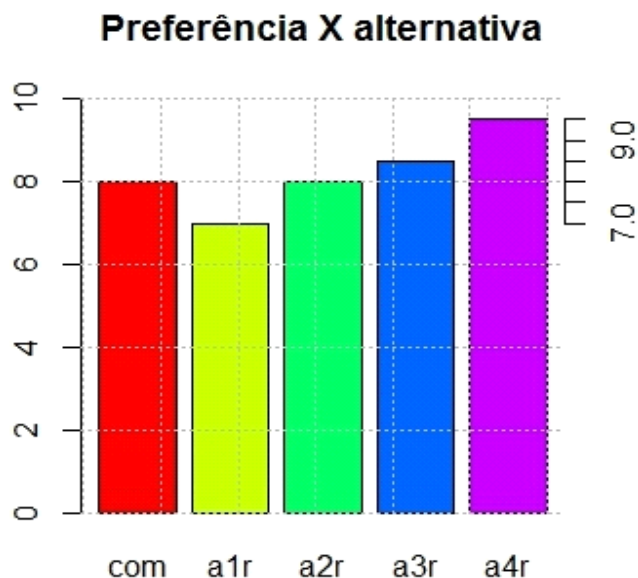


Figura 6
Evolução da preferência entre as alternativas consideradas. Fonte: Autor.

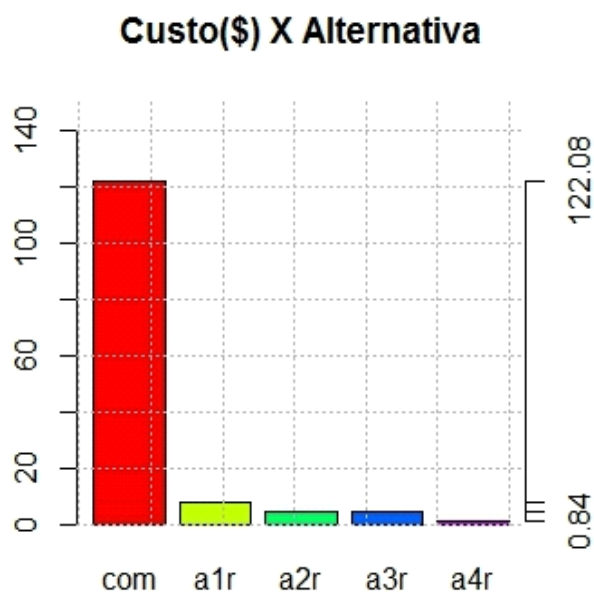


Figura 7
Evolução do custo/análise entre as alternativas. Fonte: Autor.

Ao contrastarmos a Figura 6 com a Figura 7, que representa a evolução dos custos de cada uma das alternativas, podemos verificar que o aumento no nível de interface propiciou uma redução de custo adicional à redução de custo devido à licença,

representada pela diferença da dimensão da barra vermelha, correspondente à alternativa comercial, e a barra verde correspondente a Alternativa 1-R.

Esse fato ilustra o impacto da evolução no nível de interface para aplicações na preferência por usuários finais e nos custos relacionados à operação dessa aplicação.

7. CONCLUSÕES, LIMITAÇÕES E RECOMENDAÇÕES

Este trabalho ilustrou, em uma perspectiva interdisciplinar, o uso de ferramentas de análise de decisões no gerenciamento de tecnologia da informação. Através da ferramenta de análise de decisões proposta por Howard (1988; 2004), determinou-se a melhor alternativa dentre cinco alternativas consideradas para a realização de testes de comparação de médias através de software estatístico em uma empresa de pesquisa e desenvolvimento de sementes melhoradas geneticamente.

Redução do custo relacionado ao processo de análise em cada alternativa foi evidenciada.

O trabalho demonstrou como uma função em R pode servir como interface eficiente e a influência do nível dessa interface na preferência dos usuários por diversas alternativas consideradas.

A ilustração da possibilidade do uso de plataformas livres em empresas também foi ilustrada, assim como, demonstrar como habilidades de programação em R impactaram a decisão em relação ao software mais adequado ao sistema de análise da empresa.

Dentre as limitações do trabalho podemos citar a solução de apenas uma aplicação em específico, para acomodar a realização do trabalho ao dia a dia da empresa em que foi realizado. Soluções mais robustas poderiam ser programadas contemplando a totalidade de aplicações usadas pela empresa. Tais soluções poderiam contemplar um *package* para ser carregado no R com todas as funções necessárias às aplicações, numa perspectiva próxima da proposta em Wickham, H (2014). Outra limitação seria a solução do problema em apenas uma empresa de um setor. Caso fosse possível a solução do mesmo problema para outras empresas do setor de sementes uma solução geral poderia ser proposta para o setor como um todo.

Recomenda-se que em trabalhos futuros sejam utilizadas maiores integrações na geração dos *reports* pela alternativa escolhida pela empresa que contemplem as ferramentas *Rstudio*, disponível em <http://www.rstudio.com/> e o *package Knitr* que utiliza funcionalidades da linguagem de programação *markdown* para geração de relatórios aliadas ao *package shiny* que permite a geração aplicações dinâmicas em R tornando provavelmente as notas de satisfação dos usuários descritas nos resultados ainda maiores. Os *packages* citados anteriormente são encontrados no próprio site do R.

REFERÊNCIAS BIBLIOGRÁFICAS

- Ariely, D. (2013). *The Irrational Bundle: Predictably Irrational, The Upside of Irrationality and The Honest Truth About Dishonesty*. New York. Harper.
- Bell, E. & Raiffa H. (1990). *Decision Making: Descriptive, Normative and Prescriptive Interactions*. London, Cambridge Press.

- Cormen, T. H. & Leiserson, C. E. & Rivest, R. L. & Stein (2009). *C. Introduction to Algorithms*. Cambridge MA, MIT Press.
- Cox, D. R. & Reid, N (2000). *The Theory of the Design of Experiments*. Boca Raton, Chapman and Hall.
- D. Knuth (1997). *The Art of Computer Programming: Fundamental Algorithms*. Massachusetts, Addison Wesley Longman.
- Gardener, M. (2012). *Beginning R: The Statistical Programming Language*. New York. John Wiley and Sons, Inc.
- Givens, H. G. & Hoeting, J.A. (2013). *Computational Statistics*, New York. John Wiley and Sons, Inc.
- Hinkelmann, K. & Kempthorne (2005). *O. Design and Analysis of Experiments*. New York, JohnWiley and Sons.
- Hogarth, R. (1991). *Judgement and Choice*. New York, John Wiley and Sons.
- Howard, R. (1988). *Decision Analysis practice and promise*. Management Science, 6, 679-675.
- Howard, R. (2004). *Speaking of decisions: Precise decision language*. Decision Analysis, 1, 71-78.
- Mapa (2014), *Registro Nacional de Cultivares*. Retrieved from: <http://www.agricultura.gov.br/vegetal/registros-autorizacoes/registro/registro-nacional-cultivares>.
- Mason, R. L. & Gunst, R. F. & Hess, J. L (2003). *Statistical Design and Analysis of Experiments*. New York, John Wiley and Sons.
- Micheaux, P. L. & Drouilhet, R. & Liquet, B. (2013). *The R Software: Fundamentals of Programming and Statistical Analysis*. New York. Springer.
- Montgomery, D, (2013). *Design and Analysis of Experiments*. New York. John Wiley and Sons, Inc.
- Muller, J.M (2006). *Elementary Functions Algorithms and Implementations*. Boston, Birkhäuser.
- Munchen, R.A. (2012). *The Popularity of Statistical Software*. Retrieved from: <http://r4stats.com/articles/popularity/>.
- Plessner, H. & Bletsch, C. & Bletsch, T. (2007). *Intuition in Judgement and Decision Making*. London, Lawrence Erlbaum.
- Silberberg, E. & Suen, W. (2001). *The Structure of Economics: A Mathematical Analysis*. New York, Mc Graw Hill.
- Stackelberg, H. (2014). *Foundations of a Pure Cost Theory*. Berlin. Springer.
- Teetor, P. (2011). *R Cookbook: Proven Recipes for Data Analysis, Statistics and Graphics*. Cambridge, O'Reilly.
- Wickham, H. (2014). *R packages*. Retrieved from: <http://r-pkgs.had.co.nz/demo.html>.