

# Desenvolvimento de um software estatístico web para análise de dados experimentais baseado no R: StatIFRS

Cassandro Davi Emer<sup>1</sup>, Marcos Juarez Vissoto Corino<sup>1</sup>

<sup>1</sup>Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul (IFRS)  
Campus Veranópolis – BR-470, Km 172, 6500, bairro Sapopema  
95330-000 – Veranópolis – RS – Brazil

{cassandro@gmail.com, marcos.corino@veranopolis.ifrs.edu.br}

**Resumo.** *O uso da estatística experimental é cada vez mais importante e frequente, tanto no ambiente acadêmico quanto no setor privado. Embora existam inúmeros softwares estatísticos, há uma limitação em relação ao custo das licenças, complexidade de uso e necessidade de instalação dos softwares no sistema. Foi desenvolvido um software de acesso livre baseado no R, com interface web, que permite a realização de análises estatísticas em experimentos unifatoriais, fatoriais e/ou com blocos, além de análise de correlação e elaboração de modelos de regressão. O StatIFRS pode ser de grande valia tanto para a comunidade acadêmica como para o público em geral.*

**Abstract.** *The use of experimental statistics is increasingly important and frequent not only in the academic but also in the private environment. Although numerous statistical software programs exist, there is a limitation regarding the cost of licenses, complexity of use and the need to install softwares on the system. An open-access software based on R, with a web interface, has been developed that allows for statistical analysis in single-factor, factorial, and/or block experiments, as well as correlation analysis and regression modeling. StatIFRS can be of great value to both the academic community and the general public.*

## 1. Introdução

A estatística é uma ciência matemática que se preocupa com a coleta, análise, interpretação ou explicação, e apresentação de dados [Watkins 2019, Everitt and Skron dal 2010]. De acordo com Watkins (2019), os princípios estatísticos corretamente utilizados são essenciais para conduzir qualquer pesquisa orientada por dados e, especialmente na fase de exploração, é rotineiramente uma fonte fundamental para a descoberta e inovação.

Dentro dos mais diversos campos de pesquisa ou de otimização de processos, o uso da estatística é crucial e está presente desde a concepção dos delineamentos experimentais. Dentro do meio acadêmico, é cada vez mais difícil realizar a publicação de artigos sem o uso de algum teste ou modelo [Genolini et al. 2019]. Já no setor privado, é comum o uso da estatística por seguradoras, no setor financeiro e pela indústria farmacêutica [Peterson et al. 2009], que a usa, não apenas para validar seus ensaios clínicos, mas também na otimização de processos.

Deste modo, o amparo computacional sempre foi muito importante para a estatística, sendo que existe uma grande gama de ferramentas deste segmento, desde o surgimento dos sistemas mais rudimentares. Embora muitos softwares disponíveis atualmente

sejam completos em termos de metodologias, grande parte demanda do uso de licenças (comerciais ou acadêmicas) e exigem um elevado grau de conhecimento técnico ou de programação para sua utilização. Neste aspecto, algumas iniciativas buscam sanar essa problemática, como por exemplo o R++ [Genolini et al. 2019], o JASP [Love et al. 2019], e a própria concepção do pacote Shiny [Chang et al. 2015] para o R [R Core Team 2020] baseou-se na criação de aplicativos interativos com interface web.

Porém, de uma maneira geral, os softwares estatísticos executam seu processamento localmente, requerendo que seja realizada a sua instalação no computador, o que, a princípio, parece banal, mas que pode ser um empecilho no uso de computadores em laboratórios de informática, nos quais o usuário não detém permissão para a instalação de programas, por exemplo. Outro ponto a ser considerado é o custo das licenças, no caso do uso de um software comercial em um laboratório acadêmico. Por este motivo, e como uma forma de combate à propagação de cópias ilegítimas de softwares comerciais, foi proposto o desenvolvimento de uma ferramenta estatística baseada no R, de livre acesso, totalmente web, fornecendo um protocolo simples, prático e estatisticamente sólido para a realização de análises de delineamentos experimentais usuais nas ciências agrárias.

O restante do artigo é estruturado da seguinte forma: na Seção 2, são apresentados os conceitos da fundamentação teórica; na Seção 3 são descritas as metodologias utilizadas para a concepção e desenvolvimento do software; e nas Seções 4 e 5, são dispostos uma discussão e os resultados obtidos na validação, além de exemplos da interação com a plataforma.

## 2. Referencial teórico

O modelo cliente-servidor é uma estrutura de aplicação distribuída que divide tarefas entre os provedores de um recurso ou serviço, chamados servidores, e solicitantes de serviços, chamados clientes [Reese 2000]. O servidor executa um ou mais programas de serviços, que compartilham seus recursos com os clientes. Já o cliente não compartilha nenhum de seus recursos, mas solicita conteúdo ou serviço de um servidor. Os clientes, portanto, iniciam sessões de comunicação com os servidores, que processam as solicitações recebidas e retornam os resultados.

De uma maneira geral, a abstração preocupada na apresentação durante o processo interativo, ou seja, a interface amigável ao usuário, é referida como *front end* [Pavlenko et al. 2020]. Já *back end* é o termo empregado para designar o software que realiza a manipulação de dados e lógica de negócio [Pavlenko et al. 2020]. Dentro do modelo cliente-servidor, o cliente é considerado o *front end*, e o servidor, o *back end*.

O Bootstrap é uma estrutura poderosa que fornece um conjunto de classes CSS e funções JavaScript para facilitar o processo de desenvolvimento do *front end*. Suas características de design responsivo permitem que websites sejam otimizados para visualização em telas de diversos dispositivos, sendo compatível com múltiplas plataformas e navegadores: o mesmo site funciona bem tanto em dispositivos móveis quanto em desktops [Balasubramanee et al. 2013], o que é cada vez mais importante, dado o crescente uso destes dispositivos.

Um contêiner de aplicação web é uma unidade de software que empacota o código e suas dependências para que uma ou mais aplicações sejam executadas. A comunicação

entre o cliente, tanto para ter acesso ao código que determina a renderização da interface de usuário no caso de uma interface web, como acesso aos serviços, se dá através do intercâmbio de requisições sobre o protocolo HTTP/S [Fielding et al. 1999]. Tomcat é o servidor web Java da Apache Software Foundation que atua como o recipiente de servlets [Manelli and Zambon 2020], e que permite rodar a camada de servidor a partir de código Java, na forma de um pacote WAR, gerado por construtores de projeto como o Apache Ant ou Maven.

O modelo cliente-servidor não dita que os servidores devem ter mais recursos do que os clientes. Ao invés disso, ele permite que qualquer computador de propósito geral amplie suas capacidades utilizando os recursos compartilhados de outros. A computação centralizada, no entanto, aloca especificamente uma grande quantidade de recursos para um pequeno número de computadores. Quanto mais o processamento é alocado dos clientes para os servidores, mais simples, em recursos, podem ser os clientes [Nieh et al. 2000], o que também ressoa com o conceito de computação em nuvem (*cloud computing*).

O R [R Core Team 2020], responsável pelos cálculos estatísticos e elaboração dos gráficos da ferramenta, é um software altamente produtivo para estatística e modelagem, sendo que sua implementação de procedimentos estatísticos e de gráficos é organizada na forma de pacotes. Os pacotes R são coleções de funções e conjuntos de dados desenvolvidos pela comunidade.

### **3. Métodos**

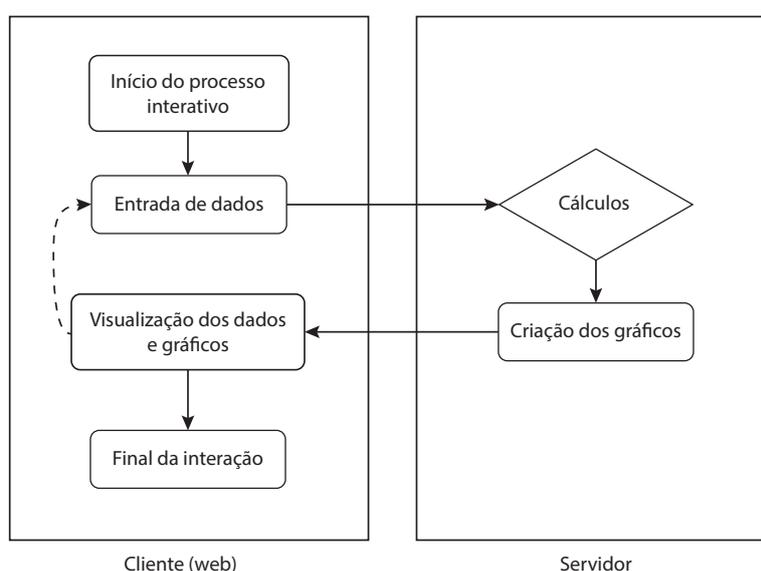
#### **3.1. Projeto de software**

Uma metodologia informal e ágil de gerenciamento de projetos foi empregada para gerir o desenvolvimento da ferramenta, através de um planejamento adaptativo e implementação iterativa, com apenas um desenvolvedor interfuncional. A medida que os módulos foram implementados, de modo cujo encapsulamento permite funcionamento sem interdependência, eram conduzidos os testes para validação das funcionalidades e integridade nos resultados do protótipo.

Foram observadas as soluções interativas de diversos softwares estatísticos para serem tomadas como base durante a etapa de análise e modelagem do sistema, visando que um conhecimento prévio do usuário pudesse ser útil durante o uso do software. Ao levar-se em conta que pretendia-se uma arquitetura cliente-servidor com interface independente de plataforma, alguns desafios pronunciaram-se, em evidência, é possível citar o processo de importação de dados. A problemática surge pois ao tornarmos a ferramenta totalmente web, surge a necessidade de uma grande flexibilidade, já que os usuários podem usar inúmeras ferramentas para organizar dados experimentais. Dentre as possíveis abordagens elencadas, optou-se pela mais simples – importação através do uso da área de transferência do sistema de dados previamente indexados, porém, a que independe da instalação de programas ou extensões no navegador, e que permite ao usuário o máximo de sigilo de dados, uma vez que ele seleciona e restringe a importação à parcela de dados que são de interesse para a análise a ser realizada. Outra abordagem poderia ser através da transferência de arquivos de ferramentas de planilhas, o que acabaria por necessitar de diálogos intermediários de interação com o usuário, além de, possivelmente, restringir a importação a determinados tipos de arquivos, e exigir a avaliação das questões de

segurança que surgem em função da inclusão de envio de arquivos ao servidor.

O modelo do software observou que o cliente ficasse somente a cargo das tarefas referentes à interação, edição e exibição dos dados de entrada, além da apresentação dos resultados e possibilidade de exportação de uma versão adequada para impressão. Enquanto que a cargo do servidor, ficou toda a interação com instâncias do software estatístico R, que se inicia com a transformação e adequação dos dados de entrada, segue com o cálculo das análises selecionadas e termina com a elaboração de gráficos. A comunicação entre cliente e servidor se dá através de uma camada de transporte sobre protocolos de segurança (SSL/TLS). Uma sequência da interação do usuário dentro da arquitetura cliente-servidor está representada na Figura 1.

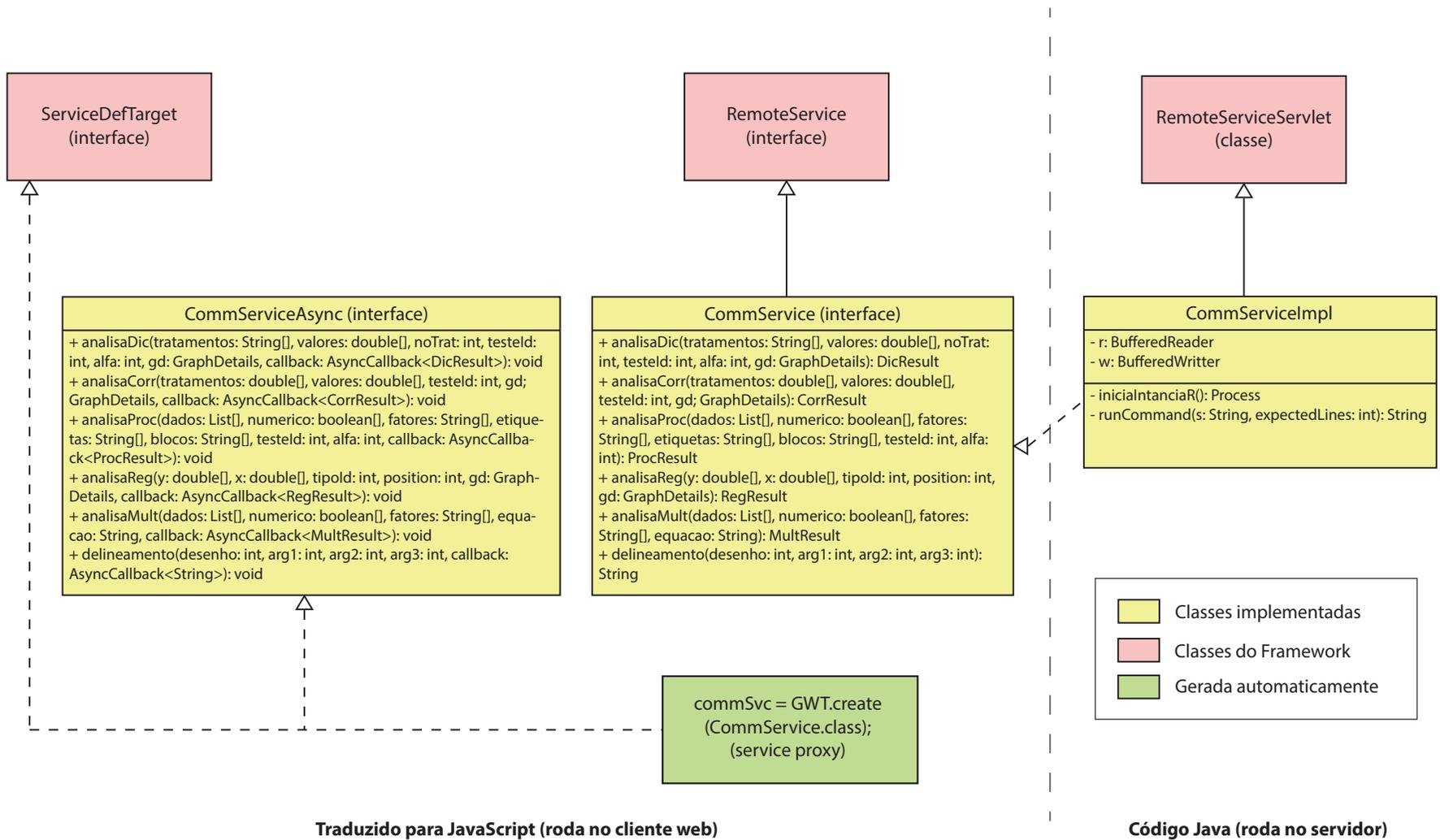


**Figura 1. Sequência da interação do usuário e fluxo de dados dentro do modelo cliente-servidor**

Com respeito ao desempenho do sistema em função de uma possível alta demanda de requisições, foi apenas dimensionado um controle de instâncias de processos, de modo a manter um pool de pendências uma vez que o número máximo de instâncias concorrentes seja atingido. Embora seja um aspecto importante, não foram testados, durante o desenvolvimento e a validação do software, a escalabilidade em função, não apenas das requisições concorrentes, mas das diferentes demandas de processamento. Foi observado, porém, que existe uma grande variabilidade no tempo de processamento decorrente do tipo de análise e número de fatores do experimento, que é incrementado, essencialmente, pelos processos de elaboração de gráficos.

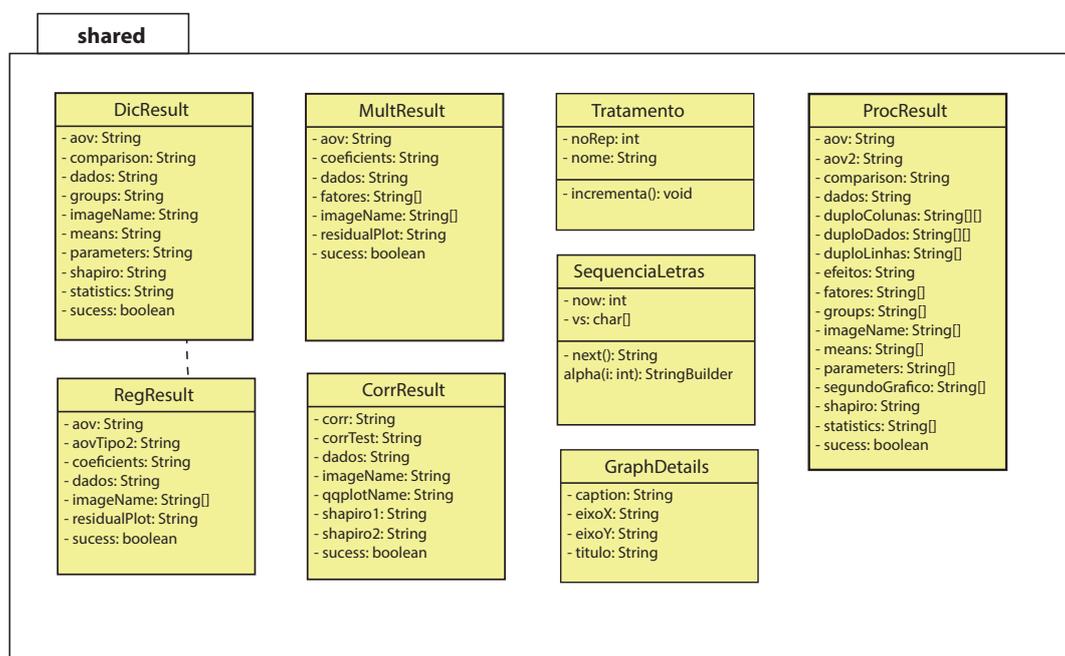
### 3.2. Desenvolvimento de software

Para o desenvolvimento do sistema foi utilizado o framework GWT (Google Web Toolkit), no qual a implementação da solução é realizada na linguagem de programação Java e, o código a ser executado no cliente é, posteriormente, traduzido para JavaScript. Nesta abordagem, pelo uso da linguagem, obtém-se a possibilidade da abstração e organização do código em classes e pacotes, favorecendo o reuso e manutenção do código.



**Figura 2. Esquema ilustrativo da interação entre as classes que implementam RPC**

Nas interações de processamento de dados estatísticos, o cliente (*front end*) irá comunicar-se com o servidor (*back end*), o qual é responsável pelas conversões de dados e interação com uma instância do processo do R, por cliente, através de RPC (*Remote procedure call* – Chamada de procedimento remoto, Figura 2). A estrutura GWT RPC torna fácil, para os componentes cliente e servidor da aplicação web, a troca de objetos Java serializados sobre HTTP/S. A implementação de um serviço GWT RPC é baseada na conhecida arquitetura de servlets Java. Dentro do código do cliente, é criada uma classe *proxy* gerada automaticamente para fazer as chamadas para o serviço. O GWT tratará da serialização dos objetos Java (Figura 3), passando entre cliente e servidor, os argumentos de métodos de chamadas e os valores de retorno.



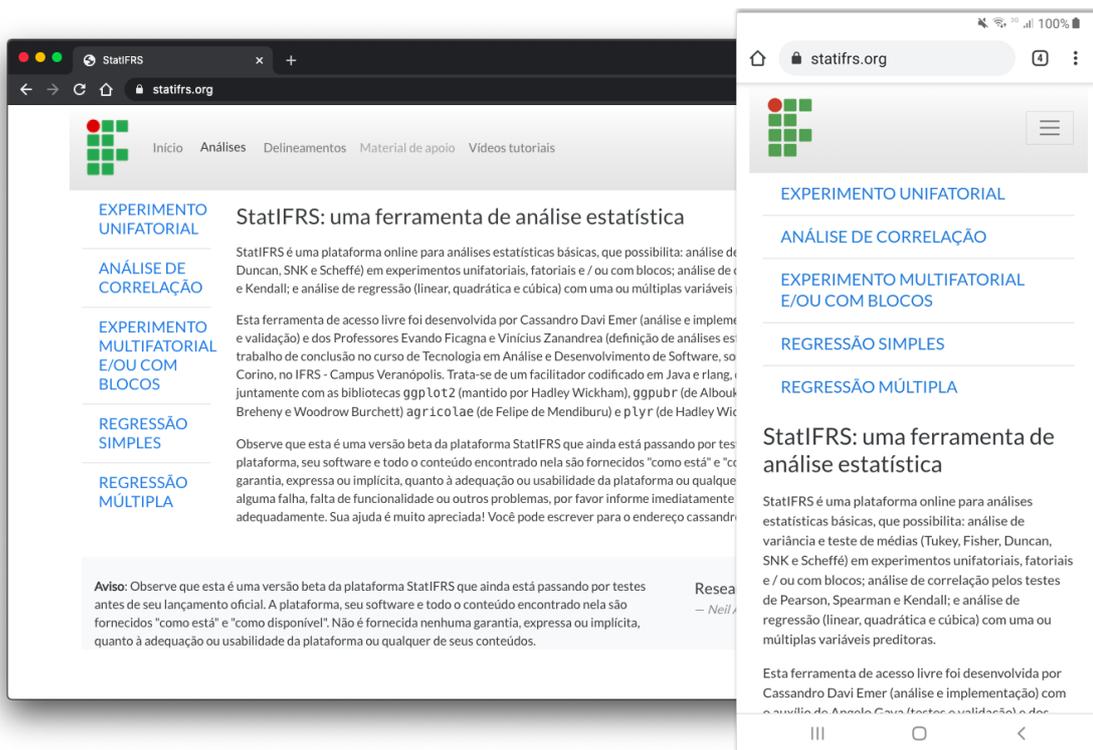
**Figura 3. Classes serializáveis comuns entre os ambientes do cliente e servidor**

As telas interativas foram todas codificadas utilizando-se do Bootstrap. A Figura 4 ilustra a captura de imagem da tela inicial do software, que embora customizada, carrega a característica visual deste framework.

O contêiner de aplicação utilizado durante a implementação e testes foi o Tomcat. Dentro do Tomcat, os cálculos estatísticos são realizados utilizando o R [R Core Team 2020], juntamente das bibliotecas ggplot2 [Wickham 2016], ggpubr [Kassambara 2020], visreg [Breheny and Burchett 2017], agricolae [Mendiburu 2020] e plyr [Wickham 2011].

Foram codificadas rotinas para a realização de análises nos delineamentos experimentais mais usuais, pela observação das análises oferecidas pelo software Assisat [Silva and Azevedo 2016] e na literatura [Montgomery 2017, Pimentel Gomes 1990, Costa 2003], separadas em diferentes módulos:

- Experimento unifatorial: interação na qual é realizada uma análise de variância e teste de médias baseando-se em uma variável contínua, em função de uma outra

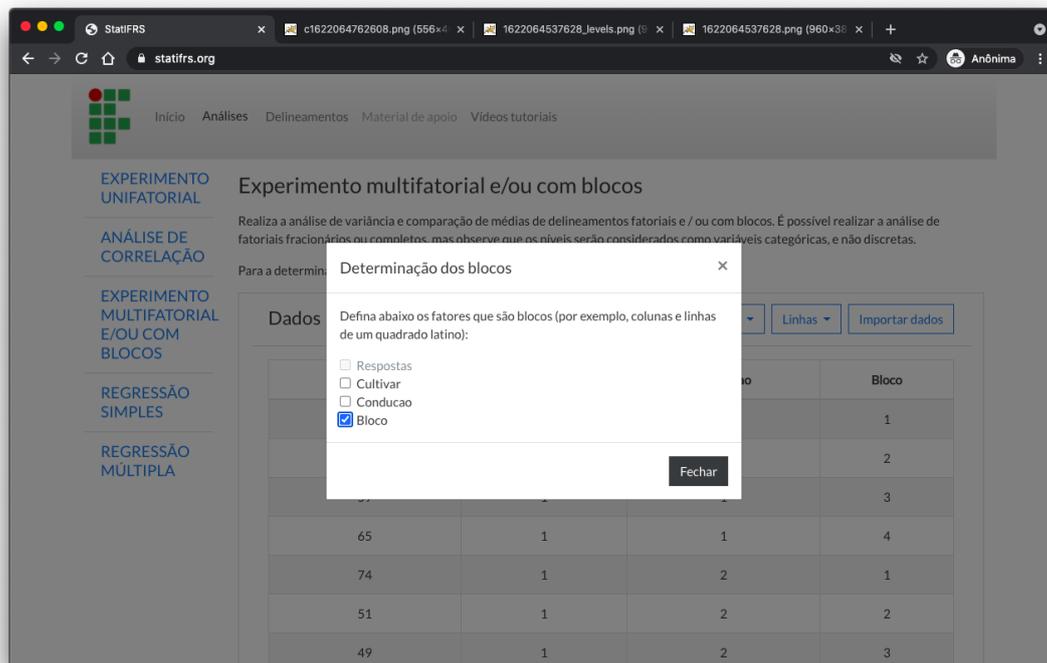


**Figura 4. Imagem da tela inicial do software**

variável discreta. O R permite que, nesta análise, existam parcelas perdidas, ou seja, tratamentos em que o número de repetições difere em relação aos demais. Ainda, é possível selecionar o teste de médias (Teste de Tukey, Teste de Scheffé, Teste Student-Newman-Keuls, Teste de Duncan e Teste de Fisher) e o valor do alfa para o teste.

- Análise de correlação: análise em que é quantificado o grau de associação entre duas variáveis contínuas. Permite o uso dos testes de Pearson, Spearman e Kendall.
- Experimento multifatorial e/ou com blocos: realiza uma análise de variância e teste de médias utilizando-se de uma variável contínua de resposta e de uma ou mais variáveis discretas. Permite ainda, a definição de blocos, situação ilustrada na Figura 5.
- Regressão simples: processo em que é realizada uma análise de variância para a determinação da significância dos coeficientes do modelo que prediz uma variável contínua (variável dependente) em função de outra (variável independente). São suportados os tipos de regressão linear, quadrática e cúbica.
- Regressão múltipla: similarmente à regressão simples, permite a avaliação de um modelo de regressão de uma variável contínua de resposta, mas em função de mais de uma variável independente. Neste caso, o componente da interação entre variáveis independentes pode ser de interesse, ficando a cargo do usuário a determinação destas possibilidades (Figura 6).

O processo de entrada de dados na plataforma pode ser realizado manualmente, através do assistente, no caso do experimento unifatorial, ou da inserção de linhas em branco nos demais. Porém, o uso pretendido é aquele onde o usuário, em posse de dados



**Figura 5. Imagem da tela de determinação de uma ou mais variáveis para blocagem**

tabulados em qualquer software de planilhas, seja operando no sistema ou web, possa importá-los diretamente à plataforma. Desta maneira, o usuário pode simplesmente transpor, utilizando-se da área de transferência do sistema, os dados previamente dispostos para que as colunas representem as variáveis no diálogo de importação, como no exemplo ilustrado na Figura 7. A possibilidade do usuário realizar a seleção parcial das colunas e linhas cujos dados são relevantes à análise sendo realizada, não só minimiza o tráfego de dados, como dá segurança ao permitir que o cliente não exponha dados desnecessariamente. O interpretador foi codificado para identificar diversos separadores de campo, e ainda, tanto o ponto como a vírgula são entendidos como divisores da fração decimal de um valor numérico. Uma vez os dados interpretados como válidos, o usuário interage com um novo diálogo (Figura 8), que se adequa conforme a análise sendo realizada, não apenas no processo, mas na nomenclatura das variáveis, de modo a facilitar o entendimento do usuário. Uma vez importados os dados, a interface permite a edição destes na própria tabela, e habilita as operações de cálculo.

### 3.3. Conformidade dos resultados

A validação dos resultados obtidos no uso da plataforma deu-se através da realização de testes e comparação às produzidas pelo software Statistica (<https://www.statistica.com>) e, quando dentro do domínio, também pelo software Assistat [Silva and Azevedo 2016], além da verificação dos resultados apresentados na literatura. Foram usados como base, exemplos apresentados por [Pimentel Gomes 1990] e [Costa 2003], além de dados experimentais próprios.

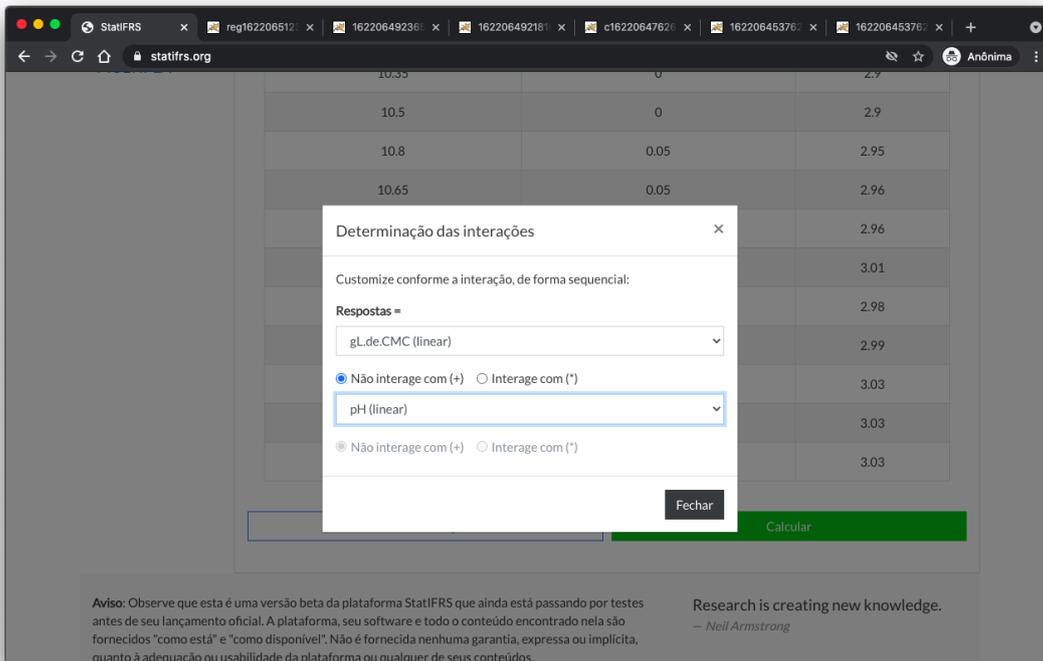


Figura 6. Imagem da tela de seleção de interação entre variáveis independentes

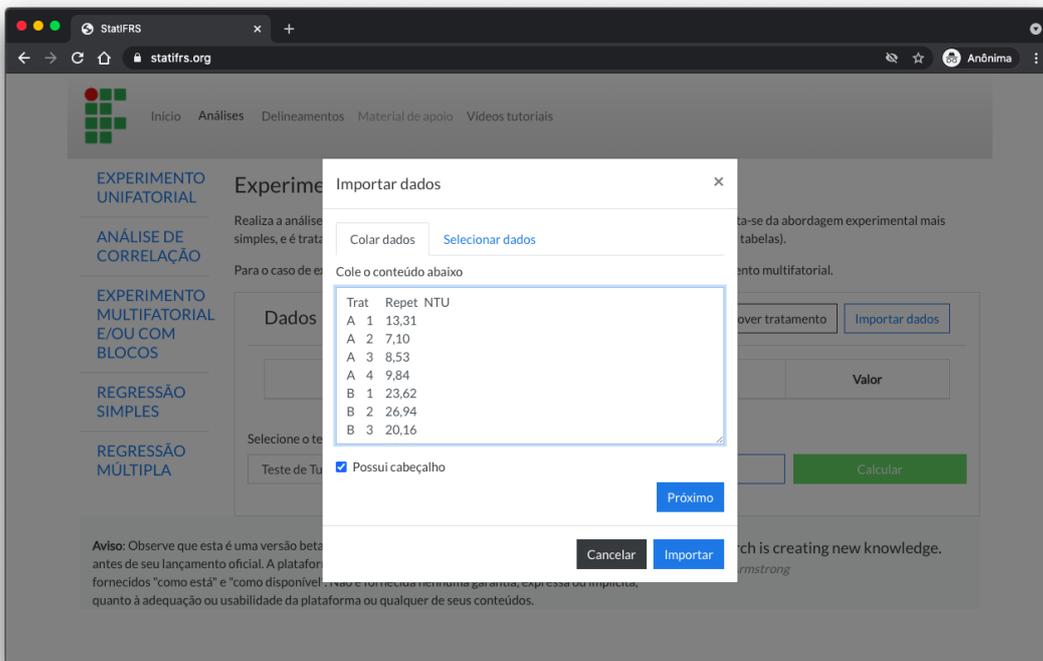
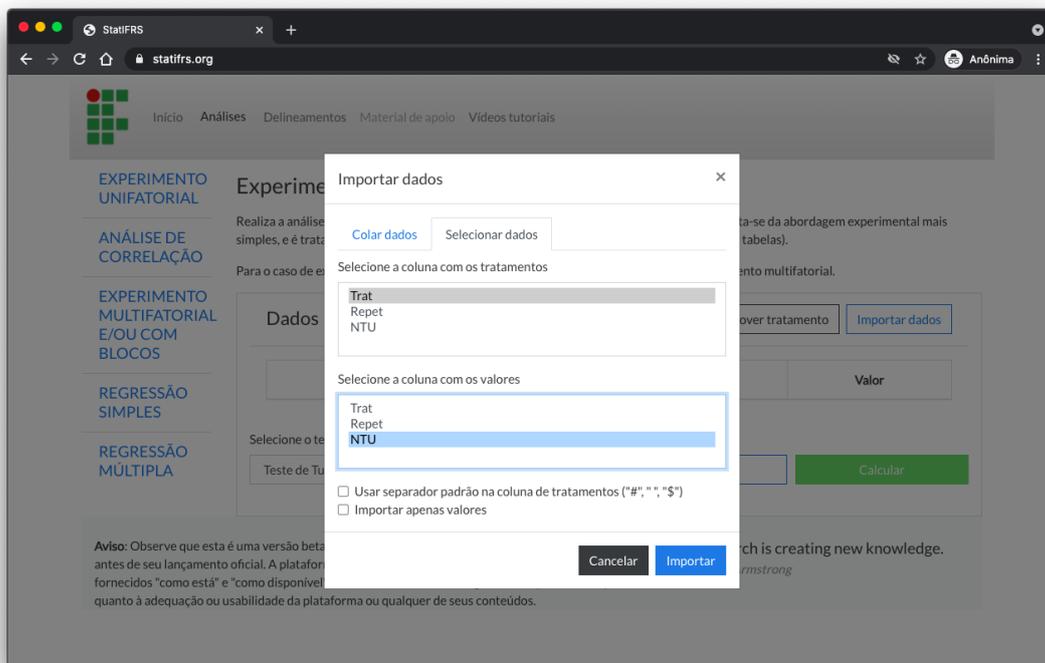


Figura 7. Processo de importação de dados da área de transferência



**Figura 8. Imagem da seleção das variáveis de um experimento unifatorial**

## **4. Resultados**

### **4.1. Validação dos resultados**

O comparativo da conformidade de resultados obtidos com os produzidos pelos outros softwares estatísticos estão dispostos na Tabela 1. Alguns delineamentos não puderam ter suas análises realizadas no Assistat pela não possibilidade da execução, dada a configuração do experimento.

É possível evidenciar a maleabilidade da ferramenta implementada, visto que permite a determinação de um ou mais blocos em fatoriais com blocagem, além da possibilidade de análises de delineamentos fracionários ou com parcelas perdidas. Estas últimas possibilidades citadas, embora úteis para análises de ensaios preliminares, como a identificação de componentes e interações, devem ser consideradas com cautela por parte do usuário. Por exemplo, é fundamental que se escolha um delineamento de Plackett Burman com número mínimo de 4 ensaios a mais que o número de variáveis, bem como a realização de, no mínimo, 3 repetições da condição de ponto central [Rodrigues and Iemma 2014].

Outro ponto importante é que, nas análises de experimentos fatoriais, os níveis são considerados variáveis discretas, portanto, os efeitos devem ser interpretados com cuidado, já que não são coeficientes diretos de valores contínuos. Em oposição, na regressão múltipla, todas as variáveis são contínuas, porém, apesar dos resultados apresentarem a análise de variância dos efeitos e interações selecionadas, a falta de análises de superfície de resposta faz com que o usuário ainda dependa de outros softwares para aprofundar-se na interpretação e apresentação destes dados.

### **4.2. Módulo de experimento unifatorial**

A análise do experimento unifatorial traz, além da tabela de variância e teste de médias com classificação em grupos de acordo com a seleção e definição do alfa pelo usuário, uma outra tabela com atributos de análise descritiva, com desvio padrão, média, mediana e primeiro e terceiro quartis, por tratamento. Ainda, traz o teste de normalidade dos dados de Shapiro-Wilk, além de parâmetros estatísticos básicos. A representação gráfica do módulo traz um gráfico de caixas (Figura 9), além de um gráfico da distribuição dos dados, por tratamento.

### **4.3. Módulo de análise de correlação**

O módulo de análise de correlação apresenta em seu resultado, além do teste selecionado, um teste de normalidade por variável, e representações gráficas, como exemplificado na Figura 10.

### **4.4. Módulo de experimento multifatorial e/ou com blocos**

A análise de um experimento fatorial no software permite a visualização da tabela de análise de variância, teste de normalidade e representação da distribuição dos dados por fator. As tabelas com testes de médias por fator e por interação são apresentados individualmente, contando ainda com uma representação gráfica (Figura 11).

**Tabela 1. Validações conduzidas através do comparativo com outros softwares estatísticos**

Delineamento	Statistica	Assistat	Observações
inteiramente casualizado	Ok	Ok	
quadrado latino	Ok	Ok	
inteiramente casualizado com parcelas perdidas	Ok	Ok	
inteiramente casualizado com blocagem	Ok	Ok	Divergências mínimas a nível de arredondamento para valor de F em relação ao Assistat.
inteiramente casualizado com parcelas perdidas e blocagem	Ok	-	
fatorial $2^k$ completo (níveis discretos)	Ok	Ok	
fatorial $3^k$ fracionário (níveis discretos)	Ok	-	Testes com 3 fatores.
fatorial $2^k$ com blocagem	Ok	Ok	Testes com 2 e 3 fatores, além do bloco.
fatorial arbitrário completo (níveis discretos)	Ok	-	Dois fatores com 2 níveis e um com 3 níveis, sem blocagem.
análise de correlação	Ok	Ok	
regressão simples	Ok	Ok	Valores de F divergem na análise de variância em relação ao Assistat, que monta a tabela usando modelos de regressão com mais coeficientes. Os valores de R e R <sup>2</sup> também divergem, já que o Assistat parece empregar apenas as médias para o cálculo destes coeficientes de determinação.
regressão simples com número de repetições divergentes	Ok	-	
regressão múltipla	Parcial	-	Apesar da conformidade dos coeficientes, ainda é necessário ao StatIFRS a implementação da discriminação do modelo e análises de superfície de resposta para que o recurso seja de interesse.

#### 4.5. Módulo de regressão simples

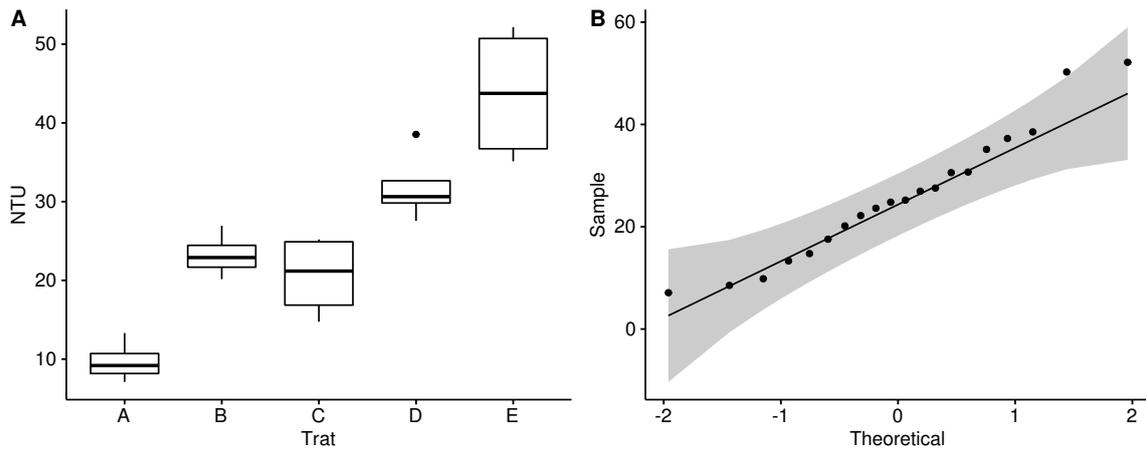
A apresentação dos resultados de um teste de regressão simples conta com a análise de variância dos modelos e dos coeficientes, separadamente. Além disso, contam com alguns parâmetros estatísticos, incluindo o  $R^2$  e  $R^2$  ajustado. A representação gráfica traz as médias com linhas verticais para os valores mínimos e máximos das repetições, além da equação do modelo e  $R^2$  ajustado calculados (Figura 12).

#### 4.6. Módulo de regressão múltipla

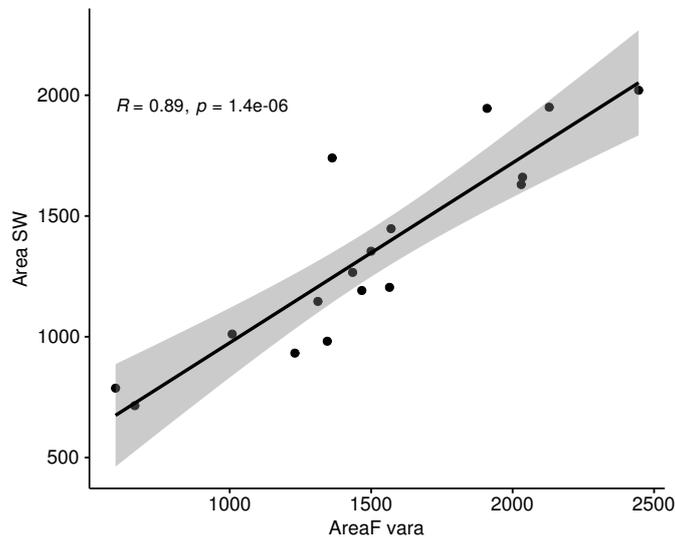
O módulo de regressão múltipla é mais limitado, uma vez que suporta apenas um modelo linear com ou sem interação entre variáveis. Em seu resultado, é apresentada uma análise de variância que permite a interpretação da significância dos coeficientes e interações.

#### 4.7. Exemplos de resultados das versões para impressão das análises

A Figura 13 traz o resultado da versão para impressão de uma análise de experimento unifatorial, em que são avaliados cinco tratamentos (A, B, C, D, E) com quatro repetições cada, através de um teste de médias. Já na Figura 14 é apresentada uma página da versão para impressão de uma análise de correlação entre duas variáveis de um experimento, avaliando a associação entre a área foliar da terceira vara de um sarmento de videira e a área foliar total do sarmento.



**Figura 9. Gráfico de caixa gerado pelo módulo de experimento unifatorial**



**Figura 10. Gráfico representando o grau de correlação linear entre duas variáveis**

Uma análise de regressão completa do modelo quadrático em um experimento do peso de sementes de uma cultivar em relação à aplicação de doses de gesso específicas ao solo, é demonstrada na Figura 15.

O resultado de um experimento fatorial é apresentado na Figura 16. As páginas estão parcialmente representadas, já que a versão de impressão no experimento fatorial permite a seleção de um fator específico, ou de todos os fatores e interações.

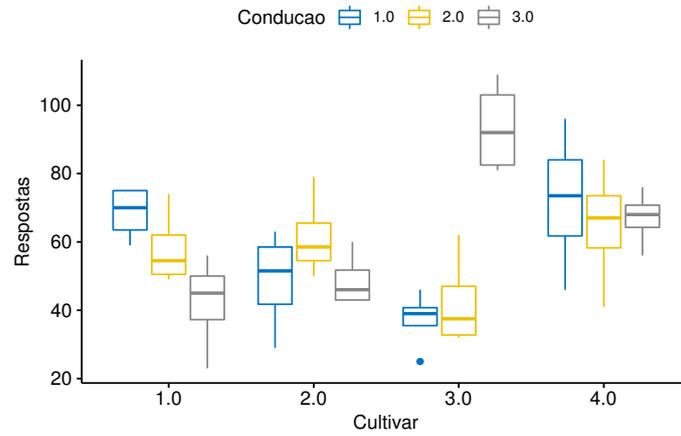


Figura 11. Gráfico de caixa de um experimento fatorial

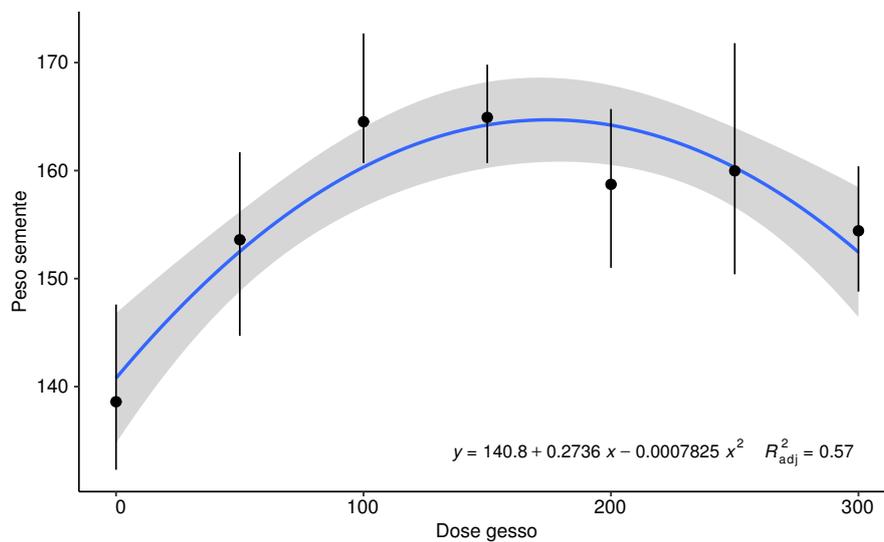


Figura 12. Gráfico representando as médias, valores mínimos e máximo e modelo de uma regressão quadrática

### Análise de variância

	GL	Soma q.	Quad. m.	Valor de F	Pr(>F)	CS*
tratamento	4	2599,7	649,9	23,02	2,95e-06	***
Residuals	15	423,5	28,2			

\* Códigos de significância: '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### Grupos

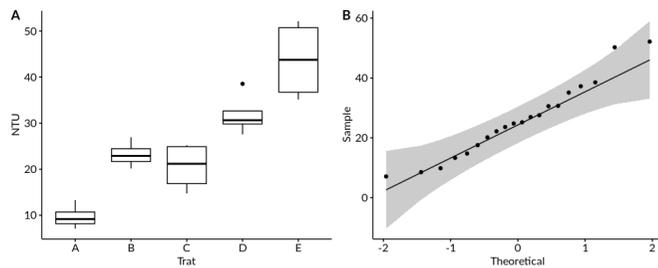
Tratamento	Média
E	43,6925 <sub>a</sub>
D	31,8425 <sub>b</sub>
B	23,2250 <sub>c</sub>
C	20,5775 <sub>bc</sub>
A	9,6950 <sub>c</sub>

Tratamentos com a mesma letra não são significativamente diferentes.

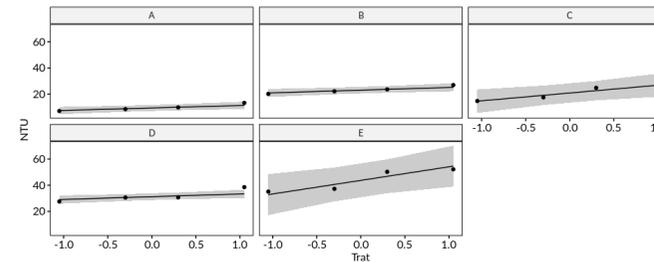
Teste	Tukey
Número de tratamentos	5
Amplitude estudentizada	4,366985
Alpha	0,05

### Teste de normalidade

Valor de W	0,95913
Nível de significância (valor de p)	0,5266



(a) Página 1



### Médias

Trat.	Média	Desvio padrão	r	Min.	Máx.	IIQ25	Mediana	IIQ75
A	9,6950	2,657097	4	7,10	13,31	8,1725	9,185	10,7075
B	23,2250	2,854441	4	20,16	26,94	21,6750	22,900	24,4500
C	20,5775	5,236477	4	14,75	25,20	16,8575	21,180	24,9000
D	31,8425	4,694624	4	27,56	38,54	29,8250	30,635	32,6525
E	43,6925	8,746879	4	35,12	52,15	36,7175	43,750	50,7250

### Estatística

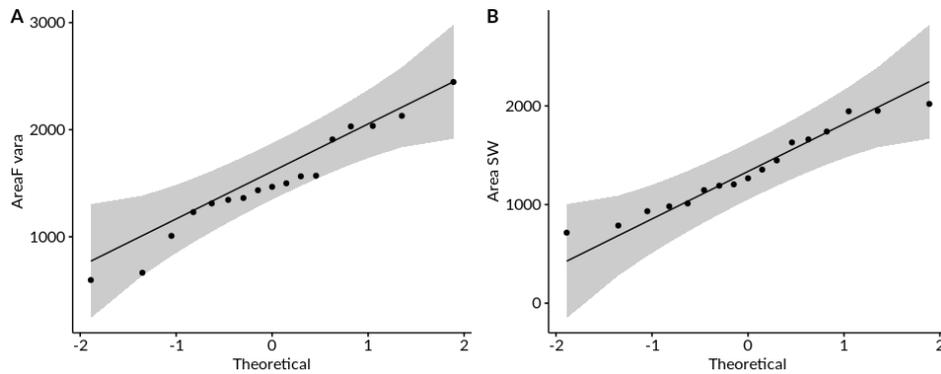
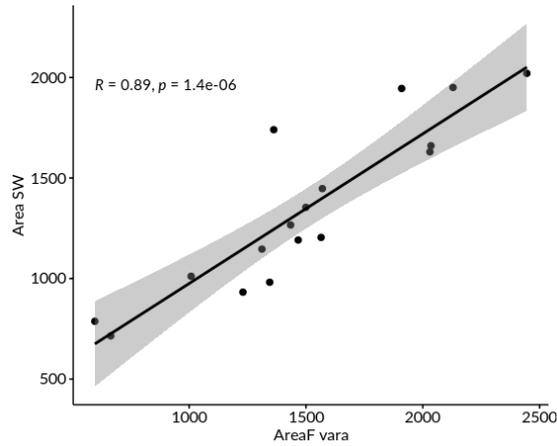
Erro quadrático médio	28,23522
Graus de liberdade	15
Média	25,8065
Coefficiente de variação	20,59048
Desvio Quadrático Médio	11,60238

(b) Página 2

Figura 13. Exemplo de resultado da versão de impressão da análise de um experimento unifatorial

## Análise de correlação

Coefficiente de correlação (r)	0,8933784
Análise	Pearson's product-moment correlation
Valor estatístico do teste t	7,7009
Grau de liberdade	15
Nível de significância do teste t (valor de p)	1,372e-06
Intervalo de confiança do coeficiente de correlação a 95%	0,7233297 a 0,9612603

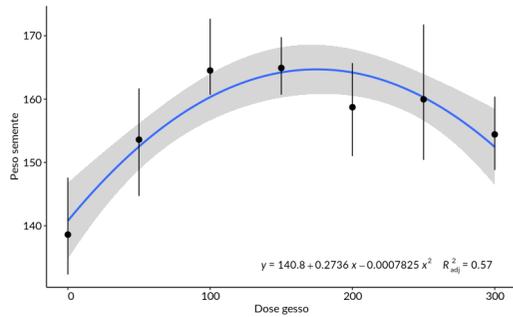


## Teste de normalidade

Valor de W (variável "A")	0,9644
Nível de significância -- valor de p (variável "A")	0,7151
Valor de W (variável "B")	0,95029
Nível de significância -- valor de p (variável "B")	0,4611

Figura 14. Exemplo de resultado da versão de impressão da análise de correlação

### Regressão simples



### Análise de variância dos efeitos

	GL	Soma q.	Quad. m.	Valor de F	Pr(>F)	CS*
Regressão linear	1	423,15	423,15	10,0258	0,004653	**
Regressão quadrática	1	1285,84	1285,84	30,4655	1,777e-05	***
Desvios	4	232,83	58,21	1,3791	0,275047	
Residuals	21	886,34	42,21			

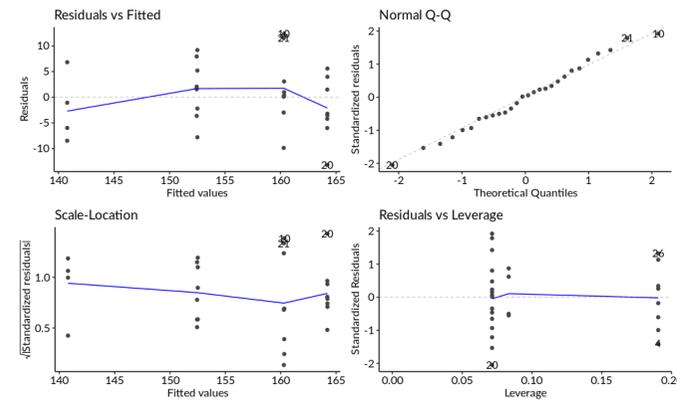
### Coefficientes

Coefficiente	Valor	Erro P.	Val. de t	Pr(> t )	CS*
Intersecção	1,408e+02	2,920e+00	48,212	2e-16	***
x <sup>1</sup>	2,736e-01	4,559e-02	6,002	2,87e-06	***
x <sup>2</sup>	-7,825e-04	1,460e-04	-5,359	1,47e-05	***

\* Códigos de significância: '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Erro padrão residual: 6,691  
 Graus de liberdade: 25  
 R<sup>2</sup> múltiplo: 0,6043  
 R<sup>2</sup> ajustado: 0,5726  
 F-estatístico: 19,09  
 Valor de p: 9,277e-06

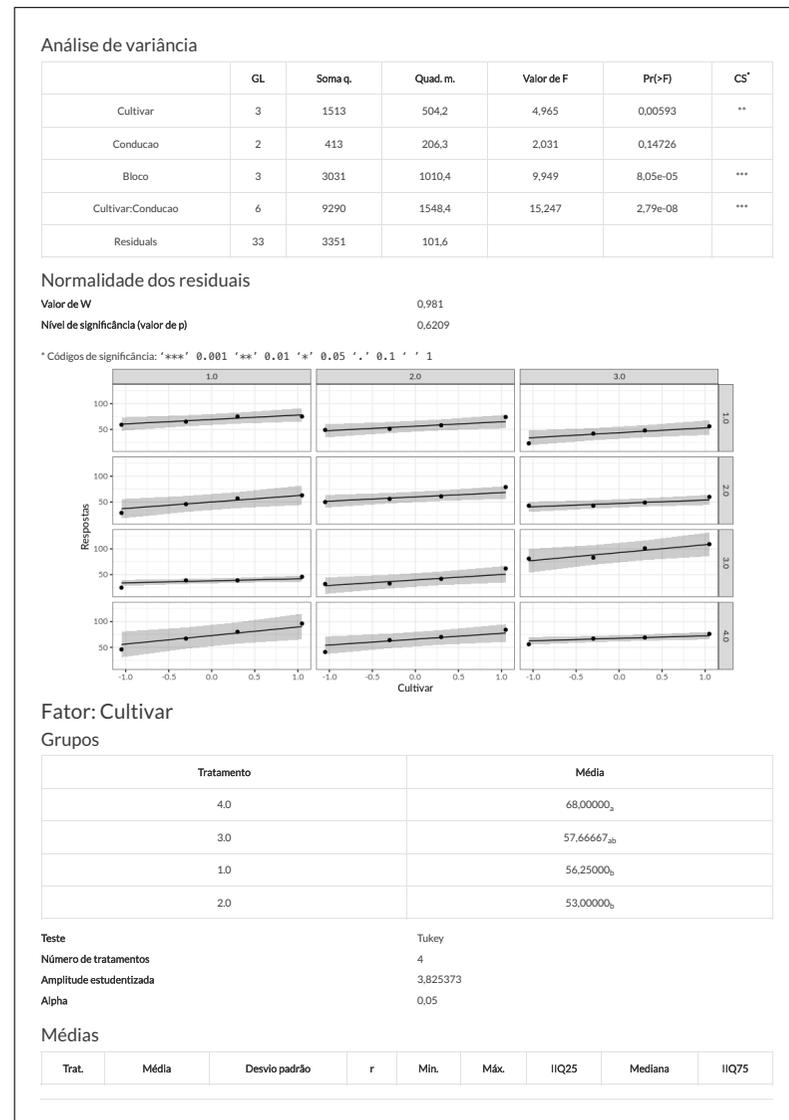
### Residuais



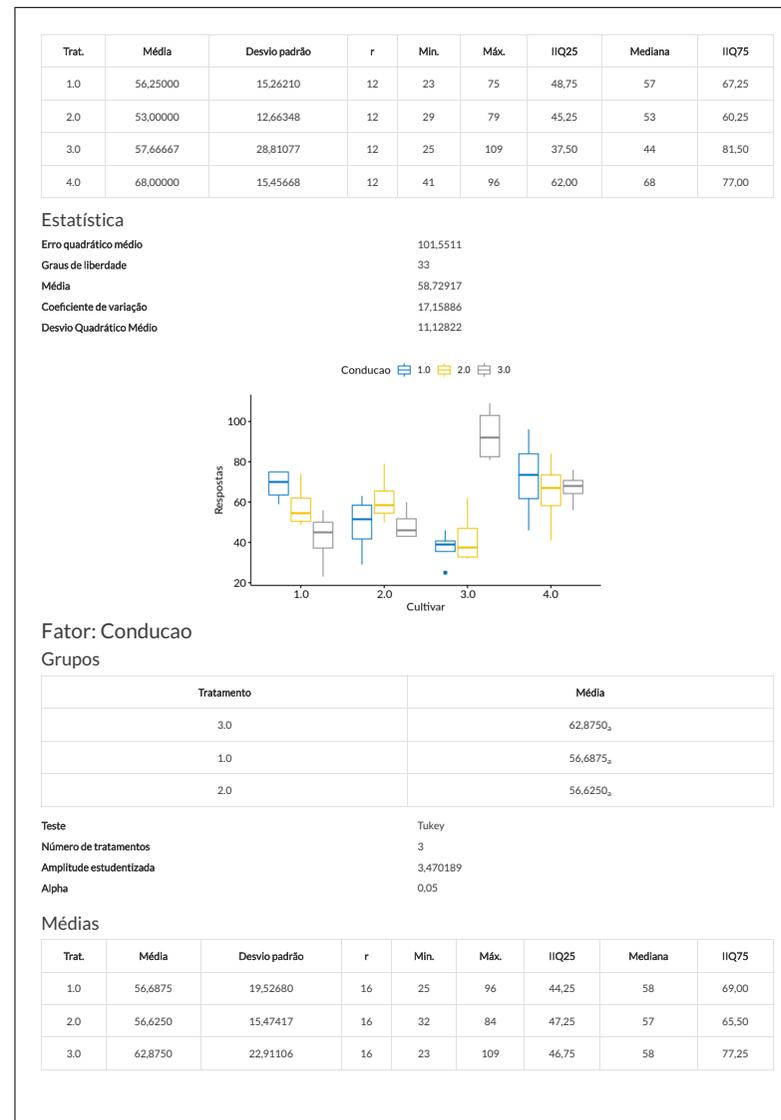
(a) Página 1

(b) Página 2

Figura 15. Exemplo de resultado da versão de impressão da análise de regressão



(a) Página 1



(b) Página 2

Figura 16. Exemplo parcial de resultado da versão de impressão da análise de experimento fatorial com blocagem

## **5. Discussão**

Foi implementado um software de análises estatísticas de livre acesso baseado no R. O StatIFRS faz uso de uma plataforma cliente-servidor que difere dos demais softwares estatísticos disponíveis, e independe de sistema operacional, arquitetura e dispensa quaisquer necessidade de instalação de aplicativos no sistema. O software ainda vai além das análises, realizando testes pertinentes e apresentando graficamente dados e resultados, permitindo ao usuário a exportação vetorial e rasterizada destes, e dispõe ainda de uma versão adequada para a impressão dos resultados. A prioridade foi fornecer um protocolo simples, prático e estatisticamente sólido para a realização de análises de delineamentos experimentais usuais do meio acadêmico e privado.

A opção dos frameworks GWT na codificação juntamente com Bootstrap para a apresentação e interação direta ao usuário se mostrou uma boa alternativa, visto que a medida que o volume de código aumentava, a possibilidade da organização deste em classes e métodos, na linguagem Java, se tornava muito útil. Além disso, ao ser incorporada toda a lógica de negócio em uma mesma linguagem, há uma diminuição da curva de aprendizado, além de conceder certo grau de independência da camada final de apresentação.

Com relação à análise de regressão múltipla, ainda há uma grande lacuna a ser preenchida pelo software StatIFRS, uma vez que diferentemente da análise de experimentos fatoriais, as variáveis de entrada são todas contínuas, e cujo modelo necessita oferecer uma maior maleabilidade. Atualmente, a ferramenta apenas analisa relações lineares e interações entre os fatores, sendo que deveriam ser considerados modelos quadráticos e exponenciais. Uma vez determinados os coeficientes significantes pelo usuário, a interação deveria permitir análises do modelo através de plotagens de superfície de resposta. Apesar de terem sido desenvolvidos os protocolos para tais abordagens, a maior demanda de processamento, a escalabilidade e alta variabilidade de opções, a nível de interação que demandam seleção por parte do usuário, impediram que, até a elaboração deste trabalho, estes recursos fossem implementados.

Em conclusão, dado seu caráter de livre acesso e independência de sistema ou instaladores, o StatIFRS pode ser útil para a realização de análises estatísticas experimentais por toda a comunidade acadêmica e público em geral.

## **6. Trabalhos futuros**

Pretende-se, antes de dar sequência à implementação de novos recursos para incrementar o suporte à análise de modelos de regressão e delineamentos mais complexos, conduzir uma validação da ferramenta junto a uma instituição de ensino com atuação na pesquisa experimental, de modo a não apenas testar a correção dos dados produzidos, mas também, observar a performance da plataforma mediante acessos concorrentes, bem como a adequabilidade em relação à interação do usuário.

## **7. Agradecimentos**

Somos gratos ao Dr. Evandro Ficanha, Me. Vinícius Zanandrea e Angelo Gava pela supervisão do desenvolvimento do software e por seus inestimáveis conselhos.

## Referências

- Balasubramanee, V., Wimalasena, C., Singh, R., and Pierce, M. (2013). Twitter bootstrap and angularjs: Frontend frameworks to expedite science gateway development. In *2013 IEEE International Conference on Cluster Computing (CLUSTER)*, pages 1–1. IEEE Computer Society.
- Breheny, P. and Burchett, W. (2017). Visualization of regression models using visreg. *The R Journal*, 9(2):56–71.
- Chang, W., Cheng, J., Allaire, J., Xie, Y., and McPherson, J. (2015). Package ‘shiny’. See <http://citeseerx.ist.psu.edu/viewdoc/download>.
- Costa, J. R. (2003). Técnicas experimentais aplicadas às ciências agrárias. *Embrapa Agrobiologia-Documentos (INFOTECA-E)*.
- Everitt, B. S. and Skrondal, A. (2010). The cambridge dictionary of statistics.
- Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P., and Berners-Lee, T. (1999). Hypertext transfer protocol–http/1.1.
- Genolini, C., Dubois, E., and Furió, D. (2019). R++, user-friendly statistical software. In *IFIP Conference on Human-Computer Interaction*, pages 605–609. Springer.
- Kassambara, A. (2020). *ggpubr: 'ggplot2' Based Publication Ready Plots*. R package version 0.4.0.
- Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., Ly, A., Gronau, Q. F., Šmíra, M., Epskamp, S., et al. (2019). Jasp: Graphical statistical software for common statistical designs. *Journal of Statistical Software*, 88(1):1–17.
- Manelli, L. and Zambon, G. (2020). *Beginning Jakarta EE web development: using JSP, JSF, MySQL, and apache tomcat for building java web applications*. Apress.
- Mendiburu, F. (2020). *agricolae: Statistical Procedures for Agricultural Research*. R package version 1.3-3.
- Montgomery, D. C. (2017). *Design and analysis of experiments*. John wiley & sons.
- Nieh, J., Yang, S. J., and Novik, N. (2000). A comparison of thin-client computing architectures.
- Pavlenko, A., Askarbekuly, N., Megha, S., and Mazzara, M. (2020). Micro-frontends: application of microservices to web front-ends. *J. Internet Serv. Inf. Secur.*, 10(2):49–66.
- Peterson, J. J., Snee, R. D., McAllister, P. R., Schofield, T. L., and Carella, A. J. (2009). Statistics in pharmaceutical development and manufacturing. *Journal of Quality Technology*, 41(2):111–134.
- Pimentel Gomes, F. (1990). Curso de estatística experimental. 13ª. *Edição. São Paulo: Nobel*.
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Reese, G. (2000). *Database Programming with JDBC and JAVA*. ”O’Reilly Media, Inc.”.

- Rodrigues, M. and Iemma, A. (2014). Planejamento de experimentos e otimização de processos. 2<sup>a</sup>. Campinas/SP.
- Silva, F. d. A. S. e. and Azevedo, C. A. V. d. (2016). The assistat software version 7.7 and its use in the analysis of experimental data. *African Journal of Agricultural Research*, 11(39):3733–3740.
- Watkins, J. C. (2019). An introduction to the science of statistics: From theory to implementation.
- Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, 40(1):1–29.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.