

**UNIVERSIDADE DE SÃO PAULO**  
**FACULDADE DE ZOOTECNIA E ENGENHARIA DE ALIMENTOS**

**PAULO AFONSO GUELDINI DE MORAES**

**Desenvolvimento de metodologia para análise de dados gráficos aplicados à avaliação da  
frequência cardíaca de equinos.**

---

Pirassununga

2021

**PAULO AFONSO GUELDINI DE MORAES**

**Desenvolvimento de metodologia para análise de dados gráficos aplicados à avaliação da frequência cardíaca de equinos.**

**(Versão corrigida)**

Dissertação apresentada à Faculdade de Zootecnia e Engenharia de Alimentos da Universidade de São Paulo, como parte dos requisitos para a obtenção do título de Mestre em Ciências no Programa de Pós Graduação em Gestão e Inovação na Indústria Animal.

**Área de Concentração:** Gestão e Inovação na Indústria Animal

**Orientador:** Prof. Dra. Roberta Ariboni Brandi

---

Pirassununga

2021

## FICHA CATALOGRÁFICA

Ficha catalográfica elaborada pelo  
Serviço de Biblioteca e Informação, FZEA/USP,  
com os dados fornecidos pelo(a) autor(a)

M827d

Moraes, Paulo Afonso Gueldini de  
Desenvolvimento de metodologia para análise de  
dados gráficos aplicados à avaliação da frequência  
cardíaca de equinos. / Paulo Afonso Gueldini de  
Moraes ; orientadora Roberta Ariboni Brandi. --  
Pirassununga, 2021.  
45 f.

Dissertação (Mestrado - Programa de Pós-Graduação  
em Mestrado Profissional Gestão e Inovação na  
Indústria Animal) -- Faculdade de Zootecnia e  
Engenharia de Alimentos, Universidade de São Paulo.

1. análises gráficas. 2. cavalos. 3. extração de  
dados. I. Brandi, Roberta Ariboni, orient. II.  
Título.

**PAULO AFONSO GUELDINI DE MORAES**

**Desenvolvimento de metodologia para análise de dados gráficos aplicados à avaliação da frequência cardíaca de equinos.**

Dissertação apresentada à Faculdade de Zootecnia e Engenharia de Alimentos da Universidade de São Paulo, como parte dos requisitos para a obtenção do título de Mestre em Ciências no Programa de Pós Graduação em Gestão e Inovação na Indústria Animal.

**Data de Aprovação:** 29/11/2021

**Banca Examinadora:**

Prof. Dra. Roberta Ariboni Brandi – FZEA/USP  
Presidente da Banca Examinadora / Orientadora

Prof. Dr. Cesar Gonçalves de Lima – FZEA/USP  
Membro Interno ao Programa

Prof. Dr. Leonir Bueno Ribeiro – UEM  
Membro Externo ao Programa

Prof. Dra. Mayara Gonçalves Fonseca – Integral AgroIndustrial  
Membro Externo ao Programa

## DEDICATÓRIA

Dedico este trabalho aos meus pais, Eliana e Paulo, minhas irmãs Anna Valéria e Liege, à minha esposa Marina e ao meu filho Caio.

## **AGRADECIMENTOS**

Aos meus pais e irmãs, pelo incentivo, apoio incondicional, pelo amor e por sempre estarem ao meu lado nos momentos que antecederam esta jornada até aqui. E a todos os meus amigos que tornaram essa caminhada mais leve e saluta.

À minha esposa Marina, pela compreensão e pelo apoio nas decisões difíceis e incertezas da vida.

Ao meu filho Caio, pelo carinho e amor incondicional.

À esta Universidade, seu corpo docente e administração que oportunizaram a janela que hoje vislumbro um horizonte superior, pela confiança e ética aqui presente.

À minha orientadora Prof.<sup>a</sup> Dr<sup>a</sup> Roberta Ariboni Brandi, por ter acreditado neste projeto, pela oportunidade dada, pelo suporte, paciência e suas correções.

Ao GPEEAC (Grupo de Pesquisas em Equinos e Animais de Companhia - FZEA), que me deu a oportunidade de vivenciar a prática do manejo de equinos.

Aos professores do PPG-GIIA e aos membros da banca examinadora por contribuírem na melhoria da pesquisa.

Aos colegas do PPG-GIIA pela troca de experiências e oportunidades de aprendizados pessoais que levarei para a vida.

O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

## **EPÍGRAFE**

“Somos o que pensamos. Tudo o que somos surge com nossos pensamentos. Com nossos pensamentos, fazemos o nosso mundo.” (Buda)

## RESUMO

MORAES, P. A. G. **Desenvolvimento de metodologia para análise de dados gráficos aplicados à avaliação da frequência cardíaca em equinos**, 2021. Dissertação (Pós-Graduação em Gestão e Inovação na Indústria Animal) – Faculdade de Zootecnia e Engenharia de Alimentos, Universidade de São Paulo, Pirassununga, 2021.

Este estudo teve como objetivo a proposição de um método para extração de dados gráficos relacionados à frequência cardíaca de equinos, obtidos através do uso do frequencímetro Polar Equine H7. As informações gráficas foram geradas pelo programa KER ClockIT<sup>®</sup>Sport. Para a escolha do programa de extração de dados gráficos levou-se em consideração atributos técnicos e funcionais, tais como a gratuidade da aplicação, sua utilização em caráter offline e a capacidade de extração automática e posterior ajuste dos dados. A extração de dados em sua forma bruta, foi realizada por meio da utilização do programa WebPlotDigitizer<sup>®</sup>. A partir dos dados brutos extraídos, foi realizado o cálculo das médias e desvios padrão de cada atividade por meio do Excel<sup>®</sup>. Adicionalmente, as atividades foram classificadas em dois tipos: "com" e "sem ruído", onde valores de frequência cardíaca iguais a zero foram verificados durante a realização da atividade equestre e considerados para efeito de classificação, sendo posteriormente realizada a comparação dos dados através do teste de correlação de Pearson, considerando-se 95% de significância. Para as atividades classificadas como "sem ruído", o valor verificado de correlação para o indicador de média foi  $r = 0,999$  e para o desvio padrão foi  $r = 0,929$ , enquanto que as atividades classificadas como "com ruído", os mesmos indicadores resultaram em  $r = 0,479$  e  $r = 0,941$  respectivamente. É possível a utilização do programa WebPlotDigitizer<sup>®</sup> como alternativa para a extração de informações gráficas, além de viabilizar a possibilidade de eliminação de dados considerados como "com ruído" através de análises de dados pormenorizados.

Palavras-chave: análises gráficas, cavalos, extração de dados.



## ABSTRACT

MORAES, P. A. G. **Methodology development for graphical data analysis applied to horse's heart rate evaluation**, 2021. Dissertação (Pós-Graduação em Gestão e Inovação na Indústria Animal) – Faculdade de Zootecnia e Engenharia de Alimentos, Universidade de São Paulo, Pirassununga, 2021.

This study aimed to propose a graphical data extraction method related to equine's heart rate measurements, obtained by the Polar Equine H7 frequencimeter usage. The graphical information has been generated by the KER ClockIT<sup>®</sup>Sport application. On the graphical data extraction software selection process, both functional and technical attributes have been considered, such as being a freeware application, its possible offline usage, an automatic feature for extraction and post data adjusts. The raw data extraction process was enabled through the use of WebPlotDigitizer<sup>®</sup> application. From the extracted raw data, the standart desviation and mean indicators were calculated using Excel<sup>®</sup>. Adicionally, the activities were classified in two diferente types: “verified noise” and “not verified noise (noiseless)”, where the heart rate frequencies equal to zero has been verified during the equestrian activity realization and considered for classification aspects, being compared through the Pearson's correlation test, considering 95% significance. For the activities classified as “noiseless”, the correlation verified value for the mean indicator was  $r = 0,999$  and for the standart desviation indicator was  $r = 0,929$ , meanwhile the activities classified as “noise verified” demonstrated the same indicators as  $r = 0,479$  and  $r = 0,941$  respectively. It's possible to use WebPlotDigitizer<sup>®</sup> as a valid alternative for the determination of heart rate extracted data, also enabling the “verified noise” data to be eliminated through a more detailed data analysis.

Keywords: graphical analysis, horses, data extraction.

## LISTA DE FIGURAS

Figura 1 – Gráfico e informações adicionais disponibilizadas pelo aplicativo ClockIT <sup>®</sup> Sport.....	23
Figura 2 – Ruído verificado nas informações coletadas pelo frequencímetro e disponibilizada pelo ClockIT <sup>®</sup> Sport – marcações com valor igual a zero entre 00:30 e 01:30.....	24
Figura 3 – Etapas iniciais do processo de extração automática de informações utilizando o WebPlotDigitizer.....	30
Figura 4 – Calibragem dos eixos X e Y para execução do algoritmo de extração de dados.....	31
Figura 5 – Delimitação de atributos para execução do algoritmo de extração de dados.....	32
Figura 6 – Delimitação de área de interesse para execução do algoritmo de extração de dados..	33
Figura 7 – Identificação dos pontos mapeados pelo algoritmo para extração de dados em sua forma bruta.....	34
Figura 8 – Visualização, opções de formatação e funcionalidades para extração dos pontos mapeados como dados em sua forma bruta.....	34
Figura 9 – Fluxograma dos principais processos vinculados à extração automática de dados gráficos através do WebPlotDigitizer.....	35
Figura 10 – Dispersão com ajuste de regressão linear para as informações de média, considerando o total de dados de amostragem do estudo. Gráfico gerado à esquerda contempla os registros que não tiveram ruído verificado (N=14), enquanto o gráfico à direita representa os registros que tiveram ruído verificado (N=7).....	38
Figura 11 – Dispersão com ajuste de regressão linear para as informações de desvio padrão, considerando o total de dados de amostragem do estudo. Gráfico gerado à esquerda contempla os registros que não tiveram ruído verificado (N=14), enquanto o gráfico à direita representa os	

registros	que	tiveram	ruído	verificado
(N=7)	.....	.....	.....	.....39

## **LISTA DE QUADROS**

Quadro 1 – Comparação de aplicações de suporte à extração de dados gráficos em termos de requerimentos técnicos e funcionais. Fonte: Autoria Própria

Quadro 2 – Consolidação dos dados referentes aos indicadores de média e desvio padrão comparados entre aplicações para validação do método proposto. Fonte: Autoria Própria

Quadro 3 – Comparativo de resultados de análise de correlação de Pearson para os indicadores de média e desvio padrão entre dados fornecidos e extraídos. Fonte: Autoria Própria

## LISTA DE SIGLAS E ABREVIATURAS

FC	Frequência Cardíaca
WPD	WebPlotDigitizer
PDF	Portable Document File (formato/tipo de arquivo de texto)
Bpm	Batimentos Por Minuto
2D	Duas Dimensões
JPEG	Joint Photographics Experts (formato/tipo de arquivo de imagem)
PNG	Portable Network Graphics (formato/tipo de arquivo de imagem)
COVID	Corona Vírus
GPS	Global Positioning System
CEUA	Comitê de Ética no Uso de Animais
BH	Raça Brasileira de Hipismo
SRD	Sem Raça Definida
EVA	Etil Vinil Acetatoé
Km/h	Quilômetros por hora
RGB	Composição de Cores Suplementares (Vermelho, Verde e Azul)

## SUMÁRIO

<b>1. INTRODUÇÃO.....</b>	<b>15</b>
<b>2. OBJETIVOS E HIPÓTESE.....</b>	<b>16</b>
2.1 OBJETIVO GERAL.....	16
2.2 OBJETIVOS ESPECÍFICOS.....	16
2.3 HIPÓTESE.....	16
<b>3. REVISÃO DE LITERATURA.....</b>	<b>17</b>
3.1 EXTRAÇÃO DE DADOS GRÁFICOS ATRAVÉS DE SOFTWARES DE APOIO ....	17
3.2 MENSURAÇÃO DA FREQUÊNCIA CARDÍACA.....	18
3.3 TESTES COMPORTAMENTAIS.....	19
3.4 COEFICIENTE DE CORRELAÇÃO DE PEARSON .....	20
<b>4. MATERIAL E MÉTODOS.....</b>	<b>22</b>
4.1 O EXPERIMENTO / BASE DE DADOS .....	22
4.2 DADOS GRÁFICOS .....	23
4.3 ANÁLISE DE DADOS.....	24
<b>5. RESULTADOS E DISCUSSÃO.....</b>	<b>26</b>
5.1 SELEÇÃO DE SOFTWARE .....	26
5.2 PROCESSO DE EXTRAÇÃO.....	29
5.3 APLICAÇÃO PRÁTICA DO MÉTODO .....	36
<b>6. CONCLUSÃO.....</b>	<b>41</b>
<b>REFERÊNCIAS.....</b>	<b>42</b>

## 1. INTRODUÇÃO

Visando o manejo seguro de equinos, é necessário conhecer sua personalidade, o que se faz através de testes de reatividade. A frequência cardíaca (FC) pode ser usada como variável auxiliar neste processo e é bem aceita por ser uma medida não invasiva de fácil obtenção. Sua mensuração pode ser feita através do uso de eletrocardiogramas ou através de frequencímetros, que permitem sua análise contínua ao longo do exercício ou teste, possibilitando sua aferição em condições de campo.

Parte dos frequencímetros empregados neste processo acabam por delimitar a possibilidade de extração e utilização dos dados coletados numa atividade realizada, limitando-se à disponibilização de informações apenas em forma gráfica, além de alguns indicadores adicionais, tais como a média e o desvio padrão, o que acaba por tornar a realização de análises mais detalhadas de FC uma atividade complexa e passível de erros.

Para que seja possível trabalhar estes dados, programas de extração de dados se fazem necessários. Uma ferramenta de digitalização de dados gráficos, de forma geral, é um software capaz de permitir que usuários possam extrair dados numéricos a partir de diferentes tipos de gráficos ou figuras. Existem diversas ferramentas de digitalização de dados gráficos atualmente que variam em preço, compatibilidade com plataformas e usabilidade geral. Para que o processo de extração seja realizado, tais programas demandam que os usuários façam a importação dos gráficos ou figuras e calibrem os eixos dos gráficos analisados por meio da determinação de dois pontos conhecidos nos mesmos para viabilização do processo de interpolação de coordenadas. A partir daí, é necessário que se delimite os elementos gráficos que permitam a identificação da função ou série de dados que se queira extrapolar como dados brutos, além da manipulação de coordenadas de forma manual para eventuais ajustes desejados. Tais aplicações permitem também que funções automáticas para digitalização de dados sejam utilizadas, permitindo que este processo ganhe velocidade e agilidade nas análises de grandes quantidades de dados gráficos e, eventualmente, estabelecendo rotinas para as análises que possuam disposição gráfica padronizada.

Partindo destas premissas, este trabalho objetivou propor uma metodologia para extração de dados a partir de gráficos gerados para avaliação da FC de equinos ao longo do tempo, de forma a realizar uma análise comparativa dos indicadores de média e desvio padrão entre as informações extraídas e as disponibilizadas pelo frequencímetro.

## 2. OBJETIVOS E HIPÓTESE

### 2.1 OBJETIVO GERAL

Este trabalho objetivou propor uma metodologia para extração de dados a partir de gráficos gerados para avaliação da FC ao longo do tempo, utilizando uma ferramenta auxiliar e aplicá-la verificando-se sua acurácia.

### 2.2 OBJETIVOS ESPECÍFICOS

1. Detalhar o processo de escolha da aplicação a ser utilizada para o processo de extração de dados.
2. Detalhar o processo para utilização da ferramenta WebPlotDigitizer como aplicação de suporte no processo de extração das informações gráficas geradas pelo uso do frequencímetro Polar Equine H7™;
3. Verificar a correlação dos dados referentes aos indicadores de médias e desvios padrões obtidos através do uso do Polar Equine H7™ com as informações obtidas através do método de extração de dados gráficos por meio da aplicação da análise de correlação linear de Pearson;

### 2.3 HIPÓTESE

A hipótese a ser confirmada na atual pesquisa, parte do pressuposto que a extração dos dados gráficos de FC é possível e aplicável com a utilização de programas computacionais e permitirá análises pormenorizadas através dos dados extraídos em sua forma bruta.



### 3. REVISÃO DE LITERATURA

#### 3.1 EXTRAÇÃO DE DADOS GRÁFICOS ATRAVÉS DE SOFTWARES DE APOIO

Existem muitos softwares de apoio para extração de dados gráficos para obtenção de valores brutos (MOEYAERT et al, 2016). Softwares que já foram estudados anteriormente (SHADISH et al., 2009; BOYLE et al., 2013; FLOWER et al., 2016; RAKAP et al., 2016) incluem o Ungraph (BIOSOFT, 2004), o GraphClick (ARIZONA SOFTWARE INC., 2010), o DataThief III (TUMMERS, 2005-2006) e o DigitizeIt (BORMANN, 2012).

Shadish et al. (2009) investigaram a confiabilidade e acuraria do programa Ungraph através da comparação dos dados extraídos de gráficos publicados em estudos com seus respectivos dados numéricos relacionados, reportados pelos autores em tabelas ou textos destas publicações selecionadas. Estes autores concluíram haver um alto nível de confiança nos dados extraídos devido à quase semelhança entre eles e os que foram reportados originalmente. Boyle et al. (2013) replicaram o estudo de Shadish et al. (2009), porém ao invés de utilizarem as informações fornecidas por estudos anteriores analisados como base para análise e validação de confiabilidade, foram gerados gráficos hipotéticos que tiveram sua correlação verificada versus os dados extraídos durante a realização deste experimento. Rakap et al (2016) investigaram a confiabilidade e a acurácia de acordo com o método proposto anteriormente por Shadish et al (2009) utilizando os programas Ungraph, GraphClick e DigitizeIt, sendo verificados níveis de correlação de 0.999 ( $P < 0,001$ ) para as três aplicações utilizadas como softwares de apoio ao processo de extração de dados gráficos.

Flower et al. (2016) focaram seus estudos em dois programas de extração de dados, sendo eles o GraphClick e o DataThief III. Em linha com os estudos anteriores (BOYLE et al., 2013; RAKAP et al., 2016; SHADISH et al., 2009), foram gerados gráficos hipotéticos (15 séries de dados) para validação deste estudo, porém também foram incluídos dados de artigos publicados como proposta de validação. Estes quatro estudos indicaram que os programas avaliados eram confiáveis e acurados (MOEYAERT et al, 2016). Entretanto, nenhum destes estudos avaliou problemas relacionados à usabilidade dos programas analisados, como o tempo demandado para a extração ou a interface de interação com usuários ((MOEYAERT et al, 2016) e, adicionalmente, nenhum deles avaliou o WebPlotDigitizer (ROHATGI, 2019).

O WebPlotDigitizer (WPD) demonstrou grande potencial como ferramenta de suporte à análise de dados. Chan et al. (2020) utilizaram esta ferramenta como suporte para

desenvolvimento de um estudo relacionado à mobilidade humana durante a pandemia de COVID desencadeada nos últimos anos, fazendo a extração de dados brutos de 5766 gráficos de mobilidade disponibilizados pelo *Google™* em formato PDF, através da aquisição de informações sobre localizações em celulares com GPS, cobrindo 131 países e 818 sub-regiões entre o período de 16 de fevereiro de 2020 a 29 de março de 2020, de forma a suportar a criação de políticas e condutas de resposta à pandemia mundial.

Malin et al. (2019) utilizaram o WPD em estudos de impactos de terremotos de forma a realizar a digitalização de sinais sismográficos de uma região da Polônia, para atividades sísmicas que foram registradas de forma analógica em uma base de dados histórica e não possuem registros digitalizados. Tais informações contribuíram para a avaliação de riscos sísmicos, especialmente para a tomada de decisões estratégicas na construção de edificações para geração de energia nuclear.

Riaz, M. (2020) utilizou o WPD como ferramenta de suporte ao estudo de componentes de microeletrônica avaliando sinais gerados por osciloscópios, digitalizando as informações referentes às formas de ondas determinadas nestes equipamentos, de forma a compará-las com as equações teóricas propostas anteriormente em estudos realizados.

A extração de dados gráfico também pode ser aplicado em sistemas biológicos, como nos estudos envolvendo a frequência cardíaca, conforme proposto por Sabela (2019), na análise de mensuração de força diafragmática relacionada à animais com insuficiência cardíaca.

### 3.2 MENSURAÇÃO DA FREQUÊNCIA CARDÍACA

O sistema cardiovascular é um sistema de transporte que consiste em uma bomba muscular, o coração e uma rede de vasos sanguíneos que contêm sangue (BONOMO et al., 2014), sendo o conjunto de eventos que ocorre entre o início de um batimento cardíaco e o início do próximo denominado “ciclo cardíaco”. Este ciclo consiste no período de relaxamento, chamado diástole, onde o coração se enche de sangue proveniente das veias, seguido pelo período de contração, chamado sístole, onde o sangue é ejetado para as artérias (GUYTON; HALL, 2011), sendo a frequência cardíaca correspondente ao número médio de ciclos cardíacos por minuto (PASCHOA; COUTINHO; ALMEIDA, 2006).

A mensuração de FC pode utilizar equipamentos como os eletrocardiógrafos, (EVANS, 2000) e como os frequencímetros (WILLIAMS et al., 2009). A utilização dos frequencímetros tem sido justificada pelo seu baixo custo e fácil aplicação em ambiente laboratorial, externo ou

ainda em condições da vida diária de um indivíduo (CRUZ et al., 2018). Frequencímetros de diferentes marcas e modelos vêm sendo utilizados em muitos estudos em modelo animal (HOPSTER et al., 1994; PARKER et al., 2010, MARCHANT-FORDE et al., 2004) devido ao seu baixo custo e fácil manuseio (ESSNER et al., 2013).

Esses aparelhos colhem informações com os animais durante o treinamento, podendo ser utilizados para se estimar o gasto de energia e também a evolução física dos animais durante o condicionamento (MANSO FILHO et al., 2012), além de inferir sobre o seu comportamento (VISSER et al., 2003). Segundo Hodgson (2014), a frequência cardíaca do cavalo em repouso depende do grau de relaxamento do cavalo individualmente e permanece na faixa de 25 a 40 batimentos por minuto (bpm).

Mensurações fisiológicas da frequência cardíaca ou mesmo da variação desta, podem ser utilizadas na avaliação da reação do animal ao manejo, condução ou a presença do humano (HAMA et al., 1996; VISSER et al., 2003).

### 3.3 TESTES COMPORTAMENTAIS

De acordo com Merkies (2014), frequência cardíaca e o temperamento são indicadores do estado comportamental em que o cavalo se encontra, sendo os testes de comportamento utilizados para desafiar os equídeos a responderem à diferentes estímulos. E através destes desafios é possível observar a conduta comportamental do animal e conseqüentemente avaliar a reatividade e o temperamento. Cada teste de comportamento possui o seu objetivo e pode avaliar uma ou mais características temperamentais e são uma forma padronizada de obtenção desta informação (LANSADE et al., 2013; OLSSON, 2008, LESIMPLE et al., 2010; KOZAK et al., 2018).

Para avaliação comportamental são utilizadas medidas como a latência (quanto tempo demorou para o animal começar a fazer um evento esperado), frequência (quantas vezes o comportamento aconteceu), duração (durante quanto tempo realizou o comportamento) e intensidade (escore de comportamento) (MILLS; NARKERVIS, 2005) e, para a avaliação da intensidade são atribuídas pontuações que indicam quão intenso foi o comportamento, e para isso, geralmente é indicada a escala de Likert para a elaboração dos escores (CALVIELLO, 2013; PIERARD; MCGREEVY; RONY, 2017).

A vantagem de usar testes comportamentais é que podem ser realizados de forma padronizada (LANSADE et al., 2013; LESIMPLE et al., 2010; KOZAK et al., 2018), sendo

que os testes comportamentais mais utilizados mundialmente são: o teste da ponte (ou teste do manuseio), teste humano passivo, teste humano ativo e o teste de novo objeto (OLSSON, 2008; VISSER et al, 2008).

O teste da ponte tem como desafio a transposição de uma superfície desconhecida e avalia o desempenho do animal guiado por um humano, onde são observados o sucesso e a reação do animal ao transpor a superfície. Na literatura, há relatos da utilização de diferentes materiais para simular o teste da ponte sendo o mais comum colchão (FUREIX et al., 2009; LESIMPLE et al., 2010; HAUSBERGER, 2011), placas de madeira e superfície de concreto (VISSER et al., 2003). Neste teste, o ser humano é limitado a apenas estimular o cavalo com uma leve pressão na guia e, observar se haverá recusa para transpor o objeto, sendo permitido no máximo três tentativas (VISSER et al., 2008) ou limitar tempo máximo de realização do teste, geralmente em 10 min (FUREIX et al., 2009).

### 3.4 COEFICIENTE DE CORRELAÇÃO DE PEARSON

O coeficiente de correlação de Pearson é uma medida de intensidade e direção de uma relação linear entre duas variáveis  $x$  e  $y$  e é obtido pela expressão (JOHNSON et al., 2009):

$$\rho_P = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}},$$

onde  $(x_i, y_i)$  são partes de  $n$  observações de variáveis  $x$  e  $y$  com médias  $\bar{x}$  e  $\bar{y}$ , respectivamente, sendo que o valor numérico de  $\rho$  varia de -1 a +1.

Quando  $\rho$  estiver próximo de -1, indica uma correlação negativa forte e, próximo de +1, uma correlação positiva forte. Um valor de  $\rho$  próximo a 0 (zero) indica que há correlação fraca ou ainda, que não há correlação (JOHNSON et al., 2009).

No caso em que as variáveis  $x$  e  $y$  são descritas por uma distribuição conjunta normal, a correlação nula implica que as variáveis são independentes linearmente, entretanto uma correlação numa (ou muito pequena) pode implicar uma relação não linear entre as variáveis (JOHNSON et al., 2009).

De acordo com Moore et al. (2004), as propriedades do coeficiente e as condições que precisam ser satisfeitas para realizar a análise de correlação de Pearson ( $r$ ) são:

- 1) O coeficiente de correlação de Pearson não diferencia entre variáveis independentes e variáveis dependentes. Dessa forma, o valor da correlação entre X e Y é o mesmo entre Y e X;
- 2) O valor da correlação não muda ao se alterar a unidade de mensuração das variáveis;
- 3) O coeficiente tem um caráter adimensional, ou seja, ele é desprovido de unidade física que o defina;
- 4) A correlação exige que as variáveis sejam quantitativas (contínuas ou discretas);
- 5) Os valores observados precisam estar normalmente distribuídos;
- 6) Faz-se necessária a independência das observações, ou seja, a ocorrência de uma observação  $x_1$  não influencia a ocorrência de uma outra observação  $x_2$ .
- 7) Deve-se observar a presença de outliers que possam afetar fortemente o coeficiente de correlação para a uma determinada amostragem analisada.

## 4. MATERIAL E MÉTODOS

### 4.1 O EXPERIMENTO / BASE DE DADOS

A geração dos dados de frequência cardíaca ocorreu em experimento aprovado pela Comissão de Ética no Uso de Animais (CEUA) da Faculdade de Zootecnia e Engenharia de Alimentos da Universidade de São Paulo sob o número do protocolo 8163230718) e realizado no Regimento de Polícia Montada 9 de Julho, na cidade de São Paulo/SP, no ano de 2015.

Durante a realização do experimento, foram utilizados 21 cavalos de policiamento de rua, das raças Brasileiro de Hipismo (BH), com idade média de  $8,3 \pm 4,3$  anos e Sem Raça Definida (SRD) com idade média de  $11,6 \pm 4,2$  anos, sendo o teste da ponte montada (VISSER et al, 2008) escolhido para coleta dos dados, realizado no picadeiro coberto com dimensões de 20 x 40 m. A simulação da ponte foi realizada utilizando-se quatro placas de EVA de 2 m de largura por 2 m de comprimento, posicionadas duas a duas, formando uma simulação de ponte com 4 metros de largura e de comprimento. A ponte foi posicionada a 10m de distância da porta de entrada do picadeiro e a 16 metros das paredes laterais.

Com os animais posicionados próximos à entrada do picadeiro, o frequencímetro foi posicionado e então ativado. O teste iniciou-se quando o cavalo adentrou no picadeiro e fez contato visual com a ponte, sendo oferecidas três oportunidades para o cavalo transpô-la. Considerou-se que o cavalo se negou a transpor a ponte, quando ele não se aproximou ou quando recuou por 3 ou mais passos. A leitura para coleta de dados do frequencímetro foi então encerrada ao término do percurso ou ao final das tentativas quando não houve sucesso na conclusão.

O modelo de frequencímetro utilizado para realização da aquisição das informações dos animais foi o *Polar Equine H7™*. Este equipamento consiste de uma cinta cardíaca com eletrodos - possibilitam a leitura em tempo real das informações e pode ser acessado por protocolo de comunicação *bluetooth* - que foi ser posicionada no dorso do animal e se comunicam com um smartwatch ou celular, que também são parte do equipamento necessário para o registro das atividades. Posteriormente, os dados coletados pelo frequencímetro são disponibilizados no aplicativo de interação com usuários, o *KER ClockIT® Sport* (<http://www.kerclockit.com/sport/>), para análises mais detalhadas.

## 4.2 DADOS GRÁFICOS

O perfil do gráfico obtido por este aplicativo é do tipo 2D (composto pelos eixos X e Y) e apresenta o layout verificado na figura 1, onde o eixo X representa a dimensão de tempo representada em segundos e que não possui referencial fixo - proposta de realização da atividade previu até três tentativas para transposição da ponte - e o eixo Y possui duas informações representadas por linhas distintas, sendo elas:

- Velocidade: representada nas informações do eixo Y posicionado à esquerda no gráfico de coloração laranja, com marcações identificadas representadas por valores que vão de 0 (zero) a 50 km/h;
- Frequência Cardíaca: representada nas informações do eixo Y posicionado à direita no gráfico de coloração roxa, com marcações identificadas representadas por valores que vão de 0 a 250 bpm e são foco das análises apresentadas neste estudo;

Algumas informações foram disponibilizadas no *ClockIT<sup>®</sup>Sport* adicionalmente, tais como a média e o desvio padrão das medidas dos eixos Y determinados no gráfico de análise das atividades.

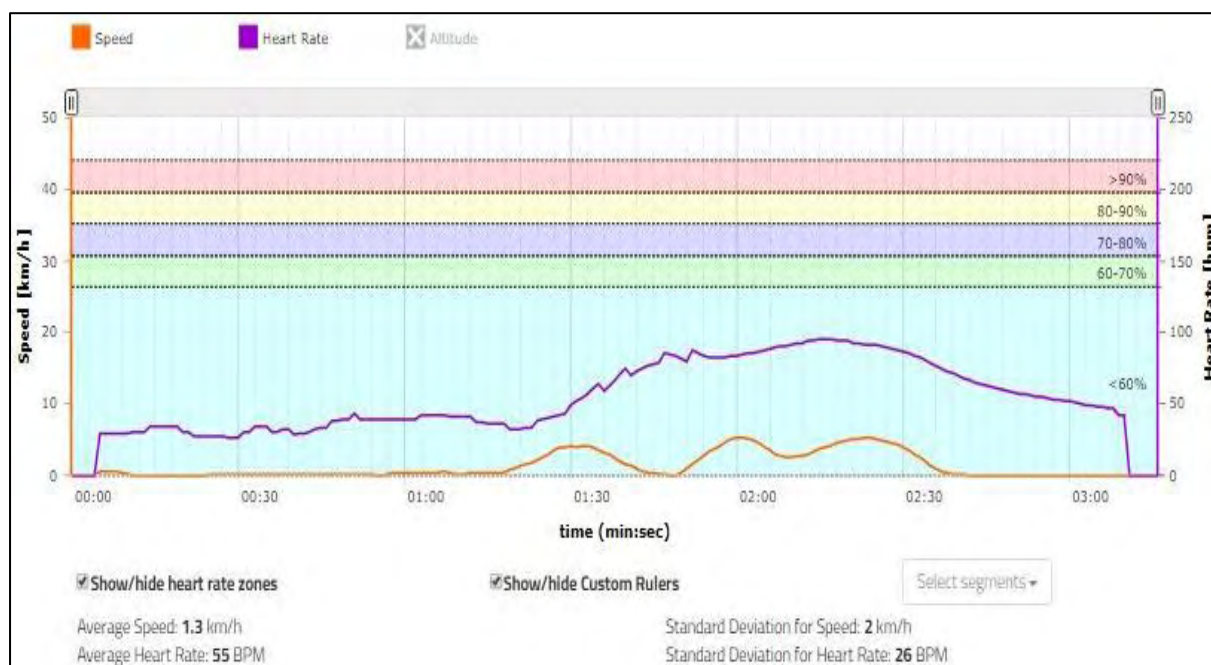


Figura 1 – Exemplo de gráfico e informações adicionais disponibilizadas pelo aplicativo ClockIT<sup>®</sup>Sport. Fonte: Autoria Própria

O *ClockIT<sup>®</sup>Sport* viabilizou os dados coletados durante determinada atividade de forma gráfica - onde as informações de velocidade e frequência cardíaca podem ser verificados por cores e eixos distintos no gráfico (figura 1) – porém, limita a extração dos dados apresentados para análises mais detalhadas, ou seja, não contempla a extração dos dados em sua forma bruta. Adicionalmente, condições ou restrições adicionais poderiam também adicionar ruídos indesejados nas informações coletadas em determinadas atividades equestres, tais como a perda de comunicação entre os dispositivos (cinta e *smartwatch*) ou mal contato entre os eletrodos da cinta e o animal, o que poderia afetar os principais indicadores analisados ao final da atividade, tal como indicado por Marjan (2010) e exemplificado na figura 2. Neste caso, para viabilizar a extração dos dados gerados em sua forma bruta através dos gráficos gerados pelo *KER ClockIT<sup>®</sup>Sport*, foi utilizada a aplicação *WebPlotDigitizer v4.2*.

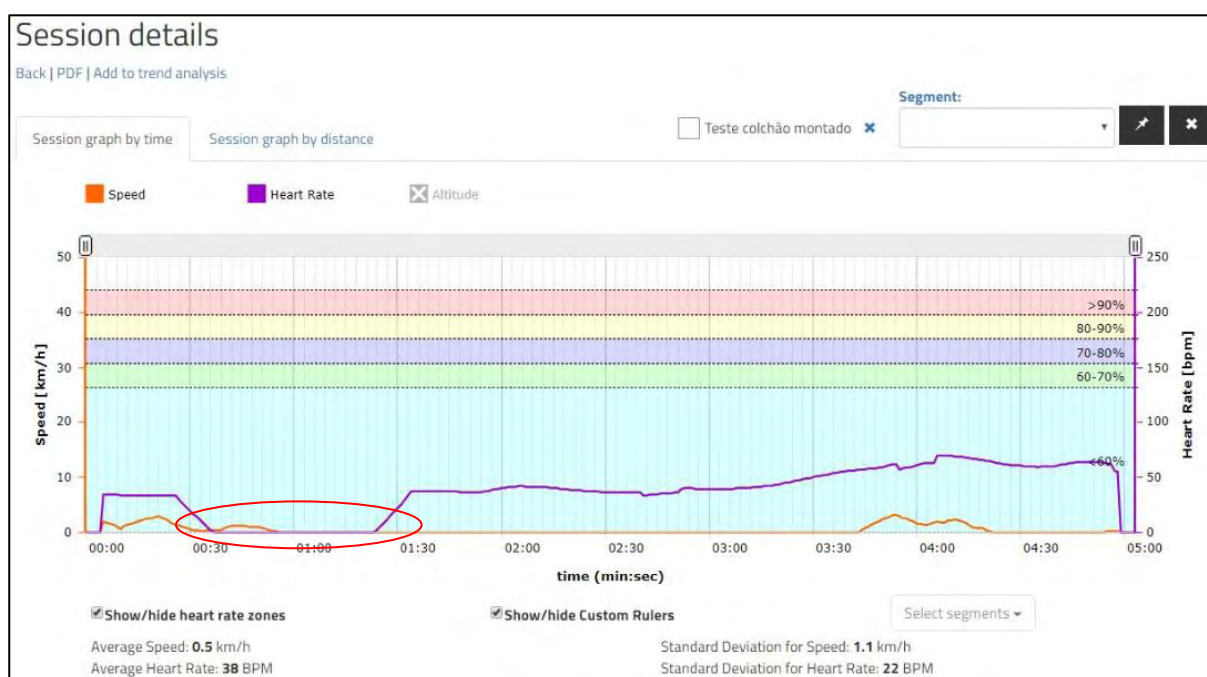


Figura 2 – Exemplo de ruído verificado nas informações coletadas pelo frequencímetro e disponibilizada pelo *ClockIT<sup>®</sup>Sport* – marcações de FC com valor igual a zero entre 00:30 e 01:30 segundos. Fonte: Autoria Própria

#### 4.3 ANÁLISE DE DADOS

O processo de extração deste experimento foi realizado apenas por um ator, ou seja, um indivíduo apenas foi responsável pelas etapas como um todo, desde a extração das informações gráficas de FC disponibilizadas no *KER ClockIT<sup>®</sup>Sport* até a geração do arquivo .CSV, que é o resultado final deste processo disponibilizado no WPD. Foi realizado um



treinamento inicial com duração de 7 horas (1 hora por dia, durante 7 dias consecutivos) para utilização do aplicativo por meio referências indicadas em seu manual (<https://automeris.io/WebPlotDigitizer/userManual.pdf>).

Após a realização do processo de extração dos dados para cada um dos 21 gráficos analisados, foram identificados os valores de FC ao longo do tempo com valor igual à zero (0) que foram gerados a partir de ruídos verificados durante a coleta de informações. Estas informações serviram como critério de classificação para as atividades, sendo elas divididas em dois grupos identificados como “sem ruído” ou “com ruído”. Para efeito de tratativa de informações, as atividades em que foram verificadas perdas de sinal na transmissão de dados de FC (valores de FC iguais a zero) tiveram estes valores identificados e então excluídos da análise, de forma a identificar o nível de impacto destes dados indevidos na comparação dos indicadores do grupo “com ruído” (identificado na tabela 2 na coluna “Ruído” com “S”). Tal classificação permitiu que os dados que não apresentaram ruído (classificados no grupo “sem ruído” - tabela 2 na coluna “Ruído” com “N”) pudessem ser utilizados de forma a confirmar a acuracidade do método de extração proposto.

A partir daí, os dados extraídos de cada atividade tiveram sua média e desvio padrão determinados pelo Excel®, por meio da aplicação das fórmulas “MEDIA” e “DESVPAD.A” para cada conjunto de dados extraídos, de forma a consolidar estes indicadores e realizar a comparação de valores entre os mesmos e as informações fornecidas diretamente pelo ClockIT®Sport.

Como parte do processo de verificação de correlação, foi realizado o teste de outlier para 3 conjuntos de dados amostrais com nível de significância de 5%, tanto para os dados de média, quanto de desvio padrão, considerando-se inicialmente a amostragem total de indivíduos (N=21) e também para os dois grupos subclassificados a partir da amostragem total (N=14/N=7).

Estas informações, por sua vez, foram então comparadas estatisticamente na aplicação *Minitab* (versão 19), por meio da aplicação do teste de correlação de Pearson com valor de  $P \geq 0,95$ .

## 5. RESULTADOS E DISCUSSÃO

### 5.1 SELEÇÃO DE SOFTWARE

O processo de extração de dados gráficos pode ser bastante desafiador e uma tarefa que demanda tempo considerável. Conforme resultados do estudo realizado por Boyle et al. (2013), existe uma relação direta entre o nível de aprendizado/expertise de um determinado ator que venha a realizar esta atividade e a acurácia verificada nos dados extraídos. Além disso, estes autores também mencionam a possibilidade de uso de rotinas para o processo de extração (por exemplo, roteiro de atividades a serem realizadas) ou de gestão dos dados (por exemplo, na aplicação de modelos de arquivos pré-formatados e predefinição de variáveis analisadas) como possibilidades de redução de tempo e melhoria da acurácia dos dados brutos extraídos, principalmente quando mais de um ator deve realizar a mesma análise para uma fonte de informação a ser verificada.

Neste sentido, este estudo buscou ferramentas auxiliares para suportar este processo, de forma a analisar a que melhor se enquadrava para que o mesmo pudesse ser realizado de forma rápida e precisa, sendo considerados critérios técnicos e funcionais a partir da realização do processo de extração para avaliação das aplicações selecionadas.

O processo para escolha das aplicações analisadas foi realizado através de pesquisas em plataformas de busca online durante 15 dias, de 01/06/2019 a 15/06/2019, sendo utilizados o *Google*<sup>™</sup> e *Bing*<sup>™</sup> como motores de busca, através do termo “*graphical data extraction tool*” ou “*graph data extraction tool*”, onde as aplicações “PlotDigitizer.com”, “WebPlotDigitizer” e “Graphreader.com” foram as melhores ranqueadas pelos resultados nas mesmas, sendo apresentadas nos primeiros resultados de pesquisas em ambas as plataformas. Adicionalmente, as aplicações em questão possuem datas recentes de atualizações de versão (últimas atualizações realizadas de 2019 em diante) e não existem estudos anteriores da literatura que fizeram comparações diretas entre as ferramentas selecionadas.

Dentre os critérios técnicos, a possibilidade de utilização de uma aplicação gratuita em detrimento à necessidade de aquisição de licença ou registro de software, além da viabilidade de utilização de hospedagem como “*client server*” em detrimento ao uso exclusivo do tipo “*web based*”, foram considerados com maior peso nesta avaliação. Tal decisão considerou a possibilidade de aplicação do método proposto sem necessidade de implicação de custos para viabilização do mesmo, além de proporcionar também a execução do mesmo em áreas remotas ou que possuam impossibilidade de uso de internet. Neste ponto, as ferramentas em questão

foram comparadas em termos de custo: WebPlotDigitizer e Graphreader.com são ferramentas gratuitas, enquanto o PlotDigitizer.com custa US\$49,00 para uma licença anual, com possibilidade de reduções de valores para cenários alternativos de uso prolongado (dois anos com valor de US\$69,00 e cinco anos com valor de US\$89,00). Para efeito de comparação sobre o tipo de hospedagem, o WebPlotDigitizer e o PlotDigitizer.com são aplicações que oferecem tanto a utilização “*web based (online)*”, quanto “*client server (offline)*”, enquanto o Graphreader.com oferece apenas a possibilidade de uso online.

Adicionalmente, o formato do arquivo de uma possível imagem a ser analisada pela aplicação escolhida também foi considerado como critério técnico, onde avaliou-se a possibilidade em se trabalhar com o formato de imagem “.JPG”, utilizado nas imagens analisadas neste estudo. Apesar de haver várias soluções possíveis atualmente para conversão de formatos de arquivos relacionados a imagens, sendo a mais simples dela a possibilidade de se “salvar como” um novo formato através de softwares de edição de imagens, esta atividade poderia demandar tempo adicional ao método proposto, sendo indesejado por este motivo, principalmente em cenários onde grandes quantidades de informações gráficas devem ser analisadas e/ou terem seus dados gráficos extraídos para a forma bruta. Neste quesito, todas as aplicações analisadas tiveram avaliação semelhante, uma vez que todas permitem a utilização do formato desejado para efeito de aplicação do processo de extração de dados gráficos.

Dentre os critérios funcionais, as análises de requisitos consideraram como critério apenas a possibilidade de uso como “sim” ou “não”. Moeyear et al. (2016) aplicaram em sua comparação de aplicações de análises gráfica o método de SUS (*System Usability Scale*), que prevê a avaliação deste critério através da aplicação de um questionário com 10 perguntas com respostas baseadas numa escala tipo Likert, indicando a força de cada enunciado por questão, que incluem tanto enunciados com declarações positivas e negativas, que por sua vez compõem um resultado final dentro de um range de 0 a 100, onde os maiores escores indicam melhor usabilidade. Tal método não foi aplicado devido ao fato do estudo em questão ter se utilizado apenas de um ator executando o processo de extração de dados gráficos durante a escolha da aplicação para este processo.

Ainda nos critérios funcionais, a possibilidade da realização do método de extração para o tipo de gráfico bidimensional (2D) foi considerado mandatório, uma vez que as análises gráficas deste estudo possuem este caráter. Sendo assim, a possibilidade em se trabalhar com outros tipos de gráficos além dos gráficos bidimensionais, não foi levada em consideração nesta avaliação. Além disso, outros quatro critérios foram considerados como desejáveis, sendo eles

“interface e usabilidade”, que avaliou a realização do processo de extração com sucesso nas primeiras 3 tentativas de uso sem referência ao manual da aplicação avaliada; “funcionalidade de autotrace”, que avaliou a automação do processo de extração sem a necessidade de intervenção manual; “painel de zoom”, que é uma funcionalidade que auxilia na ampliação da figura analisada para a correta determinação das coordenadas de referência X1, X2, Y1, Y2; e “funcionalidade para exportação dos dados”, que possibilita a análise posterior dos dados extraídos por outras ferramentas, de caráter estatístico e/ou analítico. Para as aplicações PlotDigitizer.com e WebPlotDigitizer, todos os quatro critérios funcionais são plenamente atendidos, enquanto que para a aplicação Graphreader.com, não existe disponibilizadas a funcionalidade de Autotrace e as três tentativas iniciais de uso da ferramenta para extração de dados no critério de “Interface e Usabilidade” não foram suficientes para que uma tentativa de extração de dados inicial e autônoma fosse realizada com sucesso.

O quadro a seguir lista os principais critérios avaliados, comparando as aplicações selecionadas:

CRITÉRIOS TÉCNICO / FUNCIONAIS	APLICAÇÕES AVALIADAS		
	PlotDigitizer.com	WebPlotDigitizer	Graphreader.com
Versão Online e Offline ( <i>web based</i> ou <i>client server</i> )	Sim	Sim	Não (somente online)
Software Gratuito (Freeware)	Não	Sim	Sim
Formatos de Imagem Suportados	PNG, JPG, JPEG, GIF, TIFF, BMP, SVG	JPEG, PNG, TIFF	JPEG, PNG, TIFF
Tipo de Gráfico 2D (bidirecional) suportado?	Sim (outros formatos também são suportados)	Sim (outros formatos também são suportados)	Sim (apenas gráficos do tipo XY)
Interface e Usabilidade	Sim	Sim	Não
Funcionalidade de Autotrace	Sim	Sim	Não (método manual)
Painel de Zoom (melhora da acuracidade dos pontos identificados)	Sim	Sim	Não
Funcionalidade para exportação dos dados	Sim	Sim	Sim

Quadro 1 – Comparação de aplicações de suporte para extração de dados gráficos. Fonte: Autoria Própria

## 5.2 PROCESSO DE EXTRAÇÃO

Conforme verificado nos resultados apresentados no processo de seleção de software, a partir dos critérios escolhidos e características avaliadas, o software que melhor respondeu aos requerimentos propostos neste estudo foi o WebPlotDigitizer, uma vez que os critérios funcionais são plenamente atendidos, além desta ferramenta possuir tanto o caráter de utilização gratuita (*freeware*), como também disponibilizar versão de utilização *online* e *offline*, fundamental para análises em campo, onde podem haver dificuldades no acesso à internet.

O estudo em questão propôs a combinação entre a captura do gráfico contendo as informações de FC disponibilizada pelo frequencímetro, através da aplicação *KER ClockIT<sup>®</sup>Sport* e o processamento digital de imagens utilizando o aplicativo *WebPlotDigitizer* (*WPD*), que é uma aplicação open-source (<https://automeris.io/WebPlotDigitizer/download.html>), disponibilizada para uso online ou cliente server (executável em plataformas Windows, MacOS e Linux) e é capaz de extrair dados numéricos de uma variedade de gráficos (2D, barras, dispersão) através de arquivos carregados em formatos JPEG ou PNG. Esta aplicação é bastante intuitiva, user friendly, precisa e fácil de usar. Atualmente, está disponibilizada nas seguintes opções de idioma: inglês, francês, alemão e chinês (versão simplificada) e para efeito de realização deste trabalho, utilizou-se a opção do aplicativo em inglês. Sua funcionalidade de extração automática de dados conseguiu mapear e disponibilizar diversos pontos de leitura através do mapeamento de pixels da(s) figura(s)/gráfico(s) analisado(s), além de permitir que ajustes para adição, alteração ou exclusão de pontos lidos fossem realizadas, caso necessários.

Descreve-se a seguir o processo de extração automática utilizando-se o WPD. Primeiramente, selecionou-se a imagem contendo o gráfico extraído do *ClockIT<sup>®</sup>Sport* e carregou-se a mesma através da opção '*File*' contida no menu, comando '*Load image*' (carregar imagem). Depois, selecionou-se a opção *2D (X – Y) Plot* (gráfico tipo XY) no menu '*Choose plot type*' para que a calibragem dos eixos do gráfico pudesse ser realizada. Esta ação gerou um novo evento para que os valores de referência de quatro pontos conhecidos X1, X2, Y1 e Y2 fossem identificados no gráfico carregado, de forma que as derivações das demais medidas pudessem ser realizadas pelos algoritmos de extração de dados do programa a partir destas informações. A sequência destes eventos principais para o upload da imagem necessário ao processo de extração automático é demonstrada na figura 3.

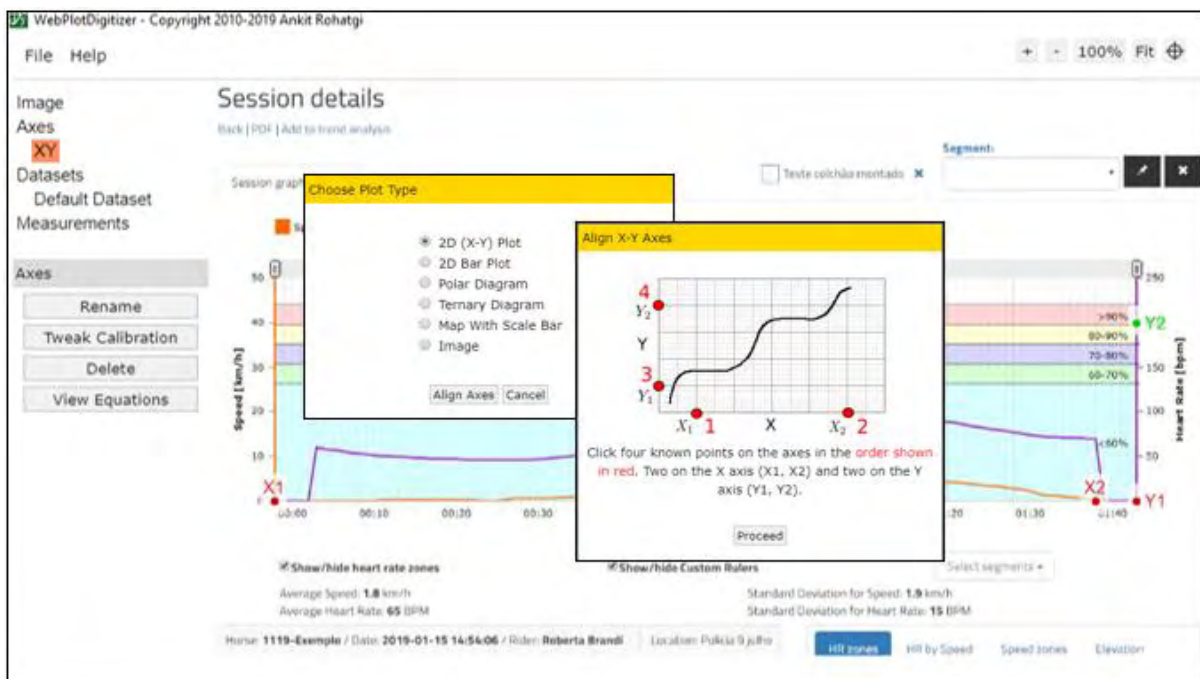


Figura 3 – Etapas iniciais do processo de extração automática de informações utilizando o WebPlotDigitizer. Fonte: Autoria Própria

Para cada imagem analisada, foi escolhida a máxima distância entre os pontos conhecidos de forma a minimizar o erro acumulado pela entrada manual de informações, já que as coordenadas em questão foram utilizadas para que o algoritmo pudesse fazer o mapeamento dos pixels da imagem carregada versus as informações selecionadas para os valores de referência. Outro detalhe importante a ser considerado foi que o eixo X possui unidade de medida referência em segundos, portanto as entradas de informações para estes pontos de referência utilizaram o formato MM:SS (minutos: segundos). A referência demonstrada na figura 4 indica a localização dos pontos inseridos numa amostra gráfica de dados do experimento, onde podem ser verificados os referenciais  $X_1=00:00$ ,  $X_2=01:40$  e  $Y_1=0$ ,  $Y_2=200$ .

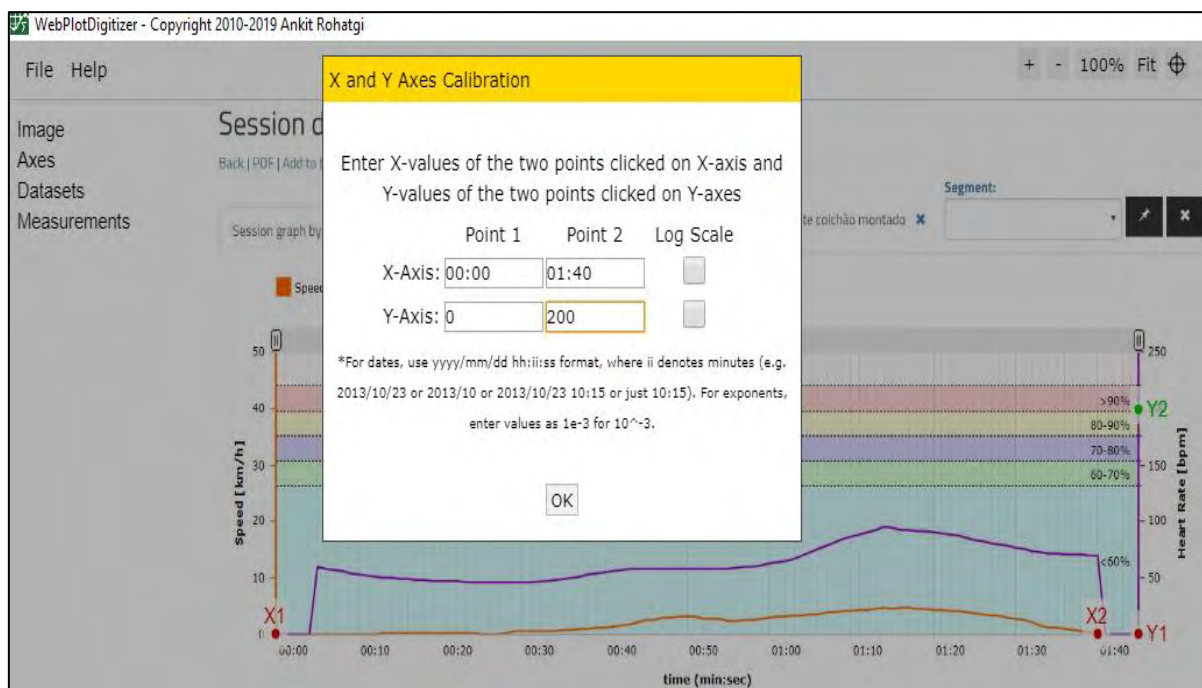


Figura 4 – Calibragem dos eixos X e Y para execução do algoritmo de extração de dados.  
Fonte: Autoria Própria

Pressionando-se o botão 'OK' determinou-se que o processo de calibração dos eixos XY foi completado e passamos ao próximo estágio, o processo de coleta de dados. Nesta etapa, iniciou-se o processo de extração automática, onde foi necessário que parâmetros adicionais fossem indicados para que algoritmo de extração de dados pudesse ser configurado de acordo com a análise desejada. O método automático de extração demanda uma diferenciação clara de cores entre as cores de fundo (por exemplo, cores indicadas para cada faixa percentual de frequência cardíaca indicada) e as cores dos gráficos desejados (neste caso, a cor roxa que representa a variação da frequência cardíaca durante a realização da atividade). Assim sendo, selecionou-se a opção 'Foreground Color' na aba 'Color' e em seguida clicou-se em na seleção de cor que se encontra à direita desta opção, sendo que automaticamente uma nova caixa de mensagem 'Specify Plot (Foreground) Color' se abriu, clicou-se no botão 'Color Picker' e então clicou-se em algum ponto em cima do gráfico a ser analisado para que a cor roxa pudesse ser selecionada (por exemplo, valores RGB indicados como R:129, G:27 e B: 175) – figura 5.

As funcionalidades 'Box', 'Pen', 'Erase' e 'View' disponibilizadas na sessão 'Mask' puderam ser utilizadas para se delimitar uma determinada região de interesse para extração dos dados no gráfico analisado. A delimitação foi realizada utilizando-se a função 'Pen', onde a





Figura 5 – Delimitação de atributos para execução do algoritmo de extração de dados. Fonte: Autoria Própria

região do gráfico pôde ser devidamente identificada através de um desenho à mão livre (identificada pela cor amarela), tentando seguir ao máximo o desenho da linha do gráfico desejado para análise (por exemplo, sobrepondo a curva de FC) – figura 6. Pressionando-se o botão ‘View’ foi possível visualizar a área delimitada por esta funcionalidade. Caso a identificação da área de delimitação não ocorresse conforme esperado, o botão ‘Erase’ permitiria que áreas indesejadas pudessem ser desmarcadas e novamente o processo de delimitação ocorresse utilizando-se a função ‘Pen’. O botão ‘Width’ permitiu que, tanto para a função ‘Pen’ ou para ‘Erase’, fosse possível ajustar a espessura do traço resultante da utilização destes comandos como um ajuste fino na seleção quando necessário.

Adicionalmente, foi utilizada o valor igual a 120 na informação ‘Distance’ como parâmetro inicial para a primeira execução, utilizada para ajustar o range de cores identificadas próximas aos valores RGB indicados, de forma a refinar os pontos encontrados pelo algoritmo. O resultado deste processo pode ser visualizado na figura 6. Por último, o algoritmo de extração utilizado pôde ser selecionado no campo ‘Algorithm’ e foi utilizado o valor ‘Averaging Window’ com valores fixos iniciais para leituras de pixels ( $\Delta X = 10$  e  $\Delta Y = 10$ ). Como ação final, pressionou-se o botão ‘Run’ de forma a finalizar o processo de extração automática.



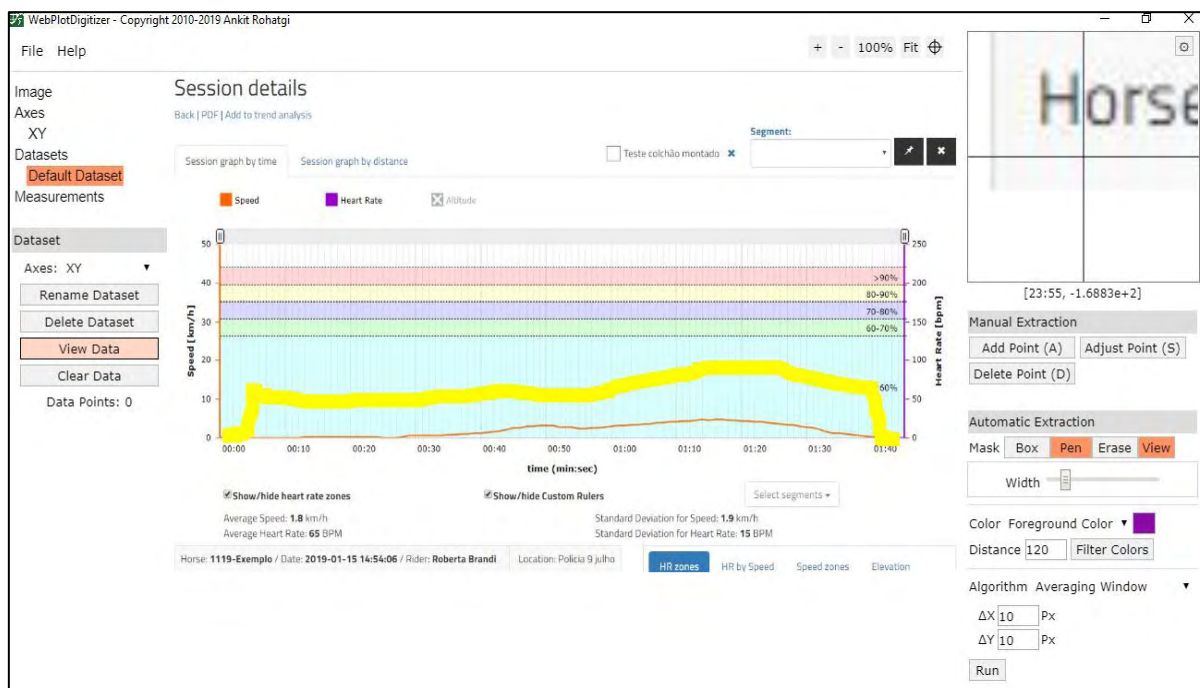


Figura 6 – Delimitação de área de interesse para execução do algoritmo de extração de dados.  
Fonte: Autoria Própria

Vale comentar que, caso o resultado na identificação dos pontos não tivesse retornado a quantidade e/ou localização esperada dos pontos no gráfico, os parâmetros ‘Distance’, ‘ $\Delta X$  e  $\Delta Y$ ’ puderam ser alterados de forma a melhorar a disposição visual dos pontos encontrados durante a execução do algoritmo. No experimento em questão, como as atividades resultaram em tempos totais de análise distintos, a quantidade de pontos desejada variou diretamente em função desta informação, onde os gráficos com mais tempo de informação para análise precisaram de menores valores de ‘ $\Delta X$  e  $\Delta Y$ ’, por exemplo. Neste momento, estes parâmetros foram alterados para  $\Delta X = 5$  e  $\Delta Y = 5$  após uma primeira execução com os dados padrão. Após esta segunda execução, os dados selecionados puderam ser então visualizados (indicados pela coloração vermelha) na imagem carregada, conforme pode ser visualizado na figura 7.

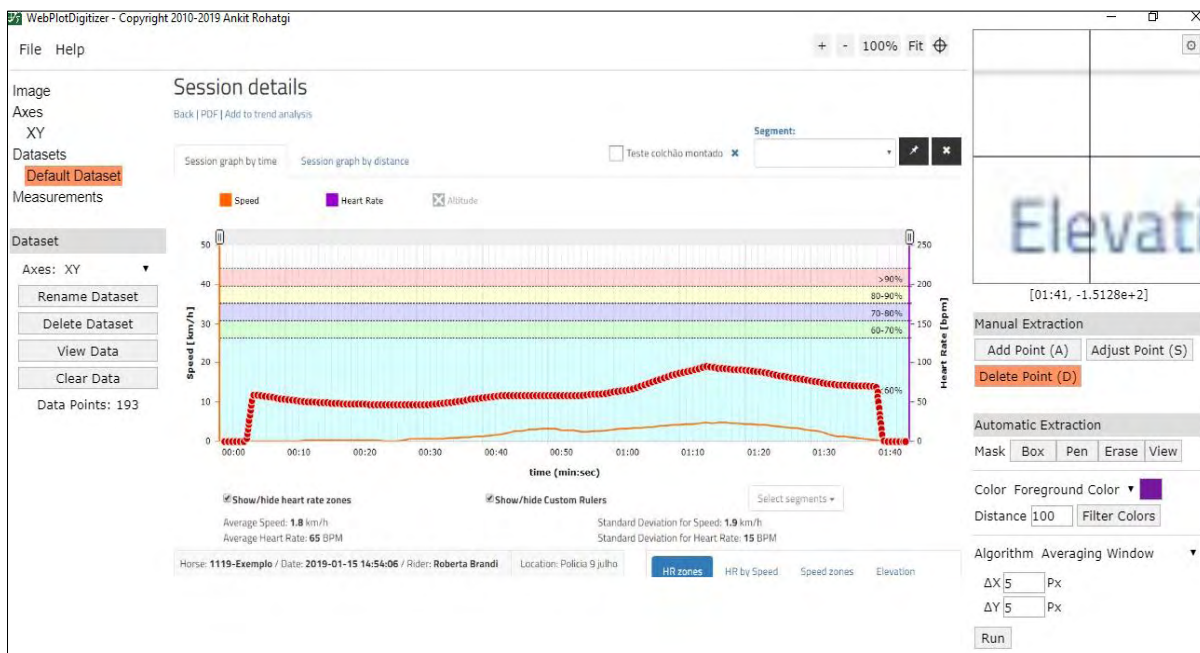


Figura 7 – Identificação dos pontos mapeados pelo algoritmo para extração de dados em sua forma bruta. Fonte: Autoria Própria

O botão 'View Data' localizado no menu 'Dataset' então foi utilizado para visualizar os dados digitalizados, conforme ilustrado na figura 8, podendo ser rearranjados por funcionalidades adicionais 'Sort' e 'Format', caso necessário. No exemplo estudado, utilizou-se as opções 'Sort by: X' e 'Order: Ascending' para que os dados disponibilizados acompanhassem a linha do tempo da atividade realizada.

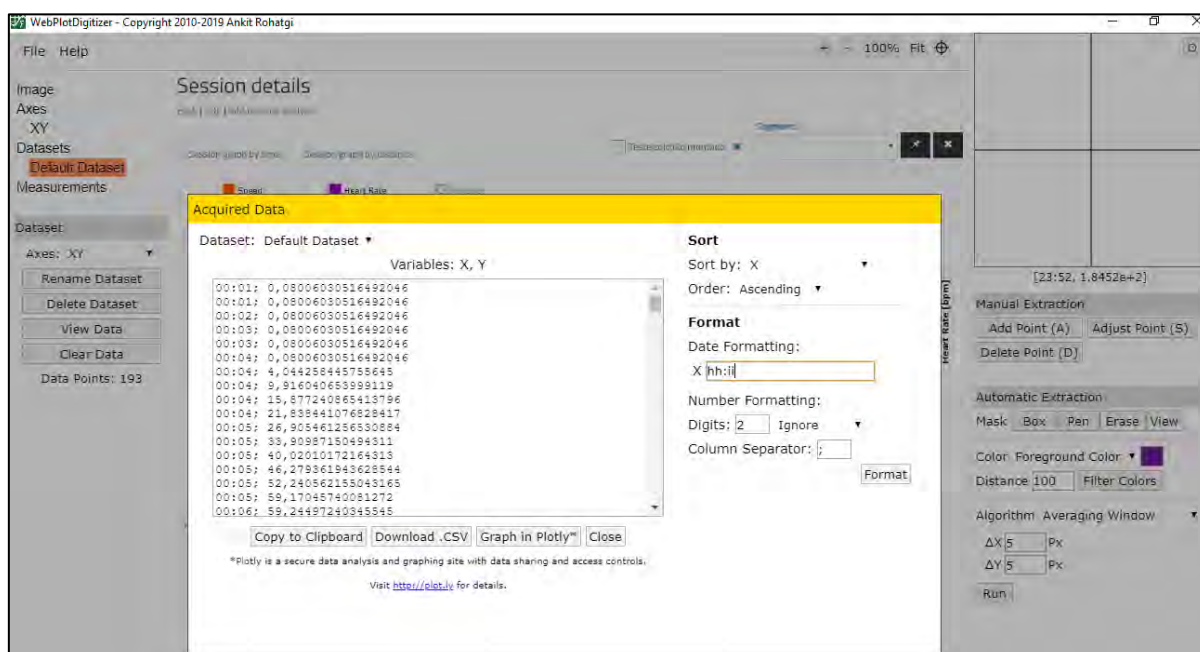


Figura 8 – Visualização, opções de formatação e funcionalidades para extração dos pontos mapeados como dados em sua forma bruta. Fonte: Autoria Própria

Para encerrar o processo de extração automática dos dados, utilizamos a funcionalidade 'Download .CSV' para análise posterior utilizando softwares com funções estatísticas como o *Excel* ou *Minitab*. Os dados inseridos nesta simulação puderam também ser gravados como um projeto, que mantém todas as informações utilizadas ao longo do processo para uso posterior no software (utilizando o menu 'File' e opção 'Load Project').

Destaca-se a versatilidade desta aplicação para eventuais ajustes manuais que se fizeram necessários (por exemplo, pontos identificados incorretamente pelos parâmetros utilizados no programa), onde as funções da sessão *Manual Extraction* nomeadas como Add Point (A), Adjust Point (S) e Delete Point (D) puderam ser utilizadas para, respectivamente, adicionar, alterar ou remover um ponto na área de interesse analisada. Neste ponto, recomenda-se que informações ainda mais detalhadas sejam consultadas no manual da aplicação (<https://automeris.io/WebPlotDigitizer/userManual.pdf>), caso necessário maior entendimento acerca dos algoritmos de extração ou funções adicionais não exploradas neste estudo.

Adicionalmente, um fluxograma contendo as principais etapas do processo descrito pode ser encontrado na figura 9:

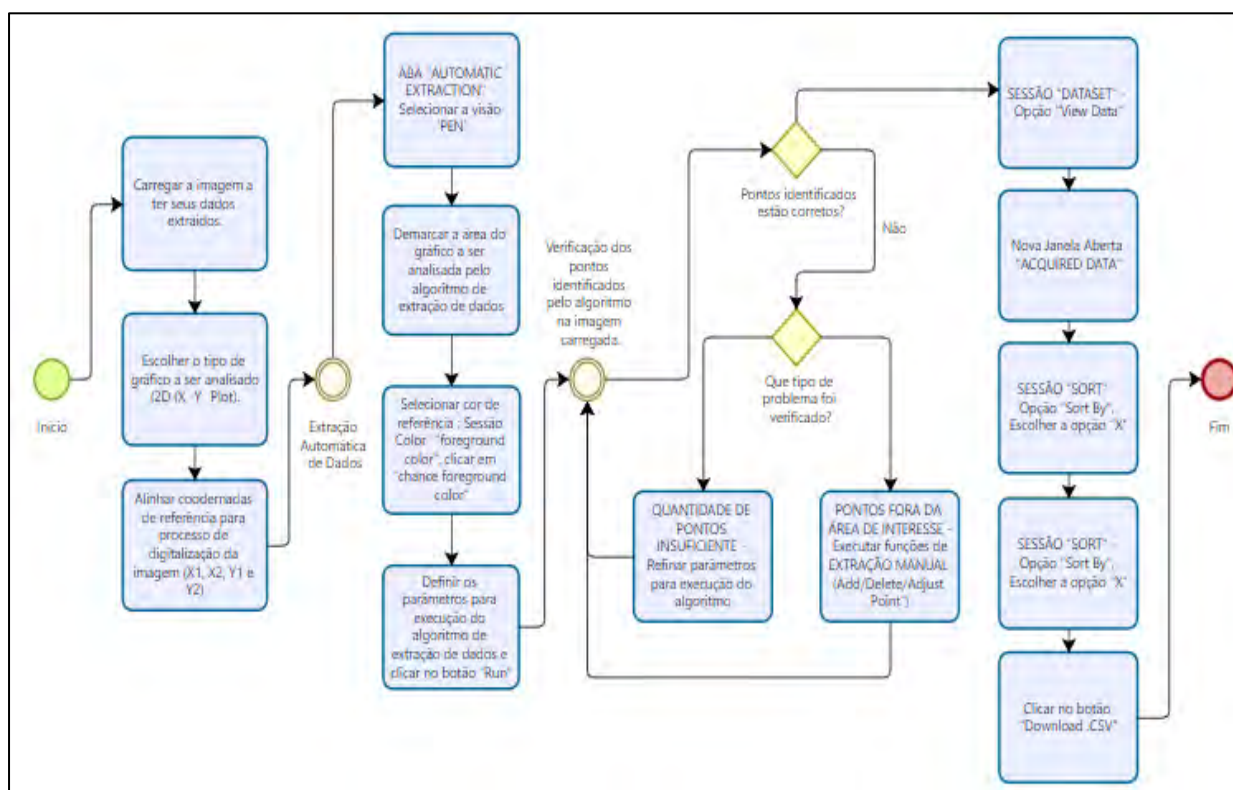


Figura 9 – Fluxograma dos principais processos vinculados à extração automática de dados gráficos através do WebPlotDigitizer. Fonte: Autoria Própria

Com relação ao tempo empregado para cada extração, o estudo em questão apresentou um tempo médio de cerca de 5 minutos para cada processo realizado. Moeyear et al. (2016) relataram em seus resultados que o tempo médio para realização de um processo de extração com o WebPlotDigitizer foi de 15.24 minutos, consideravelmente maior (aproximadamente 3 vezes), porém não existe menção clara de quais foram os principais desafios ou justificativas dos atores participantes deste estudo para esclarecer os motivos pelos quais estes valores apresentam-se mais elevados quando comparados. Uma possibilidade é que a utilização de vários atores para a realização do experimento mencionado possa ter trazido diferentes níveis de experiência ao processo, o que determinou um maior tempo médio para realização, em contrapartida à utilização de um único ator empregado neste estudo.

### 5.3 APLICAÇÃO PRÁTICA DO MÉTODO

A aplicação prática da metodologia foi realizada para os dados contidos na base avaliada de cada um dos 21 cavalos do experimento e os indicadores de média e desvio padrão para os dados aferidos pelo frequencímetro e extraídos pelo método proposto podem ser visualizados nos dados apresentados a seguir:

ID	Ruído (S/N)	KER		WPD	
		Média (bpm)	Dev Padrão (bpm)	Média (bpm)	Dev Padrão (bpm)
1	N	55	26	56	23
2	N	108	51	109	50
3	N	63	39	62	33
4	S	43	65	119	44
5	N	137	59	139	58
6	N	79	30	80	30
7	S	33	28	49	18
8	N	65	15	64	17
9	N	52	19	52	18
10	S	58	69	97	53
11	N	61	21	60	21
12	N	78	36	77	36
13	N	57	22	57	22
14	S	17	36	49	37
15	N	74	12	74	13
16	S	38	22	44	16
17	S	56	36	61	29
18	N	79	52	78	50

19	N	89	35	89	35
20	N	54	22	54	22
21	S	43	40	60	29

Quadro 2 – Consolidação dos dados referentes aos indicadores de média e desvio padrão comparados entre aplicações para validação do método proposto. Fonte: Autoria Própria

As análises gráficas que verificaram ruído ao longo do experimento foram identificadas com a letra “S” na segunda coluna. Pode-se observar que, dentre as 21 atividades realizadas, apenas 7 delas apresentaram-se com ruído. Diferentemente do que pôde ser observado para estes valores para as atividades que não apresentaram ruído, verificamos que estas informações estão com ligeira divergência de valores quando comparamos os indicadores de média e desvio padrão do método de extração e informações fornecidas diretamente pelo *KER ClockIT<sup>®</sup>Sport*, que podem ser explicados pelo impacto que os valores iguais a zero (0) determinam nestas informações para cada atividade considerada.

Durante a verificação de outliers nos dados amostrais considerados, tanto a primeira análise para N=21, quanto a análise para o grupo “sem ruído” de N=14, apontaram o item de ID “5” (cavalo Marimbondo) como outlier para o conjunto de dados de média, sendo, portanto, retirado das análises estatísticas posteriores.

Os resultados estatísticos de verificação de correlação são apresentados a seguir, considerando um intervalo de confiança de 95% para a correlação de Pearson, tanto para as informações referentes ao total da amostragem, bem como para a divisão dos grupos em “sem ruído” e “com ruído”:

Amostragem	Análise Médias		Análise Desvios Padrão	
	Correlação (r)	IC	Correlação (r)	IC
Total (N=20)	0,561	(0,561; 0,804)	0,929	(0,827; 0,972)
Atividades “sem ruído” (N=13)	0,999	(0,996; 1,000)	0,989	(0,962; 0,997)
Atividades “com ruído” (N=7)	0,479	(-0,429; 0,905)	0,941	(0,642; 0,991)

Quadro 3 - Comparativo de resultados de análise de correlação de Pearson para os indicadores de média e desvio padrão entre dados fornecidos e extraídos. Fonte: Autoria Própria

A seguir, verificou-se os ajustes de regressão linear para os gráficos de dispersão gerados para cada grupo, primeira para os dados referentes à análise de médias:

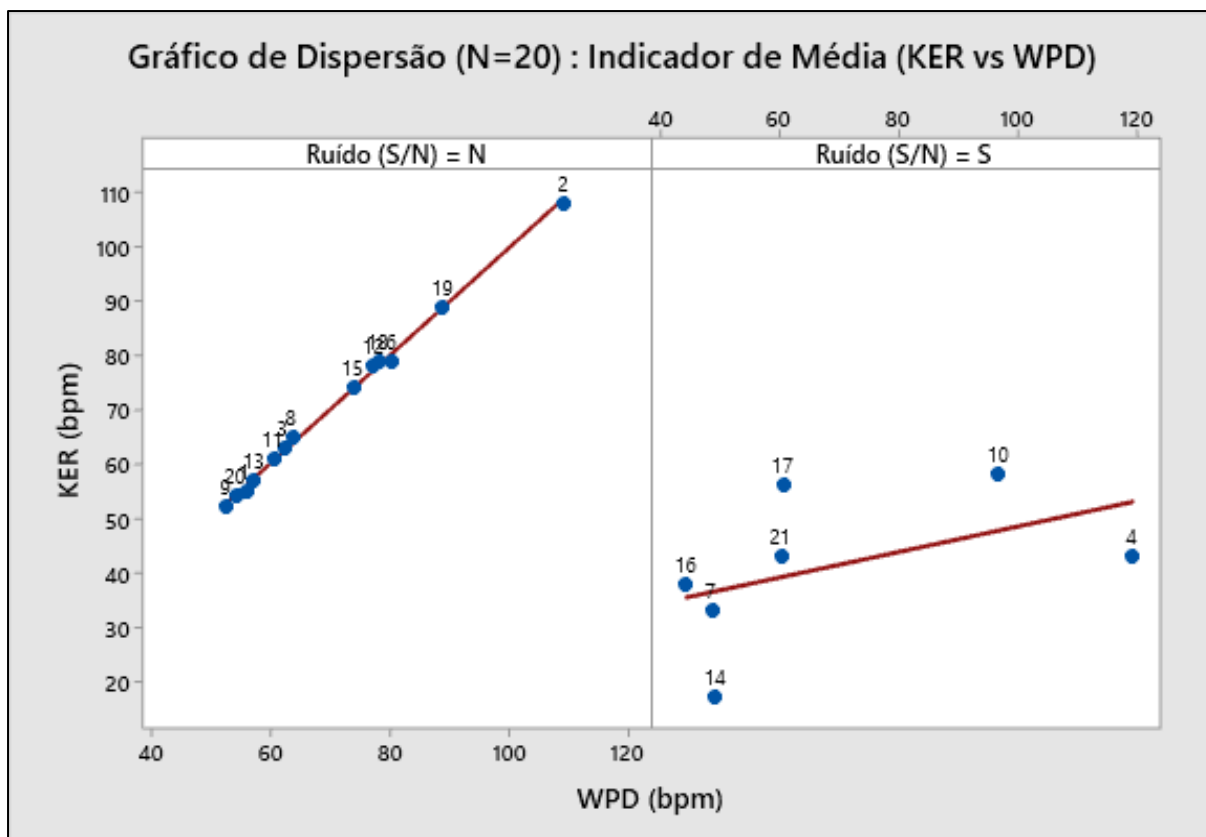


Figura 10 - Dispersão com ajuste de regressão linear para as informações de média, considerando o total de dados de amostragem do estudo. Gráfico gerado à esquerda contempla os registros que não tiveram ruído verificado (N=13), enquanto o gráfico à direita representa os registros que tiveram ruído verificado (N=7). Fonte: Autoria Própria

A análise de regressão realizada por meio do gráfico de dispersão com grupos definidos confirma graficamente a forte correlação indicada pelos dados que não apresentaram ruído, porém a baixa inclinação verificada para os dados referentes aos dados com ruído reflete a fraca correlação entre os dados de média analisados.

Da mesma maneira, a análise de regressão realizada por meio do gráfico de dispersão com grupos definidos confirma graficamente a forte correlação indicada pelos dados de desvio padrão analisados para os dois grupos, conforme demonstrado na figura 11.

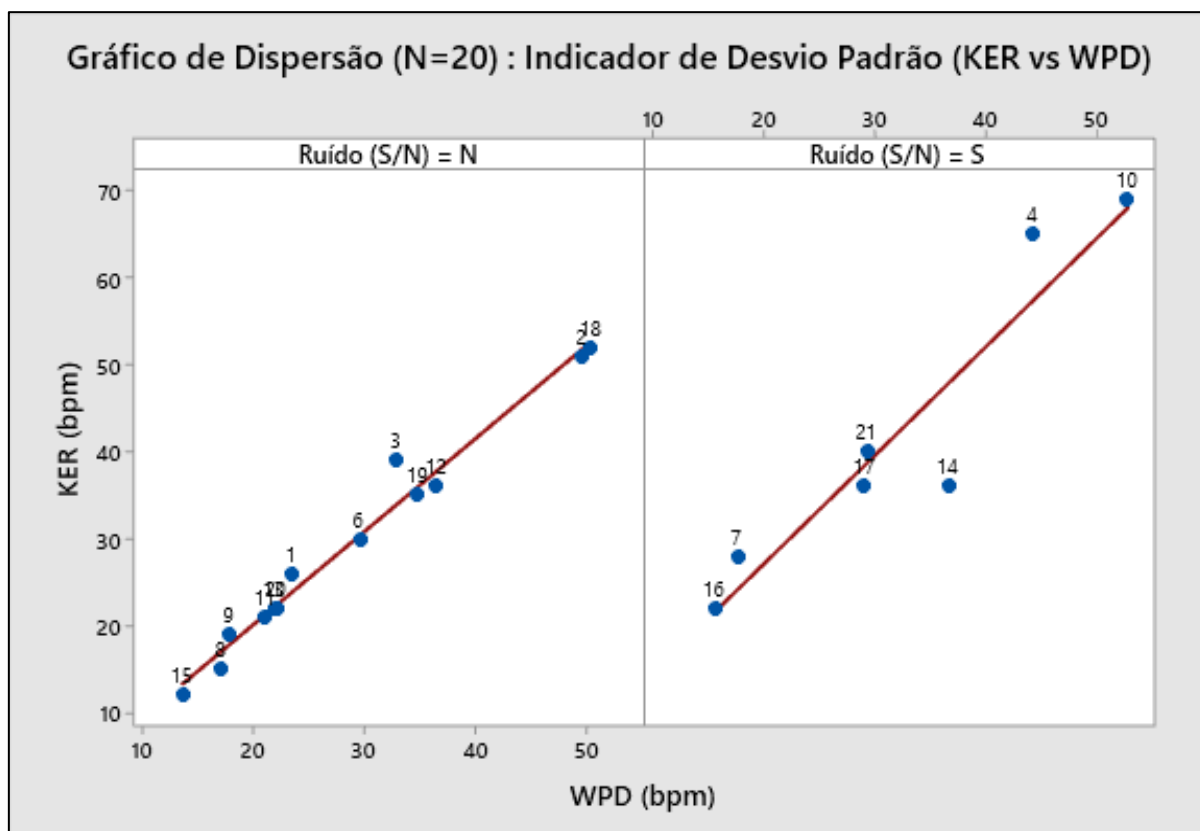


Figura 11 - Dispersão com ajuste de regressão linear para as informações de desvio padrão, considerando o total de dados de amostragem do estudo. Gráfico gerado à esquerda contempla os registros que não tiveram ruído verificado (N=13), enquanto o gráfico à direita representa os registros que tiveram ruído verificado (N=7). Fonte: Autoria Própria

Para as amostras que apresentaram ruído, verificou-se que a exclusão de valores referentes à FCs iguais a zero pode trazer alterações consideráveis em estudos que utilizam os indicadores de média e desvio padrão fornecidos diretamente pela aplicação *KER ClockIT® Sport*. A baixa correlação verificada neste cenário para os dados de média indica que dados trabalhados diretamente via aplicação do frequencímetro podem apresentar um impacto considerável em análises que não consideram a premissa de FC iguais a zero não serem possíveis de serem aferidas ao longo de uma atividade equestre. Ainda, quanto maior a quantidade de informações “com ruído”, maior o desvio verificado na correlação dos indicadores analisados.

Adicionalmente, este estudo contou com uma amostragem de N=21, o que possibilitou que o método aplicado para a extração gráfica pudesse ser gerado caso a caso de forma manual, sem a necessidade de uma função de automatização deste processo. Para um N consideravelmente maior, sugere-se que exista algum processo de automatização do método proposto, conforme descrito no estudo de Chan et al. (2020). Neste caso, é importante que as

atividades analisadas possuam um padrão na aferição/protocolo de realização, já que isto determinaria a possibilidade de simplificação no algoritmo de automatização das informações extraídas pelo método proposto. Esta observação vem de encontro à realização do experimento em questão possibilitar que o teste da ponte montada fosse realizado em até 3 tentativas, sendo que a determinação das informações totais de análise no eixo X (eixo da linha do tempo) não fossem padronizadas entre todas as atividades e, portanto, não seja possível determinar valores de referência idênticos para cada atividade realizada.



## 6. CONCLUSÃO

É possível a utilização do programa WebPlotDigitizer<sup>®</sup> como alternativa para a extração de informações gráficas, além de viabilizar a possibilidade de eliminação de dados considerados como “com ruído” através de análises de dados pormenorizados.

## REFERÊNCIAS

- ARIZONA SOFTWARE INC. **GraphClick 3.0.2**, 2010. Disponível em <<https://graphclick.en.softonic.com/mac>> e acesso realizado em 07/10/2021.
- BIOSOFT, **UnGraph for Windows (Version 5.0)**. Cambridge, UK, 2004.
- BOYLE, M.; SAMAHA, A. S.; RODEWALD, A.; HOFFMANN, A. N. **Evaluation of the reliability and validity of GraphClick as a data extraction program**. Computers in Human Behavior, 29, 1023-1027, 2013.
- BONOMO, C. C. M.; MICHIMA, L. E. S.; MIYASHIRO, P; FERNANDES W. R. **Avaliação ecocardiográfica do desenvolvimento cardíaco de cavalos atletas: comparação entre atividades físicas distintas**. Pesquisa Veterinária Brasileira. Rio de Janeiro, v.34, n.9., 2014.
- BORMANN, I. **DigitizeIt (version 2.0)**, 2012. Disponível em <<https://www.digitizeit.xyz/pt/>> e acesso realizado em 07/10/2021.
- CALVIELLO, R. F. **Avaliação da reatividade de equinos durante o manejo e na presença de estímulo desconhecido**. 2013. 75f. Dissertação (Mestrado) - Faculdade de Zootecnia e Engenharia de Alimentos, Universidade de São Paulo, Pirassununga, 2013.
- CHAN, H. F.; SKALI, A.; TORGLER, B. **A global dataset of human mobility**, CREMA Working Paper, No. 2020-04, Center for Research in Economics, Management and the Arts (CREMA), 2020.
- CRUZ, L. C.; TEIXEIRA-ARAÚJO, A. A.; MOREIRA, S. R. **O frequencímetro Polar RS800CX é sensível à manobra de alteração postural**. ConScientiae Saúde, vol. 17, núm. 1, pp. 57-64, 2018
- ESSNER, A.; SJOSTROM, R; AHLGREN, E.; LINDMARK, B.; **Validity and reliability of Polar® RS800CX heart rate monitor, measuring heart rate in dogs during standing position and at trot on a treadmill**, Physiology & Behavior, V. 114–115, 2013.
- EVANS, D. L. **Training and Fitness in Athletic Horses**. Report for Rural Industries Research and Development Corporation, 2000.
- FLOWER, A.; MCKENNA, W. J.; UPRETI, G. **Validity and reliability of GraphClick and DataThief III for data extraction**. Behavior Modification, 40, 396-413, 2016.
- FUREIX, C. et al. **How horses (Equus caballus) see the world: humans as significant “objects”**. Animal Cognition, Berlin, v. 12, n. 4, p. 643–654, 2009.
- GUYTON, A. C.; HALL, J. E. **Tratado de Fisiologia Médica**. 12ª ed. Rio de Janeiro: Elsevier, 1151 p., 2011.
- HAMA, H., YOGO, M., MATSUYAMA, Y. **Effects of stroking horses on both humans' and horses' heart rate responses**. Jpn. Psychol. Res. 38, 66–73, 1996.

HAUSBERGER, M.; MULLER, C.; LUNEL, C. **Does work affect personality? A study in horses.** PloS one, v. 6, n. 2, p. e14659, 2011.

HOGDSON, D.R.; MCKEEVER, K. H.; MCGOWAN, C. M. **The athletic horse: principles and practice of equine sports medicine**, 2nd edition. Saunders Elsevier, Philadelphia, PA, USA, pp. 9-17.

HOPSTER, H.; BLOKHUIS, HJ. **Validation of a heart-rate monitor for measuring a stress response in dairy cows.** Can J Animal Science, 74(3):465–74, 1994.

JOHNSON, R. A.; BHATTACHARYYA, G. K. **Statistics: principles and methods.** [S.I]: John Wiley & Sons, 2009.

KOZAK, A. et al. **Consistency of emotional reactivity assessment results obtained in different behavioural tests.** Applied Animal Behaviour Science, Amsterdam, v. 205, p. 54-60, 2018.

LANSADE, L. et al. **Dimensions of temperament modulate cue-controlled behavior: a study on pavlovian to instrumental transfer in horses (Equus Caballus).** PloS ONE, San Francisco, v. 8, n. 6, e64853, 2013.

LESIMPLE, A. C. et al. **Housing conditions and breed are associated with emotionality and cognitive abilities in riding school horses.** Applied Animal Behaviour Science, Amsterdam, v. 129, p. 92-99, 2010.

LIMA, R.A.S.; SHIROTA, R.; BARROS, G.S.C. **Estudo do complexo do agronegócio cavalo.** Piracicaba: CEPEA/ESALQ/USP, 2006. 251p.

MALIN, T.; GILLICH, G.; NEDELEU, D.; IANCU, V. **Digitalization of earthquake signals stored as images.** ICMS 2019 & COMEC 2019, Brasov, Romania. 2019.

MANSO FILHO, H.C.; MANSO, H.E.C.C.C.; McKEEVER, K.H. et al. **Heart rate responses of two breeds of four-gaited horses to a standardised field gaited test.** Comparative Exercise Physiology, Wageningen, v.8, n.1, p. 41-46, 2012.

MARCHANT-FORDE, R. M.; MARLIN, D. J.; MARCHANT-FORDE, J. N. **Validation of a cardiac monitor for measuring heart rate variability in adult female pigs: accuracy, artefacts and editing.** Physion Behav, 80(4):449–58, 2004.

MARJAN, J.; PRISENK, J.; VINDIS, P.; STAJNKO, D. **The art equipment for measuring the horse's heart rate.** Journal of achievements of materials and manufacturing engineering, v.41, 2010.

MERKIES, Katrina et al. **Preliminary results suggest an influence of psychological and physiological stress in humans on horse heart rate and behaviour.** Journal of Veterinary Behaviour: Clinical Applications and Research, v. 9, n. 5, p. 242-247, 2014.

MILLS, D.; NARKERVIS, K. **Comportamento equino: princípios e prática**. São Paulo: Roca, 2005.

MOORE, D. S.; McCABE, G. **Introduction to the practice of statistics**. New York, Freeman, 2004.

MOYEAERT, M.; MAGGING, D.; VERKUILEN, J. **Reliability, Validity, and Usability of Data Extraction Programs for Single-Case Research Designs**. Behavior Modification, 40(6), 874–900, 2016.

OLSSON, K. **A review of methods used to measure temperamental characteristics in horses**. Studentarbete, Skara, v. 338, p. 1-24, 2008.

PARKER, M; GOODWIN, D; EAGER, R. A; REDHEAD, E. S.; MARLIN, D. J. **Comparison of Polar heart rate interval data with simultaneously recorded ECG signal in horses**. Comp Exerc Physion, 6(04):137–42, 2010.

PASCHOA, D. C.; COUTINHO, J. F. S.; ALMEIDA, M. B. **Análise da Variabilidade da Frequência Cardíaca no exercício de força**. Revista da Sociedade de Cardiologia do Estado do Rio de Janeiro, v. 19, n. 5, p. 385-390, 2006.

PIERARD, M.; MCGREEVY, P.; RONY, G. **Effect of preceding activity and multiple testing of Belgian police horses on behavior during personality tests**. Journal of Veterinary Behavior, New York, v. 29, p. 111-119, 2019.

RAKAP, S.; RAKAP, S.; EVRAN, D.; CIG, O. **Comparative evaluation of the reliability and validity of three data extraction programs: UnGraph, GraphClick, and DigitizeIt**. Computer in Human Behavior, 55, 159-166, 2016.

ROHATGI, A. **WebPlotDigitizer**, 2019. Disponível em <https://automeris.io/WebPlotDigitizer/citation.html> e acesso realizado em 31/08/2021.

SABELA, A. K. D. **Avaliação do músculo diafragma em ratos com insuficiência cardíaca submetidos à exercício aeróbio: estudo experimental e revisão sistemática com meta-análise**. Dissertação (doutorado em ciências veterinárias), UNOESTE, 2019.

SHADISH, W. R.; BRASIL, I. C. C.; ILLINGWORTH, D. A.; WHITE, K. D.; GALINDO, R.; NAGLER, E. D., RINDSKOPF, D. M. **Using UnGraph to extract data from image files: Verification of reliability and validity**. Behavior Research Methods, 41, 177-183. 2009

SANTOS, V. P. **Variações hemato-bioquímicas em equinos de salto submetidos a diferentes protocolos de exercício físico**. Porto Alegre, RS. Dissertação (Mestrado em Ciências Veterinárias), Universidade Federal do Rio Grande do Sul, 2006.

TUMMERS, B. **DataThief III manual v. 1.1**, 2006. Disponível em <http://www.datathief.org/> e acesso realizado em 07/10/2021.

VISSER, E. K. et al. **Responses of horses in behavioural test correlate with temperament assessed by riders**. Equine Veterinary Journal 35, 176–183,2003.

VISSER, E. K. et al. **Does horse temperament influence horse-rider cooperation?** Journal of Applied Animal Welfare Science, Mahwah, v.11, n. 3, p. 267-284, 2008.

WILLIAMS, R.J.; CHANDLER, R.E.; MARLIN, D.J. **Heart rates of horses during competitive dressage.** Comparative Exercise Physiology, Wageningen, v. 6, n. 1, p. 7-15, 2009.