



## **BANCO DE DADOS EM PROCESSOS DE PESQUISA COM REDES NEURAIS ARTIFICIAIS: UMA ALTERNATIVA ÀS PLANILHAS ELETRÔNICAS**

Luan Leite Leão, Francisco Ramos de Melo  
leaoluanleite@gmail.com, francisco.melo@ueg.br

Universidade Estadual de Goiás, Sistemas de Informação, Anápolis, Goiás

**RESUMO:** Este trabalho procura apresentar um estudo geral sobre a inserção de ferramentas de manipulação de dados em ambientes de pesquisas acadêmicas. No estudo, foi constatado que grande parte dos pesquisadores utilizam planilhas eletrônicas para o trabalho. A ausência de padronização, indisponibilização e o uso incorreto de ferramentas para manipulação de dados são algumas das características iniciais destas ferramentas, gerando problemas ao pesquisador no manejo dos dados, agravando à medida que a quantidade e complexidade de dados aumenta. O estudo buscou introduzir e avaliar o emprego de ferramentas de Banco de Dados, como alternativa no processo de manipulação de grandes quantidades de dados. Foi realizado um estudo comparativo das características tanto de planilhas como de ferramentas de gerenciamento de banco de dados no que diz respeito a adequação, facilidade e aceitação de uso pelos pesquisadores. Adicionalmente, será proposta uma interface para manipulação de banco de dados que apresente facilidade, consistência e segurança na manipulação de dados pelos pesquisadores, objetivando uma substituição viável das planilhas. Com isto, serão elencados elementos que demonstram vantagens na manipulação e compartilhamento de dados, considerando o contexto do ambiente de pesquisa do programa de Mestrado em Engenharia Agrícola do CCET-UEG.

**Palavras-Chave:** Banco de dados, Manipulação de dados, Pesquisadores, Planilhas eletrônicas.

## **DATABASE IN RESEARCH PROCESSES WITH ARTIFICIAL NEURAL NETWORKS: AN ALTERNATIVE TO SPREADSHEETS**

**ABSTRACT:** This work seeks to present a general study on the insertion of data manipulation tools in academic research environments. In the study, it was found that most researchers use electronic spreadsheets for work. The lack of standardization, unavailability and incorrect use of tools for data manipulation are some of the initial characteristics of these tools, generating problems for researchers in data management, aggravating as the amount and complexity of data increases. The study sought to introduce and evaluate the use of database tools, as an alternative in the process of handling large amounts of data. A comparative study of the characteristics of both spreadsheets and database management tools was carried out in terms of suitability, ease and acceptance of use by researchers. Additionally, it will be proposed as an interface for database manipulation that



presents ease, consistency and security in data manipulation by researchers, aiming at a viable replacement of spreadsheets. With this, elements that demonstrate advantages in handling and sharing data will be listed, considering the context of the research environment of the Master's program in Agricultural Engineering at CCET-UEG.

**Keywords:** Data base, Date manipulation, Researchers, Electronic spreadsheets.

## 1. INTRODUÇÃO

O programa de Mestrado em Engenharia Agrícola (PPGEA) da Universidade Estadual de Goiás do Campus Central CET-Anápolis é formado por três linhas de pesquisa: Tecnologia da Produção Agrícola, Recursos Hídricos e Resíduos Agroindustriais. Entre outras características, as linhas apresentam um perfil de coleta, manipulação e análise de dados em médio e grande volume. Os dados da pesquisa geralmente são manipulados em planilhas eletrônicas por um número considerável dos pesquisadores, principalmente os pesquisadores mestrandos.

A ausência de padronização, indisponibilização e o uso incorreto de ferramentas para manipulação de dados pode acarretar problemas ao pesquisador e à continuidade de outras pesquisas sobre os dados. Entre os problemas podem ser listados: falta de controle sobre os dados, integridade, centralização dos dados e acesso online.

A organização das tabelas de um banco de dados se assemelha com a tabela das planilhas eletrônicas. As tabelas de um banco de dados são organizadas padrões para cada tipo de dado, de tal forma que o tipo do dado tem que ser previamente informado. Com as planilhas eletrônicas essa padronização não ocorre, por exemplo é possível ter uma coluna para datas, porém nada impede que quem esteja inserindo os dados, preencha com um valor que não seja data.

O objetivo deste trabalho é realizar um estudo que ressalte as potencialidades do uso de tecnologias de Bancos de Dados em contraposição ao uso de planilhas eletrônicas. Adicionalmente, o trabalho busca propor uma ferramenta de manipulação de dados que possa ser simples, de tal forma que um pesquisador, mesmo com pouca experiência, consiga definir, manipular e compartilhar sua base de dados.

Planilhas eletrônicas são ferramentas comumente usadas por pesquisadores no que se diz respeito à manipulação de dados, o que conduz à reflexão sobre o fato da adequabilidade de tais ferramentas para a manipulação de dados. Dell'omodarme e Valle (2006) afirmam que muitos acadêmicos fazem uso de planilhas para tratar dados. Desta forma, é compreensível que tal categoria de software, cada vez mais, vem se tornando popular como uma ferramenta para fins estatísticos. Frei (2021) complementa que as planilhas eletrônicas têm recursos interativos com layouts atraentes e o objetivo de fornecer atividades para fins diversos.

À medida que aumenta o volume de dados inseridos em uma planilha eletrônica, mais complexo se torna o processo de manipulação dos dados. Rotinas que poderiam ser realizadas manualmente ou, até mesmo, com comandos simples, tornam-se complicadas ou inviáveis, levando a necessidade de automação. Em situações como estas, é comum adotar uso de fórmulas ou, até mesmo, macros para se aplicar uma determinada lógica de programação sobre esses dados. Sobre esta situação, Dell'omodarme e Valle (2006) afirmam que “Além disso, manipulação e edição de dados não são tão simples quando se usa uma planilha”.

A limitação das planilhas eletrônicas indica a necessidade do uso de uma ferramenta apropriada para armazenar dados. Neste caso, será demonstrado que a outra ferramenta mais indicada para essa tarefa é um sistema de banco de dados. As vantagens do uso do banco de dados são diversas ao se comparar com algumas abordagens alternativas como planilhas eletrônicas. Uma comparação técnica entre as abordagens de manipulação de dados, facilmente indicará a alternativa de Banco de Dados como mais adequada. Costa (2011, p. 9) apresenta algumas vantagens:

- Maior controle dos dados: os dados ficam armazenados em um único lugar;
- Integridade: como o banco de dados trabalha com apenas uma base é menos provável que possa surgir arquivos redundantes;
- Garantia de integridades: é mais comum um dado, que está presente em dois ou mais arquivos (como as planilhas eletrônicas), ser atualizado em apenas um local. Em banco de dados é possível manter a integridade dos dados, pois os mesmos são persistidos em apenas um arquivo;
- Padronização: devido a centralização dos dados torna-se mais fácil adotar padrões, a padronização organiza e por consequência torna o uso dos dados mais eficientes.

Existe uma grande variedade de sistemas de banco de dados disponíveis. Para os objetivos deste trabalho foi considerado um banco de dados que fosse relacional. Desta forma, a escolha do banco de dados ficou entre MariaDB e PostgreSQL, por se tratar de ferramentas com um maior grau de simplicidade no manuseio.

Garrido, López, Constante (2020) concluíram que o desempenho sobre os dois bancos de dados MariaDB e PostgreSQL, apresentam semelhantes desempenhos, com resultados quase idênticos. Em todos os cenários em que foram realizados testes de consulta aos bancos de dados e testes de carregamento de informações, a margem de diferença entre os dois bancos de dados é pequena.

Para este trabalho foi selecionado o banco de dados MariaDB, pois MariaDB é muito similar ao MySQL, além de ser da categoria de software livre de Licença Pública Geral (GPL, do inglês *General Public License*), sendo mantida pela própria comunidade. Desta forma, não corre o risco de indisponibilidades, em algum momento, como pode ocorrer com o MySQL, caso a Oracle resolva privar as novas melhorias ou até mesmo limitar a ferramenta e vender licenças mais completas.



## 2. MATERIAL E MÉTODOS

Foi realizada uma pesquisa bibliográfica em trabalhos acadêmicos já publicados, disponibilizados em bancos de dados como Google Acadêmico, Scielo, Portal de Periódicos CAPES/MEC e Biblioteca Virtual: Minha Biblioteca. Especificamente, foram analisados trabalhos bibliográficos e informações que dizem respeito ao programa de mestrado da Engenharia Agrícola da Universidade Estadual de Goiás do Campus Central - sede Anápolis. Para realizar a documentação dos trabalhos acadêmicos e também para gerar e gerenciar referências dos mesmos, foi utilizado o software gerenciador de referências Mendeley.

Por meio da revisão bibliográfica, este trabalho buscou demonstrar que ferramentas de planilhas eletrônicas (como Excel, Google planilhas, LibreOffice e etc.) não são adequadas para manipulação de grandes volumes de dados. Um dos problemas observados no emprego de softwares de planilhas eletrônicas são problemas, como lentidão, decorrentes do aumento da complexidade e da quantidade de dados inseridos. Por outro lado, o banco de dados se apresenta como uma ferramenta mais adequada, por se tratar de uma ferramenta arquitetada para manipulação de dados.

As linhas de pesquisa do mestrado da engenharia agrícola realizam estudos e desenvolvimento de tecnologias. Na linha Tecnologia da Produção Agrícola, existem trabalhos empregando Inteligência Artificial, em especial a área das redes neurais, nos quais é possível verificar o uso de grandes quantidades de dados para treinamento de modelos de redes neurais.

Foi observado que grande parte dos pesquisadores, do programa de mestrado, coletavam os dados referentes a sua pesquisa e os inseriram em planilhas eletrônicas. Entretanto, essa abordagem pode resultar em alguns problemas ou dificuldades quanto a manipulação dos dados, tais como:

- Ausência de padrão: cada pesquisador com seus próprios critérios decide os padrões, formato e organização dos dados a serem utilizados nas suas planilhas;
- Centralização dos dados: Devido o pesquisador ser o autor da planilha, é comum que as mesmas fiquem retidas e sua utilização limitada ao pesquisador ou equipe do projeto. Desta forma, quando outros pesquisadores, posteriores ou externos, precisarem utilizar os mesmos dados, ficarão na dependência da disponibilidade e entendimento da planilha.
- Não disponibilização dos dados: No geral, muitos pesquisadores não têm disponibilidade de uma plataforma com os dados coletados em pesquisas anteriores. Um pesquisador que pretenda dar continuidade a uma determinada pesquisa, geralmente, terá que entrar em contato com pesquisador que desenvolveu o trabalho. Entretanto, o mesmo pode não estar disponível ou não ter mais a disponibilidade dos dados. Caso consiga uma cópia de uma planilha, ainda assim, o pesquisador beneficiado com a cópia talvez não conseguirá aproveitar os dados, por não conseguir entender a forma pela qual foram organizados os dados.

Os estudos dos requisitos do problema indicaram que a escolha de uma ferramenta para trabalhar com dados deve ser tão simples e intuitiva quanto uma planilha e, também, que tenha o poder e a versatilidade de um banco de dados.

O requisito de ser simples de ser utilizado é importante, pois a maioria dos pesquisadores do mestrado da engenharia agrícola não tem conhecimentos conceituais de banco de dados. Muito disto decorre do fato de que o banco de dados ser complexo e exigir conhecimento dos seus princípios gerais e organização. Diante desta situação, é justificada a preferência pelo Excel, que se apresenta como uma ferramenta mais simples para manusear.

É importante salientar que esse estudo não tem como premissa o desenvolvimento da ferramenta, mas sim levantar a relevância do desenvolvimento de uma ferramenta que possa armazenar dados com uso do banco de dados e uma interface interativa.

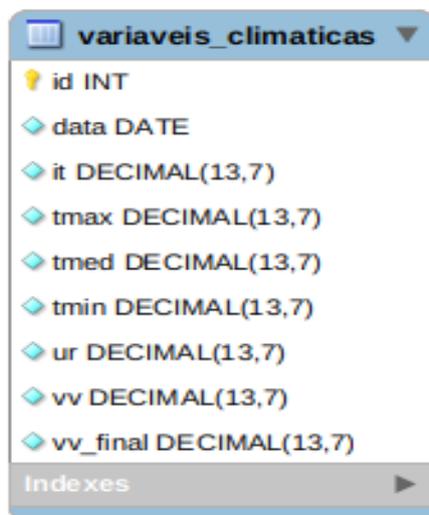
Alguns testes com planilhas eletrônicas foram realizados neste trabalho. Um deles foi a exportação da planilha de um mestrando da engenharia agrícola para a base de dados MariaDB. Uma das dificuldades foi a grande quantidade de dados. Esse mestrando em sua pesquisa precisava trabalhar com dados climáticos de estações meteorológicas da região centro-oeste do Brasil, abrangendo os estados do Distrito Federal, Goiás, Mato Grosso e Mato Grosso do Sul. Em cada Estado foram coletados dados das estações meteorológicas em um quantitativo de cidades descritos a seguir:

- Distrito Federal, 4 cidades;
- Goiás, 25;
- Mato Grosso 34;
- Mato Grosso do Sul, 28;

Foram coletados dados em 85 cidades, entre os anos de 1989 à 2019. Cada cidade contém uma planilha com esses resultados meteorológicos, cada planilha contém 17 variáveis (ou coluna) e em média mais de 11 mil linhas.

Foi criada uma tabela onde os dados das planilhas pudessem ser salvos. A Figura 1 apresenta o diagrama da tabela (Figura 1).

Figura 1 - Diagrama da tabela de variáveis climáticas.



variaveis_climaticas	
id	INT
data	DATE
it	DECIMAL(13,7)
tmax	DECIMAL(13,7)
tmed	DECIMAL(13,7)
tmin	DECIMAL(13,7)
ur	DECIMAL(13,7)
vv	DECIMAL(13,7)
vv_final	DECIMAL(13,7)

Fonte: O autor (2021).

A tabela é composto dos seguintes atributos:

- id: Identificador único é um tipo de dados Inteiro (INT);
- data: A data (DATE) que série temporal foi coletada no formato AAAA/MM/DD(padão americano);
- it: Insolação Total número em decimal;
- tmax: Temperatura Máxima do ar número em decimal;
- tmed: Temperatura Média do ar número em decimal;
- tmin: Temperatura Mínima do ar número em decimal;
- ur: Umidade Relativa média do Ar número em decimal;
- vv: Velocidade do Vento número em decimal;
- vv\_final: Velocidade do Vento Final número em decimal;

Com o auxílio da linguagem de programação Python, algumas informações, como o tipo data que foi transformado para o padrão americano(AAAA/MM/DD), foram convertidas para os padrões adotado pelo banco de dados MariaDB, posteriormente toda a planilha foi salva no banco de dados.

Para validar algumas perguntas desta pesquisa, no que se diz respeito ao uso das planilhas eletrônicas, um formulário foi elaborado e aplicado via Google Forms, por meios on-line devido a pandemia da covid-19, entre os discentes dos cursos de Arquitetura e Urbanismo, Engenharia Agrícola, Engenharia Civil, Ciências Farmacêuticas, Química Industrial, Química Licenciatura, Física Licenciatura, Matemática Licenciatura, Ciências Biológicas Licenciatura, Sistemas de Informação da Graduação e, também, para Programas de Mestrado Programa de Doutorado da Pós-Graduação.

O formulário foi aplicado no ano de 2021 e se limitou apenas ao Campus CET. O mesmo teve o objetivo de elencar informações para complementar a pesquisa. As análises dos dados

coletados ratificaram algumas expectativas na intencionalidade do formulário. Ao todo, foram coletadas respostas de 188 estudantes.

### 3. RESULTADOS E DISCUSSÃO

No estudo bibliográfico foi possível verificar que, dentre muitas opções de softwares de planilhas eletrônicas, destaca-se o software Microsoft Excel®. Segundo Barreto (2015), citado por Frei (2021, p. 11) “Dentre as planilhas eletrônicas, a planilha Microsoft Excel® é muito popular e está presente em grande parte dos computadores pessoais, assim como diferentes planilhas eletrônicas estão disponíveis em plataformas online”. Pataro (2019, p. 23) também confirma essa popularidade afirmando que “O Excel é o software de planilha eletrônica mais utilizado no mundo”.

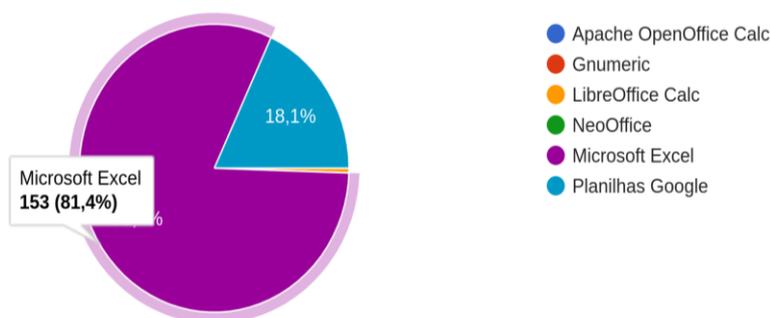
Foi realizada uma pesquisa de preferência de planilhas eletrônicas com 188 estudantes do cursos (graduação e pós graduação) do Campus CET. A Figura 2 apresenta um gráfico das principais planilhas eletrônicas utilizadas. Pode ser observado a predominância das planilhas eletrônicas Microsoft Excel® e Google, responsáveis por quase 100%. Também aparecem outras planilhas, mas com um percentual muito baixo.

Dentre os alunos pesquisados, foi observado que as planilhas eletrônicas do Google vêm se tornando populares. Dentre outras características podem ser destacadas:

- Acesso simultâneo online: Facilita trabalhos em grupos que mais de um usuário precisa editar um arquivo;
- Arquivos salvos em nuvens: Os arquivos são salvos automaticamente no Google Drive;
- Não precisa de licença: As planilhas eletrônicas do Google não necessitam de licença paga.

Figura 2 - Softwares de planilhas eletrônicas mais comuns.

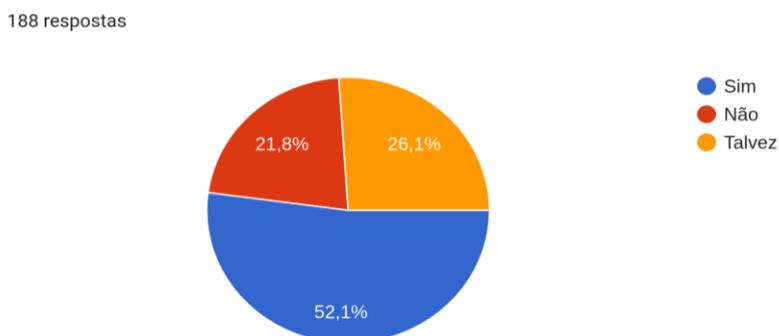
188 respostas



Fonte: O autor (2021).

Sobre a facilidade de uso da planilha Microsoft Excel®, pouco mais da metade (52,1%) dos alunos a consideram fácil de usar (Figura 3). Enquanto 21,8% dos entrevistados não consideram fácil e outros 26,1 responderam com um talvez. Tais resultados indicam que uma parcela dos estudantes entrevistados não sente facilidade no uso da planilha Microsoft Excel®. A Figura 3 apresenta um gráfico que ilustra o resultado da impressão da Facilidade de uso da planilha.

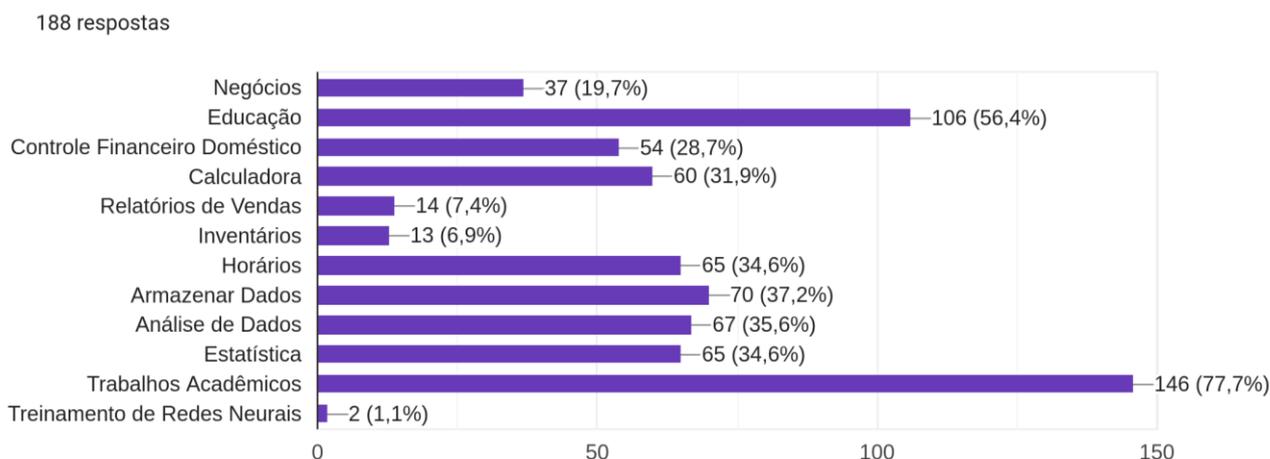
Figura 3 – Impressão da facilidade de uso da planilha Microsoft Excel.



Fonte: O autor (2021).

Sobre a finalidade de uso das planilhas, foi constatado que 77,7% usam para trabalhos acadêmicos e outros 56,4% para educação, conforme pode ser visualizado no gráfico da Figura 4. Devido o ambiente se tratar de uma Universidade, é esperado que a maioria dos entrevistados utilizem para trabalhos acadêmicos e/ou relacionados à área de educação

Figura 4 – Finalidade de utilização das planilhas eletrônicas.

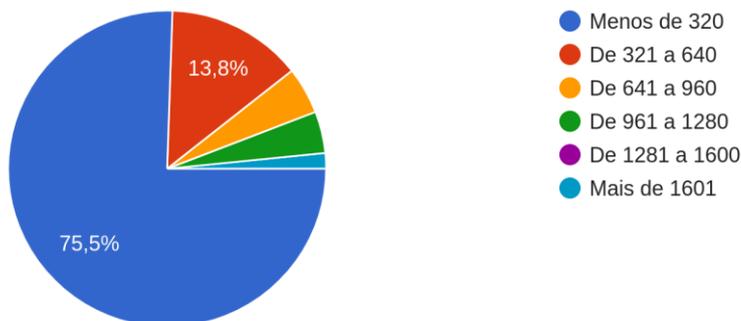


Fonte: O autor (2021).

Quanto a quantidade de linhas utilizadas no emprego das planilhas foi verificado que, para grande maioria emprega menos de 320 linhas, conforme apresentado na Figura 5.

Figura 5 - Quantidade média de linhas utilizadas nas planilhas.

188 respostas

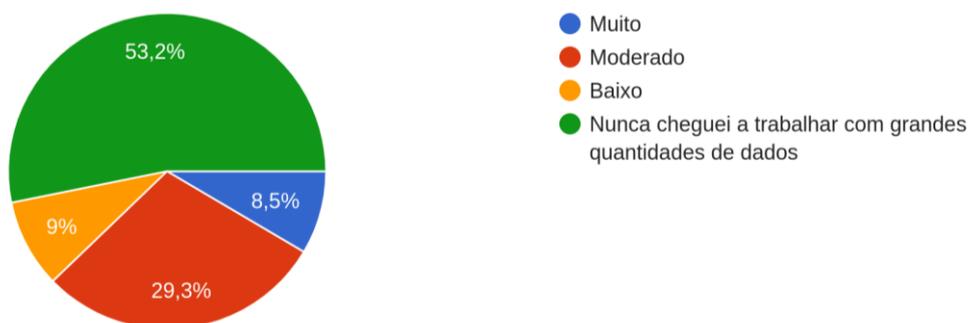


Fonte: O autor (2021).

Quanto a percepção de demora na resposta das planilhas diante de volumes maiores de dados, cerca de 37,8% acredita que as planilhas eletrônicas apresentam uma certa demora para responder entre moderado e muito, conforme Figura 6. Este cenário se mostra real a partir de quando o aluno começa a processar grandes quantidades de dados.

Figura 6 - Percepção de demora de resposta das planilhas eletrônicas com grandes quantidades de dados .

188 respostas



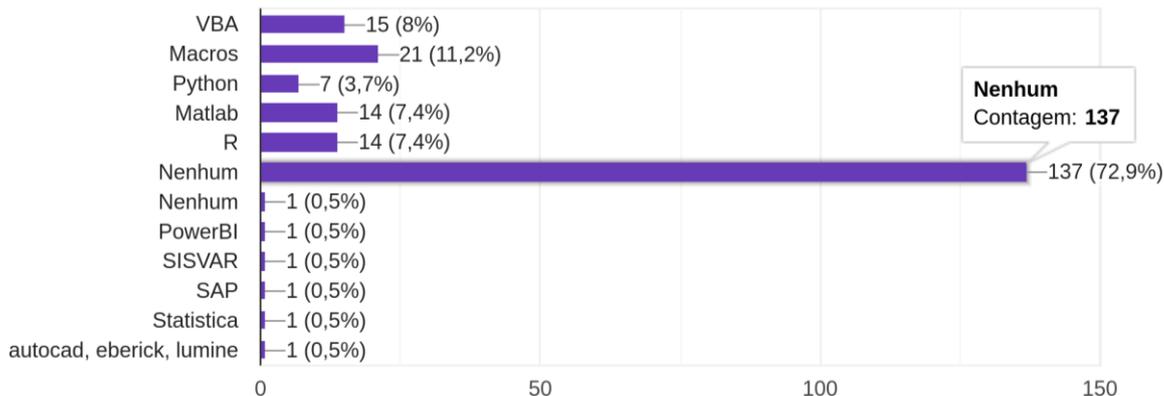
Fonte: O autor (2021).

Conforme apresentado na Figura 7, dentre a maioria dos entrevistados, não é comum o uso de algum instrumento que auxilie na manipulação dos dados. Porém é possível verificar um maior emprego de ferramentas como VBA, Macros, Python, Matlab e R.

Figura 7 - Uso de outros instrumentos para auxiliar as planilhas eletrônicas.

Anais da Semana de Iniciação Científica do Curso de Sistemas de Informação (13ª edição)

188 respostas

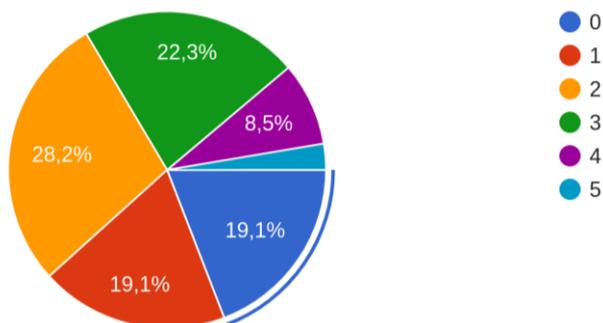


Fonte: O autor (2021).

Dentre os entrevistados, 19,1% responderam não ter conhecimento sobre banco de dados, enquanto 80,9% responderam ter um certo conhecimento. A Figura 8 apresenta um gráfico com as respostas do percentual de conhecimento/domínio sobre banco de dados. Na figura, as legendas representam uma escala de conhecimento que variam de 0 (nenhum conhecimento) a 5 (completo domínio ou equivalente) sobre banco de dados.

Figura 8: Conhecimento sobre banco de dados

188 respostas



Fonte: O autor (2021).

#### 4. CONCLUSÃO



Este trabalho apresentou um breve estudo sobre o emprego de planilhas e banco de dados em pesquisas acadêmicas. Foi realizada uma pesquisa em 188 estudantes da graduação e pós-graduação, com o objetivo de estabelecer a percepção, conhecimento e utilização das ferramentas.

Os resultados indicaram que as planilhas eletrônicas são consideradas de fácil manuseio e muito úteis para trabalhar com pequenas quantidades de dados. Quando os dados nas planilhas começam a tomar proporções maiores, a mesma não se torna mais uma ferramenta interessante para os trabalhos.

Considerando a natureza dos trabalhos de pesquisa universitários, os requisitos apontam o uso de um banco de dados como uma ferramenta mais adequada para manipulação dos dados. Entretanto, como o banco de dados não é tão simples de manusear como as planilhas, verifica-se a necessidade do desenvolvimento de uma interface mais amigável para possibilitar um trabalho com banco de dados com similaridade na simplicidade e facilidade de manuseio como ocorre com as planilhas eletrônicas.

Diante do exposto, é possível verificar que a inserção de uma ferramenta de Banco de Dados pode trazer vários benefícios institucionais no que diz respeito aos dados de pesquisa. Entretanto, essa ação deve ser acompanhada de profissionais que possam assegurar mecanismos para acesso, facilidade de uso e orientação na estruturação para manipulação dos dados.

## REFERÊNCIAS

ALVES, C. A. et al. Percalços de um estudo da prevalência de esclerose múltipla no estado do Rio de Janeiro. **Revista Brasileira de Neurologia e Psiquiatria**, v. 7, n. 3, p. 103-111, 2013.

COSTA, E. R. **Bancos de dados relacionais**. 2011. 63 f. Trabalho de conclusão de curso (Tecnólogo em Processamento de Dados) – Faculdade Tecnológica de São Paulo, São Paulo, 2011. Disponível em: <<http://www.fatecsp.br/dti/tcc/tcc0025.pdf>>; Acesso em: 18 jun. 2021.

DA SILVA, Fábio Tuoto.; DE ALMEIDA SANTOS, **Soluções em Planilhas Eletrônicas para Aplicações no Ambiente Corporativo**. Caderno PAIC, v. 16, n. 1, p. 669-682, 2015.

DELL'OMODARME, Matteo; VALLE, Giada. Teaching statistics with Excel and R. arXiv preprint physics/0601083, 2006. Disponível em: <https://arxiv.org/pdf/physics/0601083.pdf>. Acesso em: 25 ago. 2021.

DESCRIÇÃO do Programa Pós Graduação Engenharia Agrícola Mestrado. [S. l.], c2016. Disponível em: [http://www.ccet.app.ueg.br/cliente/paginas\\_cursos/programa\\_pos\\_graduacao\\_engenharia\\_agricola\\_mestrado.php](http://www.ccet.app.ueg.br/cliente/paginas_cursos/programa_pos_graduacao_engenharia_agricola_mestrado.php). Acesso em: 3 set. 2021.



FREI, Fernando. Perspectivas do uso de Planilhas Eletrônicas no Ensino de Estatística. **Revista de Ensino de Ciências e Matemática**, vol. 12, no. 1, p. 1–16, 2021. <https://doi.org/10.26843/rencima.v12n1a15>.

GARRIDO, Anabel Pilicita; LÓPEZ, Yolanda Borja; CONSTANTE, Gonzalo Gutiérrez. Rendimiento de MariaDB y PostgreSQL. **Revista Científica y Tecnológica UPSE**, vol. 7, no. 2, p. 9–16, 2020.

LIMA, Álvaro; LACERDA, Wilian; NETO, Heitor. **Seleção de características de dados utilizando Redes Neurais Artificiais**. In: SIMPÓSIO BRASILEIRO DE SISTEMAS DE INFORMAÇÃO (SBSI), 13. , 2017, Lavras. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2017 . p. 135-142. Disponível em: < <https://doi.org/10.5753/sbsi.2017.6035>.>; Acesso em: 18 jun. 2021.

LINHAS de Pesquisa. [S. l.], 4 set. 2021. Disponível em: [http://www.ppgea.ueg.br/conteudo/11844\\_linhas\\_de\\_pesquisa](http://www.ppgea.ueg.br/conteudo/11844_linhas_de_pesquisa). Acesso em: 4 set. 2021.

MENDELEY. [S. l.], 5 set. 2021. Disponível em: [https://www.mendeley.com/?interaction\\_required=true](https://www.mendeley.com/?interaction_required=true). Acesso em: 5 set. 2021.

PATARO, Adriano. **Dominando o Excel 2019**. São Paulo: Novatec Editora, 2019. 488 p.  
THE R Project for Statistical Computing. [S. l.], 5 set. 2021. Disponível em: <https://www.r-project.org/>. Acesso em: 5 set. 2021.