

ANÁLISE EXPLORATÓRIA DE DADOS: PRERROGATIVAS DE UMA INFERÊNCIA ESTATÍSTICA EFICIENTE

DANERES Lemos, Eduardo¹

RU 1395304

PADILHA, Eliandro José²

RESUMO

O presente artigo tem por objetivo confirmar a indispensabilidade da análise exploratória de dados (AED) como ferramenta de estudos estatísticos, seja em nível profissional ou acadêmico. A metodologia utilizada se referiu a uma abordagem qualitativa e quanto aos procedimentos metodológicos se tratou de uma pesquisa explicativa. Para tanto fica demonstrado o caráter heurístico da AED visto que serve de guia para o estudioso permitindo-lhe o acesso direto a informações relevantes mesmo sem conhecimentos prévios ou avançados e, deste modo, reduzindo o tempo para a tomada para decisões. Desta forma oferece um vislumbre de possíveis caminhos ou modelos, diretamente da exploração dos dados. Foram utilizados dados estruturados e feitas análises bi e multivariadas de maneira simplificada através de planilhas eletrônicas. Os recursos aplicados foram gráficos de linha e dispersão, com a obtenção de linhas de tendência, coeficiente de correlação de Pearson e coeficiente de determinação. A extração de informações por estes métodos simples apresentou-se satisfatória por servir de alicerce para estudos posteriores de inferência ou modelagem. Tal importância da AED é reconhecida, visto que, já é parte de cursos introdutórios de estatística.

Palavras-chave: Análise. Dados. Inferência. Exploratória

1. INTRODUÇÃO

A inferência estatística é um dos mais poderosos ramos da Estatística, sendo essencial para a obtenção de informações úteis a partir de amostras significativas de dados. Neste sentido parece natural que uma descrição prévia através de recursos computacionais seja imprescindível, a fim de podermos nortear qualquer trabalho de forma mais segura e assertiva. Como principal ferramenta para atingirmos este objetivo temos a análise exploratória de dados (AED) que é um conjunto de técnicas fáceis e eficientes que guiam o trabalho do pesquisador ou estatístico permitindo uma

¹ Aluno do Centro Universitário Internacional UNINTER. Artigo apresentado como Trabalho de Conclusão do Curso de Bacharelado em Matemática 2018/01.

² Professor Orientador no Centro Universitário Internacional UNINTER

visão mais ampla das informações coletadas, sem a necessidade de um modelo matemático prévio.

Este tipo de ferramenta pode ser utilizado a partir de planilhas eletrônicas, onde é aplicada essencialmente a estruturação dos dados em tabelas, o que permite a comparação direta destes através das linhas e colunas e, gráficos de dispersão e linhas, que transmitem ao observador atento as tendências e possíveis relações entre as variáveis analisadas. Na sequência, fez-se uso de outras ferramentas específicas para tratamento de dados, com intuito de embasar o que já foi observado com as técnicas anteriores, entre elas, o cálculo do índice de correlação de dados bivariados.

Seguindo o avanço nos métodos escolhidos, optou-se por traçar linhas de tendência que melhor se ajustassem aos pontos, sempre obtendo-se os respectivos coeficientes de determinação entre os diferentes tipos de ajuste possibilitados pelo software. Ao focar-se nesta parte da metodologia da AED o autor buscou superar suas próprias limitações em termos de conhecimento estatístico e avançar em direção a métodos mais conclusivos.

Embora de praxe, esta técnica é revista aqui como processo de construção heurístico tanto do profissional como do estudante de estatística. Espera-se demonstrar que, de posse destas técnicas, é possível transformar dados em informações úteis e, estas, em conhecimento, mesmo sem dominar conceitos matemáticos avançados.

Através de planilhas eletrônicas busca-se estruturar dados e antever modelos prévios sem o domínio pleno e aprofundado das técnicas avançadas da estatística inferencial. Neste momento espera-se, através da visualização gráfica e da busca de correlações, um entendimento do que os dados podem significar, ou seja, captar a transição de dados brutos para informação útil apoiado em avaliação heurística. O objetivo principal aqui é enfatizar a naturalidade da busca de soluções para problemas complexos através da solução de problemas menores sendo este caminho aberto, no contexto da análise de dados, pela AED.

2.O COMEÇO DE TUDO: A ANÁLISE EXPLORATÓRIA DE DADOS

Quando um estudante ou profissional iniciante da área estatística depara-se com um volume de dados coletados experimentalmente podem surgir perguntas

¹ Aluno do Centro Universitário Internacional UNINTER. Artigo apresentado como Trabalho de Conclusão do Curso de Bacharelado em Matemática 2018/01.

² Professor Orientador no Centro Universitário Internacional UNINTER

inevitáveis: por onde começar? Qual abordagem devo utilizar? Que informações úteis poderão ser obtidas a partir destes dados? Este não deveria ser um motivo para apreensão desde que confiássemos na nossa principal ferramenta: “nosso sistema olho-cérebro”. Neste sentido,

[...] Não existe uma ferramenta estatística única tão poderosa quanto um gráfico. Nosso sistema olho-cérebro é o mais sofisticado processador já desenvolvido, e através de exibições gráficas podemos colocar este sistema em uso para obter uma visão profunda da estrutura de dados. Uma enorme quantidade de informações quantitativas pode ser transmitida por gráficos; nosso sistema olho-cérebro pode resumir vasta informação rapidamente e extrair características salientes, mas também é capaz de focar nos detalhes.(...) (CHAMBERS, CLEVELAND, KLEINER e TUKEY 2018, p.16, tradução nossa).

Além disso,

[...]Nós, humanos,temos em nosso cérebro uma incrível capacidade para compreender padrões por meio do sentido visual. Esse potencial é tanto que, de toda a informação sensorial que temos, 80% é destinada a esse sentido [...] Por termos esta capacidade, a visualização de dados permite que nós tenhamos uma compreensão aperfeiçoada do que nos é apresentado, por meio de uma perspectiva totalmente diferente. (...) (MARQUESONI, 2017, p.190)

Logo podemos questionar: O que é a análise exploratória de dados? Assim,

[...] Trata-se de um conjunto de técnicas de tratamento de dados que, sem implicar em uma fundamentação matemática mais rigorosa, nos ajuda a fazer uma sondagem do terreno, ou seja, tomar um primeiro contato com a informação disponível. (...) (PINHEIRO, CARVAJAL e GOMES, 2009, p.19).

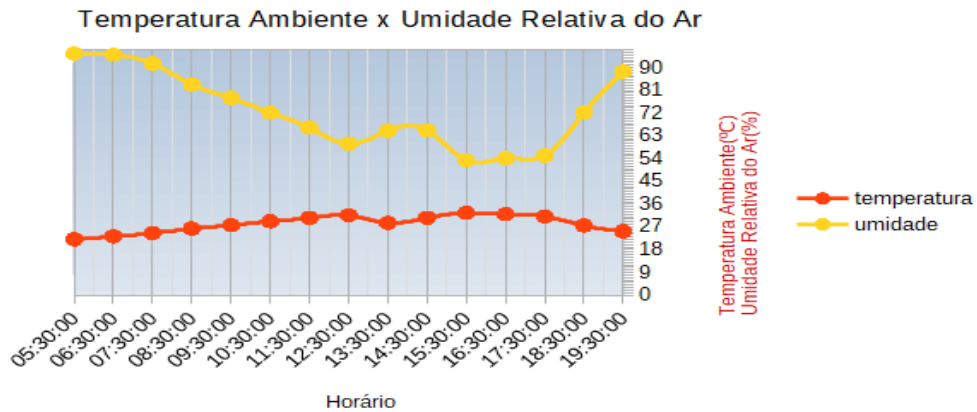
Deste modo, o autor objetiva comprovar que a base para a construção de um arcabouço de conhecimentos estatísticos iniciais é a análise exploratória de dados, primeiramente através da estruturação em tabelas e após, através do uso de recursos gráficos de diversos tipos. Por fim, as diversas ferramentas existentes oferecidas para tratamento de dados através de planilhas eletrônicas, ou, em softwares específicos, são os gráficos de linhas, diagramas de dispersão, linhas de tendência ou regressão, ajuste de mínimos quadrados e correlação, entre outras.

O gráfico abaixo ilustra brevemente algumas destas ideias:

¹ Aluno do Centro Universitário Internacional UNINTER. Artigo apresentado como Trabalho de Conclusão do Curso de Bacharelado em Matemática 2018/01.

² Professor Orientador no Centro Universitário Internacional UNINTER

Figura1:Gráfico com linhas de tendência



Fonte:Elaborado pelo autor(2021)

Podemos perceber facilmente que as duas curvas representadas acima apresentam alguma relação entre si ou com uma terceira variável. Isto é percebido pelos pontos de inflexão que são praticamente coincidentes.

Deste modo permite-se a aplicação de uma abordagem heurística através da visualização gráfica de possíveis tendências, relações de dependência e correlações utilizando ferramentas computacionais disponíveis em planilhas eletrônicas a fim de resolver o problema de extrair informações úteis.

Assim,

[...] É praticamente impossível fazer pesquisa científica sem o uso de computação. A presença de instrumentos computadorizados coletando dados o tempo todo e em todo o lugar gera dados científicos em volumes que não podem mais ser entendidos com cálculos simples, sendo necessárias muitas vezes, computações complexas. (...) (PIMENTEL, BRAGA, 2013, p. 11).

É evidente que embora tais recursos não possam substituir a análise criteriosa e aprofundada do pesquisador, são o primeiro recurso exploratório a nortear o caminho e auxiliar o estudioso em uma direção mais segura.

Afirmou-se então,

[...] O EXCEL é constituído de um conjunto poderoso de ferramentas gráficas. Há 14 tipos de gráficos padrão e mais 20 tipos personalizados, que permitem visualizar e apresentar dados de modo que você e outras pessoas

¹ Aluno do Centro Universitário Internacional UNINTER. Artigo apresentado como Trabalho de Conclusão do Curso de Bacharelado em Matemática 2018/01.

² Professor Orientador no Centro Universitário Internacional UNINTER

possam entendê-los. Um gráfico dá a você uma visão geral e rápida dos dados que não pode ser obtida utilizando a planilha numérica em si. (...) (BLOCH, 2004, p. 22).

É importante lembrar que o advento das planilhas eletrônicas ocorreu simultaneamente ao acesso a computadores pessoais. Isso ocorreu em larga escala e moldou de forma significativa o modo como processamos uma quantidade cada vez maior de dados, tanto em nível profissional quanto acadêmico. A revolução causada por estas ferramentas não apenas trouxe economia de tempo, trouxe uma ampliação de nossa capacidade cognitiva naquilo que é conhecido como inteligência aumentada.

Logo,

[...] A pesquisa em Inteligência Aumentada, por sua vez, é a programação de computadores com o intuito de promover uma base computacional para o aumento ou incremento do pensamento humano, assistindo-o, não tentando substituí-lo por simulação em máquinas. (...) (RANSDELL, 2007, p. 19).

Sendo assim, é parte inerente deste trabalho o uso intensivo de gráficos e tabelas que induzam a obtenção rápida de resultados. Veja-se a tabela e o gráfico abaixo:

Tabela 1: Dados de radiação solar e temperatura ambiente.

HORAS	RADIAÇÃO SOLAR(W/m²)	TEMPERATURA AMBIENTE
05:30:00	9,49	22,11
06:30:00	63,84	23,25
07:30:00	214,74	24,63
08:30:00	486,15	26,36
09:30:00	631,87	27,7
10:30:00	899,03	29,23
11:30:00	1038,15	30,57
12:30:00	906,42	31,62
13:30:00	339,59	28,61
14:30:00	868,72	30,65
15:30:00	735,14	32,58
16:30:00	342,3	32,04
17:30:00	256,4	31,22
18:30:00	38,05	27,56
19:30:00	6,65	25,31
COEFICIENTE DE CORRELAÇÃO DE PEARSON(ρ)		0,674632950006845

Fonte:Elaborada pelo autor (2021)

Podemos observar na tabela1 acima os valores em amarelo que representarão os pontos de máxima das curvas. Exatamente no horário das 11:30 da manhã há um registro de máxima incidência de radiação solar e, quatro horas depois, temos o registro de máxima temperatura neste dia.

O valor da correlação é calculado pela planilha eletrônica e apresentado na última linha da tabela 1 como $\rho = 0,674632950006845$. É importante destacar que não podemos inferir que a temperatura ambiente depende apenas da taxa de

¹ Aluno do Centro Universitário Internacional UNINTER. Artigo apresentado como Trabalho de Conclusão do Curso de Bacharelado em Matemática 2018/01.

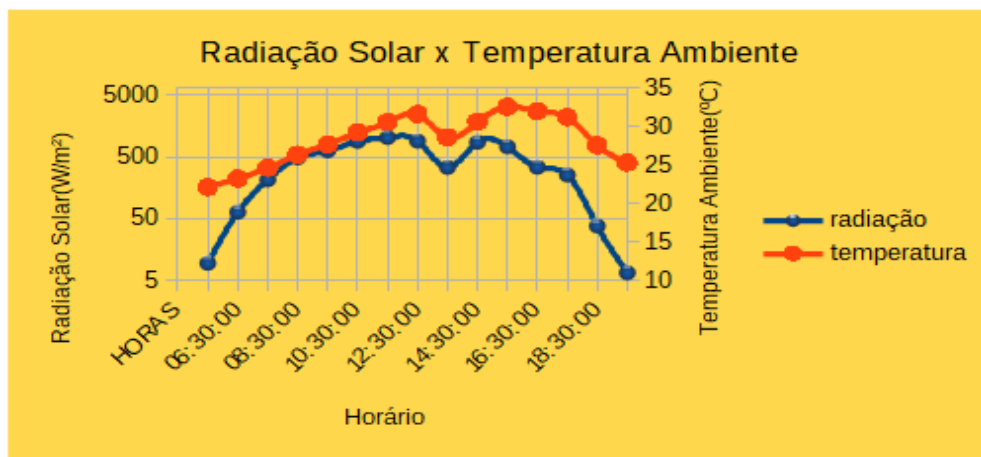
² Professor Orientador no Centro Universitário Internacional UNINTER

radiação solar, isto seria apressado. Deve-se lembrar que correlação não implica causalidade, é apenas uma associação numérica.

Ainda assim, pode-se estender este estudo e supor razoável uma análise de regressão explorando, deste modo, as ferramentas baseadas em ajuste de mínimos quadrados oferecidas por planilhas eletrônicas. Esta ferramenta de análise estatística é muito útil pois permite a escolha do tipo de regressão (linear, logarítmica, polinomial, potência, exponencial e média móvel) e mais diversas opções e funcionalidades.

Através da geração gráficos de linhas contendo as duas variáveis tabeladas dependentes do tempo, busca-se comprovar pelas imagens se o comportamento das curvas ao longo do tempo corrobora o resultado do coeficiente de correlação de Pearson. Observe o gráfico gerado com os dados da tabela 1 que vem a complementar os argumentos expostos:

Figura 2: Gráfico de linhas com os dados de Radiação solar e Temperatura ambiente.



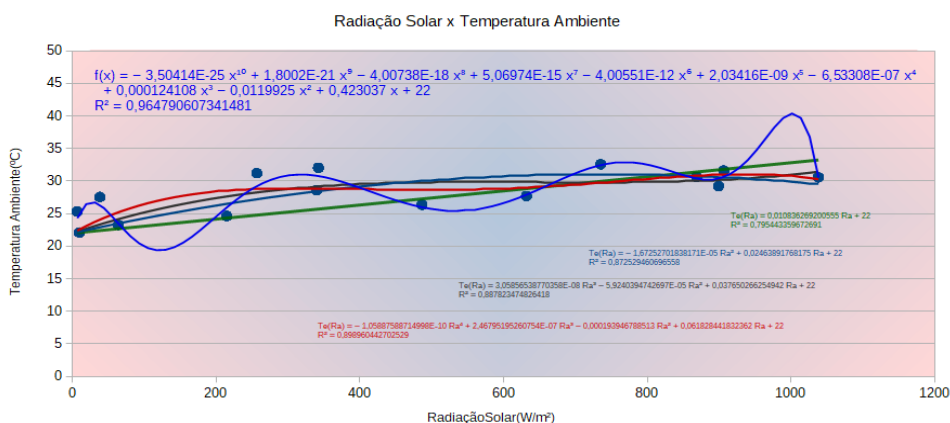
Fonte: Elaborado pelo autor (2021)

Na figura 3 abaixo temos um gráfico que correlaciona as variáveis temperatura e radiação solar. Este é um gráfico de dispersão onde o autor aplicou a ferramenta “linhas de tendência”, gerando assim curvas polinomiais de ajustes progressivos. Aqui não são discutidos os erros existentes.

¹ Aluno do Centro Universitário Internacional UNINTER. Artigo apresentado como Trabalho de Conclusão do Curso de Bacharelado em Matemática 2018/01.

² Professor Orientador no Centro Universitário Internacional UNINTER

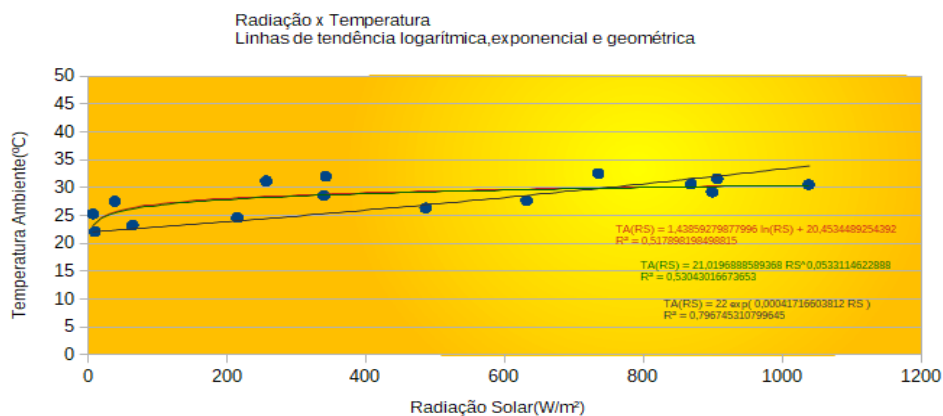
Figura 3: Gráfico com linhas de tendência polinomiais.



Fonte:Elaborado pelo autor (2021)

Na sequência logo abaixo, através das linhas de tendência observadas na figura 4, fica claro qual dos ajustes polinomiais melhor representa os dados. Neste caso específico foi fixado um ponto de partida comum à todas as curvas e isto permitiu um ajuste progressivo destes em relação aos pontos.

Figura 4: Outros ajustes possíveis



Fonte:Elaborado pelo autor (2021)

¹ Aluno do Centro Universitário Internacional UNINTER. Artigo apresentado como Trabalho de Conclusão do Curso de Bacharelado em Matemática 2018/01.

² Professor Orientador no Centro Universitário Internacional UNINTER

É importante reforçar que, o objetivo é facilitar o entendimento de como os dados variam no espaço contínuo, uma preparação apenas para uma possível inferência ou modelagem. Dentre os três tipos de ajuste utilizados acima o que mais ofereceu qualidade de aproximação foi o exponencial o que é comprovado pelo coeficiente de determinação $R^2 \approx 0,798$.

Comparativamente temos $R^2 \approx 0,965$ para um polinômio de grau 10 contra $R^2 \approx 0,798$ para um ajuste exponencial. Neste caso, sem discutirmos os erros de cada ajuste, se pretendêssemos estimar o valor da temperatura ambiente dentro do intervalo do dia fixado pela tabela 1, a melhor simulação se daria através do ajuste polinomial de grau maior do que 5.

Na sequência, a tabela 2 apresenta dados envolvendo as variáveis radiação solar e umidade relativa do ar, abrangendo o mesmo período do dia. Na própria tabela podemos perceber que as duas grandezas possuem relação inversa, ou seja, quando uma delas aumenta a outra diminui. O coeficiente de correlação de Pearson corrobora esta afirmação. Já no gráfico da figura 5 visualizamos esta relação inversa das variáveis.

Tabela 2: Dados de radiação solar e umidade relativa do ar.

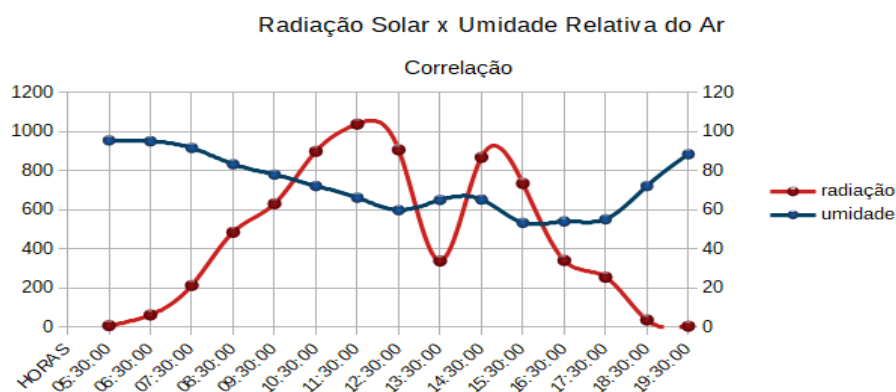
HORAS	RADIAÇÃO SOLAR(W/m ²)	UMIDADE RELATIVA DO AR (%)
05:30:00	9,49	95,52
06:30:00	63,84	95,07
07:30:00	214,74	91,65
08:30:00	486,15	83,33
09:30:00	631,87	78,03
10:30:00	899,03	72,21
11:30:00	1038,15	66,26
12:30:00	906,42	59,87
13:30:00	339,59	65,11
14:30:00	868,72	65,3
15:30:00	735,14	53,4
16:30:00	342,3	54,12
17:30:00	256,4	55,24
18:30:00	38,05	72,21
19:30:00	6,65	88,45
COEFICIENTE DE CORRELAÇÃO DE PEARSON(ρ)		-0,535402799674598

Fonte: Elaborado pelo autor (2021)

Figura 5: Gráfico de linhas dos dados de radiação solar e umidade relativa do ar.

¹ Aluno do Centro Universitário Internacional UNINTER. Artigo apresentado como Trabalho de Conclusão do Curso de Bacharelado em Matemática 2018/01.

² Professor Orientador no Centro Universitário Internacional UNINTER



Fonte:Elaborado pelo autor(2021)

Do mesmo modo que raciocinou-se com os dados da tabela 1 podemos inferir com razoabilidade que a radiação solar influencia na umidade relativa do ar. Da Física é conhecido que as moléculas de água suspensas no ar ficam mais distantes ou mais próximas devido a variação de temperatura e desta forma isto afeta os instrumentos de medida (higrômetros) que medem a concentração de umidade indiretamente. Pode-se também deduzir que a umidade relativa do ar é uma função da temperatura que, por sua vez, é função da radiação solar.

Este exemplo é utilizado aqui propositadamente para destacar um caso de correlação entre uma variável independente (radiação solar) e duas variáveis dependentes (temperatura e umidade relativa do ar). Vejamos agora os dados tabulados com três variáveis acima citadas:

Tabela 3: Dados de radiação solar, temperatura e umidade relativa do ar.

HORÁRIO	RADIAÇÃO SOLAR(W/m²)	TEMPERATURA AMBIENTE(°C)	UMIDADE RELATIVA DO AR(%)
05:30:00	9,49	22,11	95,52
06:30:00	63,84	23,25	95,07
07:30:00	214,74	24,63	91,65
08:30:00	486,15	26,36	83,33
09:30:00	631,87	27,7	78,03
10:30:00	899,03	29,23	72,21
11:30:00	1038,15	30,57	66,26
12:30:00	906,42	31,62	59,87
13:30:00	339,59	28,61	65,11
14:30:00	868,72	30,65	65,3
15:30:00	735,14	32,58	53,4
16:30:00	342,3	32,04	54,12
17:30:00	256,4	31,22	55,24
18:30:00	38,05	27,56	72,21
19:30:00	6,65	25,31	88,45

Fonte:Elaborada pelo autor(2021)

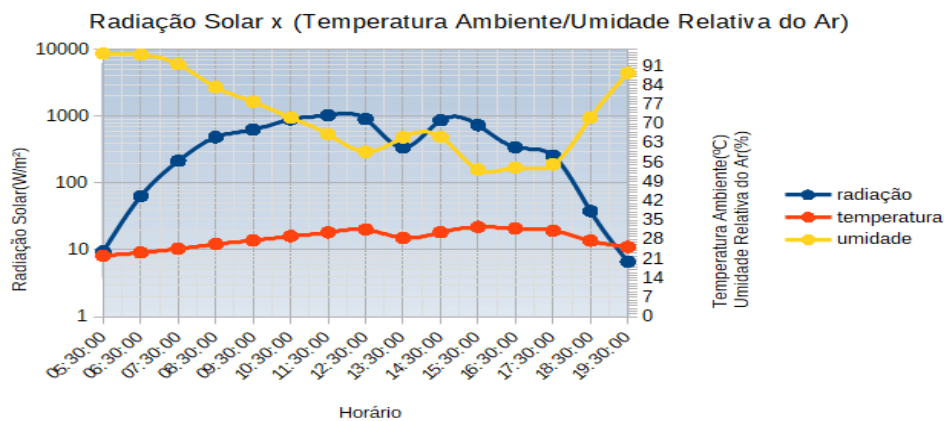
Uma análise prévia sobre a tabela 3 nos indica correlação positiva entre radiação e temperatura e negativas entre radiação/temperatura e umidade relativa do ar. Se os números não forem tão evidentes naturalmente o gráfico gerado por estes

¹ Aluno do Centro Universitário Internacional UNINTER. Artigo apresentado como Trabalho de Conclusão do Curso de Bacharelado em Matemática 2018/01.

² Professor Orientador no Centro Universitário Internacional UNINTER

dados tornarão o trabalho do analista mais simples. Através da figura 7 comprovamos esta afirmação.

Figura 6 : Três variáveis correlacionadas



Fonte:Elaborado pelo autor(2021)

No gráfico acima é perceptível que as grandezas radiação solar e temperatura (curvas em azul e vermelho) estão correlacionadas positivamente, pois as taxas de variação de ambas podem ser visualizadas através das inclinações das curvas que apresentam inclinação positiva. Já entre radiação solar x umidade relativa do ar (curvas em azul e amarelo) e temperatura x umidade relativa do ar (curvas em vermelho e amarelo) percebe-se uma relação inversa (correlação negativa).

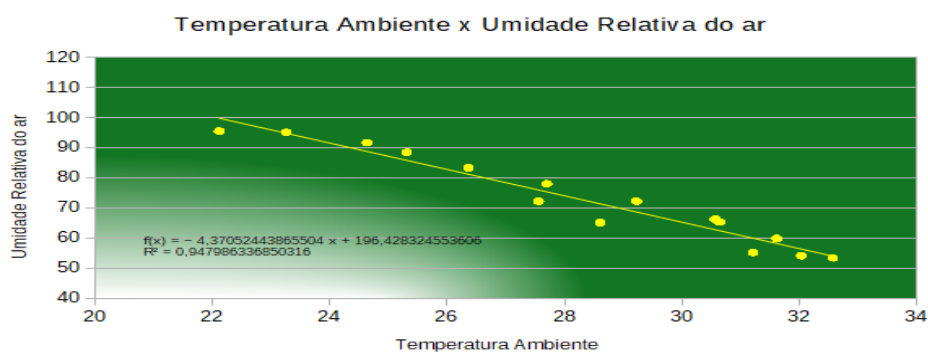
Os coeficientes de correlação para estas grandezas são iguais a $\rho = -0,535402799674598$ (umidade relativa do ar x radiação solar), $\rho = -0,973645899108252$ (umidade relativa do ar x temperatura) e $\rho = 0,674632950006845$ (radiação solar e temperatura ambiente). Estes valores comprovam a análise exploratória gráfica visualizada na figura.

¹ Aluno do Centro Universitário Internacional UNINTER. Artigo apresentado como Trabalho de Conclusão do Curso de Bacharelado em Matemática 2018/01.

² Professor Orientador no Centro Universitário Internacional UNINTER

Outra forma útil, já utilizada, é gerar retas de regressão ou utilizar a interpolação dos pontos tabelados. Para isto é feito um gráfico tipo dispersão primeiramente utilizando-se as variáveis temperatura x umidade relativa do ar. A escolha destas variáveis deu-se pelo fato de possuírem forte correlação negativa e também coeficiente de determinação, sendo do ponto de vista teórico, adequada a utilização de regressão linear. O gráfico da figura 7 apresenta este resultado:

Figura 7 : Reta de regressão com alta qualidade de ajuste



Fonte: Elaborado pelo autor (2021)

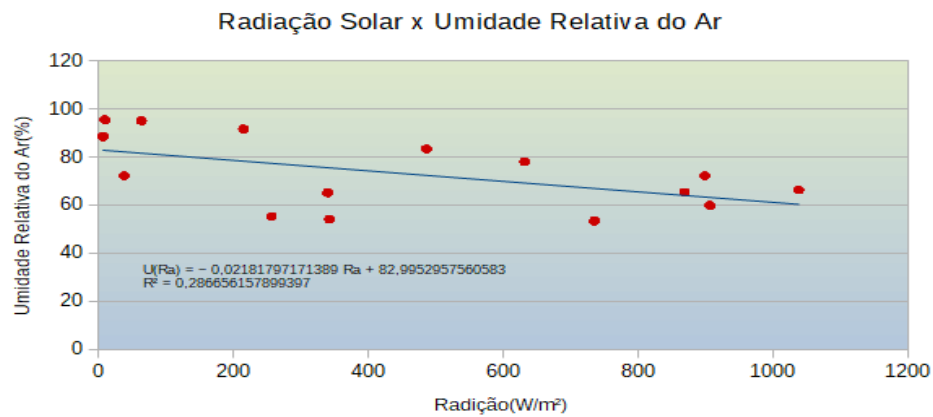
O tipo de resultado obtido acima permite razoável previsibilidade dentro desta faixa de intervalo específica. Visualmente é perceptível que a reta está muito próxima de vários pontos do gráfico o que, do ponto de vista estatístico, é corroborado pelo coeficiente de determinação, que nada mais é que o quadrado do coeficiente de correlação [$R^2 = \rho^2 = (-0,973645899108252)^2 = 0,947986336850316$].

Já a relação entre radiação e umidade não é tão fortemente correlacionada como comprovaremos novamente pelo gráfico. Veja-se:

Figura 8: Reta de regressão com ajuste de menor qualidade

¹ Aluno do Centro Universitário Internacional UNINTER. Artigo apresentado como Trabalho de Conclusão do Curso de Bacharelado em Matemática 2018/01.

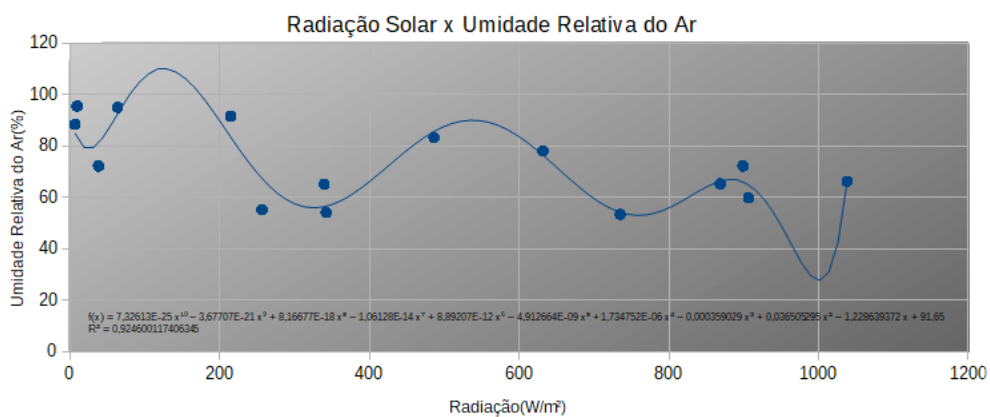
² Professor Orientador no Centro Universitário Internacional UNINTER



Fonte:Elaborado pelo autor (2021)

A situação mais recomendada para aproximar melhor os pontos e reduzir os erros seria a interpolação polinomial. O polinômio abaixo foi obtido escolhendo-se um ponto inicial determinado pelo analista, sendo que deste modo é possível avaliar o melhor ajuste visualmente. Isto significa que é possível “forçar” o ponto mais conveniente e observar o resultado através do gráfico e do coeficiente de determinação que já é obtido automaticamente bastando confirmar esta opção no software. Desta forma é possível uma avaliação prévia dos resultados antes de assumir qual melhor curva se adapta aos dados.

Figura 9: Interpolação polinomial



Fonte:Elaborado pelo autor(2021)

3. METODOLOGIA

¹ Aluno do Centro Universitário Internacional UNINTER. Artigo apresentado como Trabalho de Conclusão do Curso de Bacharelado em Matemática 2018/01.

² Professor Orientador no Centro Universitário Internacional UNINTER

Um dos momentos mais delicados e decisivos para o pesquisador iniciante é exatamente classificar, dentro de critérios bem definidos, qual o seu tipo de pesquisa, dentro de uma ampla estrutura existente, a fim de manter-se o rigor científico. Além disso,

[...] primeiro ele deve escolher, entre os diversos tipos de pesquisa, aquele que melhor se enquadra na população a ser estudada e que melhor atende aos seus objetivos; segundo, definir o melhor delineamento a ser empregado para que os objetivos possam ser alcançados. (...) (FONTENELLES, SIMÕES, FARIAS, FONTENELLES, 2009, p. 5)

Neste contexto desenvolveu-se uma pesquisa básica quanto a sua finalidade, quantitativa quanto a abordagem e descritiva quanto a sua complexidade. Ainda seguindo os critérios de definição no que diz respeito aos objetivos e procedimentos técnicos, pode-se classificá-la como explicativa.

Quanto a representação dos dados aqui utilizados o autor optou basicamente por duas formas: a representação tabular e gráfica. Estas formas de representação são as mais amplamente utilizadas devido a sua fácil construção em softwares estatísticos. Assim,

[...] Apresentação tabular numérica de dados é a representação das informações por intermédio de uma tabela. Uma tabela é uma maneira bastante eficiente de mostrar os dados levantados e que facilita a compreensão e interpretação dos dados (...) (MEDRI, 2011, p. 15).

[...] O recurso gráfico é extremamente importante para a visualização de relações numéricas. É impressionante com um simples gráfico pode ajudar na compreensão de relações (...) (BLOCH, 2004, p. 1)

Embasando-se nestes conhecimentos o autor utilizou-se de dados fornecidos por uma estação meteorológica automatizada (Datalogger) localizada no aeroporto do município de Candiota, no Rio Grande do Sul. Estes dados são coletados vinte e quatro horas e transmitidos via Internet para o Setor de Meio Ambiente de uma Usina Termoelétrica no mesmo município. Escolheu-se uma amostra referente ao dia primeiro de janeiro de 2017, contendo registros de um período de quatorze horas consecutivas deste mesmo dia.

Por conveniência, foram escolhidas as variáveis temperatura, radiação solar e umidade relativa do ar de forma a permitir o foco do trabalho na aplicação dos

¹ Aluno do Centro Universitário Internacional UNINTER. Artigo apresentado como Trabalho de Conclusão do Curso de Bacharelado em Matemática 2018/01.

² Professor Orientador no Centro Universitário Internacional UNINTER

métodos desejados. Buscou-se visualizar nestes conjuntos de dados possíveis relações de fácil demonstração do ponto de vista gráfico, pois é trivial pensar que o índice de radiação solar influencia diretamente na temperatura ambiente, na umidade relativa do ar ou ainda, que todas estas variáveis meteorológicas estejam relacionadas.

Foi utilizada uma análise bivariada onde tabulou-se os dados e calculou-se o coeficiente de correlação afim de obter-se as primeiras informações úteis. As tabelas em si servem para um primeiro reconhecimento de padrões de associação e, em seguida, os gráficos gerados a partir destas tabelas vieram a servir de apoio aos resultados numéricos aumentando, desta forma, o encadeamento lógico e o raciocínio indutivo. É importante salientar que o objetivo deste trabalho é apontar para o valor, e a importância dos métodos da AED e não a modelagem propriamente.

Com isso,

[...] As observações experimentais, quando inseridas no contexto científico, tem como objetivo fundamental permitir que se compreendam os vínculos que existem (e se de fato existem) entre as diversas variáveis que compõem um problema. O puro exercício de observação e armazenamento de dados experimentais não constitui um processo científico (...) (PINTO, SCHWAAB, 2007. p. 202).

Para complementar a abordagem foram utilizados a regressão linear e a interpolação polinomial com os correspondentes coeficientes de correlação. Estes métodos fazem parte de pacotes estatísticos existentes nas planilhas eletrônicas podendo ser pagas ou de uso livre. Todos estes recursos são de fácil utilização e encontra-se farto material explicativo na forma de tutoriais na internet. A análise iniciada através da AED dentro de um ambiente computacional, forma um conjunto básico para pavimentar o caminho do pesquisador em direção a inferência estatística ou modelagem matemática.

4. CONSIDERAÇÕES FINAIS

As ferramentas computacionais e métodos gráficos utilizados neste trabalho são de uso corrente para qualquer analista estatístico seja em nível iniciante (estudante) ou profissional. Mesmo assim, pode passar despercebido a forma indispensável que tais recursos assumem no cotidiano destes. Ficou bastante

¹ Aluno do Centro Universitário Internacional UNINTER. Artigo apresentado como Trabalho de Conclusão do Curso de Bacharelado em Matemática 2018/01.

² Professor Orientador no Centro Universitário Internacional UNINTER

evidente que, as possibilidades que surgem para o tratamento prévio de grandes volumes de dados através das ferramentas estatísticas disponíveis nas planilhas eletrônicas, são amplamente úteis e aplicáveis a casos reais.

Embora tenham sido utilizados pequenas amostras de dados e de forma bem controlada, a ênfase aqui apoia-se naquilo que os métodos gráficos podem oferecer, muito antes de partirmos para os recursos mais avançados da inferência estatística. De qualquer modo ficou claro que a simples tabulação de dados, já descortina, para um olhar atento, informações que os números em sua ordem de apresentação carregam.

Estas informações vão tomando forma a partir do momento em que os dados são dispostos em tabelas ou gráficos iluminando o caminho a frente conforme o comportamento demonstrado, específico de cada caso. Os objetivos específicos deste trabalho, aqui aplicados, nortearam corretamente a pré análise dos dados e demonstraram as correlações existentes, seja através da geração de gráficos de dispersão ou linha, seja pela utilização de ferramentas estatísticas computacionais de fácil utilização.

Ficou claro para o autor que a análise exploratória de dados do modo como é executada atualmente, ou seja, através de ferramentas estatísticas computacionais é condição *sine qua non* para todos que desejam adentrar no universo da ciência de dados. Se for um estudante que necessita do poder da estatística obterá um guia amigável que abrirá horizontes para caminhos mais avançados como o da modelagem matemática.

No caso de profissionais ou matemáticos, sem dúvida oferecerá um leque de possibilidades criativas tanto de modelos possíveis de aplicação como trará um ganho de performance simplesmente por reduzir o tempo de trabalho. A experiência e a necessidade do analista é que determinará os resultados.

REFERÊNCIAS

BARBETTA, Pedro Alberto; REIS, Marcelo Menezes; BORNIA, Antonio Cesar. **ESTATÍSTICA PARA CURSOS DE ENGENHARIA E INFORMÁTICA**. Editora Atlas SA, 2010.

BLOCH, S.C. **Excel para Engenheiros e Cientistas**. LTC, 2003.

BRUCE, Piter, BRUCE, Andrew. **Estatística Prática para Cientistas de Dados-50**

¹ Aluno do Centro Universitário Internacional UNINTER. Artigo apresentado como Trabalho de Conclusão do Curso de Bacharelado em Matemática 2018/01.

² Professor Orientador no Centro Universitário Internacional UNINTER

CHAMBERS, John M; CLEVELAND, Willian S; KLEINER, Beat; TUKEY, Paul A. **Graphical Methods for Data Analysis**. CRC Press TAYLOR & FRANCIS GROUP, 2018.
Conceitos. Essenciais, Alta Books, 2019.

FONTENELLES, José Mauro; SIMÕES, Marilda Garcia; FARIAS, Samantha Hasegawa; FONTENELLES, Renata Garcia Simões. **Metodologia da Pesquisa Científica: Diretrizes para a Elaboração de um Protocolo de Pesquisa**. Universidade da Amazônia – UNAMA, 2009.

GUIMARÃES, Paulo Ricardo Bittencourt. **Métodos Quantitativos Estatísticos**. IESDE Brasil S.A, 2008.

LANDIM, Flávia. **Análise Exploratória Bivariada**. PAPMEM-janeiro, 2020.
LAPPONI, Juan Carlos. **Estatística Usando Excel**. Elsevier, 2005.

MARIETTO, Maria das Graças Bruno; MINAMI, Mário; WESTERE, Pieter Willen (Organizadores). **Bases Computacionais da Ciência**. UFABC, 2013.

MARQUESONE, Rosângela. **Big Data : Técnicas e Tecnologias para extração de valor dos dados**. Casa do Código, 2017.

MEDRI, Waldir. **Curso de Especialização “Lato Sensu” em Estatística – ANÁLISE EXPLORATÓRIA DE DADOS**. Centro de Ciências Exatas -Departamento de Estatística, UNIVERSIDADE ESTADUAL DE LONDRINA, 2011.

MORETIN, Pedro A, BUSSAB, Wilton de O. **Estatística Básica**. Saraiva, 2004.

PINHEIRO, João Ismael D; DA CUNHA, Sonia Baptista; CARVAJAL, Santiago Ramírez. **Estatística Básica : A arte de trabalhar com dados**. Campus, 2009.
SCHWAAB, Marcio; PINTO, José Carlos. **Análise de Dados Experimentais I- Fundamentos de Estatística e Estimação de Parâmetros**. e-papers, 2007.

¹ Aluno do Centro Universitário Internacional UNINTER. Artigo apresentado como Trabalho de Conclusão do Curso de Bacharelado em Matemática 2018/01.

² Professor Orientador no Centro Universitário Internacional UNINTER