

ConEsPro: UM NOVO PROGRAMA PARA APLICAÇÕES DE FERRAMENTAS DO CONTROLE ESTATÍSTICO DE PROCESSOS

Victor Borges Zema Rosa¹ – victorbzr123@gmail.com

Sarah Arvelos Altino¹ – sarah.arvelos@ufu.br

¹ Universidade Federal de Uberlândia, Faculdade de Engenharia Química – Uberlândia, MG, Brasil

Resumo. O desenvolvimento de novas ferramentas computacionais são importantes para auxiliar os estudantes e engenheiros na temática ligado ao controle estatístico de processos. Para maior entendimento e aplicação das cartas de controle e análise estatística, o objetivo do trabalho foi o desenvolvimento de uma aplicação de código aberto, escrita na linguagem Python, com uma interface gráfica simplificada e intuitiva. A aplicação, denominada ConEsPro, foi preparada para realizar a construção das ferramentas de controle estatístico como as cartas de Shewhart, X barra e R, além das principais carta de controle por atributo, p, np, C e U, além das cartas especiais de média móvel e soma acumulada, EWMA e CUSUM. Para as análises estatísticas, foram implementados os testes de normalidade de Anderson-Darling, Shapiro-Wilk e Kolmogorov-Smirnov, além dos diagramas de dispersão, Pareto e histogramas.

Palavras-Chave: Controle estatístico de processos, Ferramentas da qualidade, Estatística para engenheiros.

1. INTRODUÇÃO

O controle estatístico de processo (CEP) é aplicado de forma ampla em diversas indústrias de manufatura. Entretanto, a aplicação de análises de capacidade (ACP) com a utilização das cartas de controle (CC) são negligenciadas ou realizadas de forma pouco aprofundada. Dessa forma, os engenheiros de controle comumente dirigem seu foco para aplicações de análises quantitativas dos dados coletados com utilização de softwares comerciais como SPSS, SAS ou Minitab. Em decorrência desse fato, a análise proativa dos dados, que apresenta grande importância para a manutenção de uma visibilidade efetiva das análises do processo, é negligenciada (CHOI; NA; YANG, 2020).

A produção e utilização de *softwares* Open Source (OSS) se apresenta como uma alternativa aos *softwares* apresentados por proverem maior flexibilidade tanto na criação, quanto na adaptação de códigos, sendo que estes são incentivados a serem liberados para domínio público (BROWN; BOOCH, 2002). Além disso, os OSS, em geral, são mais criativos, o que leva a um desenvolvimento acelerado de novos recursos, podendo ser implementados para necessidades específicas dos usuários.(PAULSON; SUCCI; EBERLEIN, 2004).

O objetivo geral do presente trabalho foi a apresentação de um *software* de código aberto capaz de aplicar e analisar automaticamente ferramentas da qualidade e realizar cálculos de estatística básica. Tendo em vista a abrangência do CEP espera-se que as rotinas sejam o ponto de partida para outros desenvolvedores para a criação e implementação de novas cartas

de controle e ferramentas estatísticas. O ConEsPro, na versão aqui apresentada (v.1.0), dispõe de cálculos de:

- Testes de normalidade de Anderson-Darling, de Kolmogorov-Smirnov e de Shapiro-Wilk;
- Cartas de controle de variáveis: para média e amplitude; para média e desvio-padrão e considerando valores individuais;
- Cartas de controle de atributos para: fração de defeituosos; número de defeituosos; número de defeitos por amostra; número de defeitos por unidade;
- Capacidade de processos observando limites de especificação bilaterais ou unilaterais;
- Gráficos de média móvel exponencialmente ponderada; de soma acumulada; em forma de histograma; de Pareto.

2. REQUISITOS DO SISTEMA E IMPLEMENTAÇÃO

ConEsPro (versão 1.0) foi desenvolvido em linguagem Python como um aplicativo com interface gráfica de usuário (GUI). Faz uso de python versão 3.0 ou superior. Python é uma ideia do programador holandês Guido van Rossum, que concebeu a linguagem na década de 1980, como um projeto de passatempo e assim chamado devido ao mesmo ser um entusiástico fã de programa de televisão “Monty Python’s Flying Circus”. Desde então, Python se tornou uma das linguagens de programação mais populares e robustas, com grande foco em legibilidade e expressividade (McCaffrey 2020). Python é tecnicamente descrita como uma linguagem de programação, “interpretada”, “digitada dinamicamente”, “de alto nível”. Por isso, trata-se de um tipo de linguagem mais simples de aprender (JONES; JONES; SIN, 2019; ROCHA; FERREIRA, 2018).

Um dos fatores que traz destaque à linguagem Python são justamente suas bibliotecas e pacotes, que proporcionam muita simplicidade as aplicações, além de garantir scripts descomplicados e eficientes. Os pesquisadores de hoje, em todas as disciplinas acadêmicas, muitas vezes precisam escrever código de computador a fim de coletar e processar dados, realizar testes estatísticos, executar simulações ou desenhar figuras. Códigos escritos em Python, linguagem a ser utilizada para o desenvolvimento deste programa são facilmente interpretados e oferecem uma maneira de criar um programa computacional que possa ser prontamente lido e replicado (RANGLES et al., 2017).

As bibliotecas utilizadas para a implementação foram: NumPy, SciPy, Pandas, sklearn e matplotlib. A biblioteca NumPy foi utilizada por fornecer um objeto de matriz multidimensional, além de uma variedade de rotinas para operações rápidas em matrizes, incluindo matemática, lógica, classificação, seleção, operações estatísticas básicas e muito mais. O pacote SciPy, por sua vez, foi responsável por adicionar um conjunto de algoritmos matemáticos e funções de conveniência construídos na extensão NumPy. Este, foi selecionado devido a sua significância em relação sessão interativa do Python, fornecendo ao usuário comandos e classes de alto nível para manipulação e visualização de dados. Além disso, foi utilizado para a implementação dos testes de normalidade dos dados.

Além destas, o pacote Pandas foi de grande importância para a aplicação pelo seu grande poder ao trabalhar com dados tabulares, como dados armazenados em planilhas ou bancos de dados. Foram utilizadas as tabelas chamadas de “DataFrames” para realizar uma gama de operações matemáticas e estatísticas. Além disso, o Pandas, por suportar a integração com muitos formatos de arquivo ou fontes de dados (csv, excel, sql, json, parquet, ...) foi utilizado para criar uma ligação direta entre o ConEsPro e os dados tabulares, utilizados para validação das ferramentas.

Já a biblioteca Matplotlib (HUNTER, 2007) foi aplicada por se basear em um conjunto de gráficos 2D para desenvolvimento de scripts interativos e geração de imagens com

qualidade de publicação em interfaces de usuário e sistemas operacionais. Dessa forma, foi utilizada nas ferramentas do ConEsPro para plotar os dados obtidos pelo uso do pacote Pandas. Por fim, a biblioteca Sklearn, ou Scikit-learn, por sua aplicabilidade em problemas de regressão e aprendizado de máquina, foi fundamental para a aplicação do diagrama de dispersão.

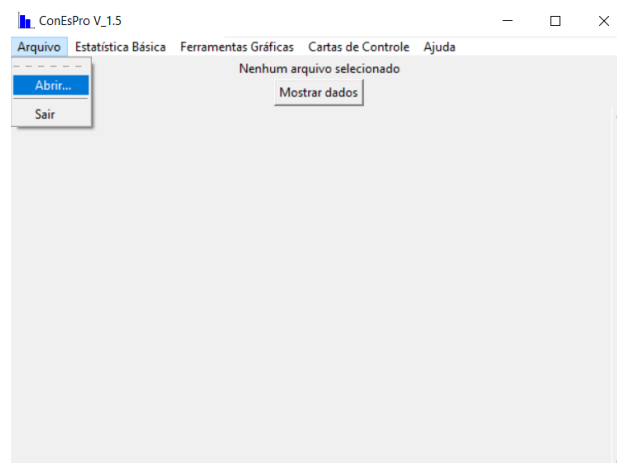
O ConEsPro usa o pacote Tkinter para criação de sua interface gráfica de usuário (GUI). A fim de facilitar e abranger o uso da aplicação por usuários não familiarizados com a linguagem de programação, a biblioteca foi utilizada na criação de uma interface gráfica intuitiva. O Tkinter foi criado para interligar a linguagem Tcl com o Python, trazendo uma variedade de ferramentas e widgets de programação orientada a objetos para a criação de aplicações com interfaces visuais de maneira prática.

O código fonte do ConEsPro está disponível em <https://github.com/victorbzr/ConEsPro.git>.

3. VISÃO GERAL DO PROGRAMA

O ConEsPro foi preparado para a entrada de arquivos em excel (.xlsx) ou em valores separados por vírgula (.csv). Para abrir um arquivo ou base de dados no programa, deve-se selecionar a opção “Arquivo” seguida de “Abrir”, como apresentado na Figura 1.

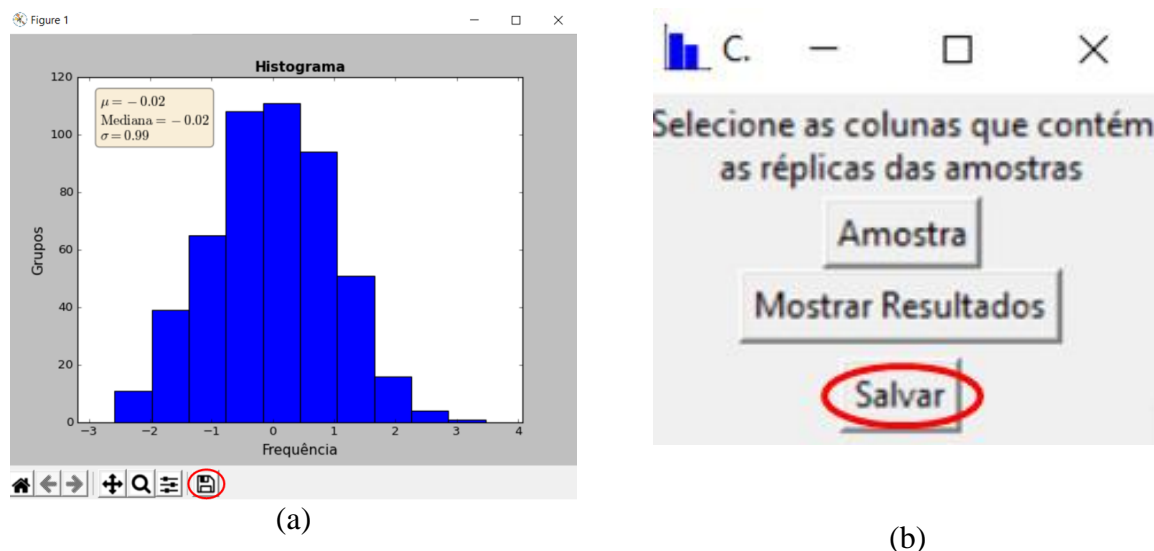
Figura 1 - Importação de dados no ConEsPro.



Após esses passos, será aberta uma nova janela onde se deve selecionar o arquivo desejado na máquina e selecionar novamente a opção “Abrir”. Os dados são convertidos em um “dataframe” da biblioteca Pandas e, assim, estará pronto para ser trabalhado pelo ConEsPro.

Foi criada uma ferramenta de salvamento para que, ao final da utilização de algumas das ferramentas, os resultados pudessem ser salvos. Além dos gráficos plotados serem salvos na própria figura, como visto na Figura 2 (a), os dados calculados podem ser salvos em um arquivo de Excel (.xlsx) ou no formato de valores separados por vírgula (.csv). Para este, deve-se selecionar a opção salvar no menu da ferramenta, procedimento demonstrado para a carta de controle \bar{X} na Figura 2 (b).

Figura 2 - (a) Botão de salvamento de figuras geradas pelo ConEsPro; (b) Botão de salvamento de dados gerados.



3. ESTUDOS DE CASO

Os estudos de caso foram realizados com a finalidade de demonstrar e validar as funcionalidades do ConEsPro versão 1.0. Dessa forma, todas as ferramentas foram utilizadas em suas configurações padrão, exceto quando especificadas as alterações.

3.1. Testes de normalidade e análise estatística básica

Para examinar os testes de normalidade foi criado um conjunto de dados de teste. Foram geradas amostras de números aleatórios extraídos de uma distribuição gaussiana. A escolha de números aleatórios gaussianos para o conjunto de dados de teste significa que esperamos que cada teste identifique corretamente a distribuição; no entanto, o tamanho pequeno da amostra pode introduzir algum ruído nos resultados.

Foram criados conjuntos de números de diferentes tamanhos: 5, 10, 20, 30, 50, 100, 300 e 500. O objetivo foi avaliar o fato de que testes de normalidade sofrem influência do tamanho amostral quanto à sua eficiência. Para todas as amostras e testes, os valores encontrados implicaram em aceitação da hipótese de normalidade dos dados.

A fim de se ilustrar o efeito de indução ao erro de um tamanho amostral reduzido, foi utilizada a ferramenta histograma do ConEsPro para gerar a representação visual da distribuição de cada uma das amostras estudadas com divisão em 5 grupos. Sendo assim, os histogramas gerados pela aplicação foram apresentados na Figura 3, para as amostras de tamanho n igual a 5 e 300. Dessa forma, evidenciou-se a dificuldade de identificação de um padrão visual de normalidade dos dados para as amostras menores.

Para realizar o teste do diagrama de dispersão, foram utilizados os dados de (Samohyl, 2009). Nesta estão dispostas relações entre volume, V , e pressão, P . A fim de descobrir se há correlação linear entre as duas variáveis, foi plotado o diagrama de dispersão apresentado na Figura 44.

Notou-se, portanto, que não houve um bom ajuste dos dados e assim, não há correlação linear entre as duas variáveis. Entretanto, pode ser feita a manipulação dessas variáveis para atingir uma correlação linear. Dessa forma, foi plotado o diagrama de dispersão para o inverso de V versus P , como apresentado na **Erro! Fonte de referência não**

encontrada., onde foi observado um bom ajuste aos dados e, assim, confirmou-se a existência de uma correlação linear entre as variáveis ajustadas.

Figura 3 – Histogramas das amostras de tamanhos $n=$ (a) 10 e (b) 300.

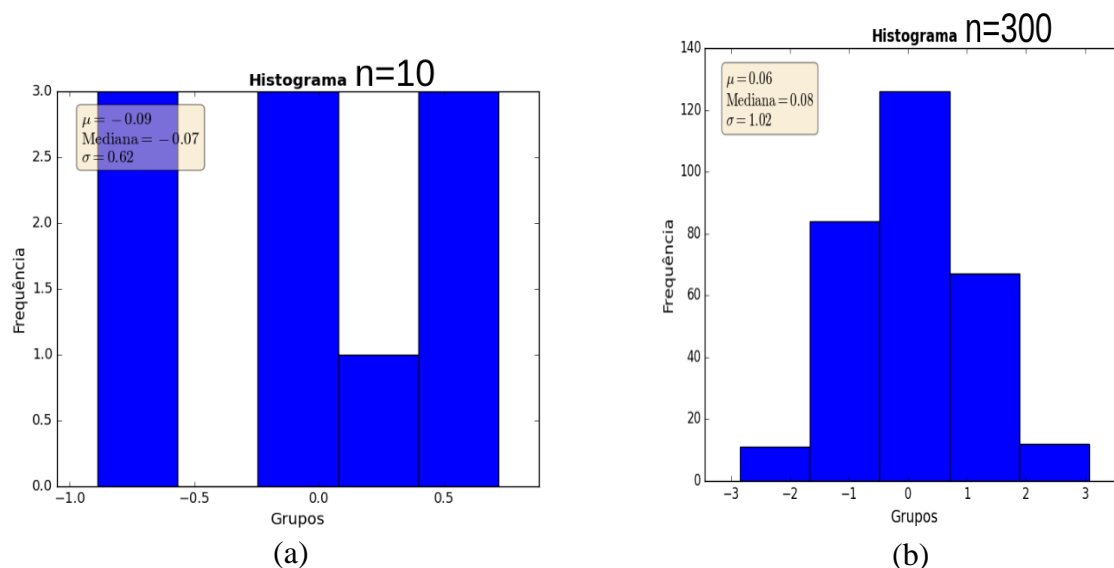
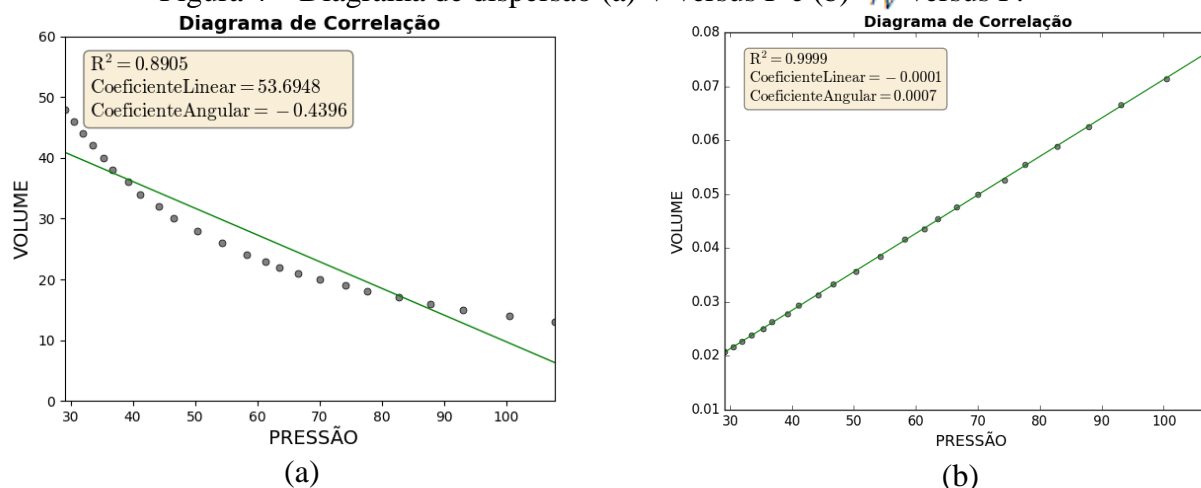


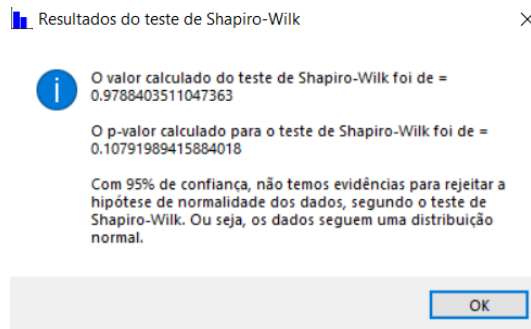
Figura 4 – Diagrama de dispersão (a) V versus P e (b) $1/V$ versus P.



3.2. Cartas de controle de variáveis

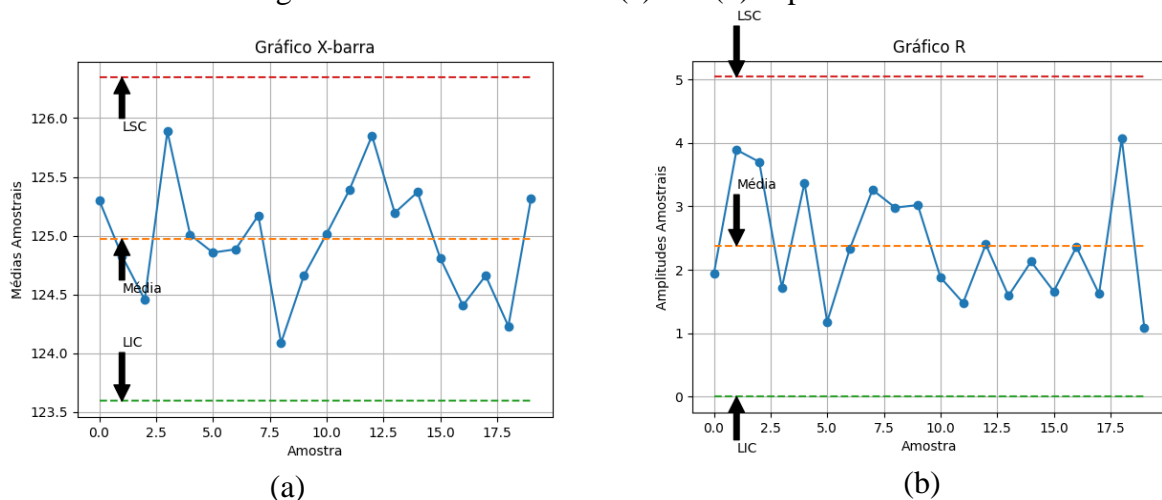
Para realizar os testes nas cartas de controle por variáveis, foram plotados os dados adaptados de Hu e Castagliola (2019), onde foram dispostos 5 medições por amostra, num total de 20 amostras. A característica da qualidade observada foi o peso do produto, considerando um processo de envase de iogurte com peso esperado de 125 g por copo. Os dados para a fase 1 foram apresentaram uma média, μ , de 124,97 g. Além disso, foi realizado o teste de Shapiro-Wilk que resultou na normalidade dos dados, como apresentado na **Erro! Fonte de referência não encontrada.5.**

Figura 5 – Resultado do teste de normalidade para a fase 1.



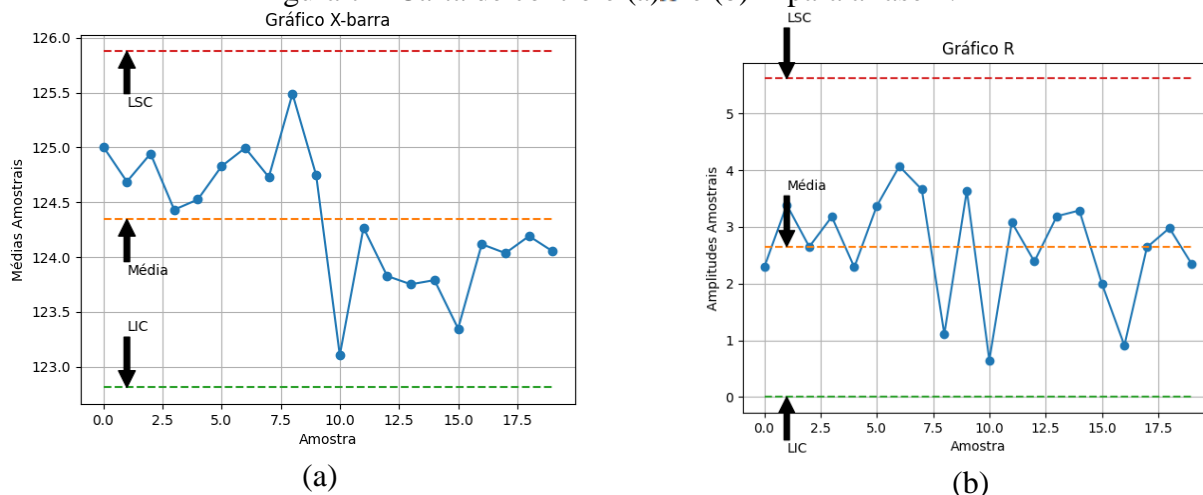
Após serem criadas as planilhas de dados tipo .csv, as informações das mesmas foram plotadas nas cartas de controle de Shewhart \bar{X} , apresentada na Figura 6 (a), e R na Figura 6 (b). Os resultados para ambas as cartas apresentam uma variabilidade possivelmente não natural, devido a presença de pontos sequenciais repetidamente no mesmo lado da reta que representa a média geral, além de tendências de crescimento e decrescimento na carta \bar{X} .

Figura 6 – Carta de controle (a) \bar{X} e (b) R para a Fase 1.



Para a fase de testes dois, as primeiras dez amostras se encontram, supostamente, sob controle e as dez amostras seguintes apresentaram um leve desvio na média. A remoção destes dados e inclusão de novas medidas (Fase 2) fez com que o teste de Shapiro-Wilk para a acusasse normalidade dos dados. A confecção dos gráficos de controle para esta nova fase podem permitir ao engenheiro a identificação visual da redução na média do processo, demonstrada por pontos sequencialmente abaixo da média para a carta \bar{X} , bem como o decremento brusco nos pesos médios das amostras a partir da décima amostra, como mostra a Figura 7(a). Para a carta R (Figura 7(b)), foi observada uma tendência de decrescimento global, indicando também a redução na média do processo. Apesar de indicar visualmente as tendências descritas, os limites de controle de três sigma foram insuficientes para detectar as anomalias no processo, demonstrando a necessidade de aperfeiçoamento das cartas implementadas.

Figura 7 – Carta de controle (a) \bar{X} e (b) R para a fase 2.

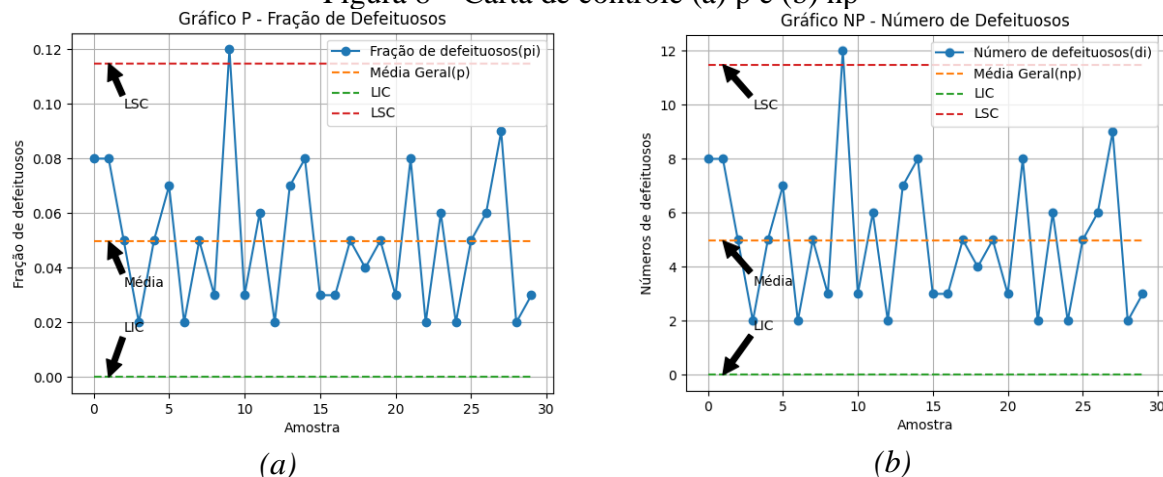


3.3. Cartas de controle de atributos

Para a validação das cartas de controle por atributos, foram utilizados os dados adaptados de Samohyl (2009), como apresentados na **Erro! Fonte de referência não encontrada.**, para as cartas p e np . Os dados representam o número de unidades defeituosas em amostras de tamanho n igual a 100, sendo que a média global dos dados foi igual a 4,97.

Os resultados foram apresentados na Figura 8 (a), para a carta de controle p , e na Figura 8 (b), para a carta de controle np . Dessa forma, ambas as cartas apresentaram o descontrole do processo através dos limites de controle de 3 sigma, ou seja, apresentaram pelo menos um ponto fora dos limites de controle.

Figura 8 – Carta de controle (a) p e (b) np

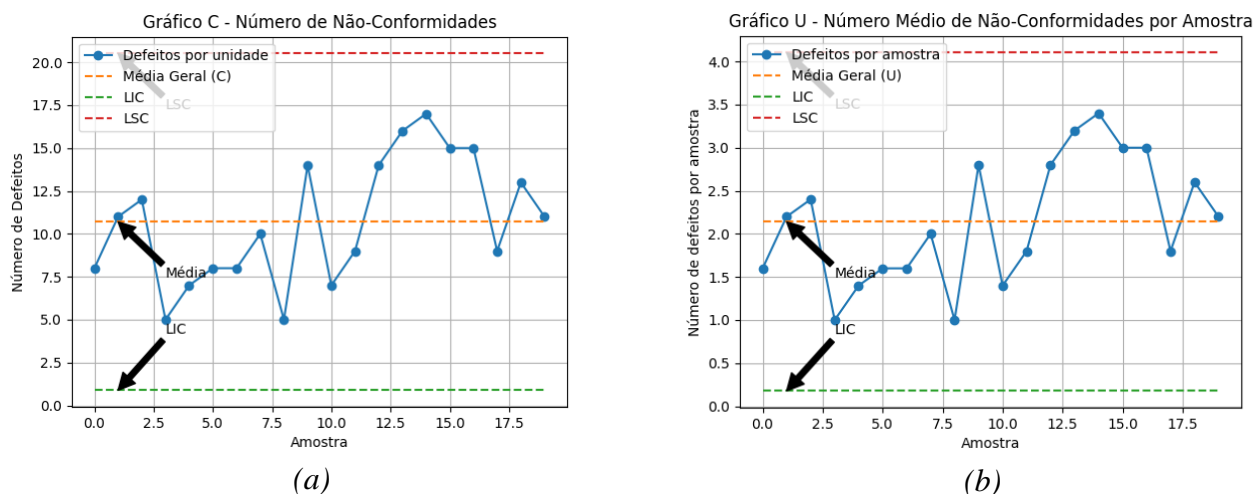


Para os testes das cartas de controle C , que necessita de dados com tamanho amostral constante (Montgomery, 2004), e U , foram utilizados os dados adaptados de Aebtarm e Bouguila (2011)

Os resultados dos testes para as cartas C , como apresentado na Figura 10 (a), e para a carta U , apresentados na Figura 10 (b), mostraram uma tendência de desvio da média para um valor mais elevado. Além disso, notou-se a presença de pontos consecutivos em um mesmo

lado da média, podendo indicar a existência de variações com causa atribuível. Apesar disso, com os limites de controle de 3 sigma não foi acusado descontrolo do processo.

Figura 9 – Carta de controle (a) C e (b) U.



3.4. Cartas de controle especiais

Para os testes das cartas de controle especiais, assim como o gráfico CUSUM, foram utilizados os dados adaptados de Wetherill (1977). Os dados representam 150 amostras de um produto de uma planta química em que ocorre uma reação exotérmica dos reagentes A e B. Para cada amostra, as eficiências da planta E(%) foram apresentadas pelo autor

Além disso, as amostras destacadas representam ocorrências fontes de variabilidade do processo em que:

- 22, 66, 100 e 138 representam uma nova batelada do reagente B;
- 42, 79 e 118 representam uma nova batelada do reagente A;
- 53 representa parada da planta;
- 126 a 134 representam um bloqueio na linha de água de resfriamento, resultando em altas temperaturas.

O gráfico CUSUM gerado pelo ConEsPro é apresentado na Figura 11 (a), onde foram observadas variações significativas para a maior parte das ocorrências descritas. Portanto, ao seguir para uma análise considerando as cartas CUSUM tabular completa (superior mais inferior) (Figura 11(b)) e MMEP (Figura 12), observa-se, principalmente na carta MMEP, os pontos onde ocorreram variações significativas no processo.

Sendo assim, foi identificado uma redução brusca da eficiência do processo com a parada da planta na amostra 53 e um aumento brusco na eficiência após as novas bateladas de reagentes, principalmente para o reagente A. Além disso, notou-se que o bloqueio na linha de água de resfriamento (126 a 134) não demonstrou ser fonte significativa de variabilidade no processo.

Figura 10 – Gráfico (a) CUSUM e (b) CUSUM tabular completa.

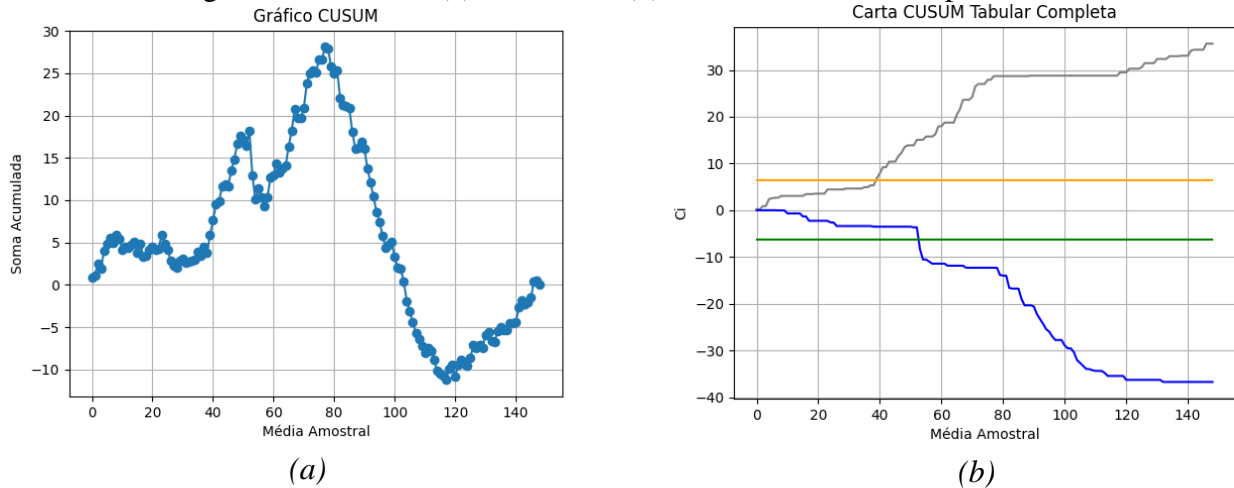
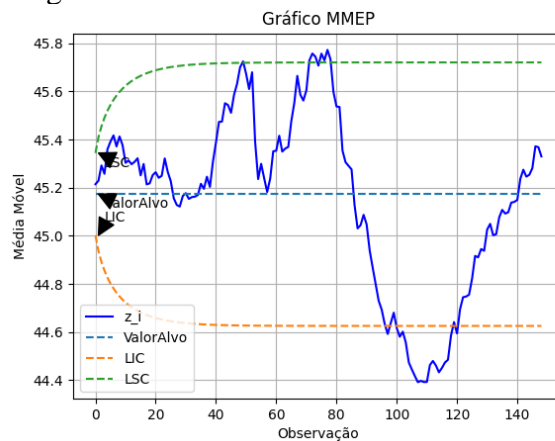


Figura 11 – Carta de controle MMEP.



CONSIDERAÇÕES FINAIS

O objetivo do presente trabalho de desenvolver um programa *Open Source* que implementasse de forma simplificada e intuitiva alguns problemas clássicos do controle estatístico de processos. Foram apresentados estudos de caso que mostra que o programa é relativamente robusto para cálculos com pequenas e médias amostras. Por se tratar de um código aberto, o ConEsPro pode ser utilizado em sua forma padrão por profissionais e universidades. Além disso, com pequenas mudanças na sua estrutura, novas ferramentas e cartas de controle podem ser implementadas para o uso em situações mais específicas, através de mudanças simples de serem implementadas no código.

REFERÊNCIAS

- Aebtarm, S.; Bouguila, N. (2011), “An empirical evaluation of attribute control charts for monitoring defects”, *Expert Systems with Applications*, vol 38, n. 6, 7869–7880.
- Hu, X. L.; Castagliola, P. (2019), “A re-evaluation of the run rules xbar chart when the process parameters are unknown”. *Quality Technology and Quantitative Management*, vol 16, n. 6, 696–725.
- Montgomery, D. C. (2004), Introdução ao controle estatístico da qualidade. 4° ed., LTC, São Paulo.
- Montgomery, D. C. (2012), “Introduction to statistical quality control”. 7° ed., Wiley.

Samohyl, R. W. (2009), Controle Estatístico de Qualidade. 1° ed., Elsevier, São Paulo.
Wetherill, G. B. (2013), “Sampling Inspection and Quality Control”. 2° ed., Springer US, Alemanha.

APPENDIX A

ConEsPro: A NEW SOFTWARE FOR STATISTICAL PROCESS CONTROL

The development of new computational tools is important to help students in the subject of statistical process control. For greater understanding and application of control charts and statistical analysis, the objective of the work was the development of an open source application, written in Python language, with a simplified and intuitive graphical interface. The application, named ConEsPro, was prepared to automatically carry out the construction of statistical control tools such as Shewhart’s \bar{X} and R charts. In addition, the main charts for attribute control such as p , np , C and U charts and the special moving average and accumulated sum charts, EWMA and CUSUM, were implemented. For statistical analyses, Anderson-Darling, Shapiro-Wilk and Kolmogorov-Smirnov formal normality tests were implemented, in addition to scatter plots, Pareto diagrams and histograms.

Keywords: *Statistical process control, Quality tools, Statistics for engineers.*