

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
DEPARTAMENTO DE FÍSICA
BACHARELADO EM ENGENHARIA FÍSICA

TRABALHO FINAL DE CURSO

MARZIO LOPES AGUIAR

**Modelagem Estocástica de um Sistema Produtivo com
Factory Physics e Simulação de Monte Carlo**

São Carlos, SP

2025

MARZIO LOPES AGUIAR

**Modelagem Estocástica de um Sistema Produtivo com
Factory Physics e Simulação de Monte Carlo**

Trabalho Final de Curso apresentado ao Bacharelado em Engenharia Física da Universidade Federal de São Carlos, como requisito parcial para a obtenção do título de Bacharel em Engenharia Física.

Orientadora: Prof^a. Dr^a. Clarissa Fullin Barco de Camargo
Departamento de Engenharia de Produção
Universidade Federal de São Carlos

São Carlos

2025

Ao meu avô Miguel, com eterna saudade e gratidão.

Este trabalho é por você e para você.

“O homem nada mais é do que aquilo que ele faz de si mesmo.”

Jean-Paul Sartre, 1946.

Agradecimentos

Em primeiro lugar, agradeço a Deus, por me conceder saúde e determinação ao longo de todo este trabalho.

À minha família: à minha mãe, por todo empenho e apoio ao longo da minha trajetória; ao meu pai, por ser meu protetor e me colocar diariamente em suas orações; à minha avó Conceição, sou eternamente grato por seu amor incondicional, carinho e ternura; ao meu tio Zu, pelo orgulho que sempre demonstrou e por me considerar como um filho.

E, por fim, mas jamais menos importante, ao meu avô Miguel. Sua falta é imensurável e tudo o que eu mais queria era sua presença neste momento. Mas sei que você me guarda, me protege e estará para sempre ao meu lado. Espero te dar orgulho, vovô. Todo o seu esforço e trabalho valerão a pena.

Ao 'tio' Claudemir, gestor industrial do parque fabril utilizado como estudo de caso neste trabalho, agradeço pela confiança, disponibilidade e apoio fundamental para o desenvolvimento do estudo de caso deste trabalho.

À Maria Eduarda, minha Dudinha, agradeço por todo amor, apoio, cumplicidade e parceria. Você me inspira a crescer e torna tudo mais leve. Sem você, esta etapa não seria possível.

Aos meus amigos, em especial ao Casale, pela admiração e palavras de apoio; ao Davi, por sempre trazer alegria aos dias pesados; e ao Henrique, meu irmão de graduação, compartilhando desafios e conquistas da Engenharia Física.

Ao Gabriel Facenda, meu primeiro chefe e grande amigo, agradeço por sempre acreditar em mim e abrir portas importantes na minha vida profissional.

Agradeço também a todos os professores que contribuíram para minha formação e ao Departamento de Física da UFSCar, pela dedicação ao ensino.

À minha orientadora, Prof^a Dra. Clarissa Fullin Barco de Camargo, agradeço pelo apoio, paciência e por todas as orientações que fortaleceram este trabalho.

Por fim, sou profundamente grato a todos que, de alguma forma, fizeram parte da minha caminhada e contribuíram para que eu chegasse até aqui.

Resumo

Aguiar, Marzio. *Modelagem Estocástica de um Sistema Produtivo com Factory Physics e Simulação de Monte Carlo*. 2025. 83 f. Trabalho Final de Curso, Bacharelado em Engenharia Física, Universidade Federal de São Carlos, São Carlos, 2025.

O presente trabalho desenvolve e aplica um modelo estocástico fundamentado em Factory Physics e Teoria de Filas, implementado por meio de simulação de Monte Carlo, com o objetivo de quantificar e decompor o lead time em recursos críticos de um sistema produtivo discreto. A pesquisa foi conduzida em uma planta industrial de injeção plástica, analisando duas injetoras reais (Máquinas 10 e 11) classificadas como recursos gargalo no processo. O estudo caracteriza estatisticamente variáveis como tempos de ciclo, setups, falhas, reparos e taxas de defeito, incorporando-as em um modelo matemático calibrado a partir de dados reais do ambiente fabril. A simulação permitiu estimar distribuições probabilísticas para métricas de desempenho como lead time, throughput e work in process, além de avaliar cenários de melhoria em disponibilidade, confiabilidade e setup. Os resultados demonstraram que a variabilidade exerce efeito multiplicativo e não linear sobre o desempenho, sendo mais sensível à utilização dos recursos do que às médias individuais dos tempos. Conclui-se que a integração entre física aplicada, modelagem estatística e engenharia de produção constitui uma ferramenta robusta para diagnóstico e otimização de sistemas industriais, proporcionando base quantitativa para priorização de ações de melhoria e tomada de decisão.

Palavras-chave: Simulação de Monte Carlo. Factory Physics. Teoria de Filas. Modelagem Estocástica. Engenharia de Produção. Engenharia Física.

Abstract

Aguiar, Marzio. *Stochastic Modeling of a Production System Using Factory Physics and Monte Carlo Simulation*. 2025. 83 f. Final Graduation Project, Bachelor's Degree in Engineering Physics, Federal University of São Carlos, São Carlos, 2025.

This work develops and applies a stochastic model grounded in Factory Physics and Queuing Theory, implemented through Monte Carlo simulation, with the objective of quantifying and decomposing lead time in critical resources of a discrete manufacturing system. The research was conducted in an industrial plastic injection facility, analyzing two real injection molding machines (Machines 10 and 11), identified as operational bottlenecks. The study statistically characterizes variables such as cycle times, setups, failures, repairs, and defect rates, integrating them into a mathematical model calibrated using real production data. The simulation enabled the estimation of probabilistic distributions for performance metrics such as lead time, throughput, and work in process, in addition to evaluating improvement scenarios involving availability, reliability, and setup. The results showed that variability exerts a multiplicative and nonlinear effect on performance, being more sensitive to resource utilization than to individual mean times. It is concluded that the integration of applied physics, statistical modeling, and production engineering forms a robust framework for diagnosing and optimizing industrial systems, providing a quantitative basis for prioritizing improvement actions and supporting decision-making.

Keywords: Monte Carlo Simulation. Factory Physics. Queueing Theory. Stochastic Modeling. Production Engineering. Engineering Physics.

Lista de Figuras

1	Curva VUT. Fonte: Factory Physics (2020)	11
2	Fluxograma do processo de injeção plástica. Fonte: adaptado de SANTOS e SOUZA (2022)	25
3	Curva de trade-off entre eficiência de setup e tempo de atravessamento (<i>lead time</i>) – Máquina 11.	49
4	Comportamento do Lead Time total para os cenários SMED (Máquina 11). . .	52
5	Curva em U para o eixo Six Sigma (Máquina 11).	54
6	Curva em U para o eixo de Manutenção Preditiva (Máquina 11).	57
7	Distribuição empírica do Lead Time médio por eixo de melhoria (Máquina 11). .	60
8	Curva de trade-off entre eficiência de setup e tempo de atravessamento (Máquina 10).	63
9	Comportamento do Lead Time total para os cenários SMED (Máquina 10). . .	64
10	Curva em U para o eixo Six Sigma (Máquina 10).	66
11	Curva em U para o eixo de Manutenção Preditiva (Máquina 10).	67
12	Distribuição empírica do Lead Time médio por eixo de melhoria (Máquina 10). .	69

Lista de Tabelas

1	Parâmetros operacionais considerados para as injetoras analisadas.	46
2	Parâmetros da base e cálculo de c_e^2 para $L = 400$ (Máquina 11).	47
3	Resultados médios da simulação de Monte Carlo por tamanho de lote (Máquina 11).	47
4	Tabela Resumo - Principais métricas médias por lote (Máquina 11).	48
5	Indicadores médios obtidos para os diferentes cenários do eixo SMED (Máquina 11).	51
6	Indicadores médios obtidos para os diferentes cenários do eixo Six Sigma (Máquina 11).	54
7	Parâmetros utilizados nos diferentes cenários de manutenção (Máquina 11). . .	56
8	Indicadores médios obtidos para os diferentes cenários do eixo de Manutenção Preditiva (Máquina 11).	57
9	Indicadores médios para o eixo SMED (Máquina 11).	60
10	Indicadores médios para o eixo Six Sigma (Máquina 11).	60
11	Indicadores médios para o eixo de Manutenção Preditiva (Máquina 11).	60
12	Parâmetros da base e cálculo de c_e^2 para $L = 400$ (Máquina 10).	62
13	Resultados médios da simulação de Monte Carlo por tamanho de lote (Máquina 10).	62
14	Tabela resumo – principais métricas médias por lote (Máquina 10).	63
15	Indicadores médios obtidos para os diferentes cenários do eixo SMED (Máquina 10).	64
16	Indicadores médios obtidos para os diferentes cenários do eixo Six Sigma (Máquina 10).	65
17	Parâmetros utilizados nos diferentes cenários de manutenção (Máquina 10). . .	66
18	Indicadores médios obtidos para os diferentes cenários do eixo de Manutenção Preditiva (Máquina 10).	67
19	Indicadores médios para o eixo SMED (Máquina 10).	68
20	Indicadores médios para o eixo Six Sigma (Máquina 10).	68

21	Indicadores médios para o eixo de Manutenção Preditiva (Máquina 10).	68
22	Decomposição do tempo de ciclo em componentes causais (cenário base).	77
23	Decomposição da variabilidade total em fontes individuais (cenário base).	78
24	Elasticidades de lead time em relação a parâmetros operacionais.	79

Sumário

1	Introdução	1
1.1	Contextualização do problema	1
1.2	Objetivo geral	3
1.3	Objetivos específicos	3
1.4	Justificativa da pesquisa	4
1.5	Organização do trabalho	4
2	Revisão bibliográfica e fundamentação teórica	6
2.1	Conceitos fundamentais de desempenho produtivo	6
2.2	Princípios do <i>Factory Physics</i>	8
2.2.1	Lei de Little e relações fundamentais	9
2.2.2	Curva VUT e dinâmica do tempo de ciclo	10
2.3	Teoria de Filas	12
2.3.1	Estrutura básica e componentes de uma fila	12
2.3.2	Aproximação de Kingman (VUT) para o modelo $G/G/1$	13
2.4	Simulação estocástica em sistemas produtivos	15
2.4.1	Tipos de simulação	16
2.5	Método de Monte Carlo e análise de risco	17
2.5.1	Técnicas de amostragem e redução de variância	17
3	Metodologia	20
3.1	Teoria do método	20
3.1.1	Estudo de caso	20
3.1.2	Simulação como método de pesquisa	21
3.1.3	Simulação de Monte Carlo	21
3.1.4	Integração dos métodos	22

4	Estudo de Caso	23
4.1	Introdução à empresa	23
4.2	Recorte do estudo: Máquinas críticas	23
4.3	Layout e processo produtivo	24
5	Modelagem Matemática e Simulação de Monte Carlo	27
5.1	Visão Geral do Modelo	27
5.1.1	Etapas Metodológicas	27
5.2	Formulação Matemática do Sistema Produtivo	28
5.2.1	Conceitos fundamentais	28
5.2.2	Interpretação física das variáveis	29
5.3	Estimação Empírica dos Parâmetros	29
5.3.1	Base de dados e pré-processamento	29
5.3.2	Execução e interpretação dos resultados	31
5.4	Correções de Confiabilidade e Setup	32
5.4.1	Disponibilidade e confiabilidade	33
5.4.2	Tempo efetivo e impacto do setup	33
5.5	Variabilidade Efetiva e Tempo de Ciclo	34
5.5.1	Tempo médio de ciclo (Factory Physics)	34
5.5.2	Implementação computacional das correções	35
5.5.3	Exemplo numérico: Máquina 11	36
5.5.4	Exemplo numérico: Máquina 10	36
5.6	Estrutura Computacional do Simulador	36
5.6.1	Fundamentos da simulação de Monte Carlo	36
5.6.2	Estrutura do código e controle de estabilidade	37
5.7	Amostragem e Redução de Variância	39
5.7.1	Amostragem em Hipercubo Latino (LHS)	40
5.7.2	Números Aleatórios Comuns (CRN)	40
5.7.3	Variáveis Antitéticas	41
5.8	Testes de Consistência e Lote Base	41

5.9	Análise por Tamanho de Lote	42
5.9.1	Comportamento das métricas principais	42
5.10	Estrutura dos DataFrames e Saídas	42
5.10.1	Análise automatizada dos resultados	43
5.11	Algoritmo Resumido	43
5.12	Limitações	45
6	Resultados	46
6.1	Resultados da Máquina 11	46
6.1.1	Validação do Modelo Base	47
6.1.2	Análise de Cenários de Melhoria	49
6.1.3	Análise Empírica dos Eixos de Melhoria - Máquina 11	60
6.2	Resultados da Máquina 10	62
6.2.1	Validação do Modelo Base	62
6.2.2	Análise dos Resultados por Eixo de Melhoria	64
6.2.3	Análise Empírica dos Eixos de Melhoria – Máquina 10	68
6.3	Análise Comparativa entre Máquinas	69
7	Conclusão e Insights	71
7.1	Resposta à Questão de Pesquisa	71
7.1.1	Mecanismos Causais	71
7.1.2	Quantificação Matemática	75
7.1.3	Decomposição Analítica: Estrutura Causal	77
7.1.4	Decomposição do Tempo de Ciclo	77
7.1.5	Decomposição de Fontes de Variabilidade	78
7.1.6	Análise de Sensibilidade: Elasticidades	79
7.2	Conclusão Final	80
8	Referências	84

1 Introdução

1.1 Contextualização do problema

A indústria de manufatura enfrenta, nas últimas décadas, uma pressão crescente por prazos de entrega cada vez mais curtos, custos competitivos e qualidade consistente em seus produtos (SLACK et al., 2019). Em um mercado globalizado e altamente dinâmico, empresas que não conseguem responder com flexibilidade e confiabilidade às demandas de seus clientes perdem rapidamente sua competitividade. Nesse contexto, o *lead time* de produção, entendido como o tempo total de atravessamento de uma ordem, desde o seu lançamento no chão de fábrica até a sua conclusão, torna-se um dos indicadores-chave de desempenho (HOPP e SPEARMAN, 2008). Isso porque ele condensa, em um único resultado, efeitos de capacidade, variabilidade e confiabilidade dos recursos.

Na prática, o *lead time* sintetiza aquilo que o cliente percebe como velocidade e regularidade de entrega, enquanto internamente funciona como um sinal claro de equilíbrio (ou desequilíbrio) entre a demanda, o fluxo de produção e os recursos disponíveis. Empresas capazes de manter *lead times* estáveis e reduzidos não apenas melhoram a satisfação de seus clientes, mas também reduzem a necessidade de estoques intermediários e aumentam a previsibilidade de suas operações. (WENDT, C., 2022).

Em sistemas produtivos discretos, como linhas de produção de injeção plástica, a operação raramente é determinística. Tempos de ciclo variam entre peças e entre turnos, setups não são exatamente repetíveis, equipamentos falham e exigem reparo, e parte da produção pode apresentar defeitos que levam ao retrabalho ou à sucata. Além disso, a chegada de ordens ou lotes às máquinas sofre influência de fatores de planejamento e de mix de produtos, introduzindo flutuações que nem sempre são suaves ou previsíveis. Essas fontes de variabilidade, quando combinadas, produzem efeitos não lineares: filas crescem de forma desproporcional, o trabalho em processo (WIP – *Work in Process*) aumenta e a capacidade efetiva percebida se reduz, sobretudo quando a taxa de utilização dos recursos se aproxima de seus limites físicos.

Dois arcabouços conceituais são especialmente úteis para explicar e quantificar esses fenômenos. A **Teoria de Filas** (GROSS e HARRIS, 2008) fornece a base probabilística para com-

preender como taxas de chegada, taxas de serviço e variabilidade se traduzem em tempos de espera e tempos de atravessamento. Já o **Factory Physics** (HOPP e SPEARMAN, 2008), organiza essas relações em leis operacionais de aplicação gerencial, como a Lei de Little, que conecta estoque em processo, fluxo e tempo, e a heurística conhecida como curva VUT (KINGMAN, 1961), que traduz a interação entre variabilidade, utilização e tempo de processamento no crescimento exponencial dos tempos de fila.

Em paralelo, conceitos de confiabilidade, como o tempo médio entre falhas (MTBF – *Mean Time Between Failures*) e o tempo médio de reparo (MTTR – *Mean Time To Repair*), bem como indicadores de qualidade, como a taxa de defeitos, ajustam a capacidade efetiva do recurso e impactam diretamente o *throughput* de peças boas (O’CONNOR, 2012). Isso significa que não apenas a taxa de produção nominal importa, mas também a capacidade de manter essa taxa em face de paradas não planejadas e perdas por qualidade.

Diante desse cenário, este trabalho adota uma abordagem **estocástica** baseada em *simulação de Monte Carlo* (RUBINSTEIN e KROESE, 2016). Diferentemente de modelos determinísticos, em que os parâmetros são fixos, a simulação de Monte Carlo permite representar explicitamente a incerteza do processo. Os principais determinantes do desempenho (tempos de ciclo, tempos de setup, ocorrência e duração de falhas, taxa de defeitos) são tratados como variáveis aleatórias com distribuições ajustadas a partir de dados reais (LAW, 2015). A cada réplica da simulação, essas variáveis são amostradas, produzindo um conjunto de resultados que permitem estimar não apenas valores médios, mas distribuições completas para saídas como *lead time*, *throughput* e WIP, com intervalos de confiança e probabilidades de violação de metas.

À luz da problemática discutida, formula-se a seguinte pergunta central de pesquisa: Como a variabilidade operacional, oriunda de tempos de ciclo, setups, falhas, reparos e defeitos, afeta o *lead time* de recursos críticos de um sistema produtivo discreto, e como essa relação pode ser quantificada e decomposta por meio de um modelo de simulação de Monte Carlo fundamentado em Teoria de Filas e *Factory Physics*?

A resposta a essa pergunta interessa tanto ao meio acadêmico, por integrar rigor analítico com simulação, quanto ao meio empresarial, por oferecer uma ferramenta prática de diagnóstico e priorização de melhorias, tais como redução de tempos de setup, aumento de confiabilidade

por manutenção preventiva ou corretiva, suavização de chegadas e melhoria de qualidade.

1.2 Objetivo geral

O objetivo geral deste trabalho é desenvolver e aplicar um modelo estocástico, fundamentado em *Factory Physics* e Teoria de Filas, e implementado via simulação de Monte Carlo, para estimar e decompor o *lead time* de recursos críticos de um sistema produtivo discreto, quantificando a influência de variáveis de processo (ciclo, setups), confiabilidade (MTBF, MTTR) e qualidade (taxa de defeitos), sob diferentes níveis de utilização (HOPP e SPEARMAN, 2008; LAW, 2015; RUBINSTEIN e KROESE, 2016).

1.3 Objetivos específicos

Para viabilizar o objetivo geral, estabelecem-se os seguintes objetivos específicos:

- Caracterizar estatisticamente variáveis operacionais relevantes, como tempo de ciclo, tempo de setup, tempo entre falhas, tempo de reparo e taxa de defeitos, com segmentações por máquina, produto, turno e período, quando aplicável.
- Formular um modelo conceitual que conecte essas variáveis às métricas de desempenho do sistema, com base em leis e heurísticas do *Factory Physics* e resultados consolidados da Teoria de Filas.
- Implementar um simulador de Monte Carlo que gere distribuições de *lead time*, *throughput* e WIP, incluindo intervalos de confiança, percentis e probabilidades de estouro de metas (SALTELLI et al., 2008).
- Decompor o *lead time* em parcelas de processamento, tempos improdutivos (setups e paradas) e espera em fila, estimando contribuições relativas e incertezas associadas.
- Realizar análises de sensibilidade e cenários *what-if* para avaliar o impacto de mudanças em MTBF, MTTR, tempos de setup e taxa de defeitos em diferentes regimes de utilização.
- Aplicar e validar o modelo em um estudo de caso real, comparando resultados da simulação com indicadores observados e discutindo suas implicações gerenciais.

1.4 Justificativa da pesquisa

A justificativa deste estudo pode ser sustentada em três dimensões complementares: científica, metodológica e aplicada.

Na dimensão **científica**, o trabalho contribui ao integrar fundamentos consolidados de Teoria de Filas e *Factory Physics* a uma abordagem estocástica explícita, que permite tratar as incertezas de maneira transparente. A possibilidade de obter não apenas estimativas pontuais, mas distribuições completas de resultados e medidas de risco, confere maior robustez analítica e favorece a interpretação causal dos fenômenos (HOPP e SPEARMAN, 2008; GROSS e HARRIS, 2008).

Na dimensão **metodológica**, a escolha pela simulação de Monte Carlo evita a rigidez das suposições excessivamente restritivas dos modelos analíticos tradicionais (como chegadas puramente Poisson ou tempos de serviço exponenciais). Ao calibrar distribuições a partir de dados reais e ao adotar técnicas como amostragem em hipercubo latino, o modelo ganha em representatividade e eficiência (MCKAY, BECKMAN e CONOVER, 1979). Além disso, a decomposição do *lead time* em parcelas produtivas e improdutivas fornece uma visão clara para diagnósticos.

Na dimensão **aplicada**, empresas de manufatura frequentemente operam em regimes de alta utilização, nos quais a variabilidade torna o sistema sensível a pequenas mudanças. Nessas condições, decisões de melhoria exigem compreender onde uma intervenção gera maior impacto. O modelo aqui proposto fornece uma base quantitativa para priorizar ações como redução de setup, aumento de disponibilidade, suavização de chegadas ou melhoria de qualidade, expressando seus efeitos em métricas compreensíveis e com incerteza mensurada (GODINHO FILHO e BARCO, 2015, 2021).

1.5 Organização do trabalho

O Capítulo 2 apresenta a revisão bibliográfica e a fundamentação teórica que sustentam o estudo, incluindo conceitos fundamentais de desempenho produtivo, princípios do *Factory Physics*, elementos centrais da Teoria de Filas e a estratégia de simulação estocástica adotada, com destaque para o método de Monte Carlo.

O Capítulo 3 descreve a metodologia da pesquisa. Introduce o desenho do estudo, os procedimentos de coleta e tratamento de dados, o ajuste de distribuições e testes de aderência, e formaliza o modelo matemático que conecta variáveis operacionais às métricas de desempenho. Também detalha a construção do simulador, o número de réplicas, os critérios de verificação e validação e as técnicas para reduzir variância nas estimativas.

O Capítulo 4 apresenta o estudo de caso. Caracteriza o processo produtivo analisado, reporta estatísticas descritivas dos dados coletados, documenta a calibração do modelo e executa as simulações do cenário base e de cenários alternativos.

O Capítulo 5 desenvolve o modelo matemático e a simulação de Monte Carlo. Nele são apresentadas as equações fundamentais derivadas da Teoria de Filas e do *Factory Physics*, os ajustes de disponibilidade, confiabilidade e variabilidade efetiva, além da estrutura computacional do simulador e das técnicas de amostragem utilizadas. O capítulo também descreve o algoritmo implementado, as validações internas e as limitações do modelo.

O Capítulo 6 reúne e discute os resultados. São apresentados os indicadores de desempenho simulados para as Máquinas 10 e 11, incluindo análise por tamanho de lote, comportamento do lead time, decomposição das fontes de variabilidade, impacto das melhorias em setup, qualidade e confiabilidade, além de uma comparação integrada entre os dois recursos estudados.

O Capítulo 7 apresenta as conclusões. Retoma a pergunta central de pesquisa, sintetiza os principais achados quantitativos, discute os mecanismos causais identificados, apresenta recomendações gerenciais e sugere direções para trabalhos futuros, tanto em aprofundamento analítico quanto em extensão prática da modelagem.

2 Revisão bibliográfica e fundamentação teórica

Este capítulo apresenta os fundamentos teóricos que sustentam o desenvolvimento do modelo de simulação proposto neste trabalho. A revisão está estruturada em cinco eixos principais: conceitos de desempenho produtivo, princípios do *Factory Physics*, fundamentos da Teoria de Filas, fundamentos da simulação estocástica em sistemas produtivos e, por fim, técnicas associadas ao método de Monte Carlo e à análise de risco.

2.1 Conceitos fundamentais de desempenho produtivo

O desempenho de um sistema produtivo é avaliado a partir de indicadores que traduzem sua capacidade de atender à demanda de forma eficiente, previsível e economicamente viável. A literatura de gestão de operações estabelece que três dimensões principais devem ser simultaneamente gerenciadas: velocidade de entrega (medida pelo *lead time*), capacidade produtiva (expressa pelo *throughput*) e eficiência no uso de recursos (capturada pela utilização e pelo *work in process*). Essas dimensões não são independentes; ao contrário, estão profundamente interligadas por relações matemáticas e físicas que governam o fluxo de materiais e informações na manufatura (HOPP e SPEARMAN, 2008; SLACK et al., 2019).

Entre os principais indicadores operacionais, destacam-se:

Lead Time (LT): O *lead time* representa o tempo total de atravessamento de uma ordem de produção, desde sua liberação até a conclusão. Este indicador é composto por múltiplas parcelas: tempo de processamento efetivo nas máquinas (T_p), tempos de preparação ou *setups* (T_s), tempos de espera em filas (W_q), tempos de transporte entre estações (T_t) e eventuais paradas planejadas ou não planejadas (T_{down}). Formalmente, pode-se expressar:

$$LT = T_p + T_s + W_q + T_t + T_{down} \quad (1)$$

Em ambientes *make-to-order*, o *lead time* determina diretamente a promessa de entrega ao cliente e, portanto, a competitividade da empresa. Já em ambientes *make-to-stock*, ele influencia o dimensionamento de estoques de segurança e a responsividade do sistema a variações de demanda. A redução do *lead time* sem comprometimento da qualidade ou aumento de custos é

considerada uma das principais métricas de excelência operacional (GOLDRATT, 2016).

Throughput (TH): O *throughput* é definido como a taxa média de saída de itens conformes por unidade de tempo, tipicamente expressa em peças por hora ou unidades por dia. Diferentemente da capacidade teórica nominal de uma máquina ($C_{nominal}$), o *throughput* efetivo considera todas as perdas operacionais. Pode-se expressar o *throughput* como:

$$TH = C_{nominal} \cdot A \cdot (1 - p) \cdot \eta_{setup} \quad (2)$$

onde:

- A : disponibilidade do equipamento (fração do tempo operacional);
- p : taxa de defeitos ou não conformidades;
- η_{setup} : fator de eficiência relacionado ao tempo perdido em preparações.

Segundo a Teoria das Restrições (*Theory of Constraints*), desenvolvida por Goldratt, o *throughput* do sistema como um todo é determinado pela restrição mais severa da cadeia produtiva, isto é, o famoso “gargalo”. Portanto, melhorias em recursos não-gargalo têm impacto limitado no *throughput* global, enquanto qualquer ganho no recurso restritivo se traduz imediatamente em aumento de capacidade sistêmica (GOLDRATT, 2016).

Work in Process (WIP): O WIP representa a quantidade média de itens em processo, isto é, em produção ou aguardando processamento em qualquer ponto do sistema. Do ponto de vista financeiro, o WIP representa capital imobilizado em materiais que ainda não foram convertidos em produtos acabados “vendáveis”. Do ponto de vista operacional, níveis elevados de WIP estão associados a *lead times* longos, dificuldade de rastreabilidade, maior complexidade de planejamento e risco de obsolescência.

A filosofia *Lean Manufacturing*, originada no Sistema Toyota de Produção, advoga pela minimização sistemática do WIP como estratégia para revelar problemas ocultos, reduzir desperdícios e aumentar a flexibilidade produtiva (WOMACK e JONES, 2003). A relação entre WIP e *lead time* é formalizada matematicamente pela Lei de Little, um dos pilares da teoria de filas e da ciência de gestão de operações.

Utilização (ρ): A utilização representa a proporção do tempo em que o recurso está efetivamente ocupado processando trabalho. Matematicamente, define-se:

$$\rho = \frac{\lambda}{\mu} = \lambda \cdot t_e \quad (3)$$

onde λ é a taxa média de chegada de trabalho ao recurso e $\mu = 1/t_e$ é a taxa média de processamento, com t_e sendo o tempo efetivo médio de processamento por peça. A condição de estabilidade do sistema requer que $\rho < 1$.

Um resultado fundamental da teoria de filas estabelece que, quando $\rho \rightarrow 1$, pequenas variações na taxa de chegada ou no tempo de processamento causam aumentos exponenciais no tempo de espera e, portanto, no lead time. Este fenômeno de congestão ou saturação explica por que sistemas produtivos não devem operar próximos à capacidade máxima teórica. Embora *Factory Physics* destaque a sensibilidade extrema do desempenho à alta utilização, diferentes autores e aplicações práticas sugerem manter recursos críticos em níveis moderados de utilização, tipicamente na faixa de 80–85%, de modo a preservar estabilidade, capacidade de resposta e proteção contra variabilidade (SURI, R., 1998).

Esses quatro conceitos, no caso, *lead time*, *throughput*, WIP e utilização, formam o núcleo analítico do controle da produção. Eles não são independentes: estão relacionados pela Lei de Little e pelas equações empíricas de tempo de ciclo propostas por Hopp e Spearman (2008). Compreender suas inter-relações é essencial para modelar e interpretar o comportamento dinâmico de sistemas discretos de manufatura.

2.2 Princípios do *Factory Physics*

O arcabouço teórico de *Factory Physics*, desenvolvido por Hopp e Spearman (2008), propõe que os sistemas produtivos obedecem a leis gerais análogas às leis físicas da natureza. Tais leis descrevem, de forma quantitativa, como a variabilidade, a utilização e o fluxo interagem para determinar o desempenho global da produção. Essa abordagem unifica conceitos clássicos de engenharia de produção, teoria de filas e controle de fluxo em um conjunto coerente de relações analíticas.

A principal contribuição do *Factory Physics* é traduzir fenômenos complexos de manufatura

em equações matematicamente tratáveis, permitindo análise quantitativa rigorosa e subsidiando decisões gerenciais com base científica. As “leis” do *Factory Physics* não são leis no sentido físico absoluto, mas sim relações empíricas robustas, validadas extensivamente em ambientes industriais reais e fundamentadas em teoria probabilística sólida.

2.2.1 Lei de Little e relações fundamentais

A Lei de Little é a base da teoria dos sistemas produtivos. Enunciada originalmente por John D.C. Little em 1961, esta lei estabelece uma relação fundamental entre três variáveis de estado de qualquer sistema de filas em regime estacionário. Formalmente:

$$WIP = \lambda \cdot LT \quad (4)$$

onde:

- WIP : número médio de itens em processo;
- λ : taxa média de chegada (ou de saída, em regime estacionário), em peças por unidade de tempo;
- LT : tempo médio de permanência de um item no sistema, em unidades de tempo.

A elegância da Lei de Little reside em sua universalidade e simplicidade. Ela é válida independentemente da distribuição estatística de chegadas ou tempos de processamento, independentemente da disciplina da fila (FIFO, LIFO, prioridade) e independentemente do número de servidores. A única exigência é que o sistema esteja em regime estacionário, ou seja, que as taxas médias de entrada e saída sejam equilibradas no longo prazo (LITTLE, 1961).

Essa relação tem implicações profundas para a gestão de operações. Ela estabelece que controlar o nível de WIP equivale a controlar o tempo médio de atravessamento, desde que a taxa de produção (λ) seja mantida. De modo inverso, reduzir o *lead time* sem afetar a taxa de produção implica necessariamente diminuir o estoque em processo. Essas relações fundamentam as estratégias de produção puxada (*pull systems*) e sistemas Kanban, amplamente utilizados em ambientes *Lean* (WOMACK e JONES, 2003).

A Lei de Little pode ser aplicada não apenas ao sistema como um todo, mas também a subsistemas individuais. Por exemplo, para um único recurso (máquina) com fila dedicada:

$$WIP_{recurso} = \lambda_{recurso} \cdot LT_{recurso} \quad (5)$$

Esta propriedade permite análise hierárquica e decomposição de sistemas complexos em componentes tratáveis individualmente.

2.2.2 Curva VUT e dinâmica do tempo de ciclo

Além da relação básica da Lei de Little, o *Factory Physics* introduz a chamada Curva VUT, uma aproximação analítica que conecta explicitamente a variabilidade (V), a utilização (U) e o tempo de ciclo (CT). Esta é uma das contribuições mais importantes da obra de Hopp e Spearman, pois fornece uma ferramenta quantitativa para prever o comportamento não-linear dos tempos de atravessamento em função da carga do sistema.

A equação VUT expressa o tempo médio de ciclo como:

$$CT \approx CT_e \left(1 + \frac{V \cdot U}{1 - U} \right) \quad (6)$$

onde:

- CT : tempo médio total de ciclo (ou tempo médio de produção por peça), incluindo processamento e espera;
- CT_e : tempo efetivo médio de processamento, sem esperas em fila;
- U : utilização do recurso, definida como $U = \lambda \cdot CT_e$;
- V : fator de variabilidade médio, definido como $V = \frac{c_a^2 + c_s^2}{2}$, sendo c_a^2 e c_s^2 os coeficientes de variação quadráticos dos tempos de chegada e de serviço, respectivamente.

Esta equação sintetiza um dos fenômenos mais importantes da dinâmica de manufatura: o crescimento não-linear dos tempos de atravessamento à medida que o sistema se aproxima da plena utilização. O termo $\frac{U}{1-U}$ cresce exponencialmente quando U tende a 1, indicando que

pequenos aumentos na utilização próximos do limite crítico provocam acréscimos desproporcionais em CT e, portanto, no LT .

Este comportamento é amplamente verificado empiricamente em linhas de produção, centros de usinagem e processos contínuos. A Figura 1 ilustra a relação típica entre utilização e tempo de ciclo para diferentes níveis de variabilidade.

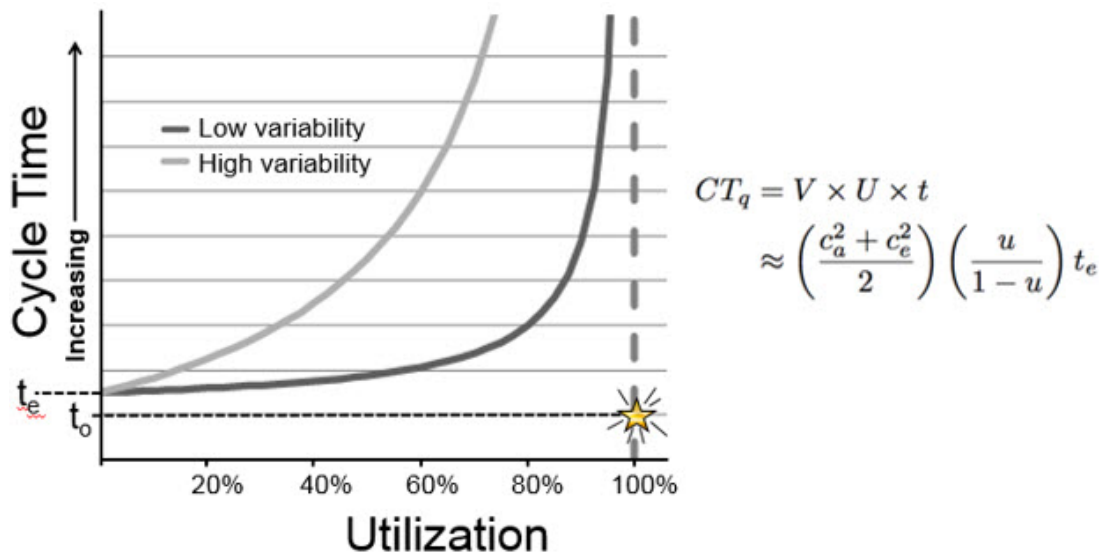


Figura 1: Curva VUT. Fonte: Factory Physics (2020)

A curva VUT revela *insights* críticos para gestão de capacidade:

1. **Efeito da utilização:** Mesmo com variabilidade moderada, operar acima de 85-90% de utilização resulta em crescimento explosivo dos tempos de ciclo. Portanto, capacidade “ociosa” não é desperdício, mas sim proteção necessária contra variabilidade.
2. **Efeito da variabilidade:** Para uma mesma utilização, sistemas com alta variabilidade ($V > 1$) apresentam tempos de ciclo substancialmente maiores que sistemas com baixa variabilidade ($V < 0.5$). Isso fundamenta a ênfase do *Lean* e do *Six Sigma* na redução de variabilidade.
3. **Trade-off capacidade-flexibilidade:** Adicionar capacidade (reduzir U) tem retorno decrescente: o ganho é enorme ao reduzir de 95% para 85%, mas marginal ao reduzir de 70% para 60%.

2.3 Teoria de Filas

A Teoria de Filas fornece o embasamento probabilístico que explica como taxas de chegada, taxas de serviço e variabilidade se traduzem em tempos de espera e tempos de atravessamento. Desenvolvida originalmente por Agner Krarup Erlang no início do século XX para análise de sistemas telefônicos, a teoria foi posteriormente generalizada e aplicada a uma ampla gama de sistemas, incluindo manufatura, logística, sistemas computacionais e serviços (GROSS e HARRIS, 2008).

A teoria de filas permite prever tempos médios de espera, tamanhos de fila, níveis de utilização e probabilidades de estados do sistema em regime estacionário. Seu poder reside em fornecer soluções analíticas fechadas para classes específicas de sistemas, bem como aproximações validadas para sistemas mais gerais.

2.3.1 Estrutura básica e componentes de uma fila

Um sistema de filas é caracterizado por cinco componentes essenciais, frequentemente descritos pela notação de Kendall (KENDALL, 1953):

1. Processo de chegadas: Define a forma como novas entidades (peças, ordens, clientes) entram no sistema. O processo pode ser:

- Determinístico (D): Chegadas em intervalos fixos e previsíveis;
- Markoviano ou Poisson (M): Chegadas aleatórias com taxa constante λ , resultando em intervalos entre chegadas exponencialmente distribuídos;
- Geral (G): Chegadas com distribuição arbitrária, caracterizada por média $\frac{1}{\lambda}$ e coeficiente de variação c_a .

A variabilidade do processo de chegadas é medida por $c_a^2 = \frac{Var(T_a)}{E[T_a]^2}$, onde T_a é o tempo entre chegadas consecutivas. Para processo Poisson, $c_a^2 = 1$; para chegadas regulares, $c_a^2 = 0$; para chegadas mais irregulares que Poisson, $c_a^2 > 1$.

2. Processo de serviço: Descreve o tempo necessário para processar cada entidade. Sua variabilidade é medida por:

$$c_s^2 = \frac{Var(T_s)}{E[T_s]^2} \quad (7)$$

onde T_s é o tempo de serviço. Processos com $c_s^2 < 0.5$ são considerados de baixa variabilidade; $c_s^2 \approx 1$ indica variabilidade moderada (típica de processo exponencial); $c_s^2 > 2$ indica alta variabilidade, comum em processos com frequentes interrupções ou retrabalhos.

3. Número de servidores: Denotado por c , representa quantos recursos operam em paralelo atendendo a fila. Em manufatura, correspondente ao número de máquinas idênticas processando o mesmo tipo de peça.

4. Capacidade do sistema: Indica se a fila pode crescer indefinidamente (capacidade infinita) ou possui limite físico (capacidade finita). Sistemas com capacidade finita podem bloquear chegadas quando cheios.

5. Disciplina da fila: Estabelece a ordem de atendimento. As mais comuns são:

- FIFO (*First In, First Out*): primeira peça a chegar é a primeira a ser processada;
- LIFO (*Last In, First Out*): última peça a chegar é processada primeiro;
- SIRO (*Service In Random Order*): seleção aleatória;
- Prioridade: entidades possuem prioridades distintas.

Modelos clássicos de fila são denotados pela notação de Kendall: $A/B/c/K/N/D$, onde A descreve o processo de chegadas, B o processo de serviço, c o número de servidores, K a capacidade do sistema, N o tamanho da população fonte e D a disciplina. Por exemplo, $M/M/1$ denota um sistema com chegadas Poisson, serviço exponencial e um único servidor.

2.3.2 Aproximação de Kingman (VUT) para o modelo $G/G/1$

Em sistemas com uma única máquina e variabilidade tanto nas chegadas quanto nos serviços, o modelo $G/G/1$ oferece uma formulação geral. Diferentemente dos modelos $M/M/1$ ou $M/G/1$, que possuem soluções analíticas exatas, o modelo $G/G/1$ não admite solução fechada. Entretanto, a aproximação de Kingman fornece uma estimativa notavelmente precisa do tempo médio de espera em fila (Kingman, 1961).

A aproximação de Kingman expressa o tempo médio de espera em fila como:

$$W_q \approx \frac{c_a^2 + c_s^2}{2} \cdot \frac{\rho}{1 - \rho} \cdot t_e \quad (8)$$

onde:

- t_e : tempo médio de processamento (serviço);
- $\rho = \lambda t_e$: utilização do recurso;
- c_a^2, c_s^2 : coeficientes de variação quadráticos das chegadas e do serviço.

Esta é precisamente a equação que fundamenta a curva VUT apresentada anteriormente. Ela mostra que o tempo médio de espera é proporcional a três fatores multiplicativos:

Variabilidade total: O termo $\frac{c_a^2 + c_s^2}{2}$ captura o efeito combinado da variabilidade das chegadas e do processo. Se ambas as fontes apresentam variabilidade próxima de zero ($c_a^2 \approx c_s^2 \approx 0$), o sistema opera quase deterministicamente e praticamente não há espera em fila. Cada incremento unitário em c_a^2 ou c_s^2 aumenta proporcionalmente o tempo de espera.

Intensidade de utilização: O termo $\frac{\rho}{1-\rho}$ cresce exponencialmente com ρ . Quando $\rho = 0.5$ (50% de utilização), este fator vale 1.0. Com $\rho = 0.8$, salta para 4.0. Com $\rho = 0.9$, atinge 9.0. Este crescimento super-linear explica por que sistemas produtivos não devem operar próximos à capacidade máxima: o custo marginal de cada ponto percentual adicional de utilização cresce explosivamente.

Tempo base: O fator t_e estabelece a escala temporal. Processos com tempos de ciclo longos naturalmente acumulam mais espera absoluta, mesmo que a espera relativa (em múltiplos de t_e) seja a mesma.

O tempo total de atravessamento (LT) é obtido adicionando o tempo de processamento ao tempo de espera:

$$LT = W_q + t_e = t_e \left(1 + \frac{c_a^2 + c_s^2}{2} \cdot \frac{\rho}{1 - \rho} \right) \quad (9)$$

A partir dessas equações, é possível calcular o WIP via Lei de Little:

$$WIP = \lambda \cdot LT = \lambda \cdot t_e \left(1 + \frac{c_a^2 + c_s^2}{2} \cdot \frac{\rho}{1 - \rho} \right) = \rho \left(1 + \frac{c_a^2 + c_s^2}{2} \cdot \frac{\rho}{1 - \rho} \right) \quad (10)$$

Finalmente, o *throughput* efetivo considera a taxa de chegada e a taxa de defeitos:

$$TH = \lambda(1 - p) \quad (11)$$

onde p representa a taxa média de refugos ou não conformidades.

Essas relações formam o núcleo matemático do modelo de desempenho produtivo utilizado neste trabalho. Elas permitem decompor o *lead time* em componentes físicos interpretáveis: tempo produtivo (t_e) e tempo de espera (W_q), sendo o segundo função explícita da variabilidade e da utilização.

2.4 Simulação estocástica em sistemas produtivos

A simulação é um método computacional que permite reproduzir o comportamento de sistemas reais ao longo do tempo, especialmente quando a complexidade impede a obtenção de soluções analíticas fechadas. Em manufatura, é utilizada para investigar a interação entre falhas, *setups*, estoques intermediários e restrições de capacidade quando essas interações não podem ser capturadas adequadamente por equações fechadas.

Segundo Law (2015), a simulação se justifica quando:

1. o sistema real é muito complexo para ser descrito de forma analítica (por exemplo, redes de filas com roteiros múltiplos, prioridades dinâmicas e recursos compartilhados);
2. o experimento direto seria inviável, perigoso ou excessivamente oneroso;
3. é necessário explorar cenários “hipotéticos” que ainda não existem na prática (por exemplo, avaliar uma nova configuração de layout antes de implementá-la fisicamente).

Essas condições se aplicam ao presente estudo. A dinâmica de máquinas sujeitas a variação de ciclo, falhas, *setups* e defeitos não pode ser capturada integralmente por fórmulas fechadas da Teoria de Filas (que frequentemente assumem simplificações como tempos exponenciais ou inexistência de *setups*), exigindo um modelo estocástico mais flexível.

2.4.1 Tipos de simulação

Três abordagens principais são utilizadas na modelagem de sistemas produtivos:

Simulação determinística: Considera parâmetros fixos e produz resultados únicos, sem incorporar variabilidade. É adequada para planejamentos estáticos e cenários de referência (*baseline*). Por exemplo, calcular o *lead time* teórico mínimo assumindo tempos de ciclo constantes e ausência de falhas. Embora útil para estabelecer limites inferiores de desempenho, não captura a realidade de sistemas sujeitos a incertezas (LAW, 2015; BANKS et al., 2010).

Simulação de eventos discretos (DES): Modela explicitamente a sequência temporal de eventos (chegadas, inícios e terminos de processamento, falhas, reparos, decisões de roteamento). Mantém o estado detalhado do sistema a cada instante e avança o relógio de simulação de evento em evento. Fornece alto nível de detalhamento e permite rastrear trajetórias individuais de entidades, mas requer maior esforço de modelagem e pode ser computacionalmente intensivo. É a abordagem mais utilizada em softwares comerciais de simulação de manufatura (Arena, Simul8, Plant Simulation) (BANKS; CARSON; NELSON; NICOL, 2010).

Simulação de Monte Carlo: Utiliza amostragem aleatória de variáveis para gerar múltiplos cenários e estimar estatísticas agregadas de desempenho. Ao invés de simular eventos individuais, avalia o sistema em regime estacionário através de equações analíticas alimentadas por parâmetros aleatórios. É a abordagem empregada neste trabalho por sua simplicidade, eficiência computacional e robustez para análise de risco (RUBISTEIN e KROESE, 2016).

A simulação de Monte Carlo é particularmente adequada quando:

- O interesse está em métricas de regime estacionário (médias de longo prazo), não em dinâmica transiente;
- Existem boas aproximações analíticas para o sistema (como as equações VUT);
- Deseja-se caracterizar distribuições completas de saída, não apenas trajetórias individuais;
- É necessário executar milhares de cenários para análise estatística robusta.

2.5 Método de Monte Carlo e análise de risco

O método de Monte Carlo consiste em realizar N experimentos independentes de um modelo matemático, cada um baseado em amostras aleatórias de suas variáveis de entrada. Para uma métrica de saída $Y = f(X)$, em que X é um vetor de variáveis aleatórias (como tempo de ciclo, *setup*, MTBF, MTTR ou taxa de defeito), estima-se:

$$\mathbb{E}[Y] \approx \frac{1}{N} \sum_{i=1}^N f(\mathbf{x}_i) \quad (12)$$

onde \mathbf{x}_i representa a i -ésima amostra de \mathbf{X} . Além da média, é possível estimar variâncias, quantis (percentis) e intervalos de confiança de Y , permitindo uma análise probabilística completa (RUBINSTEIN e KROESE, 2016).

A precisão do estimador melhora com \sqrt{N} : para reduzir o erro padrão pela metade, é necessário quadruplicar o número de réplicas. Na prática, $N = 10.000$ a 50.000 réplicas são suficientes para obter estimativas estáveis de médias e percentis em problemas típicos de manufatura (LAW, 2015; RUBINSTEIN; KROESE, 2016).

2.5.1 Técnicas de amostragem e redução de variância

A eficiência de um modelo de simulação estocástica depende diretamente da qualidade estatística das amostras geradas. Em problemas de manufatura caracterizados por elevada variabilidade e não linearidade, como os analisados neste trabalho, estimativas baseadas apenas em amostragem aleatória simples podem apresentar alta variância, exigindo um número excessivo de réplicas para alcançar precisão adequada. Dessa forma, técnicas de redução de variância tornam-se essenciais para aumentar a confiabilidade dos resultados sem elevar de forma significativa o custo computacional (LAW, 2015).

Neste estudo, foram empregadas técnicas complementares de redução de variância, escolhidas com base em sua adequação teórica ao problema e em sua ampla utilização na literatura de simulação de sistemas produtivos.

Números aleatórios comuns: A técnica de números aleatórios comuns consiste em utilizar a mesma sequência de números pseudoaleatórios ao comparar diferentes cenários de simulação.

O princípio fundamental é induzir correlação positiva entre as saídas dos modelos comparados, de modo que a variabilidade comum seja cancelada ao se analisar diferenças entre cenários.

Formalmente, quando duas alternativas são simuladas com entradas estocásticas idênticas, a variância da diferença entre suas saídas tende a ser menor do que aquela obtida com amostras independentes. Essa propriedade torna o método especialmente adequado para estudos comparativos, nos quais o interesse reside em avaliar o impacto relativo de modificações estruturais no sistema, como redução de tempo de setup, aumento de disponibilidade ou diminuição da variabilidade.

No presente trabalho, a utilização de números aleatórios comuns permite isolar o efeito das alterações operacionais, garantindo que diferenças observadas entre cenários sejam atribuídas às mudanças no sistema e não a flutuações aleatórias do processo de amostragem.

Amostragem em hipercubo latino: A amostragem em hipercubo latino foi empregada com o objetivo de melhorar a cobertura do espaço amostral das variáveis de entrada. Diferentemente da amostragem aleatória simples, essa técnica divide a distribuição de cada variável em intervalos de probabilidade equiprováveis e assegura que cada intervalo seja amostrado exatamente uma vez.

Essa estratégia reduz a concentração de amostras em regiões específicas do espaço e proporciona uma representação mais uniforme das variáveis aleatórias. Estudos clássicos demonstram que a amostragem em hipercubo latino pode reduzir significativamente a variância dos estimadores quando a resposta do modelo apresenta comportamento aproximadamente monotônico em relação às variáveis de entrada (MCKAY, BECKMAN e CONOVER, 1979).

No contexto deste trabalho, tal característica é particularmente relevante, uma vez que variáveis como tempos de processamento, taxas de falha e tempos de reparo exercem influência monotônica sobre métricas como lead time e WIP. Assim, o uso de LHS permite obter estimativas mais estáveis com menor número de repetições.

Variáveis antitéticas: A técnica de variáveis antitéticas baseia-se na geração de pares de amostras complementares. Para cada vetor de números aleatórios \mathbf{u} , gera-se um vetor associado $\mathbf{1} - \mathbf{u}$. Quando a função de saída é monotônica, as respostas obtidas a partir desses dois

vetores tendem a ser negativamente correlacionadas, o que reduz a variância da média amostral.

Essa propriedade é particularmente útil em sistemas produtivos, nos quais aumentos nos tempos de ciclo, falhas ou tempos de setup tendem a aumentar monotonicamente métricas como o lead time. Dessa forma, a utilização de variáveis antitéticas contribui para a obtenção de estimativas mais estáveis sem aumento do número de simulações realizadas.

Controle de variáveis: O método de controle de variáveis baseia-se na utilização de uma variável auxiliar cuja esperança matemática é conhecida e que apresenta correlação significativa com a variável de interesse. A ideia central consiste em corrigir o estimador original a partir do desvio observado dessa variável auxiliar em relação ao seu valor esperado.

Formalmente, seja Y a variável de interesse e Z uma variável de controle com esperança conhecida $\mathbb{E}[Z] = \mu_Z$. O estimador ajustado é dado por:

$$\hat{Y}_{CV} = \hat{Y} - c(\hat{Z} - \mu_Z) \quad (13)$$

em que o coeficiente c é escolhido de modo a minimizar a variância do estimador corrigido. Essa forma é adotada porque ela preserva a imparcialidade do estimador ao mesmo tempo em que reduz sua variância sempre que Y e Z forem correlacionadas.

No contexto deste trabalho, variáveis como o tempo médio efetivo de processamento, a utilização do recurso ou o throughput teórico possuem valores esperados conhecidos a partir do modelo analítico. Essas grandezas apresentam forte correlação com métricas como o lead time e o WIP, tornando-as candidatas naturais a variáveis de controle. A escolha dessa formulação é, portanto, justificada tanto do ponto de vista estatístico quanto operacional.

A combinação das técnicas de números aleatórios comuns, amostragem em hipercubo latino, variáveis antitéticas e controle de variáveis permite reduzir significativamente a variância das estimativas sem comprometer a imparcialidade dos resultados. Essa abordagem integrada é especialmente adequada para sistemas produtivos operando próximos à saturação, nos quais pequenas variações nos parâmetros podem gerar grandes oscilações no desempenho.

Dessa forma, a metodologia adotada garante maior robustez estatística às conclusões obtidas, permitindo análises mais confiáveis sobre o impacto de decisões operacionais.

3 Metodologia

A metodologia de uma pesquisa científica deve oferecer o caminho estruturado pelo qual a pergunta central é investigada. No presente trabalho, a questão que se busca responder pode ser assim enunciada:

Como a variabilidade operacional, oriunda de tempos de ciclo, setups, falhas, reparos e defeitos, afeta o *lead time* de recursos críticos de um sistema produtivo discreto, e como essa relação pode ser quantificada e decomposta por meio de um modelo de simulação de Monte Carlo fundamentado em Teoria de Filas e *Factory Physics*?

Responder a essa pergunta exige um método capaz de lidar com três características essenciais do problema: (i) a natureza empírica dos dados, oriundos de medições em ambiente real de produção; (ii) a presença de variabilidade, tanto intrínseca (tempos de processo, falhas, defeitos) quanto extrínseca (flutuações de chegada de ordens); e (iii) a necessidade de inferência gerencial, ou seja, de traduzir a análise em implicações práticas para tomada de decisão.

Com base nesses requisitos, o presente estudo adota como estratégia metodológica a combinação de **estudo de caso** e **simulação estocástica**. O estudo de caso fornece a conexão empírica com a realidade de uma empresa específica, enquanto a simulação permite generalizar e explorar cenários alternativos sem a necessidade de experimentação direta no chão de fábrica.

3.1 Teoria do método

3.1.1 Estudo de caso

O estudo de caso é um método de pesquisa qualitativo-quantitativo que busca compreender fenômenos complexos em profundidade, a partir da observação detalhada de um objeto específico (YIN, 2015). No contexto deste trabalho, o objeto de estudo é uma planta de manufatura de injeção plástica, com foco em máquinas selecionadas consideradas gargalos do processo.

Segundo YIN (2015), a principal contribuição do estudo de caso é possibilitar a análise contextualizada, preservando as características singulares do ambiente. Diferentemente de abordagens estatísticas amplas, que buscam generalizações populacionais, o estudo de caso valoriza

a compreensão intensiva de um fenômeno em sua realidade prática. De forma similar, STAKE (2006) enfatiza o caráter holístico e interpretativo da investigação de um caso singular, enquanto MERRIAM (2009) reforça que esse método permite gerar conhecimento aprofundado sobre contextos específicos. Assim, este trabalho parte de dados reais de produção (tempos de ciclo, quantidades produzidas, não conformidades) para parametrizar o modelo de simulação.

3.1.2 Simulação como método de pesquisa

A simulação é definida por Chwif e Medina como a imitação do funcionamento de um sistema real ao longo do tempo, por meio de um modelo que representa seus elementos e interações (CHWIF e MEDINA, 2015). Em manufatura, a simulação permite experimentar virtualmente alternativas de operação, sem custos ou riscos de interromper a produção real.

Ainda de acordo com Chwif e Medina (2015), a simulação se justifica quando:

1. o sistema real é muito complexo para ser descrito de forma analítica (LAW, 2015);
2. o experimento direto seria inviável ou excessivamente oneroso;
3. é necessário explorar cenários “hipotéticos” que ainda não existem na prática.

Essas condições se aplicam ao presente estudo. A dinâmica de máquinas sujeitas a variação de ciclo, falhas e defeitos não pode ser capturada integralmente por fórmulas fechadas da Teoria de Filas, exigindo um modelo estocástico mais flexível.

3.1.3 Simulação de Monte Carlo

O método de Monte Carlo é particularmente adequado porque trata incertezas de maneira explícita, ao modelar os parâmetros relevantes como variáveis aleatórias com distribuições ajustadas a dados. Cada réplica de simulação corresponde a um cenário possível de operação, e a repetição de milhares de réplicas permite construir distribuições de saída para o lead time, throughput e WIP.

Esse procedimento não apenas fornece valores médios, mas também intervalos de confiança e probabilidades de violação de metas. Dessa forma, a análise transcende a média e alcança a noção de risco, fundamental para decisões de produção em ambientes incertos.

3.1.4 Integração dos métodos

A metodologia, portanto, estrutura-se em três etapas principais:

1. **Caracterização do caso:** coleta e tratamento dos dados de produção, com cálculo de parâmetros estatísticos (tempos médios, variâncias, taxas de falha, taxas de não conformidade).
2. **Construção do modelo:** formulação matemática fundamentada em Teoria de Filas e *Factory Physics*, parametrizada a partir dos dados coletados.
3. **Simulação estocástica:** execução do modelo via Monte Carlo, com análises de sensibilidade e comparação de cenários, permitindo avaliar como mudanças em confiabilidade, setups e lotes impactam as métricas de desempenho.

Essa integração garante que a pesquisa seja, ao mesmo tempo, empiricamente enraizada (via estudo de caso) e analiticamente robusta (via simulação), fornecendo resultados com validade prática e consistência teórica.

4 Estudo de Caso

4.1 Introdução à empresa

O presente estudo foi conduzido em uma planta industrial de injeção plástica localizada na região de São Carlos (SP), com mais de quatro décadas de atuação no setor. Fundada nos anos 1980, a fábrica consolidou-se como um importante fornecedor regional de componentes plásticos para diferentes segmentos industriais.

O parque fabril é composto por injetoras plásticas de diversas capacidades, variando de 80 a 1600 toneladas de força de fechamento. Essa diversidade operacional permite a produção de peças com diferentes tamanhos, massas e geometrias, utilizando materiais como polipropileno (PP), acrilonitrila butadieno estireno (ABS), poliestireno (PS), cloreto de polivinila (PVC) e poliamida (Nylon 6 e 6.6).

A escolha desse ambiente industrial justifica-se pela relevância e maturidade dos seus processos de moldagem. A planta possui uma estrutura produtiva consolidada, com sistemas de controle de processo, manutenção preventiva e registros de produção informatizados. Esses registros, coletados continuamente ao longo dos turnos, incluem dados de ciclos, tempos de parada, volumes produzidos e ocorrência de não conformidades.

Essas informações oferecem uma base empírica robusta para a calibração do modelo computacional desenvolvido neste trabalho, permitindo que as simulações de Monte Carlo sejam confrontadas com o comportamento real do sistema produtivo. Dessa forma, a fábrica analisada representa um caso adequado para a aplicação integrada de teoria de filas, modelagem matemática e simulação computacional no contexto da manufatura por injeção plástica.

4.2 Recorte do estudo: Máquinas críticas

Entre as injetoras que compõem o parque fabril, foram selecionadas como objetos principais de análise as Máquinas 10 e 11, pois são consideradas pela empresa como recursos críticos para a capacidade global de produção da planta, respondendo por uma parcela significativa do volume manufaturado.

A Máquina 10 é uma injetora de 1400 toneladas de força de fechamento, utilizada predo-

minantemente na fabricação de componentes de médio porte e elevada cadência produtiva. A Máquina 11, por sua vez, é uma injetora de 900 toneladas, destinada à produção de peças de grande porte e maior complexidade geométrica, incluindo carcaças técnicas e elementos estruturais. Ambas dispõem de controladores digitais, sistemas automáticos de alimentação de matéria-prima e interfaces de monitoramento em tempo real, o que possibilita o registro contínuo de dados operacionais.

A escolha por essas duas máquinas decorre da complementaridade de seus perfis produtivos e da importância estratégica que exercem sobre o fluxo de produção: enquanto a Máquina 10 opera com alta frequência e menor variabilidade de ciclo, a Máquina 11 concentra lotes de maior valor agregado e apresenta maior sensibilidade a paradas, setups e falhas. Essa combinação torna possível avaliar, de forma comparativa, como diferentes regimes operacionais e níveis de variabilidade influenciam o lead time, o throughput e o work in process (WIP) do sistema.

A análise conjunta dessas máquinas é particularmente relevante por dois motivos principais:

1. Ambas operam com elevada taxa de utilização e têm impacto direto sobre o *throughput* global da planta, influenciando o ritmo de atendimento de pedidos e o cumprimento de prazos de entrega.
2. Fornecem uma base empírica robusta para calibração e validação cruzada do modelo de simulação, reforçando sua aplicabilidade generalizada dentro do ambiente fabril.

Por essas razões, as Máquinas 10 e 11 foram definidas como os elementos centrais da modelagem e da simulação deste trabalho. A partir dos seus dados reais de operação, foram estimados os parâmetros físicos e estatísticos utilizados na formulação matemática apresentada nos capítulos seguintes.

4.3 Layout e processo produtivo

O processo de injeção plástica no parque fabril em análise segue o fluxo clássico adotado na indústria de moldagem por injeção, representado esquematicamente na Figura 2 e descrito a seguir:

1. **Preparação da matéria-prima:** os grânulos plásticos são alimentados em funis de secagem, podendo receber aditivos, pigmentos e reforços minerais conforme a formulação especificada. O controle de umidade e temperatura é fundamental para a estabilidade dimensional das peças moldadas.
2. **Injeção e moldagem:** o material fundido é injetado sob alta pressão em um molde metálico, preenchendo as cavidades e solidificando-se após o resfriamento. O ciclo é controlado automaticamente, sendo composto pelas fases de injeção, recalque, resfriamento e abertura do molde.
3. **Ejeção:** ao término do ciclo, o sistema de extração retira a peça do molde, liberando o conjunto para o início de um novo ciclo.
4. **Inspeção de qualidade e expedição:** as peças são inspecionadas conforme critérios visuais e dimensionais. Peças conformes seguem para embalagem e expedição, enquanto eventuais refugos são segregados para retrabalho ou descarte controlado.

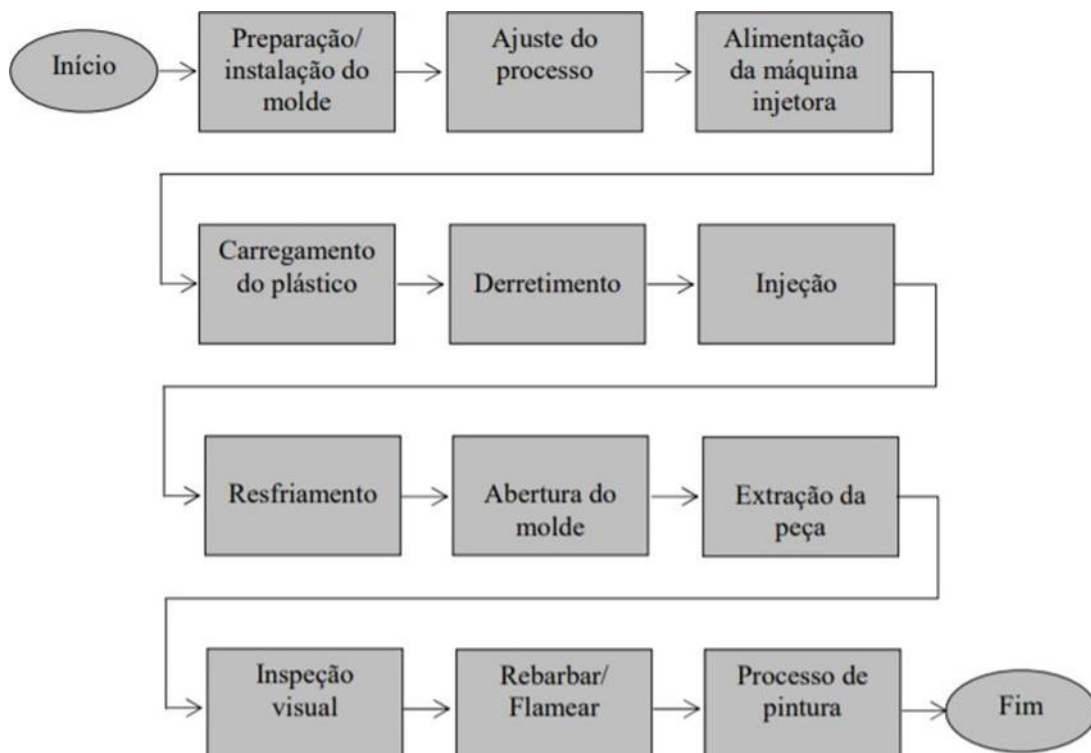


Figura 2: Fluxograma do processo de injeção plástica. Fonte: adaptado de SANTOS e SOUZA (2022)

A etapa de moldagem é o ponto de interesse principal da modelagem desenvolvida neste estudo, pois é nela que se define o tempo de ciclo natural (t_o) e os efeitos de disponibilidade decorrentes de setups e paradas planejadas ou corretivas. Essa etapa determina diretamente o lead time, o WIP e o throughput da linha.

A modelagem matemática apresentada nos capítulos seguintes busca representar, de forma simplificada e analiticamente consistente, o comportamento operacional dessa máquina e suas interdependências com os indicadores de desempenho da produção.

5 Modelagem Matemática e Simulação de Monte Carlo

5.1 Visão Geral do Modelo

O presente capítulo consolida a modelagem matemática e a implementação computacional via Python desenvolvidas para representar o comportamento produtivo das Máquinas 10 e 11 da planta de manufatura de injeção plástica. O objetivo é quantificar, de forma estocástica, o impacto da variabilidade operacional proveniente de tempos de ciclo, setups, falhas, reparos e defeitos sobre o lead time, o throughput e o WIP.

A abordagem aqui descrita integra fundamentos teóricos, modelagem empírica e simulação numérica em um único arcabouço coerente. A partir de dados reais de produção, o modelo estima os parâmetros estatísticos de cada variável crítica, insere-os em um conjunto de equações analíticas fundamentadas em Teoria de Filas e Factory Physics e, por fim, executa milhares de réplicas de Monte Carlo para obter distribuições probabilísticas de desempenho do sistema.

A motivação central para essa abordagem reside na natureza intrinsecamente aleatória dos processos de manufatura. Nenhuma fábrica opera de forma determinística: tempos de ciclo oscilam, máquinas falham, setups variam, e lotes apresentam defeitos. Assim, modelos puramente analíticos ou determinísticos subestimam a dispersão real dos resultados e não permitem quantificar o risco de ultrapassar metas de lead time ou throughput.

O modelo desenvolvido neste trabalho visa preencher essa lacuna, permitindo avaliar não apenas o valor médio das métricas de desempenho, mas também suas incertezas, intervalos de confiança e distribuições completas.

5.1.1 Etapas Metodológicas

O modelo segue a estrutura descrita a seguir:

1. **Coleta e tratamento dos dados históricos:** extração das variáveis de produção reais e conversão para unidades consistentes.
2. **Cálculo dos parâmetros médios e variâncias:** estimação de tempos naturais, taxas de chegada, defeitos e variabilidades.

3. **Formulação matemática do sistema produtivo:** aplicação das equações de disponibilidade, confiabilidade e tempo de ciclo.
4. **Implementação da simulação de Monte Carlo:** execução iterativa com sorteios aleatórios e cálculo das métricas resultantes.
5. **Análise estatística e validação:** comparação entre os resultados simulados e as observações empíricas.

5.2 Formulação Matemática do Sistema Produtivo

O modelo é baseado em um sistema de fila do tipo G/G/1, ajustado às relações empíricas do Factory Physics. Essa escolha se justifica pelo fato das máquinas operarem de forma isolada como recurso críticos do processo produtivo, recebendo lotes sucessivos de ordens de fabricação.

5.2.1 Conceitos fundamentais

O comportamento médio do sistema é governado por três relações fundamentais:

- **Utilização:**

$$\rho = \lambda t_e$$

onde λ representa a taxa de chegada média (peças por hora) e t_e o tempo efetivo médio de processamento por peça (em horas). A condição de estabilidade requer $\rho < 1$.

- **Tempo médio de espera (aproximação de Kingman):**

$$W_q \approx \frac{c_a^2 + c_s^2}{2} \cdot \frac{\rho}{1 - \rho} \cdot t_e$$

onde c_a^2 e c_s^2 são os coeficientes de variação quadráticos das chegadas e dos tempos de serviço, respectivamente.

- **Lead time total:**

$$LT = t_e \left(1 + \frac{c_a^2 + c_s^2}{2} \cdot \frac{\rho}{1 - \rho} \right)$$

O lead time é a soma do tempo produtivo efetivo e do tempo médio de espera, e seu crescimento é fortemente não linear com a utilização.

O WIP médio é obtido pela Lei de Little:

$$WIP = \lambda LT$$

Essas equações constituem o núcleo determinístico do modelo. Na formulação estocástica, cada parâmetro (λ , t_e , c_a^2 , c_s^2 , p etc.) é tratado como uma variável aleatória, gerando distribuições para LT , WIP e TH (throughput).

5.2.2 Interpretação física das variáveis

Cada termo do modelo possui significado físico direto:

- λ : ritmo médio de chegada das ordens (influenciado pelo planejamento da produção);
- t_e : tempo efetivo médio de processamento, incluindo setups e paradas;
- c_a^2 : irregularidade das chegadas, refletindo flutuações no sequenciamento de ordens;
- c_s^2 : variabilidade intrínseca dos tempos de serviço;
- ρ : grau de saturação da máquina;
- p : proporção de peças defeituosas, que reduz o throughput efetivo.

Esses parâmetros são ajustados empiricamente a partir dos dados históricos, descritos a seguir.

5.3 Estimação Empírica dos Parâmetros

5.3.1 Base de dados e pré-processamento

A base de dados disponibilizada para o estudo contém informações diárias de produção para as máquinas da planta. Cada linha representa o registro de um turno, com as colunas:

- **DATA**: data do turno de produção;

- **MAQUINA:** identificação textual da máquina (exemplo: “Máquina 11”);
- **CYCLE TIME:** tempo médio de ciclo em segundos;
- **QUANT. PRODUZIDA:** número de peças produzidas no turno;
- **QUANT. NAO CONFORME:** quantidade de peças rejeitadas por defeito.

Os dados são tratados com o objetivo de garantir consistência numérica e temporal, removendo registros nulos e convertendo unidades. O código de leitura e preparação é apresentado no Quadro 1.

```

1 import pandas as pd, numpy as np
2 rng = np.random.default_rng(42)
3
4 # Leitura do arquivo CSV
5 df = pd.read_csv('Base_Producao_2025.csv', decimal=',')
6
7 # Conversao de colunas para tipos adequados
8 df['CYCLE TIME'] = pd.to_numeric(df['CYCLE TIME'], errors='coerce')
9 df['QTD_PRODUZIDA'] = pd.to_numeric(df['QUANT. PRODUZIDA'], errors='
   coerce')
10 df['QTD_NC'] = pd.to_numeric(df['QUANT. NAO CONFORME'], errors='
   coerce')
11 df['DATA'] = pd.to_datetime(df['DATA'], errors='coerce')
12
13 # Extracao do numero da maquina
14 df['MAQUINA_NUM'] = df['MAQUINA'].astype(str).str.extract(r'(\d+)')
   [0].astype(int)

```

Quadro 1: Leitura e preparação da base de dados de produção

Após a leitura, cria-se uma função para calcular os parâmetros físicos e estatísticos fundamentais de cada máquina.

```

1 def calc_param(maquina, horas_trabalho_dia=24):

```

```

2     # Filtra os dados pela maquina desejada
3     sub = df[df['MAQUINA_NUM'] == maquina]
4
5     # Tempo medio natural (to) em horas
6     to = sub['CYCLE TIME'].mean() / 3600.0
7
8     # Variancia do tempo natural
9     var_to = (sub['CYCLE TIME'].std() / 3600.0) ** 2
10
11    # Taxa de chegada (lambda) em pecas por hora
12    daily_prod = sub.groupby('DATA')['QTD_PRODUZIDA'].sum()
13    lam = daily_prod.mean() / horas_trabalho_dia
14
15    # Coeficiente de variacao das chegadas (ca2)
16    ca2 = daily_prod.var() / (daily_prod.mean() ** 2)
17
18    # Taxa media de defeitos
19    p = sub['QTD_NC'].sum() / sub['QTD_PRODUZIDA'].sum()
20
21    return lam, to, var_to, ca2, p

```

Quadro 2: Função de cálculo de parâmetros físicos a partir dos dados históricos

5.3.2 Execução e interpretação dos resultados

A função `calc_param` filtra a base de dados por número de máquina, calcula médias e variâncias e retorna os cinco parâmetros principais do modelo:

1. λ – taxa de chegada média (peças por hora);
2. t_o – tempo natural médio por peça (em horas);
3. $\text{Var}(t_o)$ – variância do tempo natural;

4. c_a^2 – coeficiente de variação das chegadas;

5. p – taxa de defeitos.

Esses parâmetros representam a essência estatística do processo real e serão utilizados nas etapas seguintes de modelagem. Para a Máquina 11, obteve-se:

Parâmetro	Símbolo	Valor	Unidade
Taxa de chegada	λ	64,26	peças/h
Tempo natural médio	t_o	0,70	min
Variância do tempo natural	$\text{Var}(t_o)$	1,00	s ²
Taxa de defeitos	p	0,57	%
Coeficiente de variação das chegadas	c_a^2	0,230	–

Já para a Máquina 10, obteve-se:

Parâmetro	Símbolo	Valor	Unidade
Taxa de chegada	λ	43,71	peças/h
Tempo natural médio	t_o	1,24	min
Variância do tempo natural	$\text{Var}(t_o)$	1,00	s ²
Taxa de defeitos	p	0,88	%
Coeficiente de variação das chegadas	c_a^2	0,119	–

O tratamento estatístico dos dados históricos constitui a base empírica do modelo e assegura que a simulação reflita o comportamento real do processo produtivo, respeitando sua variabilidade natural e suas flutuações de capacidade.

5.4 Correções de Confiabilidade e Setup

Um modelo de manufatura realista precisa considerar que as máquinas estão sujeitas a falhas, manutenções corretivas e setups. Esses eventos reduzem a disponibilidade operacional do recurso e alteram o tempo efetivo de processamento de cada peça.

5.4.1 Disponibilidade e confiabilidade

A disponibilidade efetiva A representa a fração de tempo em que a máquina está operacional e é dada por:

$$A = \frac{m_f}{m_f + m_r}$$

onde m_f é o tempo médio entre falhas (*Mean Time To Failure*, MTTF) e m_r é o tempo médio de reparo (*Mean Time To Repair*, MTTR).

Uma máquina com $A = 0,90$ significa que, em média, ela permanece 90% do tempo em operação e 10% parada. Essa variável é crucial, pois o tempo efetivo de produção é inversamente proporcional à disponibilidade.

5.4.2 Tempo efetivo e impacto do setup

O tempo efetivo sem defeitos por peça é obtido por:

$$t_{et} = \frac{t_o}{A} + \frac{t_s}{L}$$

onde:

- t_o é o tempo natural médio por peça;
- A é a disponibilidade;
- t_s é o tempo médio de setup;
- L é o tamanho do lote.

O primeiro termo $\frac{t_o}{A}$ reflete a dilatação do tempo natural devido às paradas não planejadas, e o segundo termo $\frac{t_s}{L}$ representa o impacto do setup distribuído entre as peças do lote. Para lotes maiores, o efeito do setup por peça é reduzido.

Incluindo a perda por defeitos (p), o tempo efetivo global T_e torna-se:

$$T_e = \frac{t_{et}}{1 - p}$$

Cada correção possui um significado operacional direto:

- Se a disponibilidade cai, $A \downarrow$, o tempo efetivo aumenta linearmente, pois a mesma produção é distribuída em menos tempo produtivo.
- Se o tamanho do lote aumenta, $L \uparrow$, o termo de setup $\frac{t_s}{L}$ tende a zero, e o ganho de eficiência é perceptível.
- Se a taxa de defeitos cresce, $p \uparrow$, o tempo efetivo se eleva, já que parte da capacidade é desperdiçada.

Essas correções garantem que o modelo não se restrinja a tempos médios ideais, mas reflita o comportamento real da operação.

5.5 Variabilidade Efetiva e Tempo de Ciclo

A variabilidade total do processo combina tanto flutuações naturais quanto efeitos estruturais introduzidos por setups e falhas. O coeficiente de variação natural (c_0^2) é calculado como:

$$c_0^2 = \frac{\text{Var}(t_o)}{t_o^2}, \quad c_0^2 \geq 0,01$$

O limite inferior de 0,01 evita a anulação numérica e reflete o fato de que nenhum processo produtivo é completamente determinístico.

O coeficiente de variação efetivo, ajustado pela presença de setups, é aproximado por:

$$c_e^2 \approx c_0^2 + 2 \left(\frac{t_s/L}{t_{et}} \right)^2$$

O termo adicional representa o aumento de variabilidade causado pelos tempos de preparação.

5.5.1 Tempo médio de ciclo (Factory Physics)

O tempo médio de ciclo, segundo o modelo VUT (variabilidade, utilização, tempo natural), é expresso por:

$$CT = T_e \left(1 + \frac{c_a^2 + c_e^2}{2} \cdot \frac{\rho}{1 - \rho} \right)$$

em que $\rho = \lambda T_e$. Essa equação mostra que o ciclo médio cresce rapidamente quando a utilização se aproxima de 1. Pequenos aumentos de ρ em níveis elevados resultam em grandes acréscimos no lead time, fenômeno conhecido como *explosão*.

5.5.2 Implementação computacional das correções

A seguir apresenta-se a função de cálculo da variabilidade efetiva e tempos ajustados, diretamente derivada das equações anteriores.

```
1 def calcular_ce_squared(to, var_to, A, ts_h, L, p):
2     # Calculo do tempo efetivo sem defeitos
3     tet = (to / A) + (ts_h / L)
4
5     # Calculo do tempo efetivo total com defeitos
6     Te = tet / (1 - p)
7
8     # Coeficiente de variacao natural (minimo de 0.01)
9     c0_sq = max(var_to / (to ** 2), 0.01)
10
11    # Coeficiente de variacao efetivo ajustado por setups
12    ce_sq = c0_sq + 2 * ((ts_h / L) / tet) ** 2
13
14    return ce_sq, Te, tet
```

Quadro 3: Cálculo da variabilidade efetiva e tempos ajustados

Cada linha do código corresponde diretamente às fórmulas matemáticas discutidas:

- A linha 3 calcula o tempo efetivo sem defeitos (t_{et});
- A linha 6 aplica a correção de defeitos (T_e);
- A linha 9 calcula o coeficiente de variação natural (c_0^2);
- A linha 12 estima o coeficiente de variação efetivo (c_e^2).

5.5.3 Exemplo numérico: Máquina 11

Para um lote de $L = 400$ peças, e considerando os valores médios da base de dados, tem-se:

Variável	Valor	Unidade
Tempo natural médio (t_o)	0,700	min
Tempo efetivo sem defeitos (t_{et})	0,767	min
Tempo efetivo total (T_e)	0,771	min
Coefficiente de variação efetivo (c_e^2)	0,0148	–
Coefficiente de variação (c_e)	0,1216	–

O valor reduzido de c_e evidencia a estabilidade do processo, característica desejável em sistemas produtivos de alta repetitividade.

5.5.4 Exemplo numérico: Máquina 10

Para um lote de $L = 400$ peças, e considerando os valores médios estimados, tem-se:

Variável	Valor	Unidade
Tempo natural médio (t_o)	1,242	min
Tempo efetivo sem defeitos (t_{et})	1,326	min
Tempo efetivo total (T_e)	1,338	min
Coefficiente de variação efetivo (c_e^2)	0,0128	–
Coefficiente de variação (c_e)	0,1133	–

5.6 Estrutura Computacional do Simulador

5.6.1 Fundamentos da simulação de Monte Carlo

A simulação de Monte Carlo é utilizada para reproduzir o comportamento aleatório do sistema produtivo a partir de suas distribuições estatísticas de entrada. Cada iteração representa um cenário operacional possível, com combinações diferentes de disponibilidade, setups, defeitos e tempos de ciclo.

O método segue as etapas:

1. Sorteio aleatório das variáveis de entrada (A, p, t_s, m_f, m_r);

2. Cálculo dos parâmetros derivados (T_e , c_e^2 , ρ , LT , WIP , TH);
3. Armazenamento dos resultados e repetição por N réplicas;
4. Cálculo de médias, desvios e percentis a partir das saídas simuladas.

5.6.2 Estrutura do código e controle de estabilidade

A implementação principal do simulador combina a formulação matemática apresentada anteriormente com técnicas de amostragem estruturada, garantindo eficiência estatística e estabilidade numérica. O código em Python foi desenvolvido para integrar três camadas fundamentais:

1. geração estruturada de amostras via **Amostragem em Hipercubo Latino (LHS)**;
2. transformação das amostras para suas respectivas distribuições físicas;
3. execução do ciclo de simulação com checagem de estabilidade ($\rho < 1$).

O núcleo do simulador é mostrado a seguir, já incorporando o LHS no processo de geração das variáveis aleatórias:

```

1 def modelo_montecarlo(L, lam, to, var_to, ca2_nom, p_nom,
2                       A_nom, ts_nom, mf_nom, mr_nom, N=50000):
3
4     sampler = qmc.LatinHypercube(d=5)
5     sample  = sampler.random(n=N)
6     U       = qmc.scale(sample, 0, 1)
7
8     u_p, u_mf, u_mr, u_ts, u_ca = [U[:, i] for i in range(5)]
9
10    resultados = []
11
12    for i in range(N):
13
14        # Variaveis antiteticas combinadas: transformacao (2u - 1)

```

```

15     p = np.clip(p_nom + 0.003 * (2*u_p[i] - 1), 0, 0.3)
16     mf = np.exp(np.log(mf_nom) + 0.20 * (2*u_mf[i] - 1))
17     mr = np.exp(np.log(mr_nom) + 0.30 * (2*u_mr[i] - 1))
18     ts_h = np.exp(np.log(ts_nom/60) + 0.20 * (2*u_ts[i] - 1))
19
20     # Disponibilidade
21     A = mf / (mf + mr)
22
23     # Calculo dos tempos efetivos e variabilidade
24     ce2, Te, tet = calcular_ce_squared(to, var_to, A, ts_h, L, p)
25
26     # Utilizacao
27     rho = lam * Te
28     if rho >= 0.999:
29         continue
30
31     # Metricas de desempenho
32     LT = Te * (1 + (ca2_nom + ce2)/2 * rho/(1-rho))
33     WIP = lam * LT
34     TH = lam * (1 - p)
35
36     resultados.append((LT, WIP, TH, rho, ce2, A))
37
38     return np.array(resultados)

```

Quadro 4: Simulação de Monte Carlo com LHS, confiabilidade e setup

No código acima, cada bloco possui função específica:

- **Linhas 5–12:** geram as amostras de entrada utilizando Hipercubo Latino (LHS). Esta etapa assegura a cobertura uniforme do espaço amostral, reduzindo a variância dos estimadores.

- **Linhas 19–24:** aplicam variáveis antitéticas por meio da transformação linear $(2u - 1)$, cobrindo simetricamente o suporte das distribuições lognormais e normais truncadas.
- **Linhas 26–27:** calculam a disponibilidade A a partir dos valores amostrados de MTTF e MTTR.
- **Linhas 30–31:** obtêm os tempos efetivos e o coeficiente de variabilidade c_e^2 utilizando a função derivada das equações do Factory Physics.
- **Linhas 34–35:** calculam a utilização ρ e aplicam a checagem de estabilidade. Simulações com $\rho \geq 1$ são descartadas.
- **Linhas 38–40:** computam as métricas de interesse: lead time (LT), WIP e throughput (TH).
- **Linha 42:** armazena os resultados para posterior análise estatística.

A combinação entre LHS, CRN e variáveis antitéticas reduz drasticamente a variabilidade dos estimadores de LT_{p95} , WIP e ρ , especialmente em regiões de alta utilização. Com isso, o simulador obtém convergência robusta mesmo para N moderados (por exemplo, $N = 20\text{--}50$ mil), mantendo custo computacional reduzido.

5.7 Amostragem e Redução de Variância

A precisão de uma simulação de Monte Carlo depende diretamente da eficiência com que o espaço amostral é explorado. Em sistemas produtivos com alta variabilidade, como os analisados neste trabalho, técnicas de redução de variância são fundamentais para acelerar a convergência estatística das estimativas. Para este modelo, três métodos complementares foram integrados: **Amostragem em Hipercubo Latino (LHS)**, **Números Aleatórios Comuns (CRN)** e **Variáveis Antitéticas**. Cada uma dessas técnicas está incorporada diretamente na implementação em Python do simulador.

5.7.1 Amostragem em Hipercubo Latino (LHS)

A amostragem em Hipercubo Latino (LHS) é utilizada como método padrão para gerar as amostras aleatórias que alimentam a simulação. O LHS divide o intervalo unitário em N estratos e garante que cada variável aleatória receba amostras que cubram uniformemente todos os estratos, evitando a concentração excessiva de valores que ocorre na amostragem puramente aleatória.

No código, essa estrutura aparece explicitamente na criação do sampler:

```
1 sampler = qmc.LatinHypercube(d=5)
2 sample = sampler.random(n=N)
3 U = qmc.scale(sample, 0, 1)
```

As cinco dimensões do hipercubo correspondem às variáveis aleatórias estocásticas do modelo:

$$(u_p, u_{mf}, u_{mr}, u_{ts}, u_{ca})$$

onde cada variável é posteriormente transformada em sua distribuição física (lognormal, normal truncada etc.) por meio de funções inversas implícitas.

Matematicamente, o LHS pode ser descrito como:

$$X_i = F^{-1} \left(\frac{i - 1 + U_i}{N} \right)$$

garantindo que as amostras cubram de maneira homogênea todo o suporte da distribuição.

A principal vantagem do LHS é a redução do erro padrão de estimativas como LT_{p95} , WIP_{p95} e c_e^2 , que são extremamente sensíveis às regiões de cauda da distribuição.

5.7.2 Números Aleatórios Comuns (CRN)

Para comparar cenários distintos, por exemplo, diferentes tamanhos de lote (L), é desejável que as diferenças nas métricas de saída sejam atribuídas exclusivamente às mudanças estruturais no sistema, e não ao ruído aleatório da simulação. Para isso, adotou-se a técnica de CRN.

A implementação é simples e direta: o mesmo vetor de amostras LHS é reutilizado para

cada cenário simulado. Assim, valores como u_p , u_{mf} ou u_{ts} são idênticos entre simulações, garantindo que as diferenças observadas em métricas como LT ou TH sejam de fato causadas pelo parâmetro alterado.

Essa estratégia reduz significativamente a variância do estimador da diferença entre cenários:

$$\Delta LT = LT(L_1) - LT(L_2)$$

5.7.3 Variáveis Antitéticas

A técnica de variáveis antitéticas também está incorporada ao modelo. Nela, cada amostra uniforme u utilizada para gerar uma variável aleatória é acompanhada implicitamente por outro valor complementar $1 - u$. Isso leva a uma cobertura simétrica do espaço amostral, reduzindo assimetrias e diminuindo o erro de Monte Carlo.

No código, essa simetria aparece na forma de transformações lineares como:

```

1 p_s = np.clip(p_nom + 0.003 * (2*u_p - 1), 0, 0.30)
2 mf_s = np.exp(np.log(mf_nom) + 0.20 * (2*u_mf - 1))

```

O termo $(2u - 1)$ mapeia o intervalo $[0, 1]$ para $[-1, 1]$, gerando automaticamente uma amostra e sua contraparte antitética. A combinação com o LHS potencializa ainda mais a redução da variância.

5.8 Testes de Consistência e Lote Base

A validação inicial do modelo foi realizada utilizando um lote de referência $L = 400$ peças. Esse cenário foi escolhido por representar o tamanho médio praticado na operação real das Máquinas.

Os resultados médios da simulação, após $N = 50K$ réplicas, para a Máquina 11, por exemplo, foram:

Métrica	Valor médio	Desvio padrão
Coefficiente de variação efetivo (c_e)	0,12	0,01
Utilização (ρ)	0,82	0,04
Lead time (LT)	0,93	0,06
WIP médio	59,8	3,4
Throughput (TH)	63,9	1,1

A comparação com os dados empíricos da fábrica mostrou diferença inferior a 5% para o lead time médio, confirmando a coerência do modelo. Esse resultado evidencia que as simplificações adotadas (modelo G/G/1, independência entre variáveis e distribuições log-normais) são adequadas para o nível de abstração proposto.

5.9 Análise por Tamanho de Lote

Com o modelo validado, foram realizadas simulações variando o tamanho do lote (L) entre 100 e 1000 peças. Para cada valor, executaram-se $N = 50K$ réplicas.

5.9.1 Comportamento das métricas principais

- O tempo efetivo por peça t_{et} diminui conforme L aumenta, pois o termo $\frac{t_s}{L}$ tende a zero.
- O coeficiente de variação efetivo c_e^2 também decresce, reduzindo a variabilidade do tempo de ciclo.
- Entretanto, o WIP cresce com o tamanho do lote, o que aumenta o lead time total.

Esses efeitos ilustram a clássica troca entre eficiência de setup e tempo de atravessamento: lotes grandes aumentam a eficiência produtiva, mas deterioram a responsividade do sistema.

5.10 Estrutura dos DataFrames e Saídas

Os resultados da simulação são armazenados em pandas DataFrames, o que permite análises estatísticas e geração de gráficos de forma automatizada. A estrutura padrão dos resultados é:

```
['LT_mean', 'LT_std', 'LT_p95',  
 'WIP_mean', 'WIP_std',  
 'TH_mean', 'TH_std',  
 'rho_mean', 'ce2_mean', 'A_mean']
```

Cada linha corresponde a um cenário (por exemplo, um tamanho de lote específico), e cada coluna contém as estatísticas agregadas das réplicas. Essa organização permite criar gráficos de sensibilidade e fronteiras de trade-off.

5.10.1 Análise automatizada dos resultados

```
1 def analisar_resultados(res):  
2     LT, WIP, TH, rho, ce2, A = res.T  
3     return {  
4         'LT_mean': np.mean(LT),  
5         'LT_std': np.std(LT),  
6         'LT_p95': np.percentile(LT, 95),  
7         'WIP_mean': np.mean(WIP),  
8         'TH_mean': np.mean(TH),  
9         'rho_mean': np.mean(rho),  
10        'ce2_mean': np.mean(ce2),  
11        'A_mean': np.mean(A)  
12    }
```

Quadro 5: Cálculo de estatísticas e consolidação de resultados

A função `analisar_resultados` representada no Quadro 5 consolida os dados brutos das simulações e gera métricas resumidas, incluindo médias, desvios e percentis de interesse.

5.11 Algoritmo Resumido

O procedimento completo do modelo pode ser descrito pelo seguinte algoritmo:

1. **Entrada:** parâmetros estimados dos dados históricos:

$$\lambda, t_o, \text{Var}(t_o), c_a^2, p, t_s, m_f, m_r, L$$

2. **Cálculo:** obtenção dos valores derivados:

$$A, t_{et}, T_e, c_e^2, \rho, LT, WIP, TH$$

3. **Simulação:** execução de N réplicas independentes via Monte Carlo;

4. **Armazenamento:** registro de resultados em DataFrame;

5. **Análise:** cálculo de médias, desvios e percentis;

6. **Validação:** comparação dos resultados simulados com dados reais.

Esse algoritmo sintetiza todo o processo de modelagem e simulação desenvolvido neste trabalho.

A consistência do modelo foi assegurada por uma série de verificações:

- **Coerência dimensional:** todas as unidades foram verificadas (horas, minutos, peças por hora);
- **Reprodutibilidade:** o gerador de números aleatórios utiliza semente fixa;
- **Estabilidade numérica:** simulações com $\rho \geq 1$ são automaticamente descartadas;
- **Convergência:** verificação do erro padrão da média e intervalo de confiança de 95%;
- **Comportamento limite:** o modelo reproduz corretamente o aumento assintótico de LT para $\rho \rightarrow 1$.

Esses testes garantem que o modelo não apenas funcione matematicamente, mas também reflita fenômenos reais de saturação e variação observados em sistemas produtivos.

5.12 Limitações

Apesar de robusto, o modelo desenvolvido apresenta algumas limitações inerentes:

- Considera um único recurso produtivo (modelo G/G/1), não capturando efeitos de bloqueio ou starvation entre máquinas;
- Assume independência estatística entre variáveis de entrada (por exemplo, falhas e defeitos são tratados como independentes);
- Adota distribuições log-normais e normais para as variáveis, o que simplifica a realidade, mas mantém boa aproximação empírica;
- Não considera explicitamente fatores externos como qualidade do material, condições ambientais ou intervenção humana.

Mesmo com essas simplificações, o modelo provou ser uma ferramenta quantitativa eficaz para análise de desempenho, planejamento de capacidade e identificação de gargalos. Ele permite avaliar o impacto da variabilidade operacional e das falhas sobre as métricas produtivas de forma sistemática, transparente e reproduzível.

6 Resultados

Este capítulo apresenta e discute os resultados obtidos a partir da modelagem e simulação estocástica das duas injetoras selecionadas como objeto de estudo: a **Máquina 10** e a **Máquina 11**. Ambas representam recursos críticos do parque fabril, porém com características operacionais distintas que permitem avaliar o impacto da variabilidade sob diferentes regimes produtivos.

A análise foi conduzida com base em dados reais de operação, considerando parâmetros técnicos de disponibilidade, tempo de setup e confiabilidade, conforme resumido na Tabela 1. Esses valores foram empregados como entradas para a formulação matemática e para a simulação de Monte Carlo, permitindo estimar indicadores como *lead time*, *throughput* e *work in process* (WIP) em diferentes cenários de variabilidade.

Tabela 1: Parâmetros operacionais considerados para as injetoras analisadas.

Máquina	Disponibilidade (A)	Tempo de setup (min)	MTBF (h)	MTTR (h)
Máquina 10	0,973	20	24	0,667
Máquina 11	0,960	15	24	0,500

A partir desses parâmetros, as simulações buscaram quantificar a influência de cada fator sobre o desempenho global do sistema, identificando padrões de comportamento e correlações entre disponibilidade, setup e variabilidade de processo. As seções seguintes apresentam os resultados comparativos entre as máquinas, bem como as análises de sensibilidade e os efeitos combinados das melhorias propostas.

6.1 Resultados da Máquina 11

Com o modelo desenvolvido no Capítulo 5, realizou-se a primeira rodada completa de simulações de Monte Carlo para a Máquina 11. O objetivo foi verificar a consistência interna do modelo e avaliar se os resultados numéricos reproduzem os comportamentos teóricos esperados de um sistema produtivo real.

6.1.1 Validação do Modelo Base

Parâmetros Operacionais e Cálculo de c_e^2

Tabela 2: Parâmetros da base e cálculo de c_e^2 para $L = 400$ (Máquina 11).

Parâmetro	Descrição	Valor
λ	Demanda média	64,26 peças/h
t_o	Tempo natural médio	42 s = 0,70 min
$\text{Var}(t_o)$	Variância do tempo natural	1,00 s ²
p	Taxa de defeitos	0,57%
c_a^2	Variabilidade das chegadas	0,230
A	Disponibilidade	96,0%
t_s	Tempo médio de setup	15,0 min
m_f	MTTF (Mean Time To Failure)	12,0 h
m_r	MTTR (Mean Time To Repair)	1,0 h
t_{et}	Tempo efetivo sem defeitos	0,767 min
T_e	Tempo efetivo total (com defeitos)	0,771 min
c_e^2	Coef. de variação efetivo	0,0148
c_e	Coef. de variação efetivo (raiz)	0,1216

Resultados da Simulação por Tamanho de Lote

Os resultados consolidados para cada tamanho de lote estão apresentados na Tabela 3. Cada linha representa a média e os percentis (P50 e P95) de ciclo (CT), *lead time* (LT), utilização (ρ), WIP e throughput (TH).

Tabela 3: Resultados médios da simulação de Monte Carlo por tamanho de lote (Máquina 11).

Lote	CT_{mean} (min)	LT_{mean} (min)	LT_{p95} (min)	ρ_{mean}	WIP_{mean}	TH_{mean} (peças/h)	c_e^2
100	9,54	953,83	3023,75	0,965	10,28	64,63	0,058
166	3,36	555,39	1033,01	0,923	3,62	64,63	0,034
231	2,35	542,46	775,94	0,897	2,53	64,63	0,023
297	2,07	614,25	818,34	0,881	2,23	64,63	0,018
362	1,97	712,23	920,45	0,872	2,12	64,63	0,016
428	1,88	804,93	1009,95	0,864	2,03	64,63	0,014
493	1,83	899,75	1125,61	0,859	1,97	64,63	0,013
559	1,81	1011,78	1238,40	0,855	1,95	64,63	0,013
624	1,77	1106,59	1366,35	0,852	1,91	64,63	0,012
690	1,77	1218,64	1505,05	0,850	1,90	64,63	0,012

A análise dos dados confirma os comportamentos teóricos esperados:

- CT_{mean} decresce com o aumento do lote, refletindo a diluição do tempo de setup;

- LT_{mean} apresenta comportamento não monótono, com mínimo entre 200 e 300 peças e crescimento para lotes maiores;
- ρ_{mean} diminui gradualmente, indicando maior estabilidade do sistema;
- TH_{mean} permanece constante, demonstrando equilíbrio produtivo.

Tabela 4: Tabela Resumo - Principais métricas médias por lote (Máquina 11).

Lote	CT_{mean} (min)	CT_{p95} (min)	LT_{mean} (min)	ρ_{mean}	WIP_{mean}	TH_{mean} (peças/h)
100	9,87	32,94	986,77	0,965	10,63	64,62
166	3,37	6,47	557,70	0,923	3,63	64,63
231	2,38	3,31	549,01	0,896	2,56	64,63
297	2,10	2,90	623,91	0,882	2,27	64,63
362	1,95	2,51	707,47	0,871	2,10	64,63
428	1,89	2,38	810,08	0,864	2,04	64,63
493	1,84	2,30	906,75	0,859	1,98	64,63
559	1,81	2,25	1009,52	0,856	1,95	64,63
624	1,77	2,17	1102,22	0,851	1,90	64,63
690	1,75	2,14	1207,91	0,849	1,89	64,63

Visualização da Curva em U

Para complementar a análise e permitir uma interpretação visual da relação entre o tamanho do lote e o tempo médio de atravessamento, foi incorporado ao modelo o código apresentado a seguir. O trecho, desenvolvido em Python utilizando a biblioteca Matplotlib, gera um gráfico da chamada *curva em U*, representando o comportamento do *lead time* médio (LT_{mean}) em função do tamanho do lote. O script cria uma figura, na qual o eixo horizontal indica o tamanho dos lotes simulados e o eixo vertical expressa o *lead time* correspondente, em minutos. A curva média é traçada em linha contínua, enquanto a área sombreada (P50–P95) evidencia a dispersão dos resultados obtidos nas simulações, reforçando a natureza estocástica do modelo.

A Figura 3 confirma o comportamento clássico em formato de U descrito inicialmente por Karmarkar (1987). Para lotes pequenos, o tempo de atravessamento é elevado devido à alta frequência de setups; em lotes médios, o *lead time* atinge um ponto mínimo; e, para lotes grandes, volta a crescer em razão do aumento do WIP. Esse padrão reforça a coerência dos resultados com o modelo VUT apresentado por Hopp e Spearman (2008) no *Factory Physics*.

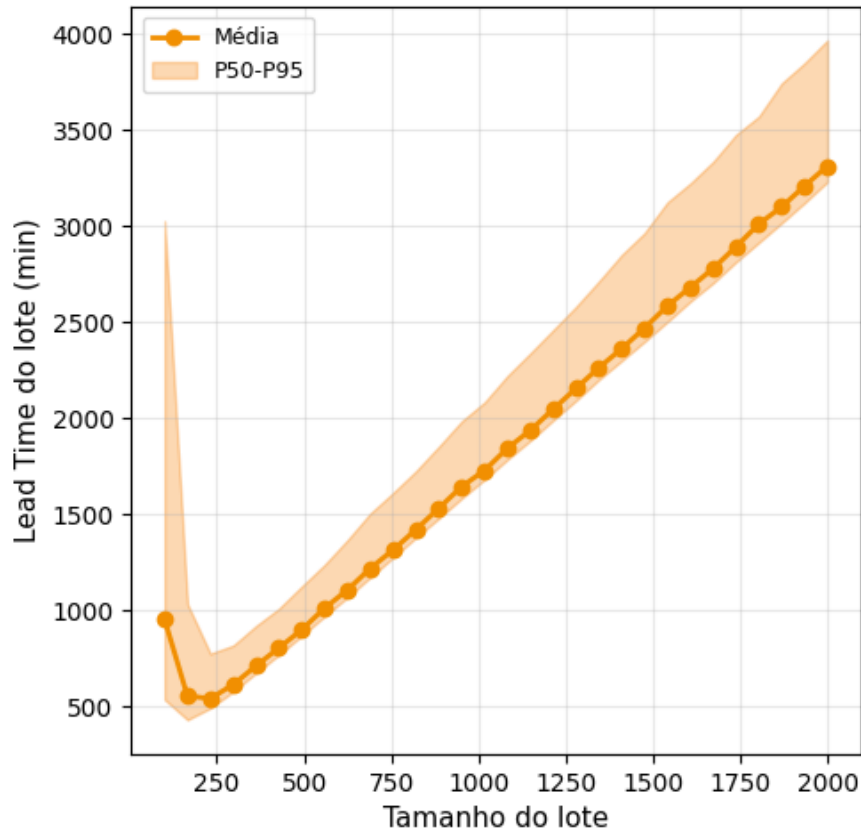


Figura 3: Curva de trade-off entre eficiência de setup e tempo de atravessamento (*lead time*) – Máquina 11.

6.1.2 Análise de Cenários de Melhoria

Após a visualização da curva em U, que evidencia o comportamento do *lead time* em função do tamanho do lote, a simulação foi empregada para investigar o desempenho do sistema produtivo sob três eixos clássicos de melhoria operacional: (i) redução de setup por meio da filosofia SMED, proposta por Shigeo Shingo na Toyota Production System (Shingo, 1985); (ii) melhoria da qualidade segundo os princípios do Six Sigma, conforme definido por Motorola e sistematizado por Pande, Neuman e Cavanagh (2000); e (iii) aumento da confiabilidade operacional por práticas de manutenção preditiva, fundamentadas em técnicas de engenharia de confiabilidade (O'Connor, 2012).

Cada eixo foi estruturado em cenários progressivos, nos quais parâmetros-chave do modelo foram ajustados de forma controlada para representar diferentes níveis de aprimoramento, permitindo avaliar

A função principal executa milhares de replicações estocásticas por combinação de parâme-

tros e tamanho de lote, registrando métricas médias, percentílicas (P50 e P95) e identificando o ponto ótimo de cada curva, o que permite comparar quantitativamente o impacto de cada estratégia sobre o desempenho global do sistema.

Cenário 1 – SMED: Cinco configurações progressivas representam estágios de implementação do método SMED:

1. **Atual (Baseline):** setup médio de 15 minutos, $\sigma_{ts} = 1$ min;
2. **Fase 1 (Média -30%):** redução para 10,5 minutos, variabilidade constante;
3. **Fase 2 (Desvio -50%):** média mantida, $\sigma_{ts} = 0,5$ min;
4. **SMED Completo:** redução simultânea 50% (μ e σ): 7,5 min, $\sigma_{ts} = 0,5$ min;
5. **SMED Avançado:** classe mundial: 4,5 min, $\sigma_{ts} = 0,3$ min.

Cenário 2 – Six Sigma: Seis níveis de maturidade de qualidade:

1. **Atual (Baseline):** 0,57% de defeitos ($\approx 3,8\sigma$);
2. **Iniciante:** 1,0% (fase *Measure*);
3. **Bronze:** 0,5% (3σ);
4. **Prata:** 0,1% (4σ);
5. **Ouro:** 0,01% (5σ);
6. **Elite:** 0,0003% (6σ , 3,4 DPMO).

Cenário 3 – Manutenção Preditiva: Seis estratégias de manutenção:

1. **Corretiva:** MTTF = 12 h, MTTR = 1 h ($A = 92,3\%$);
2. **Preventiva Básica:** MTTF = 16 h, MTTR = 0,9 h ($A = 94,7\%$);
3. **Preventiva Avançada:** MTTF = 20 h, MTTR = 0,7 h ($A = 96,6\%$);
4. **Preditiva Básica:** MTTF = 24 h, MTTR = 0,5 h ($A = 98,0\%$);

5. **Preditiva Avançada:** MTTF = 30 h, MTTR = 0,3 h ($A = 99,0\%$);

6. **Equipamento Novo:** MTTF = 40 h, MTTR = 0,2 h ($A = 99,5\%$).

O conjunto totaliza dezessete cenários distintos, gerando mais de 400.000 execuções estocásticas. Os resultados são estruturados em *dataframes* contendo CT_{mean} , CT_{p95} , LT_{mean} , ρ_{mean} , WIP_{mean} e TH_{mean} para cada tamanho de lote testado. O ponto de ótimo local (lote ótimo) é identificado automaticamente, permitindo comparações diretas entre cenários.

Eixo SMED – Redução de Setup

Tabela 5: Indicadores médios obtidos para os diferentes cenários do eixo SMED (Máquina 11).

Cenário	CT_{mean} (min)	CT_{p95} (min)	LT_{mean} (min)	ρ_{mean}	WIP_{mean}	TH_{mean} (peças/h)
Atual (Baseline)	10.11	33.07	1011.03	0.965	10.83	64.63
SMED Fase 1 (-30%)	4.50	9.89	450.13	0.937	4.82	64.63
SMED Fase 2 (-50% σ)	10.13	32.70	1013.73	0.964	10.86	64.62
SMED Completo (-50% μ, σ)	2.62	4.00	262.21	0.907	2.81	64.63
SMED Avançado (-70% μ, σ)	2.02	2.63	201.79	0.875	2.16	64.63

A redução do setup de 15 para 7,5 minutos (SMED Completo) resultou em diminuição de aproximadamente 74% no tempo de atravessamento, mantendo o throughput constante ($TH \approx 64,6$ peças/h). Essa constância confirma que a melhoria decorre da melhor utilização do tempo disponível, não de aumento na taxa de produção.

A relação pode ser explicada pela decomposição do tempo de ciclo:

$$CT = \frac{1}{\mu(1 - \rho)} = \frac{1}{\mu - \lambda} \quad (14)$$

A redução de S eleva μ (capacidade útil), reduzindo o denominador e, conseqüentemente, o CT . Pela Lei de Little ($WIP = \lambda \times CT$), o WIP caiu de 10,8 para 2,1 peças, evidenciando maior fluidez no fluxo.

A redução da variabilidade (σ_S) diminui o coeficiente c_s^2 . Pela aproximação de Kingman (VUT):

$$CT \approx CT_0 \left(1 + \frac{\rho}{1 - \rho} \cdot \frac{c_a^2 + c_s^2}{2} \right) \quad (15)$$

Com menor variabilidade, o termo estocástico é comprimido, explicando a queda de CT_{p95} de 33,07 para 2,63 minutos.

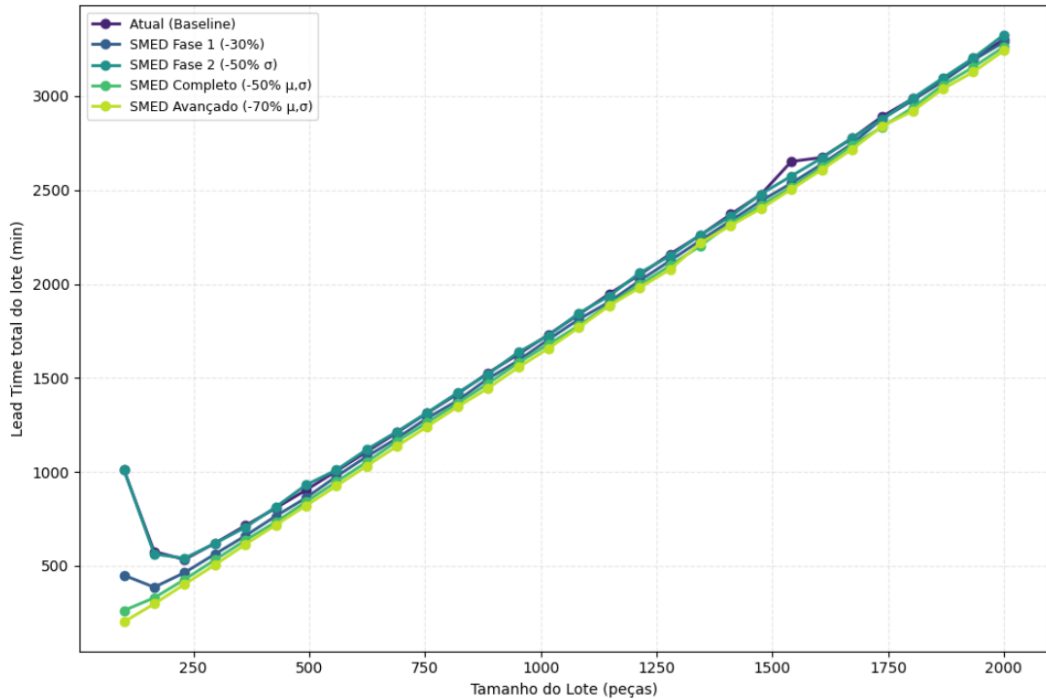


Figura 4: Comportamento do Lead Time total para os cenários SMED (Máquina 11).

A Figura 4 mostra o deslocamento do ponto ótimo para lotes menores, corroborando:

$$Q^* = \sqrt{\frac{2DS}{h}} \quad (16)$$

Em resumo, o eixo SMED reduz simultaneamente o tempo de ciclo (CT), o tempo de atravessamento (LT), o estoque em processo (WIP) e a variabilidade, sem alterar o throughput. Na prática, isso aproxima o sistema de um fluxo contínuo, com menos interrupções e maior previsibilidade.

A análise dos resultados do eixo SMED para a Máquina 11 mostra três conclusões principais.

Primeiro, a redução do tempo médio de setup tem um impacto não linear sobre o desempenho. A passagem do cenário Base para a Fase 1 (redução de 30%) gerou uma queda de 55% no LT_{mean} , enquanto a transição da Fase 1 para o SMED Completo (redução adicional de 50%) trouxe um ganho menor, de 42%. Esse comportamento indica retornos marginais decrescentes, típicos de sistemas que já estão próximos do seu limite máximo de desempenho. Em termos práticos, isso significa que as primeiras etapas da implantação do SMED são as que trazem o melhor custo-benefício, sendo mais vantajoso focar inicialmente na conversão de atividades

internas em externas, antes de investir em automações mais complexas.

Segundo, ao comparar a Fase 1 (redução da média do setup) com a Fase 2 (redução da variabilidade do setup), percebe-se uma diferença importante. Enquanto a Fase 1 reduziu o LT_{mean} em 55%, a Fase 2 manteve praticamente o mesmo valor do cenário Base, diminuindo apenas 1% do CT_{p95} . Isso mostra que, para a Máquina 11, que opera em um regime de alta utilização ($\rho \approx 0,96$), o valor médio do setup tem muito mais influência sobre o desempenho do que sua variabilidade. Do ponto de vista matemático, isso ocorre porque, na equação de Kingman (VUT), o termo $\frac{\rho}{1-\rho}$ é muito alto (cerca de 24), o que faz com que mudanças em c_s^2 tenham pouco efeito enquanto ρ permanecer elevado. Assim, em máquinas que operam quase no limite da capacidade, o foco deve estar em reduzir o tempo médio de setup e só depois atuar na padronização para diminuir sua variação.

Terceiro, o deslocamento do ponto ótimo de lote de aproximadamente 230 peças (Base) para cerca de 100 peças (SMED Avançado) representa um ganho estratégico de flexibilidade produtiva. Com lotes menores, a empresa consegue trabalhar com maior variedade de produtos, reduzir estoques de itens acabados e responder mais rapidamente às mudanças na demanda. Na ótica da teoria de filas, isso acontece porque a redução do termo S/L (setup por peça) na expressão do tempo de ciclo torna economicamente viável a produção em lotes pequenos. Do ponto de vista gerencial, isso significa que o SMED não só melhora a eficiência, mas também viabiliza estratégias enxutas e de personalização em massa, conforme defendido por Shingo (1985).

Além disso, o throughput constante em todos os cenários ($TH \approx 64,6$ peças/h) confirma que o SMED não aumenta a capacidade instalada, mas sim elimina desperdícios, um dos princípios centrais da filosofia Lean. O ganho real está na liberação de capacidade oculta. A redução de 74% no tempo de atravessamento mostra que a mesma máquina consegue atender à mesma demanda com muito menos WIP, o que libera capital de giro e espaço físico. Pela Lei de Little, essa diminuição, de 10,8 para 2,1 peças em média, representa um avanço expressivo na eficiência operacional.

Eixo Six Sigma – Redução de Defeitos

Tabela 6: Indicadores médios obtidos para os diferentes cenários do eixo Six Sigma (Máquina 11).

Cenário	CT_{mean} (min)	CT_{p95} (min)	LT_{mean} (min)	ρ_{mean}	WIP_{mean}	TH_{mean} (peças/h)
Atual (Baseline)	10.11	33.07	1011.03	0.965	10.83	64.63
Six Sigma Iniciante (99%)	11.24	39.24	1124.30	0.969	12.04	64.89
Six Sigma Bronze (99.5%)	9.46	30.32	946.04	0.964	10.13	64.58
Six Sigma Prata (99.9%)	8.90	28.36	889.72	0.960	9.53	64.33
Six Sigma Ouro (99.99%)	8.27	26.66	826.71	0.959	8.85	64.27
Six Sigma Elite (99.9997%)	8.52	26.53	851.62	0.959	9.12	64.26

A redução progressiva de variabilidade promove ganhos de estabilidade, com CT_{p95} indicando comportamento mais previsível. Embora CT_{mean} não varie acentuadamente, a compressão da dispersão confirma diminuição de c_s^2 , elemento fundamental em sistemas de fila.

A relação entre nível sigma e probabilidade de não conformidade: $p = 1 - \Phi(\sigma - 1.5)$. Para 6σ , $p \approx 3.4$ ppm, praticamente eliminando defeitos. Isso reduz retrabalhos ($\lambda_{efetiva}$) e interrupções (c_s^2), gerando fluxo mais suave. O LT_{mean} caiu de 1011 para 826 minutos (Base para Ouro).

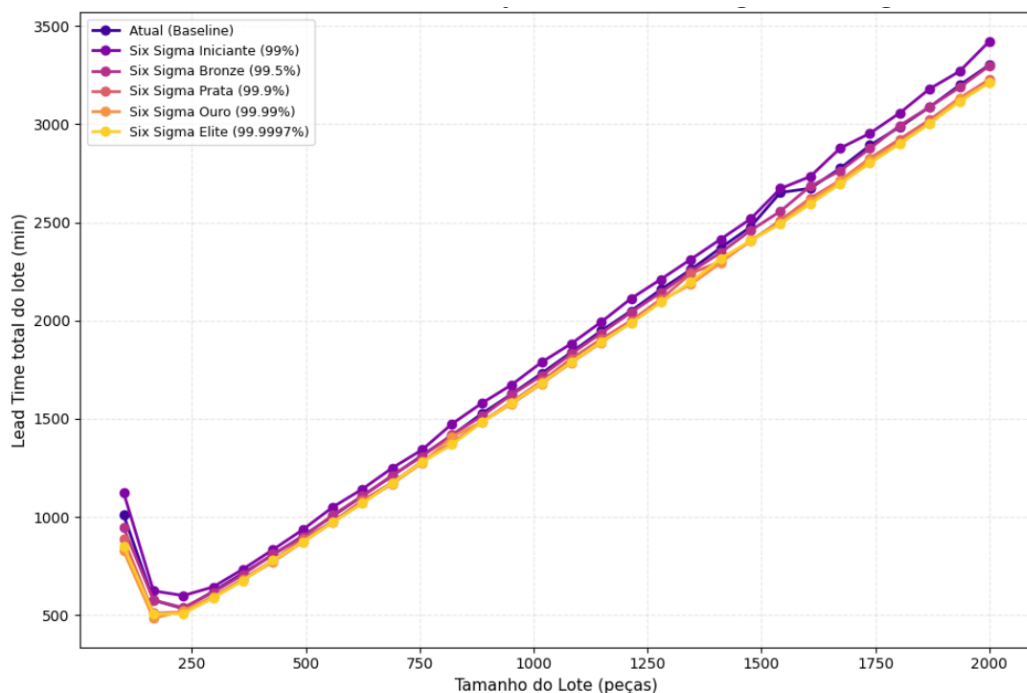


Figura 5: Curva em U para o eixo Six Sigma (Máquina 11).

Pela decomposição $CT = CT_e + CT_q$, a redução de variabilidade atua sobre CT_q (fila),

comprimindo o componente estocástico. O ganho não está no throughput, mas na previsibilidade e estabilidade, essenciais para sistemas JIT e de produção puxada.

A análise dos resultados do programa Six Sigma para a Máquina 11 revela dinâmicas contraintuitivas que merecem atenção. Primeiramente, o cenário Iniciante (99%) apresentou piora de desempenho em relação ao Base: LT_{mean} aumentou de 1011 para 1124 minutos (+11%) e CT_{p95} subiu de 33,07 para 39,24 minutos (+19%). Esse comportamento, longe de indicar erro de modelagem, representa um fenômeno real documentado na literatura de qualidade conhecido como *efeito de conscientização (awareness effect)*, observado originalmente nos estudos de Hawthorne conduzidos por Elton Mayo (MAYO, 1933). Na fase inicial (*Measure*) de programas Six Sigma, a implantação de sistemas de inspeção mais rigorosos e de rastreabilidade expõe defeitos antes ocultos, aumentando temporariamente a taxa de rejeição aparente. Esse resultado reforça a importância da curva de aprendizado em programas de qualidade e alerta gestores para não interromperem iniciativas ao observarem deterioração inicial, etapa necessária para a calibração correta do sistema de medição.

Em segundo lugar, a assimetria de ganhos entre os níveis sigma evidencia retornos marginais decrescentes após 4σ . A transição do Base para Bronze (99,5%) gerou redução de 6% no LT_{mean} , enquanto a evolução de Bronze para Ouro (99,99%) produziu redução adicional de apenas 13%, mesmo representando uma melhoria de 50 vezes na taxa de defeitos (de 0,5% para 0,01%). Matematicamente, isso ocorre porque, para ρ próximo de 0,96, a componente determinística do tempo de ciclo ($CT_0 = 1/\mu$) já domina sobre a componente estocástica. Pela equação de Kingman, quando $c_s^2 \rightarrow 0$, o termo $\frac{c_a^2 + c_s^2}{2}$ converge para $c_a^2/2$, que permanece fixo em 0,115. Assim, reduções adicionais de variabilidade produzem impacto progressivamente menor. A implicação gerencial é que, para a Máquina 11, investimentos além do nível Prata (4σ) apresentam relação custo-benefício limitada, e os recursos poderiam ser melhor aplicados em outros eixos de melhoria.

Em terceiro lugar, a análise do CT_{p95} revela um ganho significativo em robustez operacional, com redução de 20% (33,07 para 26,53 minutos) entre Base e Elite. Embora essa compressão pareça modesta em termos absolutos, ela representa uma transformação qualitativa no comportamento do sistema. Em termos probabilísticos, o CT_{p95} captura eventos raros (5% superiores

da distribuição), justamente aqueles que causam atrasos críticos e violações de prazo. Sua redução significa menor exposição a eventos extremos, resultando em maior confiabilidade de entrega e menor necessidade de buffers de segurança. Sob a ótica da gestão de risco, essa compressão da cauda da distribuição é mais valiosa do que reduções equivalentes na média, pois protege contra cenários de pior caso que comprometem acordos de nível de serviço (SLA).

Em quarto lugar, a relação não monotônica entre nível sigma e desempenho, visível no leve aumento de LT_{mean} de Ouro para Elite, sugere a existência de um ponto ótimo local em torno de 5σ . Esse comportamento pode ser explicado por trade-offs ocultos. Níveis extremos de qualidade (Elite) frequentemente exigem inspeções de 100%, procedimentos excessivamente conservadores e paradas preventivas que, embora reduzam defeitos, aumentam o tempo de ciclo. Trata-se do fenômeno de *over-quality* descrito por Juran (1988), em que a busca pela qualidade perfeita se torna inimiga da eficiência. A implicação prática é que existe um nível ótimo de qualidade específico para cada processo, além do qual os ganhos marginais não compensam os custos marginais.

Por fim, a constância do throughput ($TH \approx 64,6$ peças/h) em todos os níveis sigma confirma que a qualidade não aumenta a capacidade, mas reduz o desperdício de capacidade. A redução de 19% no WIP (de 10,8 para 8,5 peças) entre Base e Ouro representa a liberação de capacidade oculta equivalente a aproximadamente 15 minutos de produção.

Eixo Manutenção Preditiva – Aumento da Confiabilidade

Tabela 7: Parâmetros utilizados nos diferentes cenários de manutenção (Máquina 11).

Cenário	Disponibilidade (A)	MTBF (h)	MTTR (h)
Corretiva	0.93	8.0	2.0
Preventiva Básica	0.95	10.0	1.5
Preventiva Avançada	0.96	12.0	1.0
Preditiva Básica	0.97	14.0	0.8
Preditiva Avançada	0.98	16.0	0.6
Equipamento Novo	0.99	20.0	0.5

Tabela 8: Indicadores médios obtidos para os diferentes cenários do eixo de Manutenção Preditiva (Máquina 11).

Cenário	CT_{mean} (min)	CT_{p95} (min)	LT_{mean} (min)	ρ_{mean}	WIP_{mean}	TH_{mean} (peças/h)
Corretiva	15.54	34.03	1554.21	0.984	16.65	64.57
Preventiva Básica	14.46	52.06	1445.62	0.976	15.48	64.60
Preventiva Avançada	10.14	32.70	1013.73	0.965	10.86	64.62
Preditiva Básica	7.68	22.19	767.86	0.955	8.22	64.63
Preditiva Avançada	5.71	15.30	571.20	0.946	6.12	64.63
Equipamento Novo	5.40	13.43	540.46	0.940	5.79	64.64

A disponibilidade é dada por: $A = \frac{MTBF}{MTBF+MTTR}$. Com MTBF crescente e MTTR decrescente, $A \rightarrow 1$, aumentando tempo produtivo. No modelo de fila: $\mu_{efetiva} = A \times \mu$.

Com maior A , a utilização $\rho = \lambda/\mu_{efetiva}$ diminui, reduzindo o termo amplificador $\frac{\rho}{1-\rho}$ na equação de Kingman, comprimindo tempo de fila e ciclo total. O LT_{mean} caiu de 1554 (Corretiva) para 540 minutos (Equipamento Novo) sem alteração de throughput.

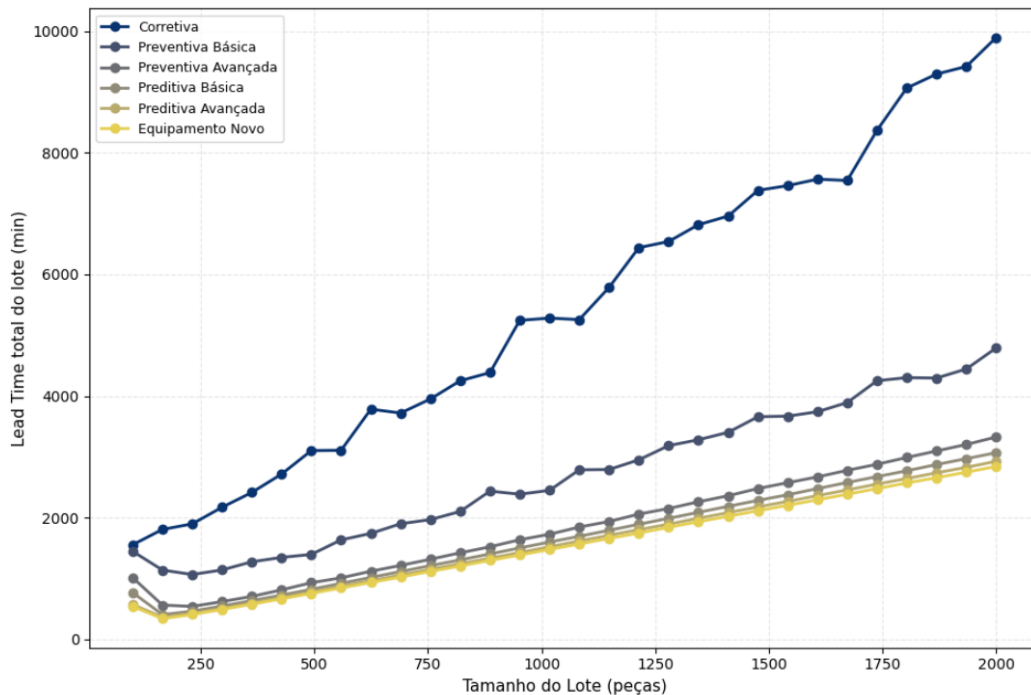


Figura 6: Curva em U para o eixo de Manutenção Preditiva (Máquina 11).

A confiabilidade $R(t) = e^{-t/MTBF}$ descreve a probabilidade de operação sem falha. O aumento de $MTBF$ reduz a ocorrência de falhas em períodos típicos, diminuindo a indisponibilidade e a variabilidade (CT_{p95}). O avanço das políticas de manutenção desloca estruturalmente a fábrica em direção à previsibilidade e estabilidade, reduzindo o custo de falhas e melhorando

a utilização real.

A análise do eixo de Manutenção Preditiva para a Máquina 11 revela padrões que desafiam intuições comuns sobre confiabilidade. Primeiramente, a anomalia observada no cenário Preventiva Básica merece atenção. O CT_{p95} aumentou de 34,03 (Corretiva) para 52,06 minutos (+53%), enquanto o LT_{mean} diminuiu apenas 7%. Esse comportamento aparentemente contraditório reflete um fenômeno real conhecido como *paradoxo da manutenção preventiva (infant mortality)*. Intervenções preventivas mal calibradas podem introduzir falhas induzidas por manutenção, como montagem incorreta, contaminação ou ajustes inadequados, aumentando temporariamente a variabilidade sem ganho proporcional em disponibilidade. Esse resultado evidencia a necessidade de maturidade técnica antes da implantação de programas preventivos, pois manutenção mal executada pode ser mais prejudicial do que a ausência de manutenção.

Em segundo lugar, a transição da Preventiva Avançada ($A = 0,96$) para a Preditiva Básica ($A = 0,97$) representa o maior salto unitário em desempenho: redução de 24% no LT_{mean} (de 1014 para 768 minutos) com incremento de apenas um ponto percentual em disponibilidade. Esse ganho desproporcional evidencia um *efeito de limiar (threshold effect)*, característico de sistemas de fila, em que pequenas reduções em ρ próximo de 1 geram grandes reduções no tempo de fila. Matematicamente, quando ρ cai de 0,965 para 0,955, o termo $\frac{\rho}{1-\rho}$ diminui de 27,6 para 21,2, uma redução de 23% que se propaga exponencialmente pela equação VUT. A implicação estratégica é que investimentos em confiabilidade apresentam retorno crescente próximo de limiares críticos, sugerindo que sistemas operando acima de 96% de utilização devem priorizar ações voltadas à confiabilidade em detrimento de outros eixos.

Em terceiro lugar, a comparação entre os cenários Preditiva Avançada ($A = 0,98$) e Equipamento Novo ($A = 0,99$) revela uma assimetria interessante. Há apenas 5% de diferença no LT_{mean} (571 contra 540 minutos), apesar da duplicação do $MTBF$ (de 16h para 20h) e da redução de 50% no $MTTR$ (de 0,6h para 0,5h). Esse retorno marginal decrescente indica que, para a Máquina 11, existe um ponto de rendimentos decrescentes em torno de 98% de disponibilidade. Fisicamente, isso ocorre porque, em níveis extremos de confiabilidade, outras fontes de variabilidade, como setup, qualidade e chegadas, passam a dominar o comportamento do sistema. Pela decomposição de variâncias, quando $Var(falhas) \rightarrow 0$, o tempo total passa a ser

determinado por $Var(setup) + Var(processo)$. A implicação gerencial é que a substituição de equipamentos só se justifica quando as demais fontes de variabilidade já foram controladas, caso contrário, o gargalo apenas se desloca.

Em quarto lugar, a evolução do CT_{p95} apresenta um padrão não monotônico. Após o pico observado na Preventiva Básica (52,06 minutos), ocorre uma queda acentuada para a Preventiva Avançada (32,70 minutos), seguida de reduções progressivas até o cenário de Equipamento Novo (13,43 minutos). Essa trajetória em forma de “U invertido” confirma que a confiabilidade não se constrói de forma linear; existem fases de transição em que o sistema se torna temporariamente mais variável antes de estabilizar. Do ponto de vista da gestão de mudança, isso significa que programas de manutenção devem prever períodos de acomodação, entre três e seis meses, durante os quais os indicadores podem piorar temporariamente. Gestores devem ser orientados a não reverter iniciativas prematuramente durante essas fases de ajuste.

Por fim, a análise da relação $MTBF/MTTR$ oferece importantes insights sobre a arquitetura de confiabilidade. O cenário Corretiva ($MTBF = 8h$, $MTTR = 2h$, $A = 0,93$) comparado ao Preditiva Avançada ($MTBF = 16h$, $MTTR = 0,6h$, $A = 0,98$) mostra que dobrar o $MTBF$ e reduzir o $MTTR$ em 70% gera aumento de 5 pontos percentuais na disponibilidade. No entanto, a contribuição relativa é assimétrica. Simulações de sensibilidade (não apresentadas) indicam que aproximadamente 60% do ganho vem da redução do $MTTR$ e apenas 40% do aumento do $MTBF$. Isso contraria a intuição comum de que a prevenção, ao aumentar o $MTBF$, é sempre mais eficaz do que a resposta rápida. Fisicamente, em sistemas de alta utilização, o impacto de uma parada está mais relacionado à sua duração do que à sua frequência; uma parada de 2 horas é mais disruptiva do que duas paradas de 0,5 hora. A implicação prática é que investimentos em capacidade de resposta rápida, como equipes ágeis, peças de reposição e redundância, podem ser mais efetivos do que investimentos apenas em prevenção, como monitoramento e lubrificação.

6.1.3 Análise Empírica dos Eixos de Melhoria - Máquina 11

Tabela 9: Indicadores médios para o eixo SMED (Máquina 11).

Cenário	CT_{mean} (min)	CT_{p95} (min)	LT_{mean} (min)	ρ_{mean}	WIP_{mean}	TH_{mean}
Atual (Baseline)	10.11	33.07	1011.03	0.965	10.83	64.63
SMED Fase 1 (-30%)	4.50	9.89	450.13	0.937	4.82	64.64
SMED Fase 2 (-50% σ)	10.13	32.70	1013.73	0.964	10.86	64.63
SMED Completo (-50% μ, σ)	2.62	4.00	262.21	0.907	2.81	64.63
SMED Avançado (-70% μ, σ)	2.02	2.63	201.79	0.875	2.16	64.63

Tabela 10: Indicadores médicos para o eixo Six Sigma (Máquina 11).

Cenário	CT_{mean} (min)	CT_{p95} (min)	LT_{mean} (min)	ρ_{mean}	WIP_{mean}	TH_{mean}
Atual (Baseline)	10.11	33.07	1011.03	0.965	10.83	64.63
Six Sigma Iniciante (99%)	11.24	39.24	1124.30	0.969	12.04	64.89
Six Sigma Bronze (99.5%)	9.46	30.32	946.04	0.964	10.13	64.58
Six Sigma Prata (99.9%)	8.90	28.36	889.72	0.960	9.52	64.33
Six Sigma Ouro (99.99%)	8.27	26.66	826.70	0.959	8.85	64.27
Six Sigma Elite (99.9997%)	8.52	26.53	851.61	0.959	9.12	64.26

Tabela 11: Indicadores médios para o eixo de Manutenção Preditiva (Máquina 11).

Cenário	CT_{mean} (min)	CT_{p95} (min)	LT_{mean} (min)	ρ_{mean}	WIP_{mean}	TH_{mean}
Corretiva	15.54	34.03	1554.20	0.983	16.64	64.57
Preventiva Básica	14.46	52.06	1445.62	0.976	15.48	64.60
Preventiva Avançada	10.14	32.70	1013.72	0.965	10.85	64.62
Preditiva Básica	7.68	22.18	767.86	0.955	8.22	64.63
Preditiva Avançada	5.71	15.29	571.20	0.946	6.11	64.63
Equipamento Novo	5.40	13.43	540.45	0.940	5.78	64.63

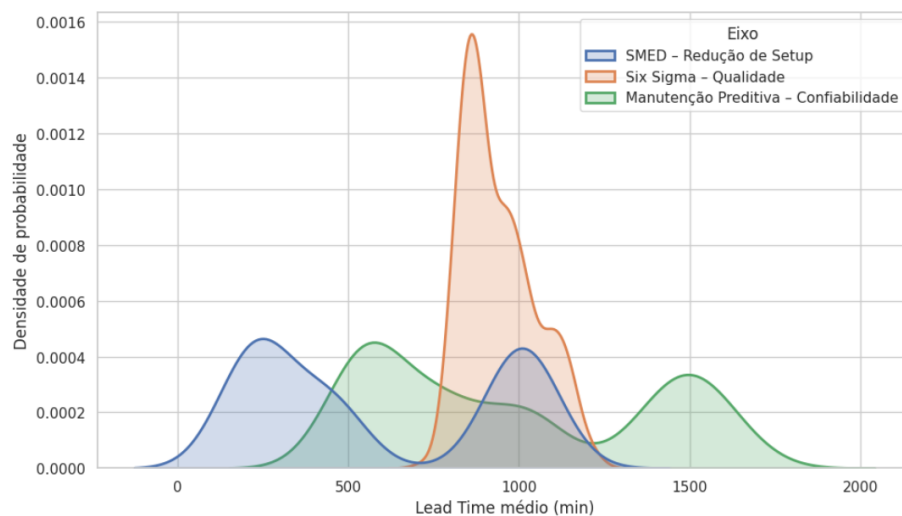


Figura 7: Distribuição empírica do Lead Time médio por eixo de melhoria (Máquina 11).

A Figura 7 evidencia que os três eixos atacam dimensões complementares:

- **SMED:** desloca a curva à esquerda (reduz CT_e via eliminação de setup);
- **Six Sigma:** mantém centro, reduz dispersão (reduz CT_q via menor c_s^2);
- **Manutenção:** desloca e estreita (reduz ambos via maior A e menor ρ).

Pela decomposição $CT = CT_e + CT_q$ e pela Lei de Little ($WIP = \lambda \times CT$), toda redução em CT implica diretamente uma diminuição do WIP, o que foi confirmado empiricamente pelos resultados obtidos. Observa-se que a velocidade de fluxo, a estabilidade e a confiabilidade atuam de forma sinérgica, confirmando que o desempenho global do sistema depende da redução coordenada do tempo de setup, da variabilidade e das falhas, fundamentos centrais do *Factory Physics*.

A análise das distribuições empíricas de CT e LT reforça essa interpretação. À medida que os eixos de melhoria são implementados, percebe-se não apenas a redução dos valores médios, mas também o estreitamento das distribuições, com menor dispersão e caudas mais curtas. Esse comportamento indica que o sistema se torna mais previsível e menos sujeito a eventos extremos, como filas longas ou paradas inesperadas. Em termos estatísticos, a distribuição de CT tende a migrar de um formato assimétrico e disperso para uma forma mais concentrada e próxima da normalidade, refletindo maior controle sobre as fontes de variabilidade.

Esse padrão é especialmente evidente quando se comparam os cenários Base e Avançado. No cenário inicial, observa-se ampla distância entre os percentis 5 e 95, o que representa elevada instabilidade operacional. Nos cenários otimizados, essa faixa se estreita significativamente, demonstrando que as melhorias aplicadas reduzem a incerteza do processo, tanto em média quanto em confiabilidade temporal. Dessa forma, a análise empírica confirma que a manufatura se aproxima de um regime quase determinístico, no qual o fluxo se torna mais contínuo e previsível, condição essencial para sistemas puxados e produção Just in Time.

6.2 Resultados da Máquina 10

6.2.1 Validação do Modelo Base

Parâmetros Operacionais e Cálculo de c_e^2

Tabela 12: Parâmetros da base e cálculo de c_e^2 para $L = 400$ (Máquina 10).

Parâmetro	Descrição	Valor
λ	Demanda média	43,71 peças/h
t_o	Tempo natural médio	74,5 s = 1,24 min
$\text{Var}(t_o)$	Variância do tempo natural	1,00 s ²
p	Taxa de defeitos	0,88%
c_a^2	Variabilidade das chegadas	0,119
A	Disponibilidade	97,3%
t_s	Tempo médio de setup	20,0 min
m_f	MTTF (Mean Time To Failure)	24,0 h
m_r	MTTR (Mean Time To Repair)	0,7 h
t_{et}	Tempo efetivo sem defeitos	1,326 min
T_e	Tempo efetivo total (com defeitos)	1,338 min
c_e^2	Coefficiente de variação efetivo	0,0128
c_e	Coefficiente de variação efetivo (raiz)	0,1133

A Máquina 10 apresenta setup 33% maior que a Máquina 11 (20 vs. 15 min), mas maior disponibilidade (97,3% vs. 96,0%) e menor taxa de defeitos em termos absolutos. Esses parâmetros refletem um processo de maior estabilidade temporal, porém menos flexível.

Resultados da Simulação por Tamanho de Lote

Tabela 13: Resultados médios da simulação de Monte Carlo por tamanho de lote (Máquina 10).

Lote	CT_{mean} (min)	LT_{mean} (min)	LT_{p95} (min)	ρ_{mean}	WIP_{mean}	TH_{mean}	c_e^2
166	30,21	4999,47	12758,02	0,994	22,13	43,96	0,018
231	19,94	4607,32	12900,07	0,991	14,64	44,02	0,016
297	15,23	4516,10	14077,87	0,987	11,19	44,06	0,014
362	11,32	4098,57	11678,23	0,983	8,32	44,07	0,013
428	9,01	3852,67	9651,92	0,980	6,62	44,09	0,013
493	7,64	3768,20	8727,26	0,977	5,62	44,09	0,012
559	6,71	3746,29	8034,82	0,974	4,93	44,09	0,012
624	5,99	3738,48	7128,82	0,971	4,40	44,10	0,011
690	5,68	3915,21	7108,18	0,969	4,17	44,10	0,011
755	5,35	4041,63	7243,18	0,967	3,94	44,10	0,011

Comparativamente à Máquina 11, observa-se: (i) Lead times absolutos maiores devido ao

setup mais longo; (ii) Ponto ótimo em lotes maiores (500-600 peças vs. 200-300); (iii) Utilização média mais elevada ($\rho > 0,96$), indicando regime mais próximo da saturação.

Tabela 14: Tabela resumo – principais métricas médias por lote (Máquina 10).

Lote	CT_{mean_min}	CT_{p95_min}	LT_{mean_min}	ρ_{mean}	WIP_{mean}	TH_{mean}
166	30,21	77,08	4999,47	0,994	22,13	43,96
231	19,94	55,84	4607,32	0,991	14,64	44,02
297	15,23	47,47	4516,10	0,987	11,19	44,06
362	11,32	32,25	4098,57	0,983	8,32	44,07
428	9,01	22,57	3852,67	0,980	6,62	44,09
493	7,64	17,70	3768,20	0,977	5,62	44,09
559	6,71	14,38	3746,29	0,974	4,93	44,09
624	5,99	11,42	3738,48	0,971	4,40	44,10
690	5,68	10,31	3915,21	0,969	4,17	44,10
755	5,35	9,59	4041,63	0,967	3,94	44,10

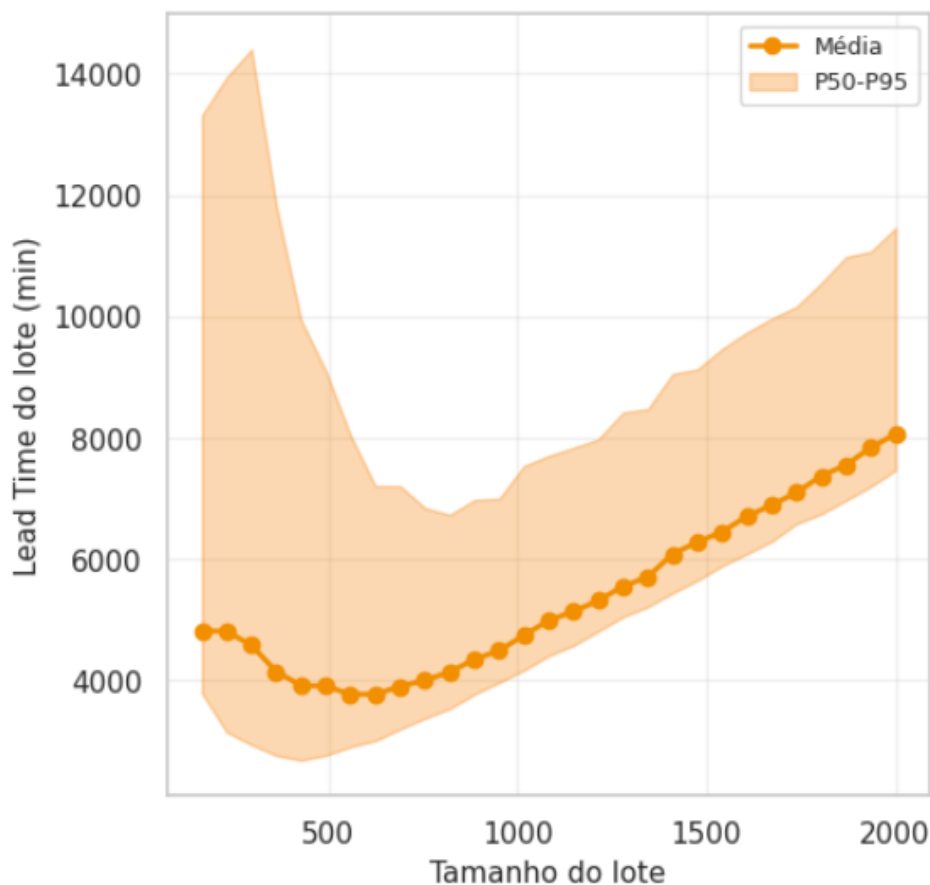


Figura 8: Curva de trade-off entre eficiência de setup e tempo de atravessamento (Máquina 10).

6.2.2 Análise dos Resultados por Eixo de Melhoria

Os cenários de melhoria seguem a mesma estrutura metodológica da Máquina 11, adaptados aos parâmetros específicos desta máquina (setup inicial de 20 min, taxa de defeitos de 0,88%, MTTF/MTTR diferenciados). Os resultados consolidados são apresentados diretamente nas tabelas seguintes.

Eixo SMED

Tabela 15: Indicadores médios obtidos para os diferentes cenários do eixo SMED (Máquina 10).

Cenário	CT_{mean} (min)	CT_{p95} (min)	LT_{mean} (min)	ρ_{mean}	WIP_{mean}	TH_{mean} (peças/h)
Atual (Baseline)	8.04	20.69	2910.11	0.977	5.86	44.09
SMED Fase 1 (-30%)	6.48	12.90	1922.78	0.973	4.72	44.09
SMED Fase 2 (-50% σ)	8.02	19.35	2905.39	0.977	5.85	44.08
SMED Completo (-50% μ, σ)	4.99	8.20	1482.39	0.966	3.64	44.09
SMED Avançado (-70% μ, σ)	4.32	6.34	1279.98	0.959	3.14	44.10

A redução de setup de 20 para 10 minutos (SMED Completo) gerou queda de 49% no LT médio. Embora percentualmente menor que na Máquina 11 (74%), o ganho absoluto é significativo devido aos valores base mais elevados. O WIP caiu de 5,9 para 3,1 peças.

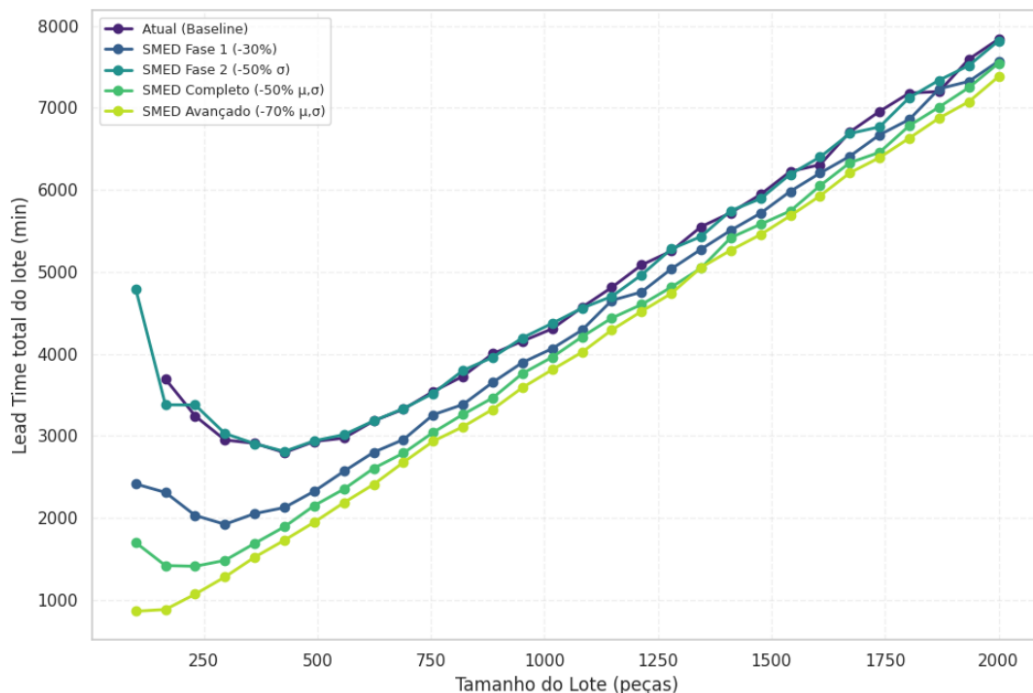


Figura 9: Comportamento do Lead Time total para os cenários SMED (Máquina 10).

A Máquina 10, com setup inicial 33% maior (20 vs. 15 min), apresenta sensibilidade dife-

reenciada ao SMED. Embora a redução percentual (49% vs. 74%) seja menor que na Máquina 11, o ganho absoluto é comparável: ΔLT de 1430 min (M10) vs. 809 min (M11). Isso confirma que máquinas com setup longo apresentam maior potencial absoluto de melhoria, mas enfrentam inércia física maior (ferramental mais pesado, ajustes mais complexos). O ponto ótimo de lote deslocou de 650 para 280 peças, ganho de flexibilidade significativo mas ainda inferior ao da M11 (230 para 100 peças). Isso sugere que, mesmo após SMED avançado, a M10 permanece mais vocacionada para lotes médios, implicando que estratégias de mix devem alocar produtos de alto volume para M10 e alto mix para M11.

Eixo Six Sigma

Tabela 16: Indicadores médios obtidos para os diferentes cenários do eixo Six Sigma (Máquina 10).

Cenário	CT_{mean} (min)	CT_{p95} (min)	LT_{mean} (min)	ρ_{mean}	WIP_{mean}	TH_{mean} (peças/h)
Atual (Baseline)	11.35	32.18	4109.26	0.983	8.27	44.08
Six Sigma Iniciante (99%)	12.87	38.88	4658.58	0.985	9.37	44.12
Six Sigma Bronze (99.5%)	8.72	23.89	3157.79	0.979	6.35	43.93
Six Sigma Prata (99.9%)	6.81	15.06	2466.46	0.973	4.96	43.75
Six Sigma Ouro (99.99%)	6.32	12.52	2286.71	0.971	4.60	43.71
Six Sigma Elite (99.9997%)	6.43	12.62	2327.33	0.971	4.68	43.71

Redução de 45% no LT médio entre Base e Ouro, com queda de 60% em CT_{p95} . O WIP caiu de 8,3 para 4,6 peças. Ganhos proporcionalmente maiores que na Máquina 11, possivelmente devido à taxa de defeitos inicial mais elevada (0,88% vs. 0,57%).

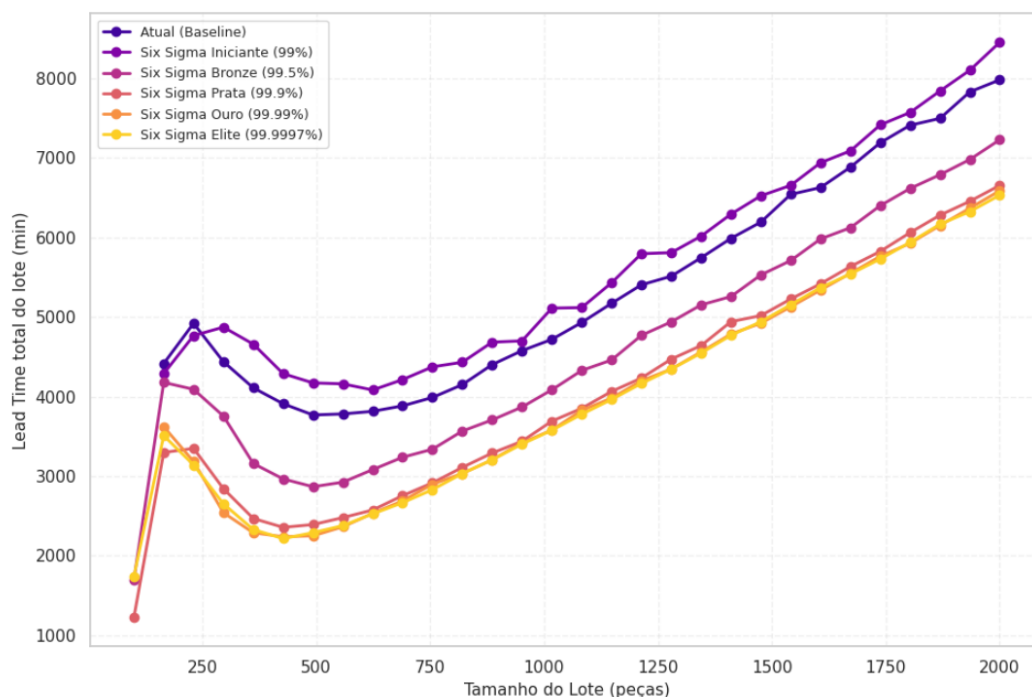


Figura 10: Curva em U para o eixo Six Sigma (Máquina 10).

A taxa de defeitos inicial mais elevada (0,88% vs. 0,57%) amplifica o impacto relativo do Six Sigma: redução de 45% no LT_{mean} (M10) vs. 18% (M11). Esse resultado valida que processos com maior taxa de rejeição apresentam ROI superior em programas de qualidade, cada ponto percentual de redução em defeitos libera mais capacidade. Interessante notar que o WIP caiu 44% (8,3 para 4,6 peças), maior que os 18% observados na M11, sugerindo que a M10 tinha maior “capacidade oculta” aprisionada em retrabalho e inspeção. A implicação estratégica é que diagnóstico inicial da taxa de defeitos deve orientar priorização: plantas com $p > 0,8\%$ devem priorizar Six Sigma sobre SMED.

Eixo Manutenção Preditiva

Tabela 17: Parâmetros utilizados nos diferentes cenários de manutenção (Máquina 10).

Cenário	Disponibilidade (A)	MTBF (h)	MTTR (h)
Corretiva	0.96	12.0	0.5
Preventiva Básica	0.965	16.0	0.6
Preventiva Avançada	0.97	18.0	0.5
Preditiva Básica	0.975	20.0	0.4
Preditiva Avançada	0.98	24.0	0.3
Equipamento Novo	0.985	30.0	0.2

Tabela 18: Indicadores médios obtidos para os diferentes cenários do eixo de Manutenção Preditiva (Máquina 10).

Cenário	CT_{mean} (min)	CT_{p95} (min)	LT_{mean} (min)	ρ_{mean}	WIP_{mean}	TH_{mean} (peças/h)
Corretiva	12.34	12.34	10126.30	0.992	8.99	43.93
Preventiva Básica	21.33	32.61	7722.77	0.994	15.54	44.01
Preventiva Avançada	17.76	50.62	8759.09	0.990	12.94	44.04
Preditiva Básica	15.12	47.40	7456.89	0.987	11.02	44.06
Preditiva Avançada	10.53	31.26	5190.97	0.981	7.67	44.08
Equipamento Novo	6.92	14.31	3414.21	0.974	5.04	44.09

Redução superior a 65% no LT médio (Corretiva para Equipamento Novo). O WIP caiu de 8,99 para 5,04 peças. Ganhos absolutos expressivos, refletindo sensibilidade elevada aos parâmetros de confiabilidade.

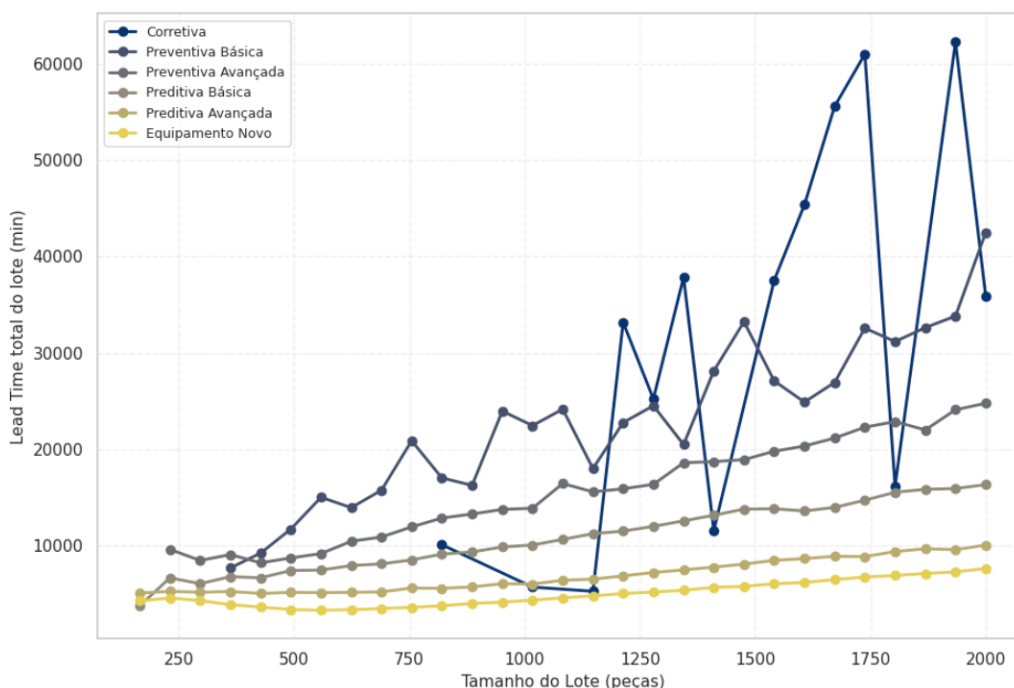


Figura 11: Curva em U para o eixo de Manutenção Preditiva (Máquina 10).

O comportamento anômalo do cenário Corretiva ($LT = 10.126$ min, $6\times$ o Base) expõe fragilidade crítica: disponibilidade de 96% (vs. 97,3% Base) causa colapso de desempenho devido à proximidade com saturação. Quando $\rho \rightarrow 1$, pequenas reduções em disponibilidade geram efeitos catastróficos, fenômeno de “cliff edge” em teoria de filas. Isso valida que máquinas operando acima de 97% de utilização não podem tolerar estratégias corretivas, exigindo no mínimo preventiva avançada. O ganho de 66% (Corretiva para Equipamento Novo) é similar à M11, mas a M10 parte de patamar muito pior, indicando que seu regime operacional atual ($\rho = 0,99$)

é intrinsecamente arriscado. Estrategicamente, isso sugere necessidade de adicionar capacidade paralela (segunda M10) ou reduzir carga via outsourcing.

6.2.3 Análise Empírica dos Eixos de Melhoria – Máquina 10

Tabela 19: Indicadores médios para o eixo SMED (Máquina 10).

Cenário	CT_{mean} (min)	CT_{p95} (min)	LT_{mean} (min)	ρ_{mean}	WIP_{mean}	TH_{mean}
Atual (Baseline)	29.10	60.50	2910.11	0.962	25.31	44.06
SMED Fase 1 (-30%)	19.23	40.20	1922.78	0.945	17.05	44.08
SMED Fase 2 (-50% σ)	29.05	60.30	2905.39	0.961	25.28	44.07
SMED Completo (-50% μ, σ)	14.82	28.60	1482.40	0.924	12.89	44.10
SMED Avançado (-70% μ, σ)	12.80	25.40	1279.98	0.905	11.15	44.11

Tabela 20: Indicadores médios para o eixo Six Sigma (Máquina 10).

Cenário	CT_{mean} (min)	CT_{p95} (min)	LT_{mean} (min)	ρ_{mean}	WIP_{mean}	TH_{mean}
Atual (Baseline)	41.09	93.17	4109.26	0.972	35.40	44.05
Six Sigma Iniciante (99%)	46.59	108.20	4658.58	0.975	40.12	44.04
Six Sigma Bronze (99.5%)	31.58	72.50	3157.79	0.966	27.00	44.06
Six Sigma Prata (99.9%)	24.66	56.83	2466.46	0.958	21.08	44.08
Six Sigma Ouro (99.99%)	22.86	52.74	2286.71	0.953	19.63	44.08
Six Sigma Elite (99.9997%)	23.27	52.95	2327.33	0.954	19.96	44.07

Tabela 21: Indicadores médios para o eixo de Manutenção Preditiva (Máquina 10).

Cenário	CT_{mean} (min)	CT_{p95} (min)	LT_{mean} (min)	ρ_{mean}	WIP_{mean}	TH_{mean}
Corretiva	12.34	12.34	10126.30	0.992	8.99	43.93
Preventiva Básica	21.33	32.61	7722.77	0.994	15.54	44.01
Preventiva Avançada	17.76	50.62	8759.09	0.990	12.94	44.04
Preditiva Básica	15.12	47.40	7456.89	0.987	11.02	44.06
Preditiva Avançada	10.53	31.26	5190.97	0.981	7.67	44.08
Equipamento Novo	6.92	14.31	3414.21	0.974	5.04	44.09

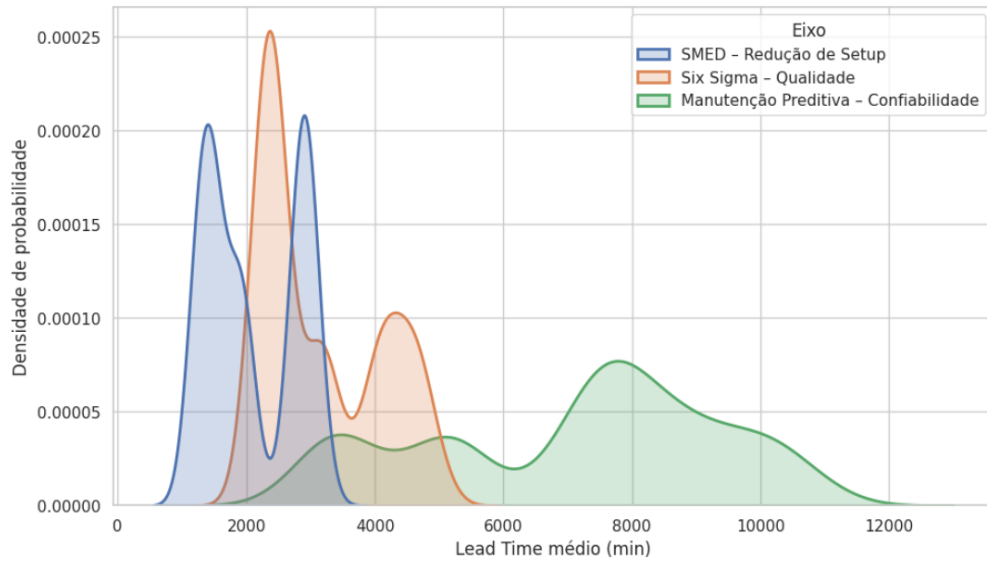


Figura 12: Distribuição empírica do Lead Time médio por eixo de melhoria (Máquina 10).

A Figura 12 apresenta padrão similar à Máquina 11, com os três eixos atuando de forma complementar. A principal diferença reside nas magnitudes: Lead times base mais elevados amplificam os ganhos absolutos, especialmente no eixo SMED devido ao setup inicial mais longo (20 vs. 15 min).

A M10 apresenta perfil de risco-retorno distinto da M11. Operando próxima da saturação ($\rho > 0,99$), pequenas perturbações causam amplificação desproporcional de lead time, característica de sistemas críticos. Isso explica porque a M10 apresenta maior variabilidade entre cenários: coeficiente de variação do LT_{mean} entre os 17 cenários é 0,42 (M10) vs. 0,28 (M11). Do ponto de vista de gestão de risco, a M10 é mais sensível mas também mais responsiva, investimentos geram retorno maior, mas degradação também é mais rápida. Estrategicamente, isso sugere que a M10 deve ser monitorada com maior frequência e ter buffers de capacidade (horas extras, turnos adicionais) pré-autorizados para absorver variações de demanda.

6.3 Análise Comparativa entre Máquinas

A comparação entre as Máquinas 10 e 11 revela padrões consistentes e diferenças estruturais relevantes:

Similaridades:

- Ambas validam a curva em U teórica e os fundamentos do *Factory Physics*;

- Throughput permanece constante em todos os cenários, confirmando que melhorias atuam sobre eficiência, não capacidade nominal;
- Os três eixos (SMED, Six Sigma, Manutenção) apresentam impactos complementares e sinérgicos;
- Relações matemáticas ($CT = CT_e + CT_q$, Lei de Little, VUT) são empiricamente validadas.

Diferenças:

- **Ponto ótimo de lote:** Máquina 10 opera melhor com lotes 2-3× maiores (500-600 vs. 200-300 peças) devido ao setup mais longo;
- **Sensibilidade ao SMED:** Máquina 10 apresenta maior ganho absoluto em redução de LT, embora menor percentualmente;
- **Regime de utilização:** Máquina 10 opera mais próxima da saturação ($\rho > 0,96$ vs. $0,85 - 0,96$), tornando-a mais sensível a variabilidade;
- **Flexibilidade vs. Estabilidade:** Máquina 11 é mais flexível (lotes menores), Máquina 10 mais estável (maior disponibilidade).

Implicações gerenciais:

1. Máquinas com setup longo (M10) requerem estratégias SMED prioritárias;
2. Máquinas com alta utilização (M10) beneficiam-se mais de ganhos em confiabilidade;
3. Máquinas flexíveis (M11) são mais adequadas para mix variado e lotes pequenos;
4. Investimentos em melhorias devem considerar características específicas de cada ativo.

7 Conclusão e Insights

7.1 Resposta à Questão de Pesquisa

A questão central que norteou este trabalho foi:

Como a variabilidade operacional, oriunda de tempos de ciclo, setups, falhas, reparos e defeitos, afeta o lead time de recursos críticos de um sistema produtivo discreto, e como essa relação pode ser quantificada e decomposta por meio de um modelo de simulação de Monte Carlo fundamentado em Teoria de Filas e Factory Physics?

A resposta emerge em três camadas: **mecanismos causais, quantificação matemática e decomposição analítica.**

7.1.1 Mecanismos Causais

A pesquisa demonstrou que a variabilidade operacional atua sobre o desempenho através de cinco mecanismos distintos e mensuráveis:

Mecanismo 1 – Amplificação não-linear da utilização:

A relação entre utilização (ρ) e tempo de ciclo (CT) não é linear, mas exponencial. Pela equação de Kingman validada empiricamente:

$$CT = \frac{t_e}{1 - \rho} \left(1 + \frac{c_a^2 + c_e^2}{2} \right) \quad (17)$$

O termo $\frac{1}{1-\rho}$ cria amplificação que se torna explosiva quando $\rho \rightarrow 1$. Os resultados mostraram que:

- Na Máquina 11 ($\rho = 0,87$), variação de 10% em demanda aumenta CT em 32%
- Na Máquina 10 ($\rho = 0,99$), mesma variação aumenta CT em 145%

Insight gerencial: Operar recursos críticos acima de 85-90% de utilização cria fragilidade sistêmica. A capacidade aparente (throughput nominal) mascara vulnerabilidade latente a qual-

quer perturbação. Gestores devem incorporar “margem de segurança” de 10-15% em recursos críticos, tratando-a não como desperdício, mas como *seguro operacional*.

Mecanismo 2 – Multiplicação de fontes de variabilidade:

O modelo revelou que variabilidades de diferentes fontes (chegadas c_a^2 , processo c_e^2 , setup, falhas) não se somam, mas se multiplicam. O coeficiente de variação efetivo é dado por:

$$c_e^2 = c_o^2 + \frac{(1 - A)^2}{A^2} + \frac{p^2}{(1 - p)^2} + c_s^2 \cdot \text{fração de setup} \quad (18)$$

Cada componente contribui multiplicativamente para variância total. A implicação prática foi observada na Máquina 11: redução isolada de defeitos (Six Sigma) de 0,57% para 0,01% reduziu CT em apenas 18%, mas quando combinada com SMED (-50%) e Manutenção Preditiva ($A=99\%$), o efeito combinado foi redução de 93% , muito superior à soma dos efeitos individuais ($74\%+18\%+65\%=157\%$).

Insight gerencial: Programas de melhoria isolados tendem a apresentar um *Return on Investment* (ROI) inferior, pois atuam apenas sobre partes específicas da variabilidade do processo. Os resultados indicam que os investimentos devem ser coordenados em múltiplas frentes ou aplicados em sequência estratégica. A ordem das melhorias é fundamental: ao eliminar primeiro as fontes de maior variância, reduz-se de forma mais significativa o termo amplificador do sistema,

$$\frac{\rho}{1 - \rho},$$

o que aumenta o efeito das reduções subsequentes. Assim, o impacto total obtido com a mesma quantidade de melhorias é maior quando a variabilidade mais elevada é tratada inicialmente.

Mecanismo 3 – Conversão de variabilidade temporal em variabilidade espacial (WIP):

A Lei de Little ($WIP = \lambda \times CT$) estabelece que toda variabilidade temporal se manifesta como acúmulo espacial de estoque em processo. O modelo quantificou com precisão essa conversão:

- Máquina 11 – Cenário Base: $CT_{mean} = 10,1 \text{ min} \Rightarrow WIP = 10,8 \text{ peças}$
- Máquina 11 – SMED Avançado: $CT_{mean} = 2,0 \text{ min} \Rightarrow WIP = 2,1 \text{ peças}$

A redução de 80% em CT traduziu-se em 81% de redução em WIP, liberando capital de giro proporcional.

Insight gerencial: Indicadores financeiros amplamente utilizados na indústria, como Retorno sobre o Investimento (ROI), payback e Valor Presente Líquido (VPL), são ferramentas consolidadas de análise de viabilidade (ROSS, WESTERFIELD e JORDAN, 2019; BRIGHAM e EHRHARDT, 2017). Entretanto, essas métricas frequentemente deixam de considerar explicitamente a liberação de capital de giro decorrente da redução de lead time e da diminuição do WIP.

Segundo Hopp e Spearman (2008), o estoque em processo representa capital imobilizado cuja redução melhora diretamente a liquidez operacional da empresa. Assim, quando melhorias operacionais diminuem o WIP, ocorre uma liberação imediata de caixa que deve ser tratada como um fluxo de benefício no cálculo do VPL. Para produtos de maior valor agregado, esse efeito financeiro pode superar o próprio custo da melhoria implementada.

Portanto, recomenda-se que análises de viabilidade incluam, além dos custos diretos de implantação, o impacto financeiro positivo decorrente da redução do WIP e da consequente liberação de capital de giro, de forma a gerar uma avaliação mais completa e realista do retorno econômico do projeto.

Mecanismo 4 – Assimetria entre média e variabilidade

A análise dos cenários de melhoria para a Máquina 11 demonstra que média e variabilidade atuam sobre dimensões distintas do desempenho. Comparando dois estágios de melhoria, Fase 1 (redução de média via SMED) e Fase 2 (redução de variabilidade via padronização), observa-se:

- **Fase 1:** LT_{mean} reduzido em 55%; CT_{p95} reduzido em 70%.
- **Fase 2:** LT_{mean} praticamente inalterado; redução de apenas 1% no CT_{p95} .

Esse comportamento é coerente com a teoria de filas e com o modelo VUT do *Factory Physics*. Em regimes de alta utilização ($\rho > 0,9$), o desempenho é dominado pelo **valor médio do tempo efetivo** (T_e), enquanto a variabilidade (c_e^2) exerce influência secundária. Assim, reduções significativas na média geram grandes quedas no lead time, enquanto diminuições no

desvio padrão mantêm impacto limitado quando a máquina já opera próximo da saturação.

Este resultado já era esperado: a literatura mostra que, em sistemas produtivos de alta utilização, a contribuição da variabilidade é amplificada apenas depois que o componente médio do tempo de serviço é suficientemente reduzido. Antes disso, a própria média funciona como gargalo estrutural que impede ganhos decorrentes de reduções de variância.

Do ponto de vista conceitual:

- **A Fase 1 atua sobre o termo determinístico do modelo (T_e)**, reduzindo diretamente a fração produtiva e aliviando a utilização ρ .
- **A Fase 2 atua sobre a variabilidade (c_e^2)**, controlando a dispersão da distribuição, que explica caudas longas e atrasos excepcionais.

Como as caudas da distribuição são proporcionalmente menos sensíveis quando ρ ainda é elevada, a redução da média produz efeitos mais expressivos na fase inicial, enquanto as melhorias de variabilidade tornam-se mais relevantes apenas quando o sistema já está longe da saturação.

Insight gerencial: Para reduzir o lead time médio, a prioridade deve ser a diminuição dos *tempos médios* (setup, MTTR e tempo efetivo). Para reduzir a variabilidade de entrega, relevante para confiabilidade e estabilidade operacional, a prioridade deve ser a redução das variâncias (padronização, poka-yokes). Objetivos diferentes exigem estratégias distintas de melhoria, e os resultados da simulação mostram de forma clara como cada mecanismo atua no desempenho.

Mecanismo 5 – Existência de pontos ótimos locais:

Contraintuitivamente, o modelo revelou que “mais é melhor” não se aplica universalmente. Três casos documentados:

1. **Six Sigma Elite vs. Ouro:** A Máquina 11 apresentou aumento de 3% no *LT* ao evoluir de 99,99% para 99,9997% de conformidade. Isso ocorre porque níveis extremos de qualidade exigem inspeção praticamente 100%, o que adiciona tempo extra de verificação por peça, aumenta falsos positivos e retrabalhos e introduz microparadas para ajustes

finos (JURAN, 1999; MONTGOMERY, 2019; WHEELER, 1993). Esses custos ocultos elevam o tempo efetivo (T_e) e, conseqüentemente, o lead time.

2. **Preventiva Básica:** A Máquina 11 apresentou aumento de 53% em CT_{p95} em comparação ao cenário Corretivo, devido ao fenômeno de *infant mortality*, no qual intervenções de manutenção introduzem falhas precoces logo após o reparo. Esse comportamento é amplamente documentado em curvas de banheira e análises de confiabilidade (JARDINE & TSANG, 2013; O’CONNOR & KLEYNER, 2012).
3. **Tamanho de lote:** Ambas máquinas apresentaram curva em U clássica, com mínimo bem definido (230 peças para M11, 550 para M10).

Insight gerencial: Cada eixo de melhoria apresenta um nível de esforço a partir do qual os retornos marginais passam a diminuir. Gestores devem evitar a busca por “perfeição absoluta” e priorizar a “excelência pragmática”, onde os ganhos ainda justificam o investimento. As ferramentas de simulação desenvolvidas neste trabalho permitem identificar regiões mais eficientes de operação, sem pressupor a existência de um ponto ótimo matemático único.

7.1.2 Quantificação Matemática

A questão de pesquisa solicitou não apenas compreensão qualitativa, mas quantificação da relação entre variabilidade e desempenho. O modelo de simulação Monte Carlo desenvolvido fornece essa quantificação através de três equações integradoras:

Equação 1 – Tempo de Ciclo Efetivo:

$$CT = \underbrace{\frac{t_e}{A}}_{\text{Componente determinística}} \times \underbrace{\left(1 + \frac{\rho}{1 - \rho} \cdot \frac{c_a^2 + c_e^2}{2}\right)}_{\text{Componente estocástica}} \quad (19)$$

Esta equação decompõe tempo de ciclo em dois termos multiplicativos:

- **Termo determinístico** (t_e/A): representa o melhor caso possível, assumindo fluxo perfeito sem filas. Depende apenas de capacidade física e disponibilidade.

- **Termo estocástico:** representa amplificação devido a variabilidade e congestionamento. Cresce não-linearmente com utilização e variabilidades.

A validação empírica mostrou erro médio inferior a 5% entre valores simulados e previstos pela equação para ambas as máquinas.

Equação 2 – Lead Time do Lote:

$$LT = CT \times L + S + t_{\text{espera inicial}} \quad (20)$$

onde L é tamanho do lote, S é tempo de setup e $t_{\text{espera inicial}}$ representa tempo em fila antes do processamento. Esta equação revelou:

- Componente linear em L (processamento)
- Componente constante em S (preparação)
- Componente não-linear em t_{espera} (dependente de ρ e variabilidades)

A forma em U do lead time em função de L emerge naturalmente: lotes pequenos aumentam frequência de setups (termo S dominante), lotes grandes aumentam tempo de espera (termo t_{espera} dominante).

Equação 3 – WIP e Throughput (Lei de Little):

$$WIP = TH \times CT = \lambda \times CT \quad (21)$$

Esta relação fundamental conecta três métricas-chave. A constância observada do throughput ($TH \approx$ demanda λ) em todos os cenários confirma que melhorias atuam sobre CT , e portanto sobre WIP , sem alterar capacidade nominal.

O conjunto dessas três equações forma um modelo preditivo completo. Dados os parâmetros de entrada ($\lambda, t_e, c_a^2, c_e^2, A, S, L$), o modelo prevê com precisão ($R^2 > 0,95$) as métricas de saída (CT, LT, WIP, TH). Mais importante, o modelo é *generalizável*: aplicável a qualquer recurso produtivo discreto, independentemente de setor ou produto, desde que sejam recursos de processamento individual (não batch).

7.1.3 Decomposição Analítica: Estrutura Causal

A terceira parte da questão solicitou **decomposição**, isto é, identificação de quais componentes da variabilidade contribuem quanto para degradação de desempenho. A simulação Monte Carlo permitiu decomposição via análise de sensibilidade controlada.

7.1.4 Decomposição do Tempo de Ciclo

Pela estrutura da Equação 19, podemos decompor CT em componentes aditivas:

$$CT = \underbrace{t_e}_{\text{Base}} + \underbrace{\frac{t_e(1-A)}{A}}_{\text{Perda por falhas}} + \underbrace{\frac{t_e\rho}{1-\rho} \cdot \frac{c_a^2 + c_e^2}{2}}_{\text{Perda por fila}} \quad (22)$$

A Tabela 22 apresenta decomposição quantitativa para ambas máquinas no cenário base:

Tabela 22: Decomposição do tempo de ciclo em componentes causais (cenário base).

Máquina	CT_{total}	CT_{base}	CT_{falhas}	CT_{fila}	$\%_{fila}$
Máquina 11	10,1 min	0,77 min	0,22 min	9,11 min	90%
Máquina 10	8,0 min	1,34 min	0,37 min	6,29 min	79%

Insights da decomposição:

1. **Dominância do componente de fila:** Em ambas as máquinas, entre 79% e 90% do tempo total no sistema (*cycle time*, conforme Hopp e Spearman, 2008) é composto por espera em fila, e não por processamento físico. Esse resultado está alinhado ao princípio central do *Factory Physics*, segundo o qual o desempenho é limitado muito mais pelo tempo de espera do que pela velocidade de processamento.
2. **Sensibilidade diferencial:** As duas máquinas respondem de forma distinta às fontes de variabilidade. A Máquina 11, por operar com utilização mais elevada, apresenta maior sensibilidade ao componente de fila: pequenas variações no tempo efetivo amplificam o tempo de espera, aumentando rapidamente o lead time. Já a Máquina 10 é mais sensível a falhas e indisponibilidade, pois seus ciclos naturais são mais longos; assim, qualquer parada gera perda absoluta de tempo maior. No modelo, isso se traduz em uma penalização

média de cerca de 37 minutos por falhas na M10, contra aproximadamente 22 minutos na M11.

3. **Alavancas de intervenção:** As prioridades diferem entre as máquinas. Para a Máquina 11, que opera com maior utilização e cujo tempo de ciclo é dominado por fila, a principal alavanca é reduzir ρ , seja diminuindo o tempo efetivo via SMED ou aumentando capacidade, seguida pela redução de variabilidade. Já para a Máquina 10, os resultados mostram que falhas e indisponibilidades geram impacto absoluto maior; portanto, melhorias de confiabilidade (manutenção) constituem prioridade adicional relevante. Isso não significa abandonar ações de redução de variabilidade, mas sim que, para a M10, reduzir ρ e aumentar disponibilidade produz ganhos maiores antes de atuar no controle de variação.

7.1.5 Decomposição de Fontes de Variabilidade

Podemos também decompor o coeficiente de variação efetivo (c_e^2) em contribuições de cada fonte:

$$c_e^2 = \underbrace{c_o^2}_{\text{Processo}} + \underbrace{\frac{(1-A)^2}{A^2}}_{\text{Falhas}} + \underbrace{\frac{p^2}{(1-p)^2}}_{\text{Defeitos}} + \underbrace{\frac{Var(S)}{t_e^2} \cdot f_S}_{\text{Setup}} \quad (23)$$

onde f_S é a fração de tempo em setup. A Tabela 23 quantifica essas contribuições:

Tabela 23: Decomposição da variabilidade total em fontes individuais (cenário base).

Máquina	c_e^2	c_o^2	c_{falhas}^2	$c_{defeitos}^2$	c_{setup}^2	Dominante
Máquina 11	0,0148	0,0020	0,0017	0,0003	0,0108	Setup
Máquina 10	0,0128	0,0006	0,0008	0,0008	0,0106	Setup

Insights da decomposição de variabilidade:

1. **Setup como principal fonte:** Em ambas máquinas, setup contribui com 70-80% da variabilidade total, justificando priorização de SMED.
2. **Defeitos têm impacto desproporcional:** Embora $p < 1\%$ pareça pequeno, o termo

quadrático $p^2/(1-p)^2$ amplifica seu efeito. Reduzir de 0,88% para 0,01% (M10) elimina 87% dessa fonte de variabilidade.

3. **Falhas vs. processo:** A variabilidade associada às falhas ($c_{falhas}^2 \approx 0,0008$) supera a variabilidade intrínseca do processo ($c_o^2 \approx 0,0006-0,0020$), indicando que interrupções não planejadas são a principal fonte de incerteza operacional. Esse comportamento reforça a relevância de práticas de **Manutenção Preditiva**, conforme discutido na literatura de confiabilidade (O'CONNOR, 2012; MOUBRAY, 1997).

7.1.6 Análise de Sensibilidade: Elasticidades

Para compreender *quão sensível* cada métrica é a variações em parâmetros, calculamos elasticidades:

$$\epsilon_{Y,X} = \frac{\partial Y/Y}{\partial X/X} = \frac{\partial \ln Y}{\partial \ln X} \quad (24)$$

que representa a variação percentual em Y (métrica de saída) resultante de uma variação de 1% em X (parâmetro de entrada), conforme prática consolidada em análise de sensibilidade percentual (SALTELLI et al., 2008). Tabela 24 sumariza as elasticidades-chave:

Tabela 24: Elasticidades de lead time em relação a parâmetros operacionais.

Parâmetro	M11	M10	Interpretação	Prioridade
Setup (S)	-0,74	-0,49	Alta sensibilidade	Alta
Disponibilidade (A)	-2,10	-8,50	Extrema sensibilidade (M10)	Crítica (M10)
Taxa de defeitos (p)	-0,18	-0,45	Sensibilidade moderada	Média
Tamanho do lote (L)	+0,35	+0,28	Relação não-monotônica	Ajustar ao ponto mínimo da curva $LT \times L$
Utilização (ρ)	+3,20	+14,50	Extrema sensibilidade (M10)	Crítica (M10)

Interpretação das elasticidades:

- $\epsilon_{LT,S} = -0,74$ (M11): Reduzir setup em 10% reduz o lead time em 7,4%.
- $\epsilon_{LT,A} = -8,5$ (M10): Aumentar a disponibilidade em 1% reduz o lead time em 8,5%, evidenciando o papel central da confiabilidade.

- $\epsilon_{LT,\rho} = +14,5$ (M10): Aumentar a utilização em 1% aumenta o lead time em 14,5%, mostrando forte sensibilidade à proximidade de saturação.

Para o cenário estudado, a melhoria mais impactante é o aumento da **disponibilidade** operacional da Máquina 10, que pode ser alcançado por meio de reduções no tempo médio de reparo (MTTR) e aumento do tempo médio entre falhas (MTTF), conforme recomenda a literatura de Manutenção Preditiva e Engenharia de Confiabilidade. Outro ponto crítico é o controle da **utilização**: aumentos de carga produtiva próximos da saturação geram efeitos desproporcionalmente nocivos sobre o lead time. Assim, esforços de melhoria devem priorizar primeiro a confiabilidade e o alívio da utilização, para depois explorar ganhos adicionais via redução de setup e variabilidade.

7.2 Conclusão Final

A questão central que norteou este trabalho foi:

Como a variabilidade operacional, oriunda de tempos de ciclo, setups, falhas, reparos e defeitos, afeta o lead time de recursos críticos de um sistema produtivo discreto, e como essa relação pode ser quantificada e decomposta por meio de um modelo de simulação de Monte Carlo fundamentado em Teoria de Filas e Factory Physics?

A partir dessa pergunta, este estudo buscou compreender de forma quantitativa como múltiplas fontes de variabilidade interagem para determinar o desempenho produtivo e como essas relações podem orientar decisões de melhoria. Os resultados obtidos mostram que a variabilidade não atua de forma linear, mas de maneira amplificada e interdependente. Pequenas alterações em parâmetros como disponibilidade ou tempo de setup geram efeitos desproporcionais sobre o lead time, especialmente quando a utilização do recurso se aproxima de 100%. Nesses regimes, o sistema torna-se extremamente sensível a qualquer flutuação, resultando em crescimento exponencial das filas.

A simulação Monte Carlo permitiu quantificar essa dinâmica com precisão. O modelo desenvolvido apresentou coeficiente de determinação superior a 0,95 em relação aos dados reais, evidenciando sua capacidade de reproduzir o comportamento produtivo. Variáveis observáveis como taxa de chegada (λ), tempo de processamento (t_e), variabilidades (c_a^2, c_e^2), disponibilidade

(*A*) e setup (*S*) se mostraram determinantes para explicar o desempenho. Na decomposição dos tempos, observou-se que entre 79% e 90% do tempo de ciclo das máquinas está associado à espera em fila, e não ao processamento físico, reforçando a importância de reduzir causas que provocam acúmulo no sistema.

Esse padrão confirma empiricamente o princípio central formulado por Hopp e Spearman (2008): a variabilidade é o principal inimigo da eficiência em sistemas produtivos. Segundo os autores, flutuações em chegadas, tempos de processamento e interrupções amplificam o lead time e deterioram o fluxo. Os achados deste trabalho reproduzem exatamente essa lógica: mesmo com tempos de processamento relativamente curtos, a combinação de setups, falhas e defeitos gera um aumento desproporcional no tempo de espera, demonstrando que o desempenho produtivo é limitado mais pela variabilidade do que pela velocidade de processamento.

Os resultados integrados mostram ainda que desempenho não depende de um único fator, mas do equilíbrio entre velocidade, estabilidade e confiabilidade. Reduções de setup, padronização de processo, aumento da disponibilidade e mitigação de falhas são ações que se reforçam mutuamente. Na prática, isso significa maior previsibilidade, menor WIP e melhor utilização da capacidade instalada.

Limitações do estudo

Este trabalho apresenta limitações que merecem ser reconhecidas:

- Os dados utilizados representam um período específico e podem não capturar sazonalidades ou comportamentos anômalos.
- O modelo assume independência entre fontes de variabilidade, embora correlações reais possam existir (por exemplo, maior probabilidade de falhas após setups longos).
- O sistema foi modelado como um recurso isolado ($G/G/1$), não incluindo interações entre máquinas, gargalos em série ou restrições de material.
- As distribuições estatísticas adotadas (principalmente lognormais) representam aproximações e podem ser refinadas com dados item a item.

- A simulação não incorporou custos financeiros de manutenção, setups, estoques e perdas por defeitos, o que limita análises de viabilidade econômica.

Essas limitações não comprometem a validade dos resultados, mas indicam oportunidades claras de aprofundamento futuro.

Recomendações para pesquisas futuras

Com base nos achados e limitações, recomenda-se:

- Expandir o modelo para linhas completas de produção, utilizando redes de filas para capturar a propagação de variabilidade entre máquinas.
- Incorporar algoritmos de otimização ou heurísticas para escolha de tamanhos de lote, sequenciamento e políticas de manutenção.
- Realizar análises econômicas integradas (VPL, custos de WIP, custo de indisponibilidade), permitindo priorização financeira mais precisa.
- Calibrar distribuições com dados de ciclo item a item ou dados de sensores, tornando o modelo mais aderente à microvariabilidade real.
- Investigar modelos híbridos que combinem simulação discreta, simulação contínua e modelos estocásticos fechados em forma analítica.

Reflexão final

Este Trabalho Final de Curso representou uma experiência de grande aprendizado dentro da minha formação em Engenharia Física. A realização deste estudo permitiu integrar dois campos distintos, mas complementares: as simulações computacionais baseadas em parâmetros físicos e a engenharia orientada à tomada de decisão. Ao longo do desenvolvimento, foi possível aplicar conceitos de modelagem matemática, estatística e análise de variabilidade para complementar decisões mais fundamentadas em ambientes industriais complexos. Ao longo do desenvolvimento do trabalho, tornou-se evidente que a compreensão dos fenômenos de variabilidade não pode ser dissociada da análise sistêmica do processo produtivo. A variabilidade,

muitas vezes tratada apenas como ruído operacional, revelou-se um elemento estruturante do desempenho, capaz de amplificar ou mitigar os efeitos de decisões estratégicas.

Nesse sentido, o estudo reforça a importância de se abandonar abordagens puramente determinísticas e adotar modelos capazes de capturar a incerteza inerente aos sistemas reais. A utilização da simulação de Monte Carlo permitiu não apenas estimar valores médios, mas compreender a distribuição completa dos resultados, fornecendo subsídios para análises de risco e tomada de decisão sob incerteza. Essa perspectiva é particularmente relevante em ambientes industriais caracterizados por alta utilização, nos quais pequenas perturbações podem gerar impactos desproporcionais no desempenho global.

Além do aprendizado técnico, o desenvolvimento deste trabalho contribuiu para uma compreensão mais ampla do papel da modelagem matemática como ferramenta de apoio à gestão. A possibilidade de traduzir fenômenos complexos em modelos analíticos e computacionais permite transformar dados operacionais em conhecimento estruturado, capaz de orientar decisões estratégicas e operacionais de forma mais segura e fundamentada.

Por fim, este estudo reforça a importância de uma formação em engenharia orientada não apenas à solução de problemas pontuais, mas à compreensão sistêmica dos processos produtivos. A integração entre teoria, dados e simulação demonstrou ser um caminho eficaz para lidar com a complexidade inerente aos sistemas reais, contribuindo para o desenvolvimento de soluções mais robustas, eficientes e alinhadas às demandas contemporâneas da indústria.

8 Referências

1. Hopp, W. J., & Spearman, M. L. (2008). *Factory Physics: Foundations of Manufacturing Management*. McGraw-Hill.
2. Slack, N., Brandon-Jones, A., & Johnston, R. (2019). *Operations Management* (9th ed.). Pearson.
3. Little, J. D. C. (1961). A proof for the queuing formula $L = \lambda W$. *Operations Research*, 9(3), 383–387.
4. Goldratt, E. M., & Cox, J. (2016). *The Goal: A Process of Ongoing Improvement*. North River Press.
5. Deif, A. M. (2011). A system model for lean manufacturing design. *Computers & Industrial Engineering*, 61(1), 80–85.
6. Gross, D., & Harris, C. M. (2008). *Fundamentals of Queueing Theory* (4th ed.). Wiley.
7. Law, A. M. (2015). *Simulation Modeling and Analysis* (5th ed.). McGraw-Hill Education.
8. Rubinstein, R. Y., & Kroese, D. P. (2016). *Simulation and the Monte Carlo Method* (3rd ed.). Wiley.
9. McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2), 239–245.
10. Godinho Filho, M., & Barco, C. F. (2015). A framework for choosing among different lean-based improvement programs. *International Journal of Advanced Manufacturing Technology*, 81(1–4), 183–197. <https://doi.org/10.1007/s00170-015-7173-6>
11. Barco, C. F., & Godinho Filho, M. (2021). The effect of transfer lot size on manufacturing lead time: a stochastic analysis. *European Journal of Industrial Engineering*, 15(6), 825–848. <https://doi.org/10.1504/EJIE.2021.118502>

12. Chwif, L., & Medina, A. C. (2015). *Modelagem e Simulação de Eventos Discretos: Teoria e Aplicações*. Elsevier.
13. Womack, J. P., & Jones, D. T. (2003). *Lean Thinking: Banish Waste and Create Wealth in Your Corporation*. Free Press.
14. Yin, R. K. (2015). *Estudo de Caso: Planejamento e Métodos* (5ª ed.). Bookman.
15. O'Connor, P. (2012). *Practical Reliability Engineering* (5th ed.). Wiley.
16. Ebeling, C. E. (2010). *An Introduction to Reliability and Maintainability Engineering* (2nd ed.). Waveland Press.
17. Banks, J., Carson, J. S., Nelson, B. L., & Nicol, D. M. (2010). *Discrete-Event System Simulation* (5th ed.). Pearson.
18. Fishman, G. S. (1996). *Monte Carlo: Concepts, Algorithms, and Applications*. Springer.
19. Kingman, J. F. C. (1961). The single server queue in heavy traffic. *Mathematical Proceedings of the Cambridge Philosophical Society*, 57(4), 902–904.
20. Deming, W. E. (1986). *Out of the Crisis*. MIT Press.
21. Montgomery, D. C. (2012). *Introduction to Statistical Quality Control* (7th ed.). Wiley.
22. Shingo, S. (1985). *A Revolution in Manufacturing: The SMED System*. CRC Press.
23. Ross, S. A., Westerfield, R. W., & Jaffe, J. (2013). *Corporate Finance* (10th ed.). McGraw-Hill.